FULL WAVEFORM INVERSION BY MODEL EXTENSION: A ROBUST METHOD TO ESTIMATE THE SEISMIC PROPAGATION VELOCITY IN THE SUBSURFACE FROM RECORDED SEISMOGRAMS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF GEOPHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Guillaume Barnier
May 2022

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Biondo Biondi) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Robert Clapp)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Eric Dunham)

Approved for the Stanford University Committee on Graduate Studies

Abstract

Seismic imaging is an effective method to produce accurate maps of the Earth's subsurface, and has been employed for decades in global seismology, hydrocarbon exploration, geothermal energy production, and more recently CO_2 sequestration and monitoring. In complex geological settings, the quality of such maps highly depends on having a reliable seismic velocity model, which can be difficult to obtain. In this thesis, I develop a novel method, namely full waveform inversion by model extension (FWIME), designed to produce accurate acoustic velocity models of the subsurface from seismic recordings when conventional methods fail. I leverage the robust convergence properties of wave-equation migration velocity analysis (WEMVA) with the accuracy and high-resolution nature of acoustic full waveform inversion (FWI) by combining these techniques into a compact, mathematically consistent, and user-friendly workflow. By doing so, I mitigate the need for accurate initial models and the presence of coherent long-offset and/or low-frequency energy within the recorded data, which are difficult and costly to acquire but often necessary for conventional methods to succeed.

The novelty of my method resides in the design of a custom loss function and the optimization strategy I develop to pair WEMVA with FWI, which is more efficient and powerful than simply applying each method separately or sequentially. My new objective function contains two components. In the first one, I modify the forward mapping of the FWI problem by adding a data-correcting term computed with an extended demigration operator, whose goal is to ensure phase matching between predicted and observed data, even when the initial model is inaccurate. The second component, which is a modified WEMVA cost function, allows me to progressively remove the contributions of the data-correcting term throughout the inversion process. The coupling between the two components is automatically and seamlessly handled by the variable projection method, which also reduces the number of adjustable hyper-parameters, thereby making my solution simple to use. For the optimization process, I devise a model-space multi-scale approach based on model reduction using spline interpolation to control and gradually increase the spatial resolution of the velocity model updates. Unlike conventional methods, FWIME simultaneously inverts the full dataset and bandwidth from the start and does not require tedious data filtering/selection.

I illustrate the potential of my proposed method by accurately inverting datasets generated by

realistic 2D benchmark models which simulate complex and challenging geological scenarios encountered in field applications. In each scenario, the dataset lacks low-frequency energy and the initial velocity model is inaccurate, which prevents conventional methods from recovering useful solutions. I develop an efficient 3D numerical implementation of FWIME with the use of general-purpose graphics processing units (GPU) to handle 3D field datasets containing tens of terabytes of information, and to recover billions of unknown parameters. I successfully apply FWIME to a 3D ocean-bottom-node dataset acquired by Shell in the Gulf of Mexico. I show that my method outperforms conventional FWI and manages to improve the velocity model and the resulting subsurface image quality.

Preface

The electronic version of this report[1] makes the input programs and applications available to the reader. The markings **ER**, **CR**, and **NR** are promises by the author about the reproducibility of each figure result. Reproducibility is a way of organizing computational research that allows both the author and the reader of a publication to verify the reported results. Reproducibility facilitates the transfer of knowledge within the Stanford Exploration Projection (SEP) and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of processing described in the paper. The author claims that you can reproduce such a figure from the programs, parameters, and makefiles input in the electronic document. The data must either be input in the electronic distribution, be easily available to all researchers (e.g., SEG-EAGE data sets), or be available in the SEP data library[2]. We assume you have a UNIX workstation with Fortran, Fortran90, C, X-Windows system and the software downloadable from our website (SEP makerules, SEPScons, SEPlib, and the SEP latex package), or other free software such as SU. Before the publication of the electronic document, someone other than the author tests the author's claim by destroying and rebuilding all ER figures. Some ER figures may not be reproducible by outsiders because they depend on data sets that are too large to distribute, or data that we do not have permission to redistribute but are in the SEP data library, or that the rules depend on commercial packages such as Matlab or Mathematica.

CR denotes Conditional Reproducibility. The author certifies that the commands are in place to reproduce the figure if certain resources are available. The primary reasons for the CR designation is that the processing requires 20 minutes or more.

NR denotes Non-Reproducible figures. SEP discourages authors from flagging their figures as NR except for figures that are used solely for motivation, comparison, or illustration of the theory, such as: artist drawings, scannings, or figures taken from SEP reports not by the authors or from non-SEP publications.

[1] http://sepwww.stanford.edu/public/docs/sep154
[2] http://sepwww.stanford.edu/public/docs/sepdatalib/toc_html/

Acknowledgments

Before I thank the people I care about, I would like to tell my story. My adventure began with a surf trip in the south west of France during the Summer of 2010. I had just quit my banking job at J.P. Morgan in London, and I had no idea what I was going to do with the rest of my life. My friend Alexandre Araman, who was working as a geophysicist for Total at the time, came to visit me at my parents' beach house in Lacanau, told me about Geophysics, and said I would like it. I was intrigued and I looked into it. A few months later, after doing some research on the subject, I realized Geophysics was going to be a great fit for me, not only for the scientific challenge, but also because I knew the energy industry would always play a critical role in the world's economy, and I wanted to be a part of it as it was happening.

I first considered applying for a Geophysics MSc program in a French institute in Paris. The head of the program kindly spent two hours with me on the phone describing the different research projects they were working on, which sounded extremely interesting. I was ready to apply. However, I was told that my profile (I was 27 years old at the time) was too "farfelu", which means "wacky" in French. I had graduated from one of the top French scientific universities in Electrical Engineering and worked for three years in financial markets at J.P. Morgan throughout the financial meltdown of 2008. I guess from a French perspective, having experience in different fields is not always rewarded.

Alexandre then mentioned the Colorado School of Mines (CSM) where he had graduated from. He introduced me to professor Thomas Davis, the head and founder of an industry-funded academic research consortium referred to as the Reservoir Characterization Project (RCP), within the Geophysics department at CSM. I remember exactly what Dr. Davis said when I called him in the Fall of 2010: "You are a banker that wants to become a geophysicist? That's awesome, we currently have a lawyer trying to do the exact same thing!" Dr. Davis advised me to apply to the program and put me in contact with André Revil, another professor in the Geophysics department specialized in Hydrogeophysics (among many other topics). I immediately loved Dr. Davis and Dr. Revil's enthusiasm and passion for their work. In the Spring of 2011, I flew to Colorado to visit CSM and to convince them to accept me to the MSc program. The day I arrived at CSM, I just knew this was where I was supposed to be, I knew I was going to fall in love with this place and with the field of Geophysics. Sometimes, in life, you just know. The following day, Dr. Davis invited me to the

RCP consortium meeting and introduced me to an audience of over 100 people from the sponsor companies. Even though I had not been accepted yet, I felt I was home. I had zero experience in Geophysics or anything related to that field and yet, Dr. Davis, Dr. Revil, and the admission committee of CSM's Geophysics department eventually accepted me to the program with a scholarship (which is what all students get when they are accepted). By doing that, they took a calculated risk which I believe perfectly illustrates the adventurous mindset with which people approach life and conduct business in the United States.

At the time I started at CSM, I was 28 years old. I had been unemployed for more than one year and living at my parents' place. Going through difficult professional situations makes you understand the value of what you are given, and it makes you hungry to succeed. I knew CSM was the opportunity of a lifetime, and I tried to maximize my experience. I gave everything I had from the first day until the last, in all disciplines. I worked hard for my Geophysics courses, I tried to ride my bike as fast as possible up Lookout Mountain on a daily basis (my record is 19'50" from the the pillar to the turn towards Buffalo Bill's grave), and I gave my very best when training with CSM's ice hockey team. I also discovered that one of the many features making CSM so unique is its world-class teaching environment. My fellow students and I learned from some of the best geophysicists in the world, with most of the time less than eight people in the class. All my professors managed to convey their knowledge and passion for Geophysics, and deeply cared about their students. I had never seen such a commitment in trying to make sure students understood everything. I guess what surprised me is that my professors were both great scientists and unbelievably-gifted teachers, which is unfortunately, a rare combination.

At CSM, I had the chance to discover fascinating fields such as seismology, optimization of geophysical inverse problems, seismo-electricity, seismic processing and interpretation, and many more. These courses were taught by the world's leading experts: Thomas Davis, Dave Hale, Walt Lynn, Misac Nabighian, André Revil, Paul Sava, Ilya Tsvankin, etc. This was the first time of my life where I actually studied not for a "grade" but rather to understand the concepts for my own benefit. Indeed, I am very competitive, so I also wanted to get the best grades I could.

Towards the end of my MSc, thanks to my advisor André Revil and his postdoctoral researcher Marios Karaoulis, I discovered what people call "research". Because of my French education, I always thought one had to be a genius to be a good researcher, and that research was just about starting with a blank sheet of paper and writing complicated math theorems out of nowhere. I am definitely not a genius, and I cannot invent nor discover new math theorems on my own. Dr. Revil and Dr. Karaoulis taught me that research could be a fun, well-guided, and challenging scientific project. They also taught me how to use a computer to numerically solve complicated physics problems. Throughout my education in France, I had only solved partial differential equations in analytical forms, or computed integrals with a piece of paper and a pencil. I did not know computers could do that numerically. However, I realized that the rigorous scientific training I received in France

allowed me to approach research problems with an extensive and powerful mathematical toolbox. I never thought that some of the abstract mathematical concepts I had previously studied would ever help me in life. This discovery opened a new realm of possibilities in my head.

My original plan was to complete my MSc at CSM, get a job in an energy company, and start a normal life. But after this research experience, I wanted to see how far I could push my ability to conduct research in Geophysics. I still was not sure what I wanted in life but I was certain that I wanted to put myself in an uncomfortable and challenging situation because I was worried about losing my mental flexibility. For me, being able to adapt to any situation in life is the most important asset I want to maintain. I also admired a lot of the PhD students from the Center for Wave Phenomena (CWP) (another Geophysics research group at CSM) such as Filippo Broggini, Esteban Dìaz, Clement Fleury, Nishant Kamath, Francesco Perrone, and Simon Luo. Eventually, I knew I wanted to work with these types of scientists, and as a game, I wanted to try to beat them by creating the best algorithm I could.

Half-way through my CSM experience, I met Chris Leader and Ohad Barak who were PhD students in the Stanford Exploration Project (SEP) at Stanford University. They introduced me to Biondo Biondi, the director of SEP, at a conference in Las Vegas in 2012. I was fascinated by SEP because many of my professors at CSM had graduated from that program. Moreover, SEP was one of the first groups that pioneered the use of computers to solve large-scale seismic imaging problems. Finally, plenty of fundamental seismic imaging algorithms had been invented at SEP. I contacted Dr. Biondi six months after the official application deadline, he flew me to Stanford, and accepted me into the PhD program.

I joined SEP in September 2013, approximately nine years ago. I never thought it would take me so long to finish, but I had not realized the vast range of learning opportunities and discoveries to be made at Stanford University. In fact, when I arrived, I thought I was fully ready to start doing research. I had obtained a perfect GPA score at CSM, so in my head, I was a big shot. Little did I know... I really overestimated myself and my ability to conduct a research project at the PhD level. Being good at taking courses is completely different than being creative and finding innovative ways to solve a new problem. In short, I had absolutely no idea of what I was doing, and I was a joke. I quickly realized that when Bob Clapp looked at my CSM transcript and crossed out all the classes I had taken, except one (my least favorite), and said they were all useless for my PhD here.

During my first month at SEP, I was told that one of the student from the group was “a genius and the smartest person I would ever meet.” Rather than being impressed, I was surprised that someone could claim such statement without knowing who I had met before. Throughout my life experiences, I noticed that people love to create myths in these types of competitive environments. What I also realized is that when you move to a new place, you have to start from scratch, no matter who you are. You are nobody unless proven otherwise. This might be obvious when I write it now, but it was not to me at that time. Nevertheless, these types of remarks fuelled my competitive

nature and made me want to succeed even more.

My first year at SEP was extremely brutal and mostly unpleasant. Daniel Blatter, Gustavo Alves (my first-year cohort) and I were quite surprised by the learning environment. The difficulty was not the scientific content in itself but finding a way to access the information. Unlike places where I had previously studied, SEP did not teach you in the conventional manner. In fact, they did not teach you at all. At first, this experience can be frustrating because most of the concepts you are supposed to learn were invented by people at SEP (the same people that are supposed to teach you). There are no "books" you can buy on Amazon where all the answers are written at the back. At some point, I thought I would never understand these important concepts. I was disappointed because I thought I was going to be spoon-fed that information from the people that invented most of these seismic imaging algorithms. In fact, if I had started Geophysics at SEP, I would have hated that field. I wanted to quit every day but I did not want to give anybody the satisfaction that I could be broken. Besides, I had been through way tougher moments during my short banking career.

Looking back at it now, I understand that this first-year experience was just an abrupt introduction to research. This job is not about taking a book written by someone and doing 100 math exercises correctly, which is what I had been trained to do. Instead (and to follow this analogy), it is about writing and gradually inventing the book itself, thereby creating your own path. For me, the key was to do whatever it took to access the information I needed, as if my life depended on it. I wanted to understand everything and every little detail, and I did not care about looking like an idiot if this mindset got me where I wanted. Eventually, with enough will power (and the help of Carlo Fortini during my first year), I succeeded. This adventure eventually gives you a lot of freedom, but there is an entry cost, and not everybody is willing to pay for it. The rest of my PhD experience at Stanford was incredible. It was not pleasant, but it was literally extraordinary. I guess no matter where you are in the world, if you just focus on one task and work at it all day, every day, it is not going to be fun.

The turning point of my PhD occurred at the beginning of my fourth year, when I started collaborating with Ettore Biondi, a fellow student from SEP (not related to our advisor, Biondo Biondi). Ettore and I rapidly realized that we enjoyed working together and gradually began cooperating on pretty much every research projects/ideas we had. We were lucky enough to have similar research interests and a shared passion for seismic imaging, optimization problems, and numerical implementations. Most importantly, we wanted to conduct scientifically clean reproducible research, and we wanted to be the best. In 2017, we found an idea of a new seismic imaging algorithm that eventually became my PhD topic. Because seismic imaging is such a mature field, it took us approximately four years (from 2013 to 2017) to understand and implement the fundamental algorithms that had been discovered by researchers in the previous decades. Our approach was to cut zero corners. We believed that in order to successfully develop and implement a new algorithm, we had to master every aspect and every line of our code. We wanted to make sure if that our method did not work,

it was due to the concept itself and not because of a poor numerical implementation.

From 2017 onward, Ettore and I worked as if we had a start-up. We convinced Dr. Biondi to give us our autonomy and an incredible amount of computational resources. He trusted our ability to deliver results on our own without much supervision. Every morning, Ettore and I would brainstorm research ideas on the board in our office, plan our daily strategy, and split the work between the two of us in order to maximize our efficiency. Once we knew what we had to do, the next phase of both of our PhD began. I had found my path, and now I just had to make it work. I think I became obsessed in making my algorithm work. I gave everything I had: I worked all day, every day, and the only thing I cared about was succeeding. In 2019, Ettore and I obtained our first compelling results, which led us to win the best student paper at the 2019 Society of Exploration Geophysicist (SEG) annual conference. Since then, we continued working together and improved our results even further, as I show in this thesis. I think what we enjoyed the most about our work was the fact that it could be applied to real-life problems encountered in the energy industry. Therefore, we always managed to get meaningful feedback from expert researchers working for our sponsor companies, which helped us advance more efficiently.

Unfortunately, when you are a PhD student, some of your family members and friends think you are just going to school, taking classes, and doing homework. People often ask you: "So, when you are done with school? It's about time to get a real job." They do not understand why a 38-year-old Guillaume would waste his time doing that. But people have no idea what you are doing and they are not even aware of their ignorance, including some very educated people. Just because they think they would not enjoy the process because "it is not normal", they conclude that it is a total waste of time and that you should also not enjoy it. I see this attitude as a total lack of respect, especially because in the United States, most PhD students tackle real-life problems that are usually extremely useful for the industry. People that have done a PhD know exactly what I am describing. My goal was not to be done with school, but rather to create the best seismic imaging algorithm I could. Indeed, "taking classes" is definitely part of the process (at least at the beginning), but my PhD experience was incredibly richer than that.

In the last nine years at Stanford, I experienced a sequence of intense (and somewhat random) adventures. Besides my PhD thesis work, professor Eric Dunham and I developed a numerical scheme to model tsunami wavefields propagation with Kalman filters, and prototype an early-warning tsunami system. I also worked in the oil and gas industry for Total, Chevron, BP, and Petrobras on real-life and impactful problems. In 2017, I was caught in the middle of Hurricane Harvey in Houston. I was evacuated on a boat in the middle of the city, I appeared in the local news TV channel, and I was able to come back to my apartment only three weeks later. I picked up my belongings, put them on a Safeway kart, and left. On the way back to California, I picked-up my friend François in Albuquerque, and we went one of the most epic trail-running road trips I have ever done in my life, passing through Arizona, Utah, Nevada (with a solid pit stop in Las Vegas), and California.

I also enrolled in a Hebrew class at Stanford for one year (for a girl), which was really fun but very challenging. I would wake up at 5am every day, work on my Hebrew lesson for two hours, go to Hebrew class, and then start my "normal" PhD day. Eventually, it did not work out with that girl, but at least I managed to learn Hebrew. For fun, Ettore and I took classes in the mathematics department on stochastic processes and stochastic partial differential equations. In 2108, I decided to learn how to code in C++/CUDA and become an expert in general-purpose graphics processing units (GPU), focusing on developing efficient finite-difference code to simulate seismic wave propagation. My advisor let me do that during the summer, and I worked at it all day, every day. It turns out that the algorithm I created for my PhD thesis would have never been able to work if I had not developed these computational skills. Then, I went to burning man where I ran the ultra-marathon. Some of the runners were naked and people made me take tequila shots during the race. That same year, I was invited by Petrobras in Rio de Janeiro (Brazil) for one month, in order to collaborate with their seismic imaging research team. I took this opportunity to travel and trail-run through Patagonia for another month.

In the Bay Area, I discovered trail-running. I became obsessed by it and overly competitive. I trained hard and I won three races (the overall classification, not just my "age group"). I won two half-marathons in Mount Diablo and Montara Mountain, and one Spartan Race in Mount Diablo.

In 2019, Ettore and I initiated a collaboration with Paul Williamson and Mohamed Dolliazal at Total to investigate the use of our algorithm to potentially solve some of the problems encountered in their hydrocarbon-producing fields. They invited us to consult and work with them in Houston at multiple occasions, which was extremely fruitful. Thanks to our advisor, Ettore and I traveled and presented our work at conferences in Denmark, Italy, France, the UK, etc.

In the last few years, during the pandemic, I decided to fully investigate the field of machine learning (ML). My advisor allowed me to enroll in multiple classes such as machine learning, deep learning, and reinforcement learning. These courses were taught by some of the best ML scientists in the world, such as Andrew Ng. In January 2021, I initiated a deep learning project with professor Elizabeth Tong from the neurology department at the Stanford School of Medicine. We built a team of four students (which included Ettore), and developed an algorithm that could predict brain-tissue damage for ischemic stroke patients. We are all still currently working hard on this project and we just managed to file a patent for our proposed methodology. This project is particularly meaningful for me because it may have a direct impact on how patients recover from strokes.

Throughout my PhD, I moved 15 times and I lived in Stanford, San Francisco, Mountain View, Houston, Redwood City, Palo Alto, Roquefort les Pins (France), Cannes (France), Nice (France), and Berkeley. I also had the opportunity to work with scientists from around the world, from various nationalities and many different cultures.

Indeed, the above list is merely a subset of all the stories and experiences I had the chance to live throughout my PhD life at Stanford, while still working full time on my main thesis. Some moments

were pleasant, some were less fun, and some were horrible. Most importantly, this had little to do with school or school work. My goal was never to just "finish" my PhD to get a good job (even though it is important), but rather to enjoy, learn, and grow as much as I could from the process. I worked hard to become the best researcher and scientist I could be, so that I would have no regret at the end of the adventure.

This story would have never been possible without the help of others. In fact, every meaningful goals I reached in my life were achieved because I was part of a good team. I never thought I would be able to replicate this template during my PhD, but Ettore, Stuart Farris, and I managed to create an amazing teamwork environment. In most graduate programs, I believe the scientific difficulty is not the main issue. The main hurdle is the mental aspect. Therefore, sharing these tough moments with good teammates and office mates can make the experience more bearable, and sometimes even fun.

That being said, I now would like to acknowledge the people that helped me throughout this journey. First, I want to thank Alexandre Araman. Besides being a great friend and a wise mentor, Alexandre introduced me to Geophysics, put me in contact with Tom David, found me an internship at Total after my first year at CSM, and finally convinced the recruiters at Total to offer me a job with a great salary. I can never thank him enough for all that.

I want to thank the Colorado School of Mines Geophysics department as a whole. If it was not for these people, I would have hated Geophysics and I would have never finished my PhD at SEP. They gave me the best opportunity of my life, trusted me without knowing much about me, and gave me the passion for Geophysics. In particular, Tom Davis and André Revil were the ones that had faith in me from the beginning. Their were great mentors during my time at CSM. When I had questions regarding my career choices, Tom Davis was the first one I consulted. André Revil was not only my advisor, but he also became my friend. I loved doing research with him because he possesses such an innovative, creative and dynamic mind. Moreover, I enjoyed his honesty and his bluntness, which were quite refreshing. I miss our morning conversations while eating a muffin at Café 13.

I also want to thank all my professors and mentors at CSM: Dave Hale, Ken Larner, Yaoguo Li, Walt Lynn, Misac Nabighian, Paul Sava, John Stockwell, Ilya Tsvankin, and Terry Young. I think Dr. Tsvankin was the most influential professor I had at CSM. He gave me the passion for seismology and introduced me to the world of seismic anisotropy. Indeed, my friends and fellow colleagues played an important role in my life at CSM: Andrew Muñoz, Chris Steinhoff, Clement Fleury, Esteban and Carla Dìaz, Filippo Broggini, Francesco Perrone, Heather Davey, Jieyi Zhou, JJ Bishop, Kjetil Haavik, Loren Ziegler, Luiz Martins, Marios Karaoulis, Nishant Kamath, Oscar Jarillo Michel, Satyan Singh, and Chi Zhang.

My friend Imad Atshan and his wife Farah were also part of the reason I made it to (and through) Stanford University, and I want to tell them I am grateful for their friendship. Imad and I started

together at Colorado School of Mines, and stayed friends since then. We had no idea what Geophysics was about and I remember walking on the exhibition floor at the SEG 2011 annual meeting in San Antonio. We were like kids at Disney Land trying to ride acquisition trucks, and we did not go to one single technical presentation. Even though Imad is younger than me, he has always been more mature and down to Earth. When I applied to Stanford, he re-read my letter to Dr. Biondi and corrected most of it. He added that I had a 4.0 GPA. He was also there when I needed him the most. Bottom line, among all their qualities, I really value Imad and Farah's generosity, intelligence and kindness.

I thank all the SEP sponsors for their financial and scientific support over the years, even during the challenging times encountered by the energy industry in the past decade. None of our work at SEP would be possible without such solid partners. I also thank Shell Exploration and Production for providing the 3D dataset presented in this thesis and allowing me to present my results.

In 2015, I had the chance to work at Chevron and meet some great mentors and scientists such as Dimitri Bevc, Mike Hoversten, Mark Meadows, and John Washbourne. I learned a lot from their expertise and technical skills. Mark Meadows played a particular role in the success of my algorithm. His curiosity, advice, interest, and questions related to my work helped me mitigate some challenging bottlenecks.

At BP, I deeply thank Imtiaz Ahmed, Andrew Brenders, Joe Dellinger, John Etgen, Elizabeth L'Heureux, David Meaux, Scott Michell (now at TGS), Xukai Shen, and Madhav Vyas. More specifically, I thank David for being a great mentor. David and I did not work together for long but I learned a lot from him and from his way of approaching seismic imaging. Before working at BP, I never took the time to step back and analyze the importance of what research and development teams bring to the table from a business perspective. For me, seismic imaging was just cool, so I did it. But David helped me reflect on this important aspect, which also influenced the way I approached my scientific presentations. Finally, I also have a lot of respect for his management style: he is able to give people some "slack" while still ensuring that his teams are very efficient and productive.

At Total, I appreciate the time spent with Paul Willimason and Mohamed Dolliazal. Ettore, Paul, Mohamed and I had some interesting discussions not only about Geophysics, but about life in general. Paul and Mohamed believed in us and in our algorithm. They invested considerable resources and time to mentor us, they helped us improve our algorithm, and they invited us on multiple occasions to their Houston office to collaborate. I am thankful for their generosity and patience. All the memories I had while working with them were fantastic. At Petrobras, I thank Gustavo Alves and André Bulcão for inviting me to present my research and work with them in Rio de Janeiro. We had a great time collaborating, though half of the time was spent trying to install SEPLIB on their machines. At Exxon, I thank Anatoly Baumstein. I have always been impressed and inspired by Anatoly's work. More recently, Anatoly helped me through some difficult times, and I will never forget that.

I thank my advisor Biondo Biondi for accepting me to SEP and supporting me for almost one decade. One of the main qualities I discovered about Biondo is his loyalty to people, and more specifically to his students. Biondo will always have your back, no matter what. I also admire Biondo's long-term vision for the group and his ability to maintain SEP as a leading academic research consortium over the last few decades. His wisdom and deep understanding of the energy industry enabled him to anticipate the needs from our sponsor companies and stir students' research projects accordingly. Biondo also gave me the opportunity to thrive at Stanford. He did not "teach" me or "advise" me in the conventional way. He did not tell me what to do. Instead, he let me grow as an independent research scientist and allowed me to design the path I wanted for my PhD. Even though this may look great on paper, it was not always pleasant or easy. In fact, hard work rarely is, there is no magic. But I will always be thankful for the scientific freedom he gave me. Finally, I value the that fact that Biondo always tries to push his students to try new research areas and investigate new scientific topics, even if they do not directly align with the group's research focus. I think Biondo believes that innovative ideas can be inspired by various techniques employed in other disciplines, and also by collaborating with different scientific fields. This approach also materializes by Biondo's recruiting philosophy: most of the students from SEP come from very different scientific backgrounds.

Bob Clapp is the closest person to an advisor I have had for the past nine years at SEP. Bob is extremely intelligent, creative, messy, and straightforward. If you need to know the truth about how you are performing, Bob will tell you straight away. It may sound very cliché, but Bob taught me how to approach and solve research problems, and to relax when things do not work, because they never do at first. I really enjoyed our weekly meetings where we talked about everything. These meetings were very refreshing, and I will really miss them. Finally, I appreciate Bob's genuine passion for Geophysics and for seismic imaging. He advised me during the entire development of my thesis work, and I could see that he thought it was cool. This feeling was one of the best I had at SEP.

Indeed, I thank my defense committee: Eric Dunham, Dustin Schroeder, and Elizabeth Tong. I had the pleasure to work with Eric Dunham on my secondary project, which focused on numerical simulation of tsunami wavefields to create early-warning systems. In addition to Eric's intelligence and rigorous approach to research, I really appreciate the fact that he takes the time to meet with all of his students, every single week. Elizabeth Tong and I started collaborating on a machine learning project in January 2021, which aimed at creating an algorithm predicting brain-tissue damage for stroke patients. Working with Dr. Tong was just incredible. I loved her optimism, generosity, curiosity, expertise, dynamism, commitment, and intelligence. I feel very lucky for this experience.

Rachael Madison and Liliane Pereira are both kind and extremely competent. They work behind the scenes to make sure everything goes seamlessly for students and professors, and they always have a smile on their face. They also both care a lot about the students. I cannot count the number of

times they helped me because I had forgotten to fill out a form, or I had not registered to a class with the right amount of credit. Unfortunately, it is very easy to overlook the work they do every day. I am very thankful and lucky I had the chance to work with them, and I will never forget them.

I also want to thank Alex Maurel, Ariel Lellouch, Arnaud Pladys, Aurelien Bocquet, Aurore Vidal, Carlo Romero, Charles Limouse, Chris Haubrich, Guillaume Emmenegger, Guillaume "le legionnaire" Sivadier, Julien and Rachel Bacal, Justin Tarpey, Laurene and the Petitjean family, Laurent Hautefeuille, Maï Dang Vu, Nick Noone, Olivié Charbonneau, Paul Planeix, Samia Naciri, Sheila Bhardwaj, Stanislav Kubacek, Stephen MacCabe, and Vincent Seguin. These people really counted (and still do) in my life. I acknowledge my fellow SEP and Geophysics students, namely Rustam Akhmadiev, Gader Alalli, Ohad Barak (and his wife Sylvie), Milad Bader, Alejandro Cabrales, Jason Chang, Thomas Cullison, Eliza Dawson, Miguel Ferrer, Julio Frigerio, Fantine Huot, Chris Leader, Dongzhuo Li, Elita Li, Musa Maharramov, Eileen Martin, Andreas Mavrommatis, Ossian O'Reilly, Sjoerd De Ridder, Rahul Sarkar, Kevin Seats, Yi Shen, Joe Stitt, Jonathan Voyles, Mandy Wong, and Yang Zhang. A special thanks to Thomas Cullison who not only guided me during the development of my GPU code, but also helped me get my dream job at NVIDIA.

My friends at Amity were part of the reason I stayed somewhat sane during these difficult years. There is no way I could have made it without such a great group of athletes and friends. I thank Aaron, Avo, Hyoung, Jamie, Leah, Marissa, Markus, Mira, and Travis. I also thank NCFIT, including Emily, Franky, Jason, Liz, Nikki, Rae Lee, and Zach.

Jean Virieux and Bill Symes were both incredible mentors and sources of inspiration for Ettore and myself. In addition to many fruitful conversations I had with Jean, I would have never been able to make my algorithm work without him. In May 2019, he came to Stanford as a visiting professor, and met with us on a daily basis. At that time, I was stuck on a problem with my method, and he gave me an idea that solved my problem. I would still be stuck at that point if it was not for Jean Virieux. Bill Symes is also someone Ettore and I admire. We love his research (my thesis algorithm builds on his work), his approach to seismic imaging, and his rigor.

The first-year mafia was composed of Daniel Blatter, Gustavo Aves, and myself. Daniel and I had no idea about what we were supposed to learn and do for our first nine months at SEP. Gustavo managed to navigate the first year a touch better than us, probably because he had more experience in seismic imaging. I was honored to share my first year with these two guys. We suffered together trying to translate in human language what Jon Claerbout was trying to say, and eventually, we gave up. Fun times. Gustavo and I continued to work together over the years, and he invited me to Brazil in 2018 to collaborate. He and his wife Caro showed me around Rio de Janeiro, and introduced me to their friends. A great memory.

Guillaume "el ladron" Barnier sends a big up to the full TVX team: Alex "the general" Araman, Florian "cash flow" Graillot, Thibault "the manager" Chabrelie, Frederic "fchavano" Chavanon, Henri "riton" Lecieux, Paul "maverick" Wilczynski, Rodolphe "the frog" Jenatton, Frederic "DJ

Rico" Mereau, Thomas "the patriot" Volmer, and Will "Willou" Wierzowski. I really owe Rodolphe a lot. In the past two years, he went out of his way to help me get a job and I cannot forget that. I will also miss Tom's visits to the Bay Area and camping trips. We had some really fun times. I hope we can continue the tradition in Texas.

I will miss Joe Stefani. He is one of the smartest, most humble, and "chill" person I know. Joe is passionate and curious about Geophysics, Physics, and pretty much any scientific topic. We both shared a passion for trail running (and racing), which led us to spend countless hours running together. He showed me the beauty and the greatness of Marin county, a world-class playground for trail-running. I remember when he and I ran the Mount Tamalpais hill climb race on Labor day in 2015. That was one of shortest but most intense suffer fest I have ever participated in. I also remember when I would pick him up from his place at 7am in San Francisco in the weekend. We would drive up to Marin county, park the SUV, and run for hours in the wilderness of Marin. These are probably the best memories I keep from my time at Stanford.

My friend Carlo Fortini helped me survive my first year at SEP. He taught me all the fundamental concepts of seismic imaging, and most importantly, how to translate them into code. He and I worked really hard during my first summer. We re-derived all the seismic operators along with their adjoint, and we wrote our own code from scratch. After that, I knew I was ready and that I would eventually get my PhD.

I want to thank François and his wife, Marion. They are my friends, and there is nothing much more to say. They know they were always there for me, for the good and the bad times. I will miss our Sunday lunches at Zareen's after the traditional (and mandatory) farmers' market. Also, if it wasn't for Francois, I would still be at the bottom of the Grand Canyon.

Roy Aviner is just one of those rare people who are extremely successful, smart, kind, humble, and loyal. I met Roy when I was working at J.P. Morgan. One day in January 2009, my boss told me he was going to fire me. When everybody at my desk left for the day, I stayed and I cried in front of my computer, like an idiot, in the middle of the trading floor. Roy saw me and had some very kind words, and I think our friendship started that day. There are so many good things I could say about Roy. Among them, Roy is the kind of friend that calls you at 2am from Dubai when you are 38 years old, still a PhD student, alone, depressed, and unsure of yourself. While other people make fun of you, he encourages you and tells you he admires what you are doing. Sometimes, you just need to hear that sentence to make it through the desert, and Roy is the type of person that will help you in these difficult moments. A true friend.

Stuart Farris is one of my main pillars of my life at Stanford. Stuart, Ettore and I shared the same office for about five years. There are so many epic moments we all had together that it would be a letdown to try to describe them. One thing I can say is that I would not have made it without him. He is very intelligent, logical, calm, down to Earth, and patient. I do not know how he managed to stay sane while being in the office with me during horrible and frustrating times. I admit, I may

have a slight tendency to swear in French when the black and white Unix terminal does not do what I want it to do. All these qualities make him a powerful researcher, and a great friend. I just hope that one day, I will be able do more push-ups and pull-ups than him. Thankfully, I still dominate him on the assault bike, and I plan on keeping it that way.

Mandana was part of my life for a while, and even if we went our separate ways, I will never forget her. She is a kind, caring, sensitive, beautiful and very generous person. She also makes the best Tiramisu, but she is not a great swimmer. I was really lucky to meet her as we shared some unforgettable moments. I hope to see her on television at the Olympics one day (Paris 2024?).

To Niels, Eleonore, Mamat, Julie, Guillaume, and Anne-Sophie: my friends, we were not able to see each other much in the past decade, but I always think about you and I know you will always support me no matter what. I cannot wait to be back into your lives. Mamat, thanks for having my back, for visiting me, and for encouraging me during the toughest moments.

Vincent, Elise, Isa, Manu, Michele, Lena, Aline, Clement, and Romy are like family to me. They never miss a birthday or an occasion to call. Vincent and I grew up together and he is one of my oldest and closest friends. We met when we were six years old, went to school together, played soccer against each other, played ping pong, NBA JAM, pretended we were part of some basketball dunk contests (Jordan versus Shaq), and spent most of our childhood and teenage years next to each other. I admire Vincent and Elise, they created a great universe for their family and children. They are loyal, kind, smart, and logical. Being around them makes you happy. Gianni and Candy are very similar. Loyal, caring and genuinely interested in knowing what is going on in their friends' lives. I am also thankful for my long-time friend Cyrielle (and her newborn baby girl, Roxane).

Indeed, I want to thank my family. My uncle Bernard and his partner, Amin. Bernard was always there to guide me through life and share his wisdom. My cousin Philippe who is always there for me. Among many good memories we have together, Philippe invited me to Switzerland to trail-run for one week in the snow. It was an epic - but cold - trip. I also thank Maxime, Nina, Alya, and Nyssa. They always think about me. I hope to be more involved in their lives in the coming years. Of course, I thank my dear aunt Anne, who is one of the most generous and selfless person I know. And finally, my grand mother Claudette. And even though Michele Seguin is not technically part of my family, it really feels like she is. She met me when I was 15 years old and gave me some Latin lessons while I was recovering from a serious ski accident. She has been a mentor since then.

By now, I think it made it pretty clear that in the past eight years, Ettore and I made the best team I have ever formed in my life. Together, we went to Hell, but we eventually made it back on Earth. I have no word to describe what we went through together at SEP and in our personal lives. Ettore brought me (and the people in SEP) so much knowledge and wisdom. He always has a positive and optimistic attitude, I have no idea how he does that. I have always said that there is not one bad thing I can say about Ettore, and I really hope I will never stop on working with him.

Finally, I must dedicate my thesis work to my parents, Catherine and Laurent. They did not

always understand me, but they always supported me with whatever I decided to do in my life. I really do not know what I would have done if I had a son as intense, weird and annoying as myself, and whether I would have had the same patience they had for me. They offered me everything, they enabled me to travel around the world, live abroad, discover new countries and cultures, learn five different languages, go on very long surf trips at the end of the world without hearing from me for months. They have always had an unlimited patience and generosity towards me, and I feel very lucky to have them. They have been there with me for the full ride, and I owe them everything.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Similar to sonar, seismic imaging (or reflection seismology) is a high-resolution non-invasive technique in which seismic waves, more commonly known as sound waves, are employed to produce accurate maps of the subsurface. When seismic waves propagate through the ground, they interact with the internal structure of the Earth, such as interfaces between different rock formations. Such interactions may cause seismic waves to be reverberated and then potentially measured by sensors placed at the surface or in boreholes. Once processed, these measurements can provide useful information to identify and localize subsurface features of various scales, such as the core-mantle boundary location, giant meteor craters, earthquake source locations, and geologic traps for hydrocarbon reservoirs. For the past 50 years, seismic imaging has been the most exploited and effective method for hydrocarbon exploration. Energy companies heavily rely on having good-quality seismic images to ensure safety and efficiency during exploration, drilling and production.

However, reliable images of the subsurface can be quite challenging to generate in certain regions of the world where complex geological settings are encountered. The main difficulty within the imaging process is to obtain an accurate estimate of the speed at which seismic waves propagate in the ground, referred to as a velocity model. Moreover, small errors in velocity models may lead to large errors in seismic images, making them untrustworthy for exploration. Therefore, energy companies and academic groups have invested considerable amounts of resources in research and development to improve velocity-model building algorithms. As a result, this topic has become one of the most competitive and active areas of research within the seismic imaging community. Many promising techniques have been proposed and much progress has been made in the past two decades, but so far, no one has been able to design a robust, user-friendly, and computationally-tractable method. In this thesis, I tackle this challenging task by developing a novel velocity-model building technique referred to as full waveform inversion by model extension (FWIME).

In this chapter, I provide an overview of the oil and gas industry's conventional seismic imaging process for hydrocarbon exploration and I highlight the limitations of current techniques. Then, I

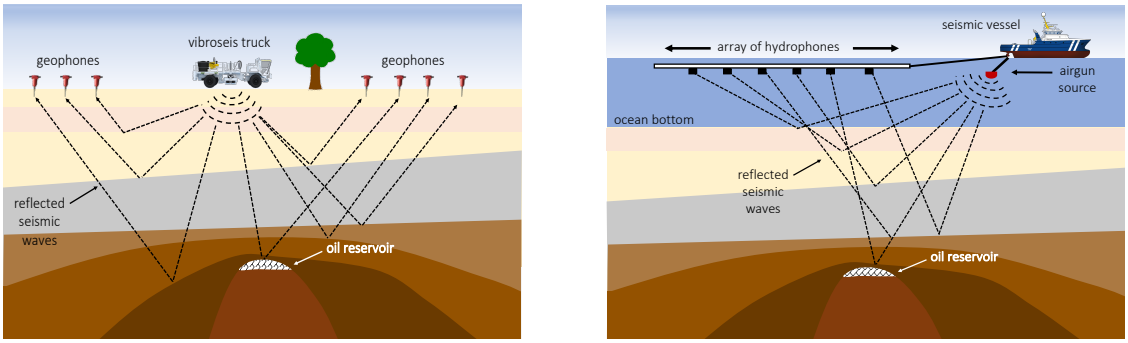
briefly explain the novelty and the contributions that my dissertation work brings to the field of seismic imaging. Finally, I summarize the outline of my thesis.

1.1 Conventional imaging: a three-step process

The industry's conventional imaging process is conducted in three stages: (1) seismic data acquisition, (2) velocity model estimation, and (3) migration. In this section, I give a high-level overview of each step.

1.1.1 Step 1: Seismic data acquisition

Seismic data are acquired on the field by conducting a seismic survey, in which seismic waves are generated by controlled active sources and propagated into the ground. When these waves encounter an interface between two different rock layers (an abrupt change of rock properties), part of their energy is reflected back to the surface, and recorded by an array of sensors. These recordings are referred to as seismograms and are similar to the ones measured by seismic stations that monitor earthquakes. For land surveys (onshore acquisitions), seismic signals are generated with vibroseis trucks or dynamite explosions, and are recorded by geophones, as shown by the schematic diagram in Figure 1.1a. Marine seismic surveys (offshore acquisitions) use airgun sources deployed behind the seismic vessel, and an array of hydrophones (also towed behind the ship) record the water-pressure variations generated by the reflected energy (Figure 1.1b).

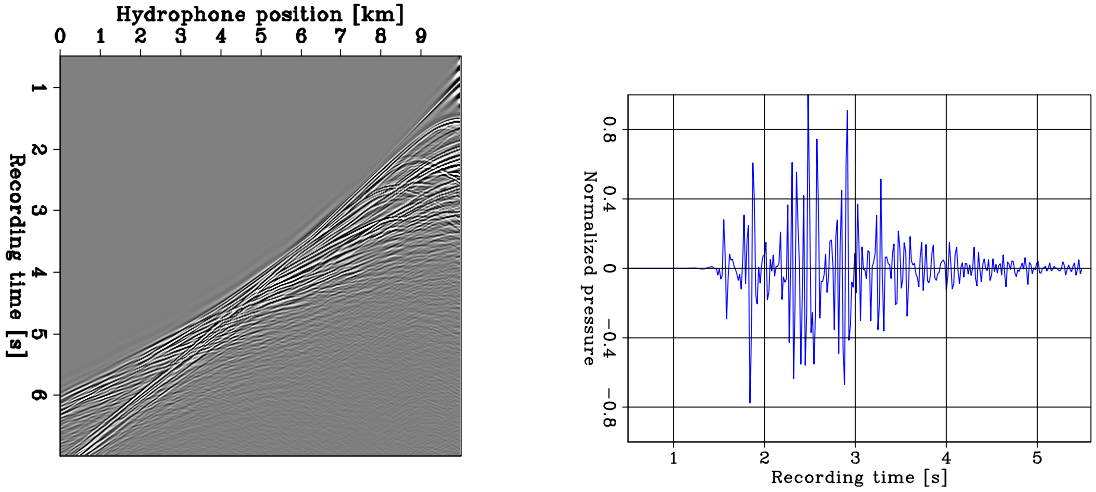


(a) (b)

Figure 1.1: Schematic diagrams representing seismic acquisition surveys. (a) Onshore survey. (b) Offshore survey. Black arrows represent seismic waves propagating through the subsurface. The various features shown in the diagrams are not drawn to scale. [**NR**]

Figure 1.2 shows an example of synthetic offshore seismic data numerically generated with a source placed at $x_s = 10$ km, and a horizontal array of hydrophones. Such panel is referred to as a common shot gather (one source, multiple sensors). The vertical axis corresponds to the recording time of the sensors, and the horizontal axis displays the hypdrophones' horizontal position (usually

towed a few tens of meters below the surface). Black/white colors correspond to the presence of pressure changes recorded by the geophones (arrival of a seismic wave), and gray indicates that no pressure change is observed. Figure 1.2b shows the normalized pressure changes as a function of time extracted from the hydrophone located at $x_r = 9$ km in Figure 1.2a, and is commonly referred to as a seismic trace. In order to gather enough information about the subsurface, seismic datasets are composed of thousands of shot gathers generated with sources placed at different locations. The goal is to "illuminate" or "interrogate" the subsurface from a wide range of positions and angles. Usually conducted by service companies, seismic surveys may last for months and cost tens of millions of dollars. Eventually, these datasets are stored numerically and may contain tens of terabytes of information.



(a) (b)

Figure 1.2: Representative seismic data recorded during an offshore acquisition survey with a source placed at $x_s = 10$ km. (a) Shot gather. (b) Pressure signal (seismic trace) extracted from panel (a) for a hydrophone placed at $x_r = 9$ km. [**ER**]

1.1.2 Step 2: Seismic velocity estimation

Velocity estimation is usually the most challenging step of the seismic imaging workflow, and is the focus of my dissertation work. Due to the high-dimensional nature of the model space (the unknown velocity values at each position in the ground) and the large amount of seismic data to process, this estimation step is recast as a numerical optimization problem and usually solved on large computer clusters.

In the context of oil and gas exploration, the subsurface area of interest is usually a few tens of kilometers in width and can be as deep as 20 km. The acoustic wave speed highly depends

on the type of rocks in which sound propagates, and typically ranges from 1.5 km/s for water to 7 km/s for volcanic rocks (compared to 0.3 km/s in the air). In the last decade, one particular velocity-model building algorithm referred to as acoustic full waveform inversion (FWI) has become the industry standard for velocity estimation (largely due to the increased computational power) and many case studies have shown its potential at recovering accurate and high-resolution solutions (Lailly and Bednar, 1983; Tarantola, 1984; Virieux and Operto, 2009; Shen et al., 2018). This algorithm requires two inputs: (1) seismic data, and (2) an approximate initial guess of the velocity model (referred to as the initial model). FWI is then applied in an iterative manner as follows:

1. Using an estimated velocity model, a seismic survey (similar to the one acquired on the field) is simulated on a computer.

2. The numerically simulated seismic data are then compared to the data measured on the field (field data).

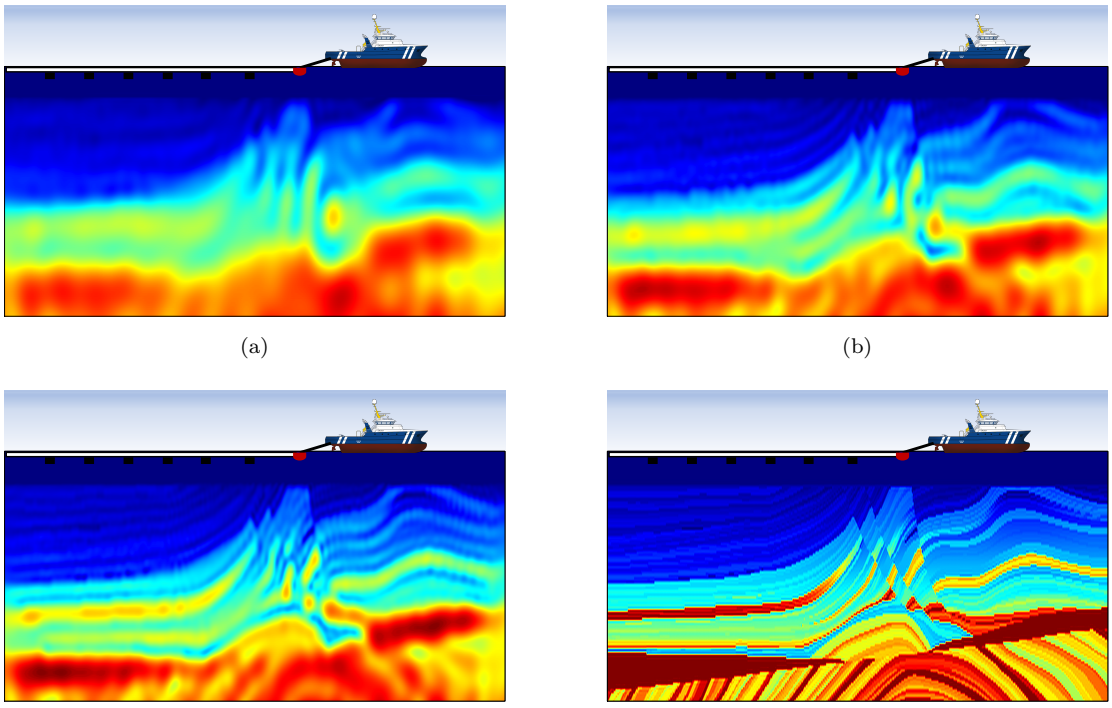
3. If the two datasets are not "similar" enough according to a predefined metric, it indicates that the estimated velocity model is inaccurate (not a good representation of the Earth). In that case, the lack of similarity between the two datasets (also called error, misfit, or residual) is translated into an update of the velocity model to gradually improve its accuracy. For machine learning applications, this step is usually referred to as the "back-propagation".

4. This iterative procedure is performed until the simulated seismic data becomes "similar enough" to the field data. That is, when the data misfit is minimized. The underlying assumption on which this process is based is that if the datasets are similar, the estimated velocity model has also become accurate.

I illustrate a successful application of FWI on the Marmousi2 model (Martin et al., 2006), a well-known 2D synthetic test used in the seismic imaging community to benchmark the performance of velocity model building algorithms. The true velocity model is shown in Figure 1.3d, and the vertical/horizontal axes represent depth and horizontal position. The model is approximately 17 km wide and 3.5 km deep. Warm/cold colors correspond to high and low velocity values, respectively (dark blue corresponds to 1.5 km/s, red is 4.5 km/s). Figures 1.3a-c show the sequence of velocity models obtained by FWI at various iterations. The initial model (Figure 1.3a) is low resolution and lacks many details. After applying FWI, the final recovered model (Figure 1.3c) is excellent and captures more high-resolution features from the true Earth.

1.1.3 Step 3: Seismic migration

In the third stage, the estimated velocity model and the recorded data are used as inputs to an algorithm called migration, which aims at producing higher-resolution images than in the previous



(a) (b)

(c) (d)

Figure 1.3: Panels showing 2D synthetic velocity models of the Earth recovered during a successful FWI process. The vertical and horizontal axes represent depth and horizontal position, respectively. The model is approximately 17 km wide and 3.5 km deep. (a) Initial model. (2) FWI model obtained after 40 iterations. (c) Final FWI model obtained after 80 iterations. (d) True model. The various features are not drawn to scale. [**CR**]

step. Unlike velocity-estimation techniques, the goal of migration is to accurately model the sharp interfaces between different rock layers (e.g., sandstone, shale, basalt), rather than to obtain the value of the speed of sound. Migration can be formally defined as the process by which seismic recordings (such as the one shown in Figure 1.2) are geometrically re-located in space to the location where the event occurred in the subsurface (Chen et al., 2015). Intuitively, if a sensor records a wave at the surface that was reflected by a point (or an interface) in the subsurface, migration tries to identify the subsurface location where the reverberation occurred.

To illustrate this concept, I design a simple experiment where the subsurface is composed of two horizontal homogeneous rock layers (Figure 1.4). I assume that the velocity v_1 in the top layer (drawn in yellow) is known, and I wish to find the depth of the interface between the two layers, $\mathbf{L}_1$. A wave is sent into the ground at $t = t_0$, propagates through the first layer of unknown thickness $\mathbf{L}_1$, is reflected by an interface, and gets recorded at the surface at $t = t_0 + \tau$ by a receiver. The quantity τ corresponds to the total time it took for the wave to travel from the source to the interface, and from the interface to the receiver. If the source-receiver distance $\mathbf{D}_{SR}$ and the propagation speed v_1 in the first layer are known, it is straightforward (using Pythagorean theorem) to infer the position of the interface $\mathbf{L}_1$ that gave rise to the reflection,

$$\mathbf{L}_1 = \sqrt{\tau^2 v_1^2 - \mathbf{D}_{SR}^2}. \tag{1.1}$$

For more complex geological settings, this equation does not hold, but numerical methods can be employed, such as reverse-time migration (RTM) (Baysal et al., 1983).

I re-visit the Marmousi2 example (Figure 1.3) and I generate a RTM image (Figure 1.5a) using 140 shot gathers such as the one shown in Figure 1.2a, and the velocity model obtained with the FWI scheme (Figure 1.3c). Black and white colors indicate the presence of interfaces, while gray colors correspond to homogeneous regions. The resolution of this map is higher than the one of the FWI velocity model and more details about the Earth can be observed.

Finally, a team of experts will interpret the geological features of the subsurface by examining this map, and use this additional information to infer the presence/location of hydrocarbon reservoirs (as shown by the colored annotations in Figure 1.5b). Hence, the accuracy and quality of the final migrated image will have a huge impact on the planning of hydrocarbon reservoir identification, on the drilling and production phases (an offshore well can cost hundreds of millions of dollars), and on the overall business-decision process.

1.1.4 Limitations of conventional techniques

As the industry is exploring regions with increasing geological complexity, velocity estimation becomes a very difficult task, and conventional methods are hampered by three main challenges.

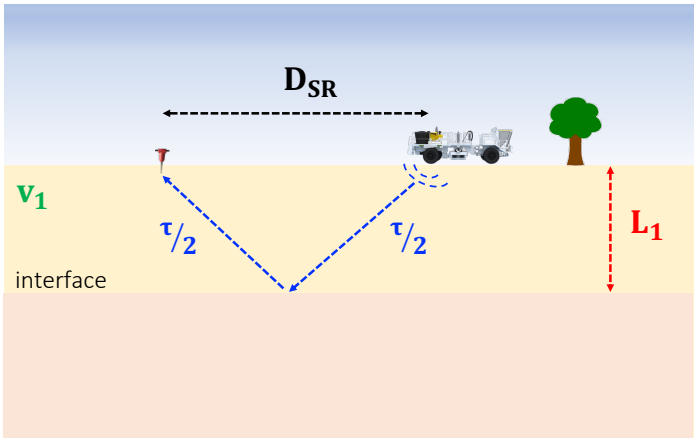
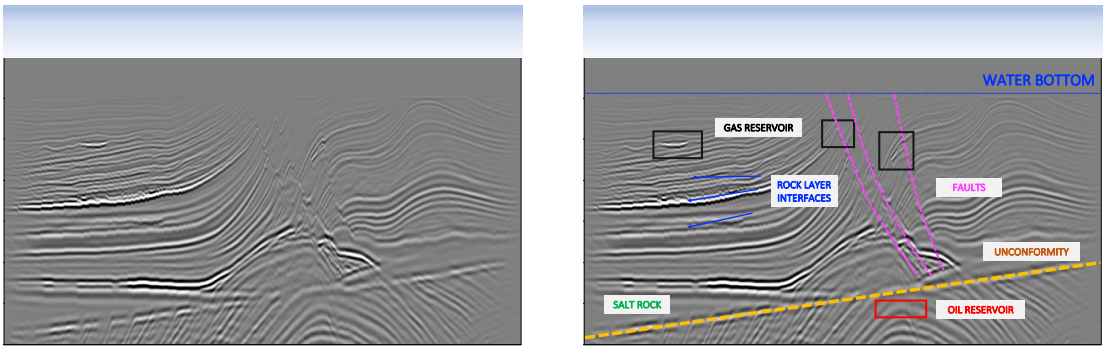


Figure 1.4: Schematic diagram illustrating an onshore acquisition in a simple geological setting. [**NR**]



(a) (b)

Figure 1.5: RTM images computed with the dataset composed of 140 shot gathers such as the one shown in Figure 1.2, and with the FWI velocity model from Figure 1.3c. The vertical and horizontal axes represent depth and horizontal position, respectively. The model is approximately 17 km wide and 3.5 km deep. (a) Raw image. (b) Interpreted image. [**NR**]

The large number of unknown parameters

The area within the subsurface to be characterized needs to be spatially discretized into a 3D array of voxels, and then stored on a computer. The (unknown) value of each voxel represents the speed of sound at a given location, and must be numerically estimated. For 3D field applications, the number of unknown parameters can be extremely large. For instance, a survey area with a horizontal extent of 10 km x 10 km with a maximum depth of 10 km and discretized with a spatial sampling of 10 m (which is quite common for industry standards) requires the estimation (and storage) of as many as 10^9 parameters.

The computational cost

Due to the increasingly-large amount of data to process (in the order of terabytes) and the dimension of the unknown model space, velocity-model building algorithms require the use of iterative numerical methods based on minimizing a loss function with gradient-descent optimization schemes. At each iteration, thousands of seismic wave propagations must be simulated on a computer, usually by numerically solving a partial differential equation (PDE). This requirement, combined with the high-dimension nature of the problem, make these algorithms very computationally intensive, even with the use of general-purpose graphics processing units (GPU). In addition, unlike most supervised deep learning problems, the computational cost of velocity-estimation techniques is so high that a thorough hyper-parameter search is often computationally intractable. For instance, one single forward pass of the modeling scheme could last as long as a couple of days on a modern industry-sized computer cluster. Consequently, there is great demand to create easily-applicable methods for which the number of adjustable hyper-parameters (and the need for human input) is reduced.

The ill-posedness of the inverse problem

The most challenging issue comes from the fact that conventional seismic velocity estimation (using reflection data) is a mathematically and numerically ill-posed problem. The quality of the estimated solution is very sensitive to (1) the accuracy of the initial guess and to (2) the presence of long-offset low-frequency energy recorded in seismic datasets, which may either be too costly or just impossible to acquire. Moreover, seismic surveys might not be able to send and record enough coherent signal from certain regions of the subsurface, which makes the properties at these locations difficult to recover. Therefore, if the initial guess is too far from the true solution, most gradient-descent optimization schemes tend to recover non-geologically realistic (and thus non-useful) models. From an optimization standpoint, this phenomenon can be explained by the presence of multiple spurious local minima within the loss function that is being minimized.

I illustrate this issue by applying FWI on the Marmousi2 example but starting with a less accurate initial model (not shown here) than the one in Figure 1.3a. The recovered model is displayed in

Figure 1.6, and is far from the true Earth (Figure 1.3d). Figure 1.7 shows the RTM image using the same seismic data as in the previous example (Figure 1.2) but with the inaccurate FWI inverted model. As expected, the quality deteriorates and the image becomes difficult to interpret from a geological aspect.

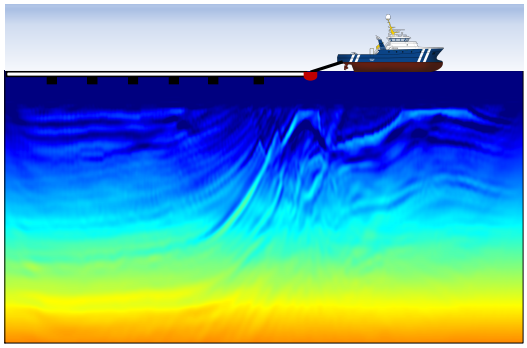


Figure 1.6: 2D panel of the FWI inverted model starting from an inaccurate initial guess (not shown here). The vertical and horizontal axes represent depth and horizontal position, respectively. The model is 17 km wide and 3.5 km deep. [**CR**]

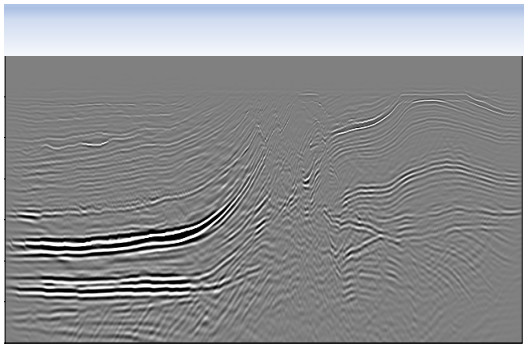


Figure 1.7: RTM image computed with the dataset composed of 140 shot gathers such as the one shown in Figure 1.2, and with the FWI velocity model from Figure 1.6. The vertical and horizontal axes represent depth and horizontal position, respectively. The model is 17 km wide and 3.5 km deep. [**ER**]

1.2 Thesis contributions

The goal of this thesis is to develop a new robust, compact, and computationally-feasible method in order to produce accurate 3D acoustic velocity models of the subsurface when existing methods fail. My proposed solution is referred to as full waveform inversion by model extension (FWIME). The contributions of my work can be summarized and decomposed into four main components.

- **A new theoretical formulation**: There exists two main families of algorithms for velocity-model building, namely tomography and FWI. Wave-equation migration velocity analysis (WEMVA), a specific type of tomography, tends to be more robust and less sensitive to the accuracy of the initial guess than FWI but produces low-resolution results (Symes and Kern, 1994; Stolk and Symes, 2002; Sava and Biondi, 2004). Alternatively, FWI can recover high-resolution features of the Earth but is strongly affected by inaccurate initial guesses. The first novelty of this thesis comes from the design of a hybrid cost function that successfully combines WEMVA and FWI into a new compact efficient formulation. This combination is more powerful than applying each technique individually or consecutively. By using the variable projection method to consistently handle the coupling between the two techniques, I reduce the number of adjustable hyper-parameters and I make my algorithm simple to use, even for non-expert geophysicists. Finally, I devise a novel general optimization framework based on re-parametrizing the velocity model on spline grids. This new framework is crucial for the success of my method.

- **Successful applications to realistic industry-benchmark tests**: I demonstrate the potential of my technique by applying it on four challenging 2D synthetic examples, which are well-known benchmark tests within the seismic imaging research community. These four cases simulate some of the most problematic geological scenarios encountered in hydrocarbon exploration. In each case, conventional methods fail whereas FWIME recovers an excellent solution.

- **An efficient industry-quality 3D GPU implementation**: My new loss function design increases the computational cost by at least one order of magnitude compared to conventional methods. To make FWIME feasible for 3D field-data applications, I develop a fully-fledged 3D GPU solution in Python, C++, and CUDA that can be seamlessly deployed on high-performance computing (HPC) platforms.

- **A successful 3D field-data application**: I successfully apply my method to a deep-water node-acquisition 3D field survey acquired in the Gulf of Mexico, and I show that FWIME recovers a much more accurate solution than the one obtained with conventional FWI.

1.3 Thesis overview

- **Chapter 2: Seismic modeling operators**. I present formal mathematical definitions and derivations of the four main acoustic seismic modeling operators employed in FWIME. I explain the concept of extended modeling (and extended operators) which is the most important tool that makes my method successful. For each operator, I provide its physical interpretation, and I explain how it can be efficiently implemented on a computer.

- **Chapter 3: Theory and design**: I define the FWIME objective function, and I show that it is the sum of two terms: (1) a data-fitting component (modified from a conventional FWI problem), and (2) an annihilating component that possesses similar features as a WEMVA objective function. The coupling between the two components is handled by the use of the variable projection method (Golub and Pereyra, 1973). I thoroughly analyze each term and I compute them on simple numerical examples.

- **Chapter 4: Optimization: A model-space multi-scale approach**. I describe the optimization process of FWIME. I derive and discuss the main properties of the FWIME gradient. Then, I devise a model-space multi-scale optimization workflow that improves the convergence properties of my method by creating more convex descent paths towards the optimal solution. I apply FWIME on three 2D synthetic tests where conventional FWI converges to non-physical solutions. Each test is designed to generate simple datasets containing only one wave mode (transmission, reflection or refraction) in order to demonstrate that FWIME can automatically invert any type of seismic data using the same framework and without the need for intensive hyper-parameter tuning.

- **Chapter 5: Application of FWIME to realistic benchmark tests**. I apply FWIME on four realistic 2D synthetic examples to show its ability at automatically and successfully inverting datasets composed of all types of waves. In each scenario, the inverted dataset lacks low-frequency energy and the initial model is inaccurate. Such conditions are quite common in field-data applications, and lead standard methods to fail at recovering useful Earth models. I give practical details on how to successfully use my method and select optimal hyper-parameter values. FWIME's excellent performance in the four tests suggests that the method has the potential to significantly disrupt the conventional imaging workflow of the oil and gas industry.

- **Chapter 6: 3D field-data application**. I apply FWIME to a 3D ocean bottom node (OBN) field dataset provided by Shell Exploration Inc. (Shell), acquired in the Gulf of Mexico. My analysis is composed of two steps. First, I use an accurate initial model provided by Shell which required intensive pre-processing work and human input, and I apply a conventional 3D FWI scheme. The inverted model is excellent and geologically consistent. In the second step, I purposely design an inaccurate model (which would typically demand less preliminary processing). In this case, conventional FWI is unable to recover a useful solution. Then, I apply FWIME and I recover a model whose quality is similar to the one obtained with FWI from the first step (using the accurate initial model).

Chapter 2

Seismic modeling operators

In this chapter, I provide a theoretical description of the four fundamental seismic wave-equation modeling operators used in FWIME, and more generally in seismic imaging and velocity model-building algorithms. I refer to them as forward modeling, Born modeling, extended Born modeling, and tomographic modeling. For each operator, I derive its analytical expression, I provide a physical interpretation and connection with its use in seismic exploration, and I describe its numerical implementation. Throughout this thesis, I will consider the Earth 's subsurface as an acoustic, isotropic and constant-density medium. Even though inaccurate, this approximation is commonly used in acoustic velocity-model building algorithms in order to reduce their computational costs inherent to the large amount of seismic data to process. Furthermore, I will use the following conventions,

- $\Omega \subset \mathbb{R}^3$ is the spatial-integration domain (i.e., the volume of interest I wish to characterize within the subsurface),
- $\mathcal{T} \subset \mathbb{R}^+$ is the time-integration domain (i.e., the time interval during which the pressure field is recorded),
- $\mathcal{F}(\mathcal{S}_1, \mathcal{S}_2)$ is the set of squared-integrable functions (over $\mathcal{S}_1$) mapping elements of the set S_1 to elements of the set S_2.
- $\mathcal{M}^{N_r \times N_c}$ is the set of real-coefficient matrices with N_r rows and N_c columns,
- $\mathcal{M}^N$ is the set of real-coefficient square matrices with N rows and N columns.

Moreover, all vectors and operators are displayed in bold font. Finally, for some of the derivations shown in this chapter, I will purposely omit to specify the integration bounds (when unambiguous) in order to simplify notations.

2.1 Wave-equation modeling

I begin by writing the partial differential equation (PDE) satisfied by a pressure field in an acoustic isotropic and constant-density medium. Then, I define the acoustic wave-equation operator (referred to as the forward modeling operator) and show that it can not only be interpreted as a nonlinear operator with respect to the unknown velocity map (the set of parameters I ultimately want to recover in my thesis), but also as a linear mapping with respect to the seismic source term. When I consider this linear case, I derive the analytical expression of its adjoint operator. Finally, I provide a description of its numerical implementation.

2.1.1 The acoustic wave equation

The PDE satisfied by a pressure field u in an acoustic isotropic constant-density medium at position $\mathbf{x}$ and time t is given by

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla^2 u(\mathbf{x},t) = s(\mathbf{x},t), \tag{2.1}$$

where $u \in \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R})$ is the pressure field, $s \in \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R})$ is the source field, $m \in \mathcal{F}(\Omega, \mathbb{R}^+)$ is the velocity map (assumed to be invariant with time), ∇^2 is the Laplacian operator, and the initial time-boundary conditions are given by

$$u(\mathbf{x}, t=0) = \frac{\partial u}{\partial t}(\mathbf{x}, t=0) = 0. \tag{2.2}$$

Using the representation theorem, the solution of equation 2.1 can be expressed by

$$u(\mathbf{x},t) = \int\limits_{\mathbf{x}' \in \Omega} g_m(\mathbf{x}, t, \mathbf{x}', 0) * s(\mathbf{x}', t)\, d\mathbf{x}', \tag{2.3}$$

where $*$ indicates a convolution in time, and g_m is the acoustic isotropic Green's function generated from a point source $\mathbf{x}'$ at $t = 0$, propagated in a medium characterized by the velocity map m, and recorded at position $\mathbf{x}$ and time t (Aki and Richards, 2002). In seismic exploration, s is commonly assumed to be located at a known point source in space $\mathbf{x}_s$. In such case, $s(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_s)\, w(t)$, where w is the source's time signature and δ is the spatial Dirac delta function. In this case, equation 2.3 can be further simplified to

$$u(\mathbf{x},t) = g_m(\mathbf{x}, t, \mathbf{x}_s, 0) * w(t). \tag{2.4}$$

In practice, the pressure field cannot be recorded everywhere in the subsurface. It is sampled by recording devices (also called "receivers") such as geophones or hydrophones that are usually located at the surface of the Earth, at the bottom of the ocean, or even inside a borehole. The set of recordings for all sources and receivers is referred to as the seismic data d. For a given source s, the subset of d corresponding to the recording of a single receiver at position $\mathbf{x}_r$ is referred to as a seismic trace $d(t; s, \mathbf{x}_r)$ and can be extracted from the pressure field as follows:

$$\begin{aligned} d(t; s, \mathbf{x}_r) &= \int_{\mathbf{x}\in\Omega} \delta(\mathbf{x} - \mathbf{x}_r)\, u(\mathbf{x}, t)\, d\mathbf{x} && (2.5)\\ &= u(\mathbf{x}_r, t) && (2.6)\\ &= \int_{\mathbf{x}'\in\Omega} g_m(\mathbf{x}_r, t, \mathbf{x}', 0) * s(\mathbf{x}', t)\, d\mathbf{x}', && (2.7) \end{aligned}$$

Table 2.1 summarizes the units and dependencies of the wave-equation variables introduced in this section.

Variable	Name	Dependency	Units
u	Pressure field	Space and time	Pa
d	Seismic data	Space and time	Pa
m	Acoustic velocity	Space	km/s
s	Source field	Space and time	Pa/km^2

Table 2.1: Summary of the units and dependencies for the wave-equation variables.

2.1.2 Forward modeling operator

Given a source field s and a receiver r, equation 2.7 represents a nonlinear mapping $\mathbf{f}$ between a velocity map m and a seismic trace d,

$$\begin{aligned} \mathbf{f} : \mathcal{F}(\Omega, \mathbb{R}^+) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) && (2.8)\\ m &\mapsto d = \mathbf{f}(m). \end{aligned}$$

Alternatively, for a fixed velocity map m, equation 2.7 represents a linear mapping $\mathbf{F}_m$ between a source field s and a seismic trace d recorded at receiver r,

$$\begin{aligned} \mathbf{F}_m : \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ s &\mapsto d = \mathbf{F}_m s, \end{aligned} \tag{2.9}$$

where

$$\boxed{(\mathbf{F}_m s)(t; \mathbf{x}_r) = \int\limits_{\mathbf{x}' \in \Omega} g_m(\mathbf{x}_r, t, \mathbf{x}', 0) * s(\mathbf{x}', t)\, d\mathbf{x}'.} \tag{2.10}$$

Additionally, I can express $\mathbf{F}_m$ as the composition of two linear operators,

$$\mathbf{F}_m = \mathbf{E}\mathbf{P}_m, \tag{2.11}$$

where $\mathbf{P}_m$ is defined by

$$\begin{aligned} \mathbf{P}_m : \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) &\mapsto \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) \\ s &\mapsto u = \mathbf{P}_m s. \end{aligned} \tag{2.12}$$

$\mathbf{P}_m$ corresponds to the propagation operator of the source field s into a pressure field u, and its analytical expression is given by equation 2.3. $\mathbf{E}$ is a spatial extraction operator that samples the pressure field u at the receiver position $\mathbf{x}_r$. It is defined by

$$\begin{aligned} \mathbf{E} : \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ u &\mapsto d = \mathbf{E}u, \end{aligned} \tag{2.13}$$

and its analytical expression is given by equation 2.5.

2.1.3 Adjoint modeling operator

As I will describe in the next chapter of my thesis, the optimization scheme I use requires the computation of the adjoint operator of $\mathbf{F}_m$ (referred to as $\mathbf{F}_m^*$), which is defined by

$$\begin{aligned} \mathbf{F}_m^* : \mathcal{F}(\mathcal{T}, \mathbb{R}) &\mapsto \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) \\ d &\mapsto s = \mathbf{F}_m^* d, \end{aligned} \tag{2.14}$$

and satisfies

$$\mathbf{F}_m^* = \mathbf{P}_m^* \, \mathbf{E}^*. \tag{2.15}$$

To derive $\mathbf{P}_m^*$, I first define an inner product on $\mathcal{U} = \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R})$,

$$< u_1, u_2 >_{\mathcal{U}} = \int\limits_{\mathbf{x} \in \Omega} \int\limits_{t \in \mathcal{T}} u_1(\mathbf{x}, t) \; u_2(\mathbf{x}, t) \; dt \; d\mathbf{x}. \tag{2.16}$$

The adjoint operator $\mathbf{P}_m^*$ is the unique operator defined by

$$\begin{aligned} \mathbf{P}_m^* : \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) &\mapsto \mathcal{F}(\Omega \times \mathcal{T}, \mathbb{R}) \\ u &\mapsto s = \mathbf{P}_m^* \, u \end{aligned} \tag{2.17}$$

that satisfies

$$< \mathbf{P}_m s, u >_{\mathcal{U}} = < s, \mathbf{P}_m^* \, u >_{\mathcal{U}} . \tag{2.18}$$

for all s and u in $\mathcal{U}$. To obtain the analytical expression of $\mathbf{P}_m^*$, I follow a similar derivation as the one described in Tarantola (1984). The left side of equation 2.18 is given by

$$< \mathbf{P}_m s, u >_{\mathcal{U}} = \int\limits_{\mathbf{x}} \int\limits_{t} (\mathbf{P}_m s)(\mathbf{x}, t) \; u(\mathbf{x}, t) \; dt \; d\mathbf{x}, \tag{2.19}$$

$$= \int\limits_{\mathbf{x}} \int\limits_{\mathbf{x}'} \int\limits_{t} g_m(\mathbf{x}, t, \mathbf{x}', 0) * s(\mathbf{x}', t) \; u(\mathbf{x}, t) \; dt \; d\mathbf{x}' \; d\mathbf{x}. \tag{2.20}$$

Furthermore,

$$\int_{t\in\mathcal{T}} g_m(\mathbf{x},t,\mathbf{x}',0) * s(\mathbf{x}',t)\ u(\mathbf{x},t)\ dt = \int_{t\in\mathcal{T}} s(\mathbf{x}',t)\ g_m(\mathbf{x},-t,\mathbf{x}',0) * u(\mathbf{x},t)\ dt, \tag{2.21}$$

where $g_m(\mathbf{x},-t,\mathbf{x}',0)$ is the anti-causal Green's function. Hence, the term $g_m(\mathbf{x},-t,\mathbf{x}',0)*u(\mathbf{x},t)$ corresponds to the propagation of field u backward in time (Tarantola, 1984). I can now re-write equation 2.20 as

$$< \mathbf{P}_m s, u >_{\mathcal{U}} = \int_{\mathbf{x}}\int_{\mathbf{x}'}\int_{t} s(\mathbf{x}',t)\ g_m(\mathbf{x},-t,\mathbf{x}',0) * u(\mathbf{x},t)\ dt\ d\mathbf{x}'d\mathbf{x} \tag{2.22}$$

$$= \int_{\mathbf{x}'}\int_{t} s(\mathbf{x}',t) \int_{\mathbf{x}} g_m(\mathbf{x}',0,\mathbf{x},t) * u(\mathbf{x},t)\ d\mathbf{x}\ dt\ d\mathbf{x}'. \tag{2.23}$$

Equation 2.23 is obtained using both the time-invariance and reciprocity properties of the Green's function g_m for an acoustic isotropic medium (Aki and Richards, 2002). That is, $g_m(\mathbf{x},-t,\mathbf{x}',0) = g_m(\mathbf{x},0,\mathbf{x}',t) = g_m(\mathbf{x}',0,\mathbf{x},t)$. The right side of equation 2.18 can be written as follows:

$$< s, \mathbf{P}_m^*\ u >_{\mathcal{U}} = \int_{\mathbf{x}\in\Omega}\int_{t\in\mathcal{T}} s(\mathbf{x},t)\ (\mathbf{P}_m^*\ u)(\mathbf{x},t)\ dt\ d\mathbf{x}. \tag{2.24}$$

Using equations 2.18, 2.23, and 2.24 (and the uniqueness property of the adjoint operator), I can now obtain the analytical expression of the application of $\mathbf{P}_m^*$ on an input field u,

$$(\mathbf{P}_m^*\ u)(\mathbf{x},t) = \int_{\mathbf{x}'\in\Omega} g_m(\mathbf{x},0,\mathbf{x}',t) * u(\mathbf{x}',t)\ d\mathbf{x}'. \tag{2.25}$$

Moreover, it can easily be shown that the application of $\mathbf{E}^*$ on a seismic trace d recorded at receiver r is given by

$$(\mathbf{E}^* d)(\mathbf{x},t) = \delta(\mathbf{x}-\mathbf{x}_r)\ d(t;\mathbf{x}_r). \tag{2.26}$$

Thus, I obtain the analytical expression of $\mathbf{F}_m^*$:

$$(\mathbf{F}_m^*\ d)(\mathbf{x},t) = g_m(\mathbf{x},0,\mathbf{x}_r,t) * d(t;\mathbf{x}_r). \tag{2.27}$$

If multiple recorded traces are employed in the process, equation 2.27 becomes

$$\boxed{(\mathbf{F}_m^*\, d)(\mathbf{x},t) \;=\; \sum_{j=1}^{N_r} g_m(\mathbf{x},0,\mathbf{x}_{r^{(j)}},t) * d(t;\mathbf{x}_{r^{(j)}}),} \tag{2.28}$$

where N_r is the number of receivers and $r^{(j)}$ refers to the j^{th} receiver. Equations 2.27 and 2.28 correspond to the adjoint process of the one described in equation 2.10. As explained in Tarantola (1984), it can be interpreted as the propagation backward in time of the seismic trace(s) d injected at position $\mathbf{x}_{r^{(j)}}$ into the medium characterized by the velocity map m.

2.1.4 Numerical implementation

I explain how I numerically simulate the propagation and recording of pressure fields that satisfy PDE 2.1 for a single source/receiver pair (the extension to a collection of source/receiver pairs is straightforward). I discretize the PDE in time and space and I solve it using an explicit finite-difference (FD) approach. I compute the discretized forward/adjoint modeling operators defined in equations 2.10 and 2.27, and I show that their corresponding matrices are large but very sparse. Finally, I provide an efficient implementation in which the operators' coefficients are not stored in memory and only their application onto vectors is computed.

Discretization

After discretization, the recorded trace $\mathbf{d}$ is an element of $\mathbb{R}^{N_t}$, the source field $\mathbf{s}$ and pressure field $\mathbf{u}$ are both elements of $\in \mathbb{R}^{N_w}$, and the velocity map $\mathbf{m} \in \mathbb{R}^{N_m}$. Here, $N_w = N_m \times N_t$ is the size of the wavefield, N_m is the model size with $N_m = N_x \times N_y \times N_z$, where N_j is the number of spatial samples in the j^{th}-direction. N_t is the number of time samples for the simulation. The discretized forward modeling operator $\mathbf{F_m}$ (equation 2.9) and its adjoint $\mathbf{F_m^*}$ (equation 2.14) are now defined as

$$\begin{aligned} \mathbf{F_m} : \mathbb{R}^{N_w} &\mapsto \mathbb{R}^{N_t} \\ \mathbf{s} &\mapsto \mathbf{d} = \mathbf{F_m}\mathbf{s}, \end{aligned} \tag{2.29}$$

and

$$\begin{aligned} \mathbf{F_m^*} : \mathbb{R}^{N_t} &\mapsto \mathbb{R}^{N_w} \\ \mathbf{d} &\mapsto \mathbf{s} = \mathbf{F_m^*}\mathbf{d}. \end{aligned} \tag{2.30}$$

Using a central-difference scheme, equation 2.1 can now be written as follows,

$$\left[\mathbf{M}^{-2}\mathbf{D}_t^2 - \boldsymbol{\nabla}^2\right]\mathbf{u}[i] \quad = \quad \mathbf{s}[i], \tag{2.31}$$

where $\mathbf{u}[i]$ and $\mathbf{s}[i]$ are the time slices of the pressure and source fields at time index i. $\mathbf{D}_t^2 \in \mathcal{M}^{N_n}$ is a second-order central-difference time-derivative operator, $\boldsymbol{\nabla}^2 \in \mathcal{M}^{N_m}$ is a high-order Laplacian operator, and $\mathbf{M} \in \mathcal{M}^{N_m}$ is a diagonal operator that scales its input by the velocity vector $\mathbf{m}$. The initial time-boundary conditions are given by $\mathbf{u}[1] = \mathbf{u}[2] = 0$.

Forward propagation

The discretized forward modeling operator can be expressed as $\mathbf{F_m} = \mathbf{EP_m}$ (equation 2.10). Therefore, the main challenge is to efficiently implement the application of $\mathbf{P_m}$ on an input source $\mathbf{s}$ (the numerical implementation of the extraction operator $\mathbf{E}$ is straightforward). Due to the large problem sizes encountered in field applications, it is usually not possible to store the sparse matrix $\mathbf{P_m}$ on computers' random access memory (RAM). Hence, I compute the values of $\mathbf{u}$ iteratively with a forward-substitution scheme using the following recursive equation,

$$\mathbf{u}[i] \quad = \quad \mathbf{A}\ \mathbf{u}[i-1] + \mathbf{B}\ \mathbf{u}[i-2] + \mathbf{C}\ \mathbf{s}[i-1]. \tag{2.32}$$

Operators $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{I}_d$ are all elements of $\mathcal{M}^{N_m}$, defined by

$$\mathbf{A} \quad = \quad \mathbf{m}^2\boldsymbol{\Delta}\mathbf{t}^2\boldsymbol{\nabla}^2 + 2\mathbf{I}_d, \tag{2.33}$$

$$\mathbf{B} \quad = \quad -\mathbf{I}_d, \tag{2.34}$$

$$\mathbf{C} \quad = \quad \mathbf{m}^2\boldsymbol{\Delta}\mathbf{t}^2\ \mathbf{I}_d. \tag{2.35}$$

$\mathbf{I}_d$ is the identity operator, and $\mathbf{m}^2\boldsymbol{\Delta}\mathbf{t}^2$ is a diagonal operator that scales its input by $m^2\Delta t^2$, where Δt is the time-sampling rate for used for the FD scheme. Additionally, by writing equation 2.32 for all time steps, I obtain the following linear system,

$$\mathbf{Q}\ \mathbf{u} \quad = \quad \mathbf{D}_c\ \mathbf{s}, \tag{2.36}$$

where $\mathbf{Q} \in \mathcal{M}^{N_w}$ is an invertible lower-triangular (sparse) matrix whose entries are given by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & -\mathbf{A} & \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{B} & -\mathbf{A} & \mathbf{I}_d & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{B} & -\mathbf{A} & \mathbf{I}_d & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{B} & -\mathbf{A} & \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{B} & -\mathbf{A} & \mathbf{I}_d \end{bmatrix}, \tag{2.37}$$

and $\mathbf{D}_c \in \mathcal{M}^{N_w}$ is a block-diagonal invertible matrix,

$$\mathbf{D}_c = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{C} \end{bmatrix}. \tag{2.38}$$

From equation 2.36, I can explicitly express $\mathbf{u}$ as

$$\mathbf{u} = \mathbf{Q}^{-1}\mathbf{D}_c\,\mathbf{s}. \tag{2.39}$$

Hence,

$$\mathbf{P_m} \;=\; \mathbf{Q}^{-1}\mathbf{D}_c, \tag{2.40}$$

and

$$\mathbf{F_m} \;=\; \mathbf{E}\mathbf{Q}^{-1}\mathbf{D}_c, \tag{2.41}$$

where $\mathbf{Q}^{-1}$ is a lower triangular matrix (it is the inverse of a lower-triangular matrix). Therefore,

the application of the discretized forward modeling operator on an input source field defined by

$$\mathbf{d} \quad = \quad \mathbf{F_m}\, \mathbf{s} \tag{2.42}$$

can be efficiently implemented by applying the following chain of linear operators,

$$\mathbf{d} \quad = \quad \mathbf{E}\mathbf{Q}^{-1}\mathbf{D}_c\, \mathbf{s}. \tag{2.43}$$

By using the fact that the large matrices involved in equation 2.43 are sparse, this numerical implementation avoids the need to store them entirely on RAM.

Adjoint propagation

The adjoint process of $\mathbf{F_m}$ is given by

$$\mathbf{F}^*_{\mathbf{m}} = \mathbf{D}^*_c(\mathbf{Q}^{-1})^*\, \mathbf{E}^*, \tag{2.44}$$

where $(\mathbf{Q}^{-1})^*$ is an upper-triangular matrix. The main difficulty resides in finding the recursive relation to compute the application of $(\mathbf{Q}^{-1})^*$ on a input field $\mathbf{u}_{in}$ giving rise to an output field $\mathbf{u}_{out}$,

$$\begin{aligned} \mathbf{u}_{out} &= (\mathbf{Q}^{-1})^*\mathbf{u}_{in} && (2.45)\\ &= (\mathbf{Q}^*)^{-1}\mathbf{u}_{in} && (2.46)\\ \mathbf{Q}^*\mathbf{u}_{out} &= \mathbf{u}_{in}, && (2.47) \end{aligned}$$

where the upper-triangular matrix $\mathbf{Q}^*$ is given by

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{I}_d & -\mathbf{A}^* & -\mathbf{B}^* & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d & -\mathbf{A}^* & -\mathbf{B}^* & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_d & -\mathbf{A}^* & -\mathbf{B}^* & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_d & -\mathbf{A}^* & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_d & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_d & -\mathbf{A}^* \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_d \end{bmatrix}. \tag{2.48}$$

I can now write the adjoint recursive relation,

$$\mathbf{u}_{out}[i] \quad = \quad \mathbf{A}^*\mathbf{u}_{out}[i+1] + \mathbf{B}^*\mathbf{u}_{out}[i+2] + \mathbf{u}_{in}[i+1], \tag{2.49}$$

which defines the backward-substitution scheme used for the numerical implementation of the adjoint modeling operator. In a similar fashion as the forward propagation, none of these matrices need to be entirely stored on RAM, making this implementation efficient for large-scale problems.

2.2 Linearization of the wave-equation

In this section, I analyze and characterize the linear variations of the pressure field u (and the recorded data d) when the velocity map m is perturbed, which corresponds to finding an expression of the first-order differential of $\mathbf{f}$ (defined in equation 2.8) with respect to m. I derive the analytical expression of the Fréchet derivative operator of $\mathbf{f}$ (also known as the Jacobian matrix in the finite-dimension case) as well as its adjoint operator. Then, I provide a physical interpretation on the mechanism of these operators using some schematic illustrations. Finally, I describe how I implement them numerically. Throughout this section, my derivations are performed assuming one source/receiver pair, and then generalized for multiple source/receiver pairs. Hence, I do not always explicitly write the dependency of each operator on s and r.

2.2.1 Born modeling operator

In seismic imaging, the Fréchet derivative operator of $\mathbf{f}$ is written $\mathbf{B}_m = \frac{\partial \mathbf{f}}{\partial m}\Big|_m$ and referred to as the Born modeling operator evaluated at m, also known as the demigration operator. It is the linear

mapping of a velocity perturbation p to a recorded seismic trace d extracted at receiver r from a pressure field q, given a seismic source field s and a velocity map m:

$$\begin{aligned} \mathbf{B}_m : \mathcal{F}(\Omega, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ p &\mapsto d = \mathbf{B}_m\, p. \end{aligned} \tag{2.50}$$

I obtain the expression of the Born modeling operator by perturbing the velocity map in equation 2.1. For notation compactness, I write $u_m(\mathbf{x}, t) = u(\mathbf{x}, t; m)$. Recall that u_m satisfies

$$\frac{1}{m^2(\mathbf{x})} \frac{\partial^2 u_m(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u_m(\mathbf{x}, t) = s(\mathbf{x}, t). \tag{2.51}$$

By perturbing the velocity map by p, the solution $u_{m+p} = u_m + q$ satisfies

$$\frac{1}{(m(\mathbf{x}) + p(\mathbf{x}))^2} \frac{\partial^2 u_m(\mathbf{x}, t) + q(\mathbf{x}, t)}{\partial t^2} - \nabla^2 \big(u_m(\mathbf{x}, t) + q(\mathbf{x}, t)\big) = s(\mathbf{x}, t). \tag{2.52}$$

By combining equations 2.51 and 2.52,

$$\frac{1}{m^2(\mathbf{x})} \frac{\partial^2 q(\mathbf{x}, t)}{\partial t^2} - \nabla^2 q(\mathbf{x}, t) = s_{sec}(\mathbf{x}, t) + \epsilon(\mathbf{x}, t). \tag{2.53}$$

Here, s_{sec} is referred to as the secondary source, expressed by

$$s_{sec}(\mathbf{x}, t) \;=\; 2 \frac{p(\mathbf{x})}{m^3(\mathbf{x})}\, \ddot{u}_m(\mathbf{x}, t), \tag{2.54}$$

where $\ddot{}$ symbolizes the second-order time-derivative operator, $\ddot{u}_m$ is sometimes referred to as the source wavefield (written u_{sou}), and the function ϵ is composed of the higher-order terms in p. The wavefield perturbation q satisfies a similar PDE as the one in equation 2.1 but with a modified source field $s_{sec} + \epsilon$. Hence, I can write

$$q(\mathbf{x}, t) \;=\; \int\limits_{\mathbf{x}' \in \Omega} g_m(\mathbf{x}, t, \mathbf{x}', 0) * \big(s_{sec}(\mathbf{x}', t) + \epsilon(\mathbf{x}', t)\big)\, d\mathbf{x}' \tag{2.55}$$

$$\;=\; \int\limits_{\mathbf{x}' \in \Omega} g_m(\mathbf{x}, t, \mathbf{x}', 0) * s_{sec}(\mathbf{x}', t)\, d\mathbf{x}' + r_\epsilon(\mathbf{x}, t), \tag{2.56}$$

with

$$r_\epsilon(\mathbf{x},t) \quad = \quad \int\limits_{\mathbf{x}'\in\Omega} g_m(\mathbf{x},t,\mathbf{x}',0) * \epsilon(\mathbf{x}',t)\; d\mathbf{x}'. \tag{2.57}$$

By defining an inner product on $\mathcal{P} = \mathcal{F}(\Omega,\mathbb{R})$ as

$$< p_1, p_2 >_{\mathcal{P}} \quad = \quad \int\limits_{\mathbf{x}\in\Omega} p_1(\mathbf{x})\; p_2(\mathbf{x})\; d\mathbf{x}, \tag{2.58}$$

and using the norm $||\cdot||_{\mathcal{P}}$ associated with this inner product, it is straightforward to show that $r_\epsilon = o\big(||p||_{\mathcal{P}}\big)$. Therefore, equation 2.56 can be re-written as follows:

$$u_{m+p}(\mathbf{x},t) \quad = \quad u_m(\mathbf{x},t) + \int\limits_{\mathbf{x}'\in\Omega} g_m(\mathbf{x},t,\mathbf{x}',0) * s_{sec}(\mathbf{x}',t)\; d\mathbf{x}' + o\big(||p||_{\mathcal{P}}\big). \tag{2.59}$$

I now extract the wavefields at the receiver position $\mathbf{x}_r$ by applying the extraction operator $\mathbf{E}$ to both sides of equation 2.59, and I obtain

$$\mathbf{f}(m+p) \quad = \quad \mathbf{f}(m) + \mathbf{L}p + o\big(||p||_{\mathcal{P}}\big), \tag{2.60}$$

where $\mathbf{L}$ is a linear operator with respect to p. From equations 2.59 and 2.60 (and using the uniqueness property of the Fréchet derivative operator), I can deduce that $\mathbf{B}_m = \mathbf{L}$ with

$$\big(\mathbf{B}_m p\big)(t) \quad = \quad \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_r,t,\mathbf{x},0) * s_{sec}(\mathbf{x},t)\; d\mathbf{x} \tag{2.61}$$

$$\quad = \quad \int\limits_{\mathbf{x}\in\Omega} \frac{2p(\mathbf{x})}{m^3(\mathbf{x})}\; g_m(\mathbf{x}_r,t,\mathbf{x},0) * \ddot{u}_m(\mathbf{x},t)\; d\mathbf{x}. \tag{2.62}$$

For seismic imaging problems, the process described in equation 2.62 is usually conducted multiple times using various source fields $s^{(i)}$ generated by point sources with different spatial positions and time signatures. For each source, a set of receivers $r^{(j)}$ are also employed (for simplicity, I assume that the set of receivers is the same across all sources). Therefore, for a source field $s^{(i)}$ and receiver positioned at $\mathbf{x}_{r^{(j)}}$, equation 2.62 can be more formally written as

$$\boxed{\big(\mathbf{B}_m p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) \quad = \quad \int\limits_{\mathbf{x}\in\Omega} \frac{2p(\mathbf{x})}{m^3(\mathbf{x})} g_m(\mathbf{x}_{r^{(j)}}, t, \mathbf{x}, 0) * \ddot{u}_m(\mathbf{x}, t; s^{(i)})\, d\mathbf{x},} \tag{2.63}$$

where $\ddot{u}_m(\mathbf{x}, t; s^{(i)})$ is the source wavefield generated by the source $s^{(i)}$. As previously mentioned, the process described in equation 2.63 is referred to as a demigration and can be decomposed into four steps: (1) the source wavefield is propagated into the medium with velocity m. (2) The source wavefield interacts with the model perturbation p, thereby creating a secondary source. (3) A scattered wavefield is generated by the secondary source and propagated into the medium with velocity m. (4) The data perturbation (also referred to as Born modeled data) is recorded at the receivers positions.

2.2.2 Adjoint Born operator

In seismic imaging, the adjoint Born operator $\mathbf{B}_m^*$ is commonly referred to as either the migration operator, or the reverse-time migration (RTM) operator (Baysal et al., 1983). To obtain its analytical expression, I first define an inner product on $\mathcal{D} = \mathcal{F}(\mathcal{T}, \mathbb{R})$,

$$< d_1, d_2 >_{\mathcal{D}} \quad = \quad \int\limits_{\mathbf{t}\in\mathcal{T}} d_1(t)\; d_2(t)\; dt. \tag{2.64}$$

$\mathbf{B}_m^*$ is the unique operator that satisfies

$$< \mathbf{B}_m p, d >_{\mathcal{D}} \quad = \quad < p, \mathbf{B}_m^* d >_{\mathcal{P}}, \tag{2.65}$$

for all $p \in \mathcal{P}$ and $d \in \mathcal{D}$.

$$\begin{aligned}
< \mathbf{B}_m p, q >_{\mathcal{D}} \quad &= \quad \int\limits_{t\in\mathcal{T}} d(t)\; \big(\mathbf{B}_m p\big)(t)\; dt && (2.66)\\
&= \quad \int\limits_{t\in\mathcal{T}} d(t) \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * s_{sec}(\mathbf{x}, t)\; d\mathbf{x}\; dt && (2.67)\\
&= \quad \int\limits_{\mathbf{x}\in\Omega} p(\mathbf{x}) \int\limits_{t\in\mathcal{T}} d(t)\; g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * \frac{2\ddot{u}_m(\mathbf{x}, t)}{m^3(\mathbf{x})}\; dt\; d\mathbf{x} && (2.68)\\
&= \quad \int\limits_{\mathbf{x}\in\Omega} p(\mathbf{x}) \int\limits_{t\in\mathcal{T}} \frac{2\ddot{u}_m(\mathbf{x}, t)}{m^3(\mathbf{x})}\; g_m(\mathbf{x}, 0, \mathbf{x}_r, t) * d(t)\; dt\; d\mathbf{x}. && (2.69)
\end{aligned}$$

Equation 2.69 is obtained by applying the same property shown in equation 2.21. Furthermore,

$$< p, \mathbf{B}_m^* \, d >_{\mathcal{P}} \quad = \quad \int\limits_{\mathbf{x}\in\Omega} p(\mathbf{x}) \big(\mathbf{B}_m^* d\big)(\mathbf{x}) \, d\mathbf{x}. \tag{2.70}$$

By uniqueness of the adjoint operator, I obtain the expression of the application of $\mathbf{B}_m^*$ on a seismic trace d,

$$(\mathbf{B}_m^* \, d)(\mathbf{x}) \quad = \quad \int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x},t) \; g_m(\mathbf{x},0,\mathbf{x}_r,t) * d(t) \; dt \tag{2.71}$$

$$= \quad \int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x},t) \; u_{rec}(\mathbf{x},t) \; dt. \tag{2.72}$$

Equation 2.72 indicates that the application of $\mathbf{B}_m^*$ on a seismic trace is the zero time-lag cross-correlation between the (scaled) source wavefield and u_{rec}, where u_{rec} is the wavefield generated by the seismic trace d, injected at $\mathbf{x}_r$, back-propagated in time. It is commonly referred to as the receiver wavefield, and is given by

$$u_{rec}(\mathbf{x},t) \quad = \quad g_m(\mathbf{x},0,\mathbf{x}_r,t) * d(t). \tag{2.73}$$

Furthermore, for an experiment conducted multiple times with various sources, it can easily be shown that

$$\boxed{(\mathbf{B}_m^* p)(\mathbf{x}) \quad = \quad \sum_{i=1}^{N_s} \int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x},t;s^{(i)}) \; u_{rec}(\mathbf{x},t;s^{(i)}) \; dt} \tag{2.74}$$

where N_s is the total number of sources employed in the experiment, and $u_{rec}(\mathbf{x},t;s^{(i)})$ is the receiver wavefield generated by the collection of N_r recorded traces (from the i^{th} source) propagated backward in time,

$$u_{rec}(\mathbf{x},t;s^{(i)}) \quad = \quad \sum_{j=1}^{N_r} g_m(\mathbf{x},0,\mathbf{x}_{r^{(j)}},t) * d(t;s^{(i)},\mathbf{x}_{r^{(j)}}). \tag{2.75}$$

Equation 2.74 corresponds to the migration or imaging process and its output is usually referred to as an image. Its mechanism can be decomposed as a sequence of four steps: (1) the source

wavefield is propagated forward in time into the medium with velocity m. (2) The input data are injected at the receivers positions and propagated backward in time to generate the receiver wavefield. (3) For each point in the subsurface, the image is obtained by computing the zero time-lag cross-correlation between the (scaled) source and the receiver wavefields. (4) The final image is obtained by summing the contribution from each source $s^{(i)}$.

2.2.3 Geophysical interpretation

I provide physical insight on the Born forward and adjoint linear modeling operators. To do so, I consider the following experiments illustrated by the schematic diagrams in Figures 2.1 and 2.2. I place a source and a receiver at the surface and I assume the Earth is characterized by a uniform velocity map m, commonly referred to as the background velocity model.

Forward Born modeling (demigration)

As I showed in the previous section, the application of $\mathbf{B}_m$ on a velocity perturbation p (represented by the green dot in Figure 2.1b) generates a data perturbation d. According to equation 2.63, the source wavefield is generated at $t = t_0$ from the source location and propagated forward in time into the subsurface with the (known) velocity map m. This process is illustrated in Figure 2.1c, where the arrow indicates the direction of propagation with increasing time values. At $t = t_1$, the interaction of the source wavefield with the velocity perturbation p creates a secondary source s_{sec}. The convolution of the Green's function g_m with this secondary source indicates that a secondary wavefield (referred to as the scattered wavefield u_{scat}) is created, propagated forward in time, and recorded at the receiver location at time $t = t_2$ (blue arrow in Figure 2.1d).

Adjoint Born modeling (migration)

For a single source/receiver pair, the application of $\mathbf{B}_m^*$ on an input data trace d (equation 2.72) can be re-written as

$$p(\mathbf{x}) \quad = \quad \frac{2}{m^3(\mathbf{x})} u_{sou}(\mathbf{x},t) \otimes u_{rec}(\mathbf{x},t), \tag{2.76}$$

where $\otimes$ denotes the zero time-lag cross-correlation. In equation 2.76, the image p is non-zero only if both source and receiver wavefields coincide in time and space, which is the well-known imaging condition first described by Claerbout (1985). The migration (or RTM) process $\mathbf{B}_m^*$ corresponds to the transformation of recorded seismic data d into seismic images p. Under certain conditions, it can be used to produce accurate images of the interfaces between the different rocks layers within the subsurface.

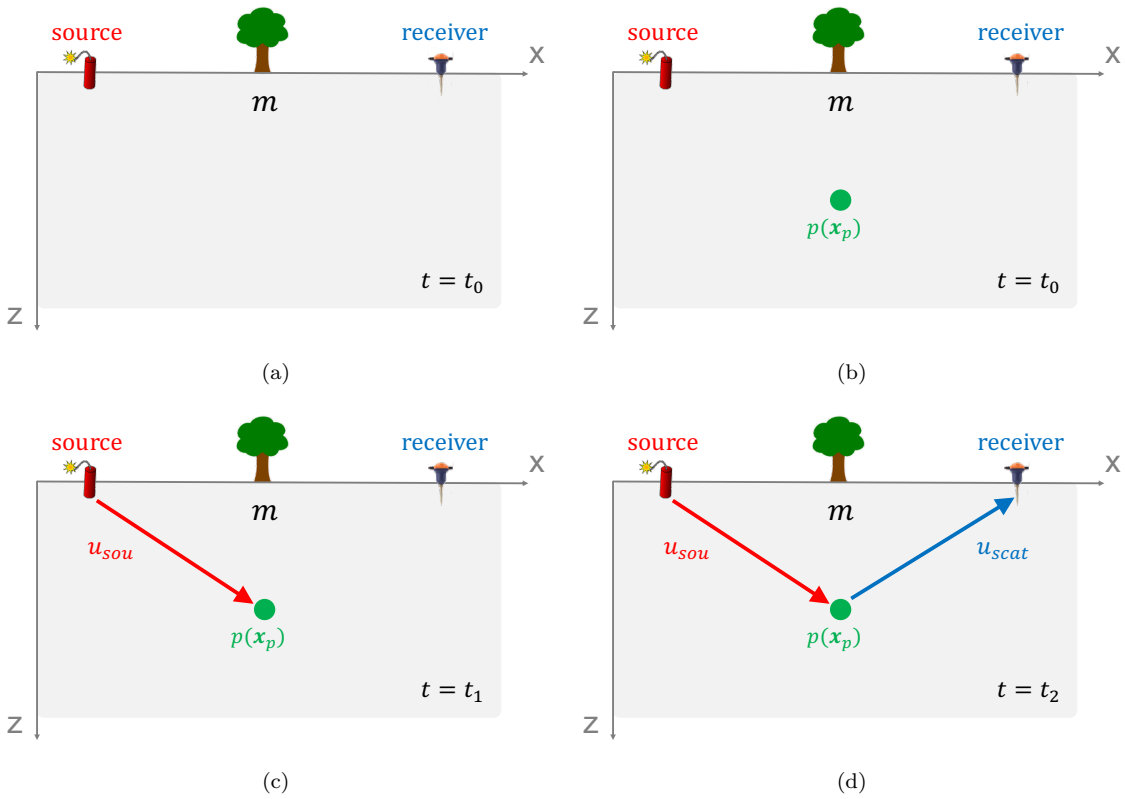


Figure 2.1: Schematic diagrams illustrating the mechanism of the Born forward modeling operator (i.e., the demigration process). The red and blue arrows represent the source and scattered wavefields propagating in the medium. (a) Initial state. (b) Model perturbation p (green dot). (c) Propagation of the scaled source wavefield u_{sou} forward in time (red arrow) generated by the seismic source. (d) Propagation of the scattered wavefield u_{scat} forward in time (blue arrow), recorded by the receiver, giving rise to a data perturbation d (not shown on the diagram). [**NR**]

Figure 2.2 illustrates a sequence of snapshots of the source (red arrow) and receiver (dark-blue arrow) wavefields at three time steps of the adjoint process. In the same manner as for the forward modeling, the source wavefield is propagated forward in time, while the receiver wavefield is generated at the receiver position and propagated backward in time. To gain better insight on the behavior of the receiver wavefield, I assume that the seismic trace injected at the receiver contains only one event at $t = t_2$. That is, $d(t) = \delta(t - t_2)$. Hence, the propagation of d backward in time implies that this wavefield should satisfy the anti-causal wave-equation and reach the receiver position at $t = t_2$.

Figure 2.2b shows that at $t = t_1$, the source and receiver wavefields coincide in time and space, which in turn gives rise to a model perturbation p, as illustrated in Figure 2.2c. Finally, at $t = t_2$, the receiver wavefield reaches the receiver position (Figure 2.2d).

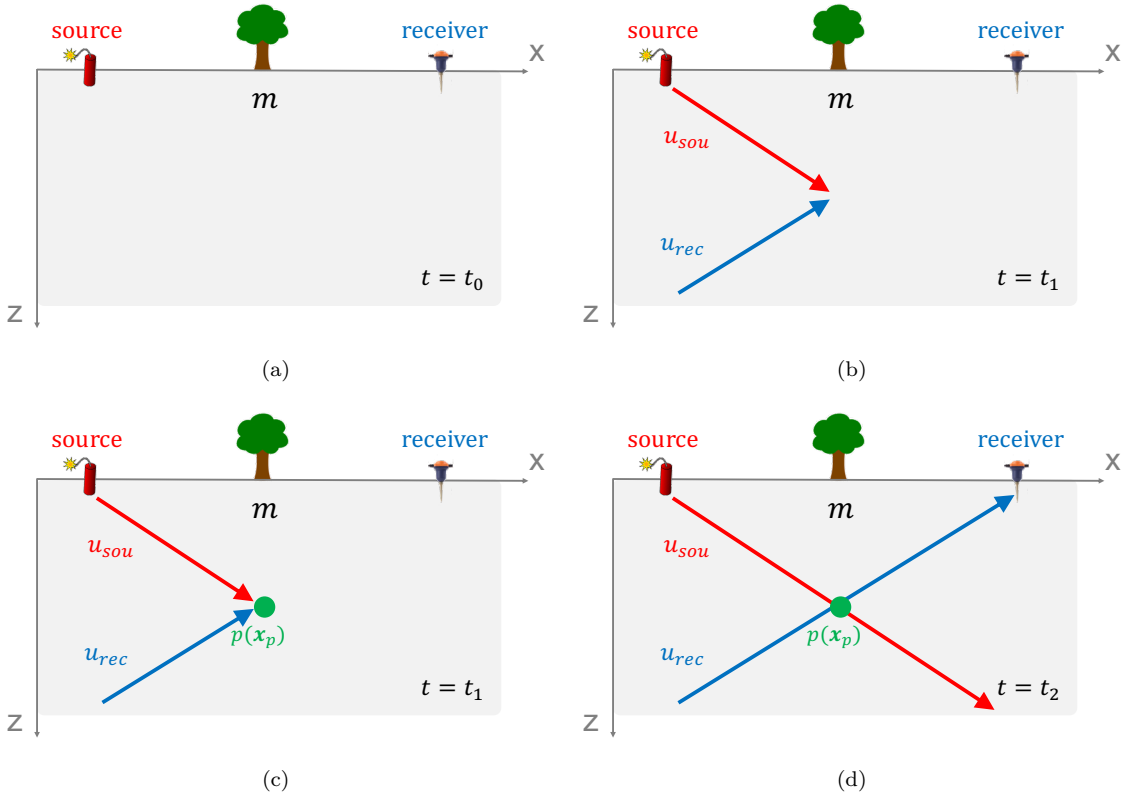


Figure 2.2: Schematic diagrams illustrating the mechanism of the adjoint Born modeling operator (i.e., the migration process). The red and blue arrows represent the source and receiver wavefields, respectively. (a) Initial state. (b) Schematic snapshot of both wavefields at propagation time $t = t_1$. (c) At $t = t_1$, both wavefields coincide in time and space, which gives rise to a model perturbation p. (d) At $t = t_2$, the receiver wavefield reaches the receiver location from where it was generated. [**NR**]

2.2.4 Numerical implementation

The discretization and efficient implementation of the Born modeling operator $\mathbf{B_m}$ can be achieved by applying a sequence of linear operators analogous to the ones used for the forward modeling modeling $\mathbf{F_m}$ (equation 2.43). Recall the application of $\mathbf{B_m}$ on a model perturbation $\mathbf{p}$,

$$\mathbf{d} \quad = \quad \mathbf{B_m}\,\mathbf{p}, \tag{2.77}$$

where $\mathbf{d} \in \mathbb{R}^{N_t}$, $\mathbf{p} \in \mathbb{R}^{N_m}$, and $\mathbf{B_m} \in \mathcal{M}^{N_t \times N_m}$. As I show in equation 2.61, the output $\mathbf{d}$ can be also expressed as

$$\mathbf{d} \quad = \quad \mathbf{F_m}\,\mathbf{s}_{sec} \tag{2.78}$$

where the secondary source $\mathbf{s}_{sec} \in \mathbf{R}^{N_w}$ is created from the scattering of the scaled source wavefield by the model perturbation $\mathbf{p}$. Moreover, equation 2.54 indicates that this secondary source is linear with respect to $\mathbf{p}$ and can be expressed as

$$\mathbf{s}_{sec} \quad = \quad \mathbf{U}_{sou}\,\mathbf{p}, \tag{2.79}$$

where $\mathbf{U}_{sou} \in \mathcal{M}^{N_w \times N_m}$ is given by

$$\mathbf{U}_{sou} = \begin{bmatrix} \mathbf{u}_{sou}[1] \\ \vdots \\ \mathbf{u}_{sou}[N_t] \end{bmatrix}. \tag{2.80}$$

Here, $\mathbf{u}_{sou}[i] \in \mathcal{M}^{N_m}$ is a diagonal matrix that multiplies the input by the values of the scaled source wavefield at time index i. For each source, the application of $\mathbf{U}_{sou}$ on an input velocity perturbation $\mathbf{p}$ requires the pre-computation (and possibly the storage) of the source wavefield. Assuming this task can be achieved, $\mathbf{B_m}$ and $\mathbf{B_m^*}$ can both be expressed as a sequence of discretized linear operators and efficiently implemented on a computer:

$$\mathbf{B_m} \quad = \quad \mathbf{F_m}\,\mathbf{U}_{sou} \tag{2.81}$$

$$\mathbf{B_m^*} \quad = \quad \mathbf{U}_{sou}^*\,\mathbf{F_m^*}. \tag{2.82}$$

2.3 Extended (Born) modeling

Conventional seismic modeling operators can be extended by increasing the dimension of their input space (thereby providing additional degrees of freedom) in order to model/fit/predict more complicated waveforms (e.g., highly nonlinear effects present in the observed data) than their non-extended counterparts, regardless of the accuracy of the velocity map. In FWIME, I follow the work of Biondi and Almomin (2014) and I extend the Born modeling operator by adding a non-physical axis to the model perturbations, which can either be subsurface offsets or time lags. In this section, I present the main motivation for developing such tools in the context of seismic waveform-inversion algorithms and I derive the analytical expression of the extended Born modeling operator (and its respective adjoint) for both extensions. Finally, I provide geophysical insight on the potential and effectiveness of this type of modeling at fitting complex waveforms in the data without the need of an accurate estimation of the velocity map.

2.3.1 Motivation

Symes (2008) first introduced and formalized the concept of extended modeling in an attempt to design a new acoustic-velocity model-building optimization scheme by combining the two main families of techniques described in the previous chapter, namely FWI and WEMVA. The need for such type of modeling originated from one specific limitation inherent to the conventional Born modeling operator (equation 2.63). As stated by the author, "the linearized [Born] modeling operator is not surjective in general, unless the data is consistent kinematically with the velocity model v. This occurs generically only for very special acquisition geometries, for example single shot or offset gathers. In general a poorly chosen velocity prevents accurate fitting of multioffset data."

This claim can be first understood from a practical standpoint. Consider the following scenario where (1) some multi-offset data d (i.e., data recorded or modeled for many source/receiver pairs) are generated with an unknown velocity map m_{true}, such that $d = \mathbf{f}(m_{true})$ (as shown in equation 2.8), and (2) an estimated velocity map m is given. The author's statement indicates that if m is too different from m_{true}, there may not exist a model perturbation p that simultaneously satisfies

$$d - \mathbf{f}(m) \approx \mathbf{B}_m p \tag{2.83}$$

for the *entire collection* of source/receiver pairs. From a mathematical perspective, the non-surjectivity of $\mathbf{B}_m$ can be explained by the fact that this operator is merely the first-order differential of the modeling operator $\mathbf{f}$. Hence, it may not be able to model some of the nonlinear terms of the data-difference $d - \mathbf{f}(m)$. Unfortunately in many model-building techniques (including FWIME), ensuring the existence of such p regardless of the accuracy of m seems to be a necessary (but not sufficient) condition to recover an accurate estimate of m_{true}.

In order to address this issue, Symes (2008) proposed to modify the operator $\mathbf{B}_m$ by allowing its input p to have an additional degree of freedom, referred to as an extension (or extended axis). Model extensions have been developed and successfully used in the context of image-domain wavefield tomography (Symes and Carazzone, 1991; Sava and Biondi, 2004; Shen, 2005; Sava and Fomel, 2006; Yang, 2013), and for hybrid waveform-inversion methods such as the one proposed in this thesis (Fleury and Perrone, 2012; Van Leeuwen and Herrmann, 2013; Biondi and Almomin, 2014; Huang et al., 2017; Barnier and Biondi, 2020). Following a similar philosophy as the one developed in Biondi and Almomin (2014), I choose to extend the Born modeling operator (and the model perturbation) using either subsurface offsets or time lags, depending on the problem configuration. Although no theoretical proof has been proposed to my knowledge yet, there are plenty of numerical evidence showing that such extensions are able to satisfy the surjectivity property (equation 2.83), even when very complex events are recorded in the data and the estimate of the velocity map is inaccurate (Symes, 2008; Almomin, 2016; Barnier and Biondi, 2020).

2.3.2 Extended model representation

In the following, I use the $\sim$ symbol to denote all extended operators and model perturbations (also referred to as extended images). I use $\mathbf{h}$ and τ to denote the subsurface-offset and time-lag variables, respectively. Figure 2.3 shows schematic illustrations of an extended image $\tilde{p}$ in a two-dimensional scenario. The extended axis (corresponding to either subsurface offsets or time lags) is represented by the green arrow. The physical plane of $\tilde{p}$ is defined as the set of points whose coordinates on the extended axis are zero (Figure 2.3b). A common image gather (CIG) is the set of points of $\tilde{p}$ that share the same horizontal position (Figure 2.3c). CIGs for subsurface-offset and time-lag extensions are referred to as subsurface-offset CIGs (SOCIG) and time-lag CIGs (TLCIG), respectively. Finally, the extended depth plane is the set of points within $\tilde{p}$ sharing the same depth value (Figure 2.3d).

2.3.3 Subsurface-offset extension

For one source/receiver pair and a fixed velocity model, the Born modeling operator extended in subsurface offsets is a mapping from an extended perturbation $\tilde{p}$ to a seismic trace d:

$$\begin{aligned} \tilde{\mathbf{B}}_m : \mathcal{F}(\Omega \times \tilde{\Omega}, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ \tilde{p} &\mapsto d = \tilde{\mathbf{B}}_m \tilde{p}, \end{aligned} \tag{2.84}$$

where $\tilde{\Omega} \subset \mathbb{R}^3$ is the subsurface-offset extension domain. The application of $\tilde{\mathbf{B}}_m$ on $\tilde{p}$ is defined as the seismic trace extracted from a field q at a receiver position $\mathbf{x}_r$, where q is the solution of the following PDE (Symes, 2008),

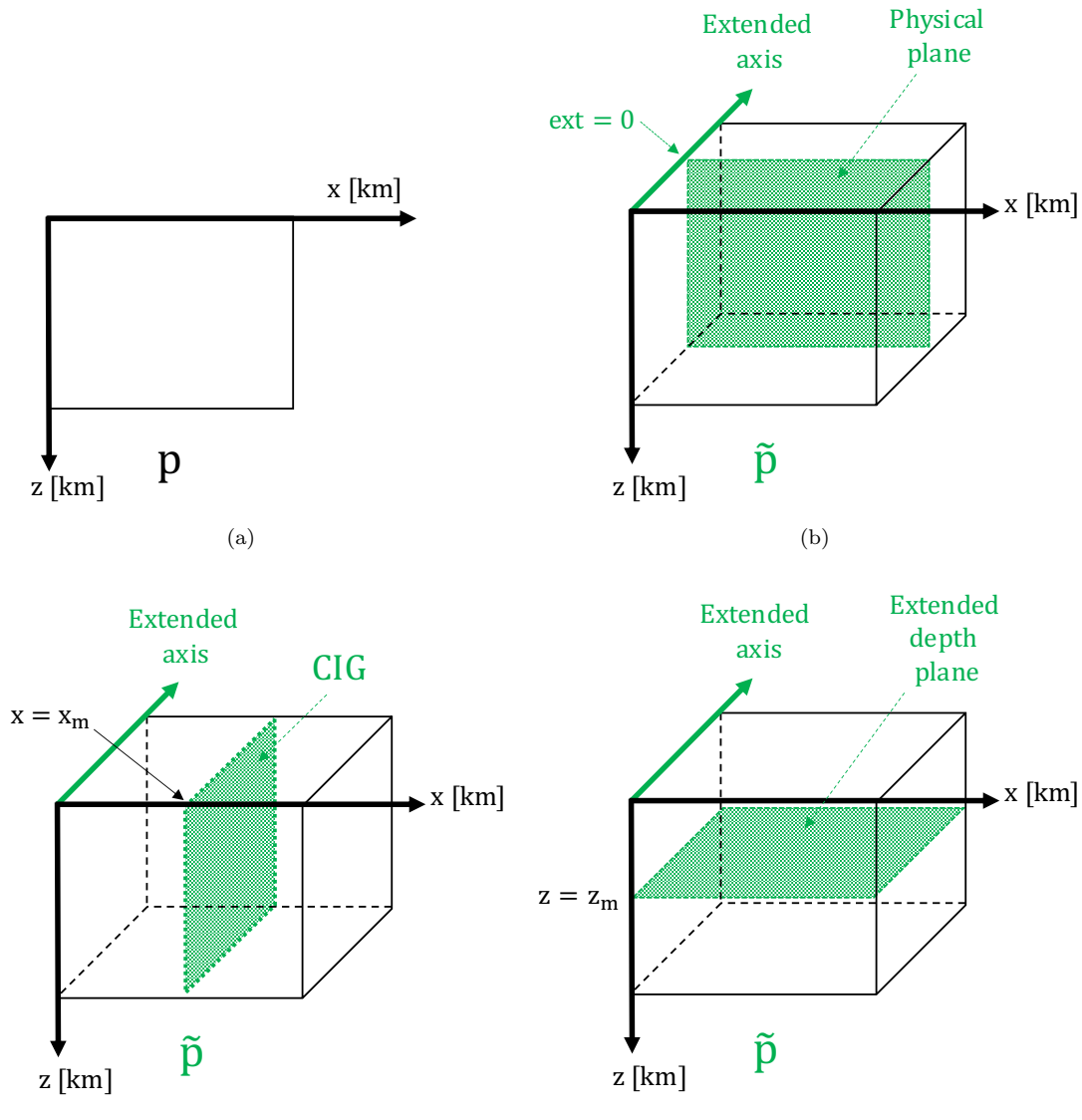


Figure 2.3: Schematic representations of a model perturbation in a two-dimensional setting. (a) Conventional (non-extended) perturbation p. (b) Physical plane of an extended perturbation $\tilde{p}$. (c) Common image gather (CIG) extracted at $\mathbf{x} = \mathbf{x}_m$ from $\tilde{p}$. (d) Extended depth plane extracted at $\mathbf{z} = \mathbf{z}_m$ from $\tilde{p}$. [**NR**]

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 q(\mathbf{x},t)}{\partial t^2} - \nabla^2 q(\mathbf{x},t) \quad = \quad \int\limits_{\mathbf{h}\in\tilde{\Omega}} 2\frac{\tilde{p}(\mathbf{x}-\mathbf{h},\mathbf{h})}{m^3(\mathbf{x}-\mathbf{h})}\ddot{u}_m(\mathbf{x}-2\mathbf{h},t)\; d\mathbf{h}, \tag{2.85}$$

with the same initial time boundary conditions as in equation 2.2, and where $\mathbf{h} \in \tilde{\Omega}$. Equation 2.85 is a similar PDE as equation 2.53, but with a modified source term $\tilde{s}_{sec}$:

$$\tilde{s}_{sec}(\mathbf{x},t) \quad = \quad \int\limits_{\mathbf{h}\in\tilde{\Omega}} 2\frac{\tilde{p}(\mathbf{x}-\mathbf{h},\mathbf{h})}{m^3(\mathbf{x}-\mathbf{h})}\ddot{u}_m(\mathbf{x}-2\mathbf{h},t)\; d\mathbf{h}. \tag{2.86}$$

$\tilde{s}_{sec}$ is the output of the convolution (over the extended-axis variable $\mathbf{h}$) of the scaled source wavefield with the extended image. For multiple source/receiver pairs, the mapping expressed in equation 2.84 is given by

$$\boxed{\big(\tilde{\mathbf{B}}_m\tilde{p}\big)(t;s^{(i)},\mathbf{x}_{r^{(j)}}) = \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_{r^{(j)}},t,\mathbf{x},0) * \int\limits_{\mathbf{h}\in\tilde{\Omega}} 2\frac{\tilde{p}(\mathbf{x}-\mathbf{h},\mathbf{h})}{m^3(\mathbf{x}-\mathbf{h})}\ddot{u}_m(\mathbf{x}-2\mathbf{h},t;s^{(i)})\; d\mathbf{h}\; d\mathbf{x}.} \tag{2.87}$$

The adjoint extended Born operator (for a single source/receiver pair) is defined as

$$\begin{aligned} \tilde{\mathbf{B}}_m^* : \mathcal{F}(\mathcal{T},\mathbb{R}) &\mapsto \mathcal{F}(\Omega\times\tilde{\Omega},\mathbb{R}) \\ d &\mapsto \tilde{p} = \tilde{\mathbf{B}}_m^* d, \end{aligned} \tag{2.88}$$

and by conducting an analogous derivation as for the non-extended case, I can show that its expression (for multiple source/receiver pairs) is given by

$$\boxed{(\tilde{\mathbf{B}}_m^*\; d)(\mathbf{x},\mathbf{h}) \quad = \quad \sum_{i=1}^{N_s}\int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})}\ddot{u}_m(\mathbf{x}-\mathbf{h},t;s^{(i)})\; u_{rec}(\mathbf{x}+\mathbf{h},t;s^{(i)})\; dt,} \tag{2.89}$$

using the same definition for u_{rec} as in equation 2.75. Similarly as for the non-extended case, equation 2.89 is the zero time-lag cross-correlation between the spatially-shifted source and receiver wavefields (Rickett and Sava, 2002). Due to the high computational cost of the subsurface-offset extension for three-dimensional field-data applications, $\mathbf{h}$ is usually limited to the horizontal plane, in which case the extended image $\tilde{p}$ becomes a five-dimensional hypercube.

Forward modeling with a subsurface-offset extension

I provide some physical intuition (rather than a theoretical proof) on why the subsurface-offset extension has the potential to satisfy the surjectivity condition expressed in equation 2.83 for multiple source/receiver pairs. For that, I use a similar experiment and schematic diagrams as for the non-extended case. Figure 2.4 illustrates the extended demigration process given by equation 2.87 for a conventional surface acquisition geometry. As shown in Figure 2.4b, I place an extended perturbation $\tilde{p}$ in the subsurface parametrized by

$$\tilde{p}(\mathbf{x},\mathbf{h}) \quad = \quad \delta(\mathbf{x}-\mathbf{x}_p)\,\delta(\mathbf{h}-\mathbf{h}_p), \tag{2.90}$$

where $\mathbf{h}_p$ is a horizontal vector. The source wavefield u_{sou} is propagated from the source location into the subsurface. According to equation 2.87, u_{sou} interacts with the extended perturbation at time $t=t_1'$ when its wavefront reaches $\mathbf{x}=\mathbf{x}_p-\mathbf{h}_p$ (Figure 2.4c), and instantly generates a scattered wavefield as if the scatterer (i.e., the secondary source) was positioned at $\mathbf{x}=\mathbf{x}_p+\mathbf{h}_p$ (Figure 2.4d). Finally, this scattered wavefield is propagated and eventually recorded at the receiver position at time $t=t_2'$. Here, both source and scattered wavefields are propagated with the same velocity map m.

This extended demigration process allows the source wavefield to interact with the model perturbation at different spatial positions than its actual physical location, thereby potentially introducing a time shift compared to the "physical" (i.e., non-extended) process. In Figure 2.4b, $\tilde{p}$ is located at $\mathbf{x}_p$, but the source wavefield behaves as if $\tilde{p}$ was located at $\mathbf{x}_p-\mathbf{h}_p$. Intuitively, one can expect that by "displacing" the model perturbation along the extended axis (while keeping its physical location $\mathbf{x}_p$ fixed), it is possible to generate time-shifted scattered wavefields (using the same source wavefield and velocity m) that can illuminate the receivers locations with different apparent traveltimes and incidence angles. Compared to its non-extended counterpart, this process provides the ability and freedom to linearly model events in the (multi-offset) data that possess features associated with an apparent velocity map very different than m (by demigrating the same extended image for all source/receiver pairs), which is key to satisfy the surjectivity condition expressed in equation 2.83.

To illustrate this property, I design a simple experiment (Figure 2.5) where I show how extended demigration can create simple time shifts in the modeled data (for a single source/receiver pair) simply by adjusting the value of the model perturbation along its extended axis. I generate a source wavefield at $t=t_0$ from the source location with a time signature represented in Figure 2.5a. I then demigrate $\tilde{p}(\mathbf{x},\mathbf{h})$ (defined equation 2.90) for three different values of $\mathbf{h}_p$, and I record the modeled data d at the receiver location. Figure 2.5b shows the recorded data modeled with $\mathbf{h}_p=\mathbf{0}$. The data generated with the extended and non-extended operators are identical. Figure 2.5c shows the recorded data when a similar value of $\mathbf{h}_p$ as the one drawn in Figure 2.4c is used. In this configuration,

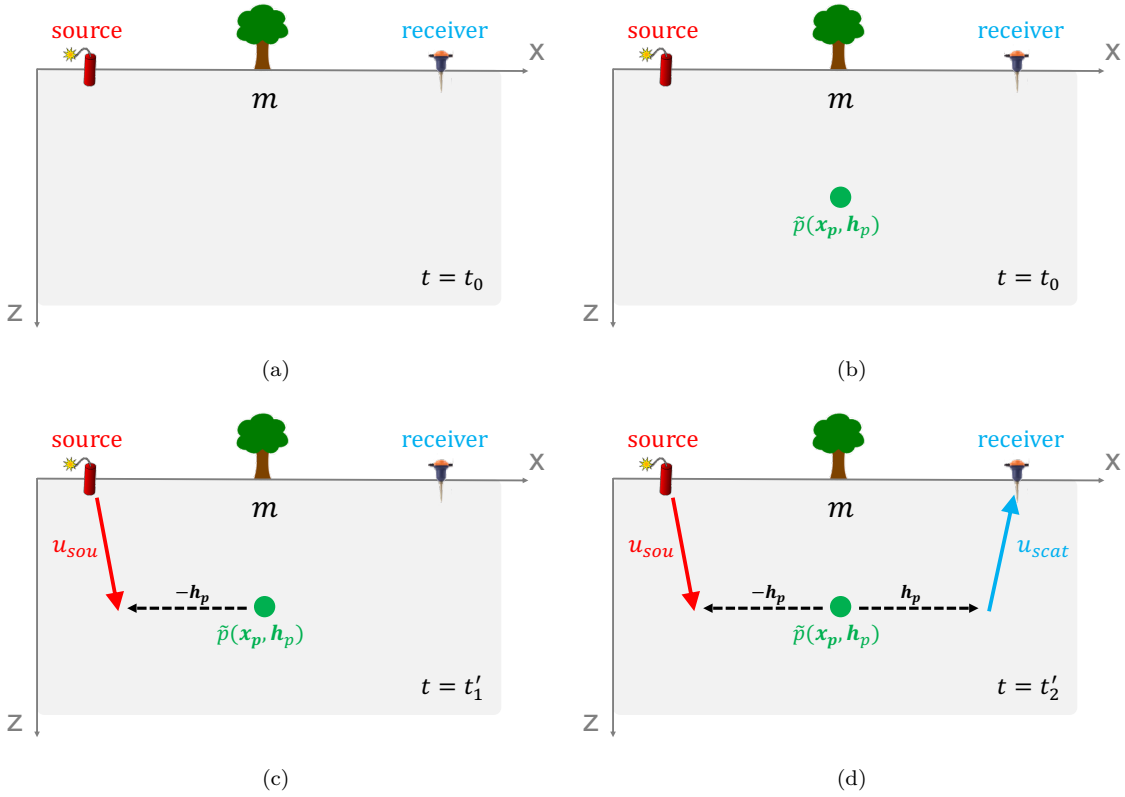


Figure 2.4: Schematic diagrams illustrating the mechanism of the extended Born forward modeling operator using a subsurface-offset extension. This process is also referred to as an extended demigration. The red and light-blue arrows represent the source and scattered wavefields, respectively. (a) Initial state. (b) Extended model perturbation $\tilde{p}$ (green dot). (c) Snapshot of the source wavefield at propagation time $t = t_1'$ as its wavefront reaches $\mathbf{x} = \mathbf{x}_p - \mathbf{h}_p$. At the same time, the interaction of the source wavefield with the extended perturbation generates a scattered wavefield from position $\mathbf{x} = \mathbf{x}_p + \mathbf{h}_p$. (d) At $t = t_2'$, the scattered wavefield reaches the receiver location and is recorded. [**NR**]

both source and scattered wavefields travel a smaller distance, and the resulting recorded event (light-blue curve) precedes the one from the non-extended operator (dark-blue curve). For an observer placed at the receiver location, the velocity of the medium seems higher. Conversely, by changing the sign of $\mathbf{h}_p$, the traveltime of the recorded trace can be increased, as shown in Figure 2.5d.

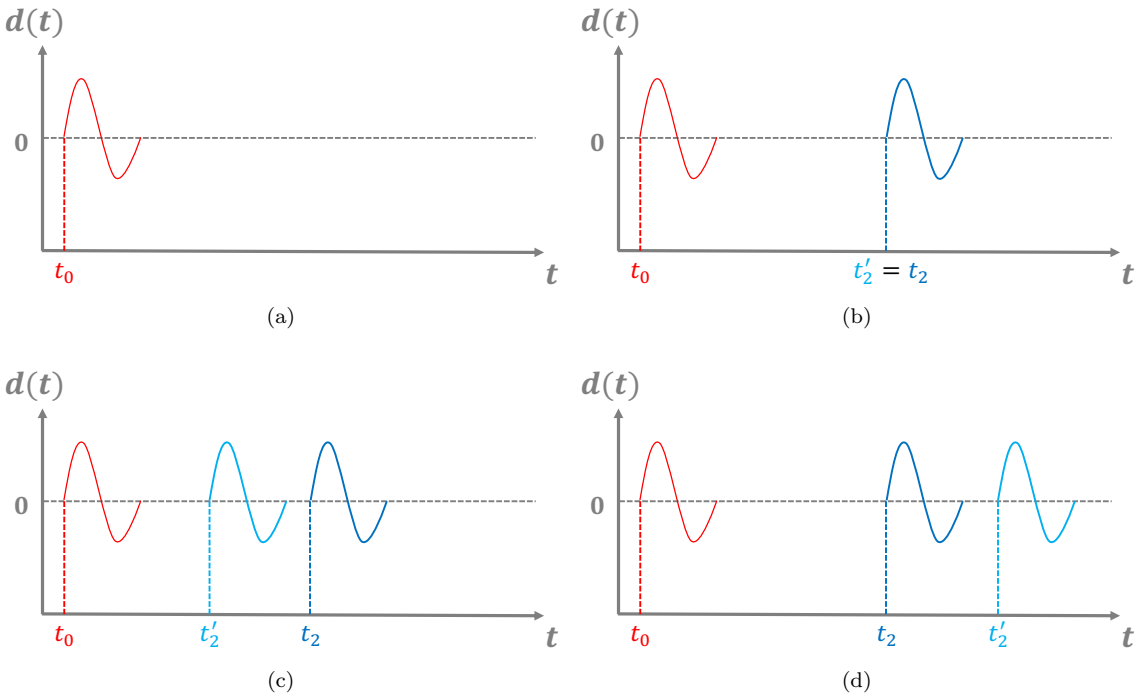


Figure 2.5: Schematic diagrams representing the source signature (red curves), the recorded seismic traces at the receiver position modeled with different value of $\mathbf{h}_p$ (light-blues curves), and the recorded seismic trace using a non-extended Born modeling operator (dark-blue curves). The source signature is only displayed for reference, and is not recorded by the receiver. For simplicity purposes, I also neglected all amplitude-decay effects related to geometrical spreading. (a) Time signature of the source field. (b) Modeled data using an extended perturbation with $\mathbf{h}_p = \mathbf{0}$. Both light- and dark-blue curves are identical. (c) Modeled data using a similar value of $\mathbf{h}_p$ as the one drawn in Figure 2.4c. (d) Modeled data using the same value of $\mathbf{h}_p$ as in (c) but with opposite sign. [**NR**]

The ability of extended modeling at linearly generating data with complex waveforms (even with smoothly varying velocity maps) can be further appreciated when multiple receivers are employed. Not only does it provides the flexibility to produce "non-physical" time shifts (as we saw in the previous paragraph), but it also offers the possibility to model and fit events in the data with flexible moveout signatures. Figure 2.6a shows an extended demigration process with a similar configuration as the one in Figure 2.4, but using an array of receivers. The modeled data for this experiment is represented in Figure 2.7a. If the distance between the scattering location $\mathbf{x} + \mathbf{h}$ and the receiver array is much greater than the lateral extended of array itself, the local (hyperbolic)

moveout signature of the modeled data is characterized by an apparent velocity $m_a \approx m/sin(\theta)$, where θ is the incidence angle of the scattered wavefield (for simplicity, m is assumed to be uniform). Furthermore, by shifting the physical location of the extended model perturbation from $\mathbf{x}$ to $\mathbf{x}'$ and by adjusting its value on the extended axis from $\mathbf{h}$ to $\mathbf{h}'$ (Figure 2.6b) such that the traveltime recorded by leftmost receiver r_1 is unchanged (Figure 2.7a), the moveout of the modeled data is now characterized by a different (lower) apparent velocity $m'_a \approx m/sin(\theta')$, as shown in Figure 2.7b.

One can notice that if only a single source s is employed, the seismogram shown in Figure 2.7b can likely be obtained without the use of an extended perturbation and extended modeling, by simply shifting the position of the non-extended perturbation p within the physical plane. However, for multiple sources (and receiver arrays), the surjectivity condition from equation 2.83 is met only if

$$\cap_{i=1}^{N_s} \mathcal{S}_{s^{(i)}} \neq \varnothing, \tag{2.91}$$

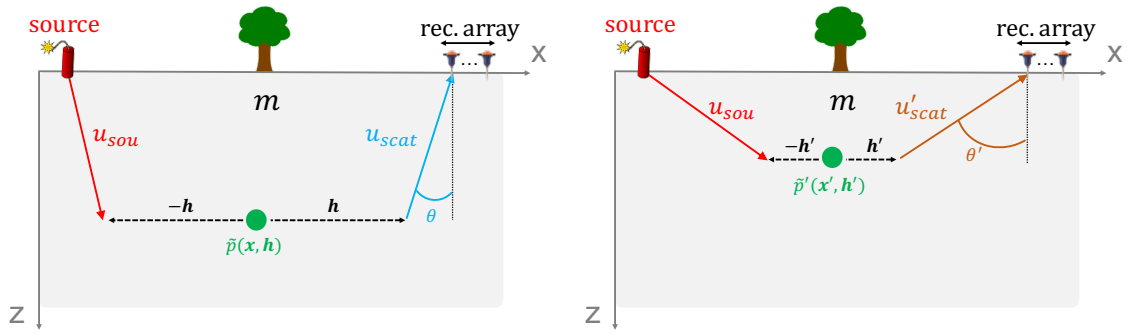
where N_s is the number of sources, and $\mathcal{S}_{s^{(i)}}$ corresponds to the set of non-extended perturbations satisfying equation 2.83 for the i^{th} source. As mentioned by Symes (2008), it has been numerically observed that for a large number of sources, equation 2.91 is usually not satisfied. Additionally, Almomin (2016) explains that this phenomenon is caused by the averaging and destructive interference of the data components in the non-extended perturbation space. On the other hand, extended modeling using a subsurface-offset (or time-lag) extension provides a larger set of solutions for $\tilde{p}$. That is,

$$\mathcal{S}_{s^{(i)}} \subset \tilde{\mathcal{S}}_{s^{(i)}}, \tag{2.92}$$

such that

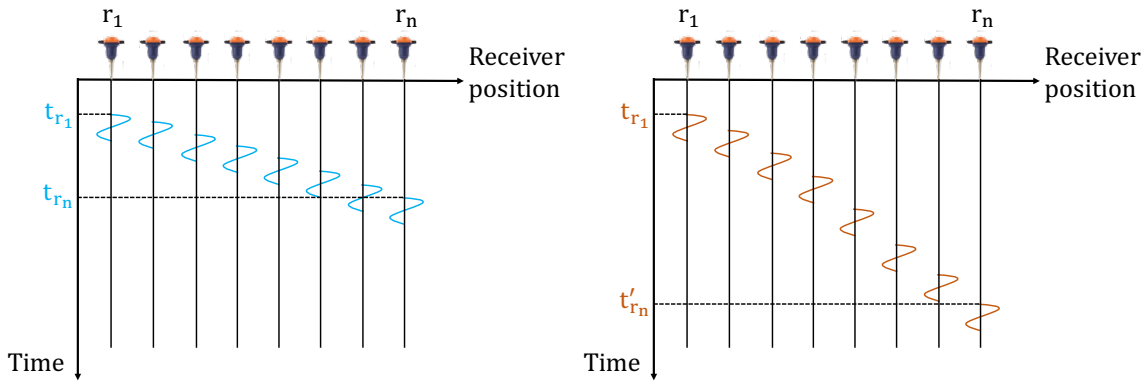
$$\cap_{i=1}^{N_s} \tilde{\mathcal{S}}_{s^{(i)}} \neq \varnothing, \tag{2.93}$$

where $\tilde{\mathcal{S}}_{s^{(i)}}$ corresponds to the set of extended perturbations satisfying equation 2.83 for the i^{th} source. Even though not proven here, this additional set of solutions seems to ensure the existence of a (well-chosen) extended perturbation that can satisfy both equations 2.83 and 2.93. In the next chapter, I show how to compute such solution in the context of FWIME, and I provide plenty of successful numerical examples where I make use of this powerful tool.



(a) (b)

Figure 2.6: Schematic illustrations of extended demigration using two extended perturbations (with a constant velocity map m). The red arrow represents the source wavefield, while the blue and brown arrows represent the scattered wavefields. The modeled data is recorded by a receiver array at the surface. (a) Demigration of the initial extended perturbation $\tilde{p}$. (b) Demigration of the shifted extended perturbation $\tilde{p}'$. [**NR**]



(a) (b)

Figure 2.7: Schematic representations of the recorded pressure data generated by the extended demigration processes shown in Figure 2.6 using (a) the initial extended perturbation $\tilde{p}$, and (b) the shifted extended perturbation $\tilde{p}'$. [**NR**]

Migration with a subsurface-offset extension

Migrated images extended in subsurface offsets have been thoroughly studied and successfully used for image-domain tomography (e.g., WEMVA) where the goal is to recover the optimal velocity model that maximizes the focusing of space-lag gathers (Sava and Biondi, 2004; Yang, 2013; Díaz Pantin, 2016). In this section, I illustrate the extended migration process (equation 2.89) with schematic diagrams shown in Figure 2.8 that represent a sequence of snapshots of the source (red arrow) and receiver (blue arrow) wavefields at three time steps. In the same manner as for the non-extended case, the source wavefield is propagated forward in time, while the receiver wavefield is generated at the receiver position and propagated backward in time. Figure 2.8a shows the initial state. At $t = t'_1$ (and according to equation 2.89), the zero time-lag cross-correlation of the spatially shifted source and receiver wavefields gives rise to an extended model perturbation $\tilde{p}(\mathbf{x}, \mathbf{h})$ as shown in Figure 2.8c. Finally at $t = t'_2$, the receiver wavefield reaches the receiver location from where it was generated (Figure 2.8d).

2.3.4 Time-lag extension

For one source/receiver pair, the Born modeling operator extended in time lags is a mapping from an extended perturbation $\tilde{p}$ to a seismic trace d:

$$\begin{aligned} \tilde{\mathbf{B}}_m : \mathcal{F}(\Omega \times \tilde{\mathcal{T}}, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ \tilde{p} &\mapsto d = \tilde{\mathbf{B}}_m \tilde{p}, \end{aligned} \tag{2.94}$$

where typically $\tilde{\mathcal{T}} = [-\tau_{max}, \tau_{max}]$, and the application of $\tilde{\mathbf{B}}_m$ on $\tilde{p}$ is defined as the seismic trace extracted from a field q at a receiver position $\mathbf{x}_r$, where q satisfies the following PDE,

$$\frac{1}{m^2(\mathbf{x})} \frac{\partial^2 q(\mathbf{x}, t)}{\partial t^2} - \nabla^2 q(\mathbf{x}, t) = \int_{\tau \in \tilde{\mathcal{T}}} 2 \frac{\tilde{p}(\mathbf{x}, \tau)}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x}, t - 2\tau)\, d\tau. \tag{2.95}$$

with the same initial time boundary conditions as in equation 2.2, and where $\tau \in \tilde{\mathcal{T}}$. Equation 2.95 is a similar PDE as equation 2.85, but with a modified source term $\tilde{s}_{sec}$:

$$\tilde{s}_{sec}(\mathbf{x}, t) = \int_{\tau \in \tilde{\mathcal{T}}} 2 \frac{\tilde{p}(\mathbf{x}, \tau)}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x}, t - 2\tau)\, d\tau. \tag{2.96}$$

$\tilde{s}_{sec}$ is the output of the convolution (over the extended-axis variable τ) of the scaled source wavefield with the time-lag extended image. The mapping expressed in equation 2.94 for a collection

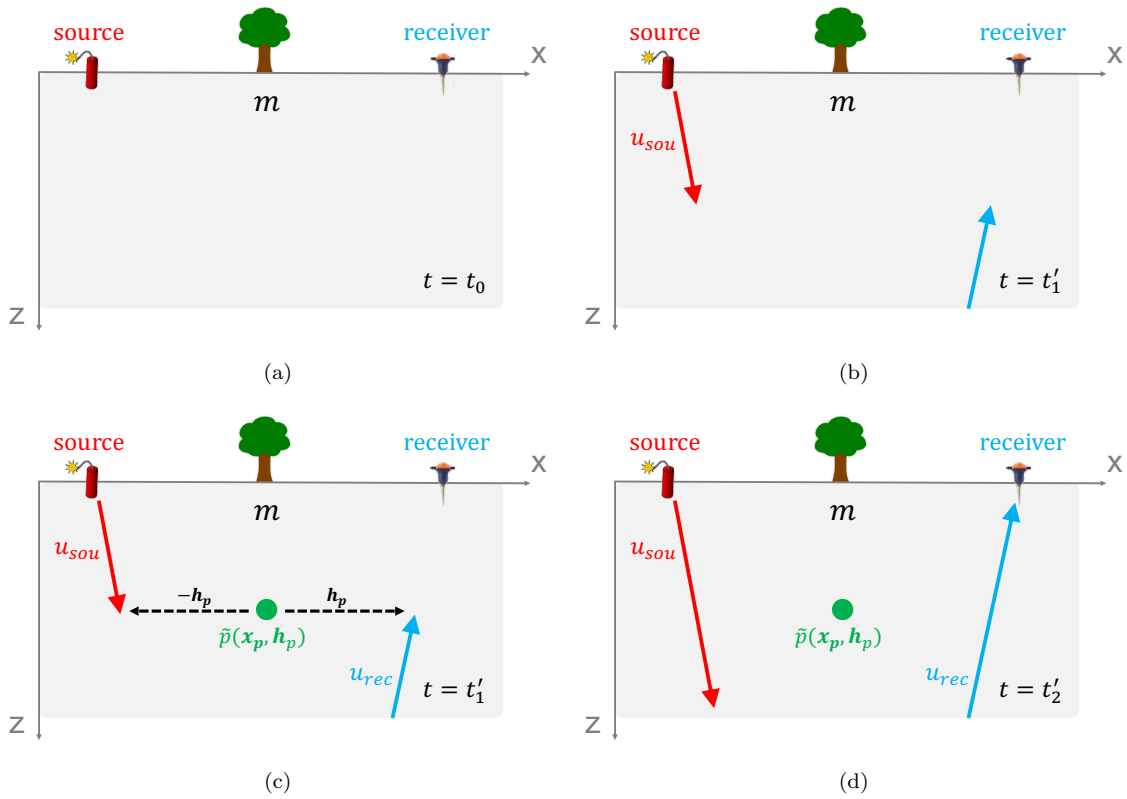


Figure 2.8: Schematic illustration of a subsurface-offset extended migration process. The red and blue arrows represent the source and receiver wavefields, respectively. (a) Initial state. (b) At $t = t_1'$, the source and receiver wavefields reach subsurface points $\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$, respectively. (c) At $t = t_1'$, the zero time-lag cross-correlation of the spatially-shifted source and receiver wavefields gives rise to an extended model perturbation $\tilde{p}$. (d) At $t = t_2'$, the receiver wavefield reaches the receiver location from where it was generated. [**NR**]

of source/receiver pairs is given by

$$\boxed{\big(\mathbf{B}_m\tilde{p}\big)(t;s^{(i)},\mathbf{x}_{r^{(j)}}) \quad = \quad \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_{r^{(j)}},t,\mathbf{x},0) * \int\limits_{\tau\in\tilde{\mathcal{T}}} 2\frac{\tilde{p}(\mathbf{x},\tau)}{m^3(\mathbf{x})}\ddot{u}_m(\mathbf{x},t-2\tau;s^{(i)})\,d\tau\,d\mathbf{x},} \tag{2.97}$$

and its adjoint process is

$$\boxed{(\tilde{\mathbf{B}}_m^*\,d)(\mathbf{x},\tau) \quad = \quad \sum_{i=1}^{N_s}\int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})}\ddot{u}_m(\mathbf{x},t-\tau;s^{(i)})\;u_{rec}^{(i)}(\mathbf{x},t+\tau;s^{(i)})\,dt.} \tag{2.98}$$

Equation 2.98 is the well-known extended imaging condition based on time shifts between source and receiver wavefields, first proposed by Sava and Fomel (2006). Since this time-lag extension requires only one additional axis even in 3D, it can be more computationally attractive and memory efficient than subsurface offsets. Moreover, it has been numerically observed that the time-lag extended Born operator is more effective at modeling events generated by steeply-dipping reflectors as well as diving waves. However, its downside is th lack of spatial directionality for 3D applications as well the number of wavefield time slices needed to be loaded in memory during the modeling process (Almomin, 2016; Barnier et al., 2018).

Forward modeling with a time-lag extension

Figure 2.9 describes the time-lag extended Born demigration process using the same scenario as for the subsurface-offset extension shown in Figure 2.4. I place an extended perturbation parametrized by

$$\tilde{p}(\mathbf{x},\tau) \quad = \quad \delta(\mathbf{x}-\mathbf{x}_p)\;\delta(\tau-\tau_p). \tag{2.99}$$

For this specific example, I set $\tau_p < 0$ (Figure 2.9b). According to equation 2.97, the scaled source wavefield u_{sou} is propagated from the source location into the subsurface, and interacts with the extended perturbation at time $t = t_1 - 2|\tau_p|$ before its wavefront reaches the perturbation's physical location $\mathbf{x}_p$ (Figure 2.9c). A scattered wavefield is generated from $\mathbf{x}_p$, propagated, and recorded at the receiver location at $t = t_2 - 2|\tau_p|$ (Figure 2.9d), which corresponds to an earlier arrival time than for the non-extended Born operator. Alternatively, a positive value of τ_p would have created a delay in the modeled data. Therefore, as for the subsurface-offset extension, it is possible to create time shifts in the modeled data by moving the position of the extended perturbation on the time-lag axis, and without modifying the wavefield propagation velocity m. Although not shown here, a similar

analysis as the one proposed for subsurface offsets could be conducted to show the ability of time-lag modeling at fitting complex events in the data by taking advantage of the extended axis. In the next chapter, I provide plenty of numerical examples that support this claim.

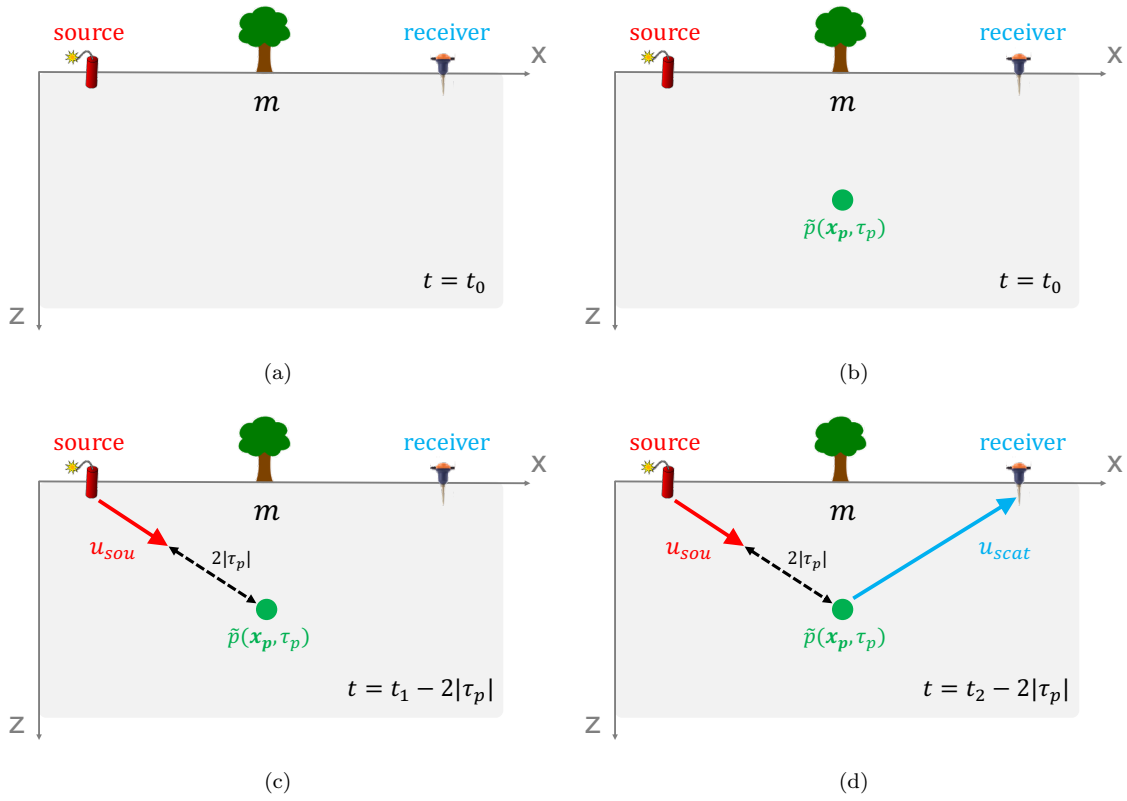


Figure 2.9: Schematic diagrams illustrating the mechanism of the extended Born forward modeling operator using a time-lag extended perturbation. The red and light-blue arrows represent the source and scattered wavefields, respectively. (a) Initial state. (b) Extended model perturbation $\tilde{p}$ (green dot) arbitrarily chosen with $\tau_p < 0$. (c) At $t = t_1 - 2|\tau_p|$, the interaction of the source wavefield with the extended perturbation generates a scattered wavefield at $\mathbf{x}_p$. (d) The scattered wavefield reaches the receiver location and is recorded at $t = t_2 - 2|\tau_p|$. [**NR**]

Migration with a time-lag extension

The mechanism of time-lag extended migration is summarized in Figure 2.10. The source wavefield is propagated forward in time, while the receiver wavefield is generated at the receiver position and propagated backward in time. Figure 2.10b shows a snapshot of the source wavefield at $t = t_1 - \tau_p$, while Figure 2.10c shows the receiver wavefield at $t = t_1 + \tau_p$. The time-lag imaging condition from equation 2.98 implies that the cross-correlation of these two time-shifted wavefields produces an extended image $\tilde{p}(\mathbf{x}_p, \tau_p)$ (Figure 2.10d).

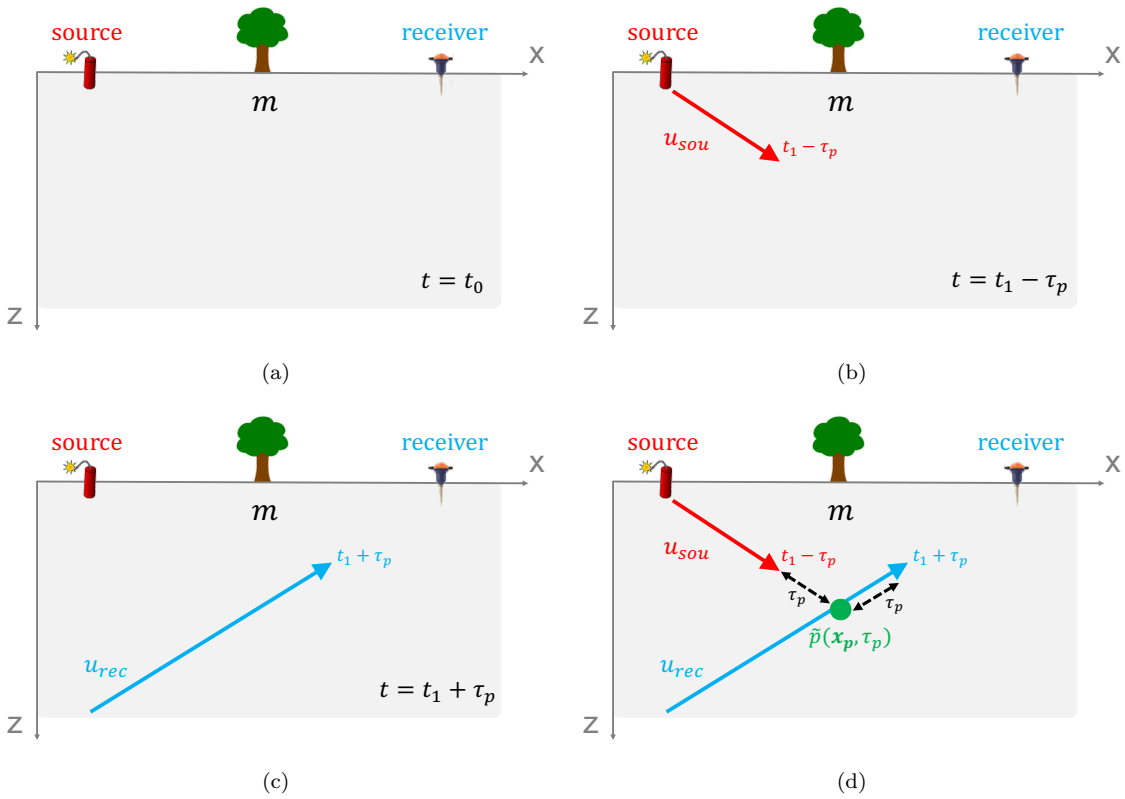


Figure 2.10: Schematic illustration of a time-lag extended migration process. The red and blue arrows represent the source and receiver wavefields, respectively. I illustrate the migration process for $\tau_p > 0$. (a) Initial state. (b) Snapshot of the source wavefield at $t = t_1 - \tau_p$. (c) Snapshot of the receiver wavefield at $t = t_1 + \tau_p$. (d) The cross-correlation of the time-shifted source and receiver wavefields gives rise to an extended image $\tilde{p}$. The dashed-black arrows indicate that the propagation time between the wavefronts of each wavefield and $\mathbf{x}_p$ is τ_p. [**NR**]

2.3.5 Numerical implementation

The discretization and numerical implementation of the extended Born forward and adjoint operators is conducted in a similar manner as for the non-extended case. The input $\tilde{\mathbf{p}}$ is now an element of $\mathbb{R}^{N_{\tilde{p}}}$ where $N_{\tilde{p}} = N_m \times N_{ext}$, and N_{ext} is the size of the extension domain. For the subsurface-offset extension, $\mathbf{h}$ is discretized with the same sampling as for the spatial dimensions. For the time-lag extension, τ is discretized with the same time-sampling rate as the modeled/recorded data. Equations 2.81 and 2.82 become

$$\tilde{\mathbf{B}}_{\mathbf{m}} = \mathbf{F}_{\mathbf{m}}\, \tilde{\mathbf{U}}_{sou} \tag{2.100}$$

$$\tilde{\mathbf{B}}^*_{\mathbf{m}} = \tilde{\mathbf{U}}^*_{sou}\, \mathbf{F}^*_{\mathbf{m}}, \tag{2.101}$$

where $\tilde{\mathbf{B}}_{\mathbf{m}}$ and $\tilde{\mathbf{U}}_{sou}$ are elements of $\mathcal{M}^{N_w \times N_{\tilde{p}}}$, and

$$\tilde{\mathbf{U}}_{sou} = \begin{bmatrix} \tilde{\mathbf{u}}_{sou}[1] \\ \vdots \\ \tilde{\mathbf{u}}_{sou}[N_t] \end{bmatrix}. \tag{2.102}$$

The matrices $\tilde{\mathbf{u}}_{sou}[i] \in \mathcal{M}^{N_m \times N_{\tilde{p}}}$ perform a convolution (over the extended variables $\mathbf{h}$ or τ) between the extended input model $\tilde{\mathbf{p}}$ and the (scaled) source wavefield at time index i, as described in equation 2.86 and 2.96. Consequently, the application of operator $\tilde{\mathbf{U}}_{sou}$ on an input model is more computationally intensive than its non-extended counterpart $\mathbf{U}_{sou}$ (equation 2.80).

2.4 Tomographic operator

In this section, I analyze the tomographic modeling operator, first proposed by Biondi and Almomin (2014). This operator is closely related to the first-order differential of the Born modeling operator with respect to the background velocity map, and is necessary for the optimization scheme I develop in this thesis. Then, I derive its analytical expression, I provide some physical insight to better understand its mechanism, and I describe how I implement it numerically.

2.4.1 Motivation

To give some context on the use of this tomographic operator, I consider the following scenario where I am given a function $\mathcal{B}_r$ that maps a velocity function to a seismic recording (assuming a single source/receiver pair) as follows:

$$\begin{aligned} \mathcal{B}_r : \mathcal{F}(\Omega, \mathbb{R}^+) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) && (2.103) \\ m &\mapsto d = \mathcal{B}_r(m), && (2.104) \end{aligned}$$

and whose analytical expression is given by

$$\big(\mathcal{B}_r(m)\big)(t) = \int\limits_{\mathbf{x}\in\Omega} r(\mathbf{x})\; g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * \ddot{u}_m(\mathbf{x}, t)\; d\mathbf{x}, \tag{2.105}$$

where u_m corresponds to solution of PDE 2.1,

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 u_m(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u_m(\mathbf{x}, t) = s(\mathbf{x}, t). \tag{2.106}$$

In equation 2.105, I assume that r is a known function (referred to as the reflectivity function), independent of m, and measured in $\mathrm{s}^2/\mathrm{km}^2$. Note that in this section, r does not refer to the receiver. The application of $\mathcal{B}_r(m)$ is equivalent to applying the Born modeling operator (equation 2.63) to a scaled model perturbation r. As such, it is linear with respect to the reflectivity function r, but nonlinear with respect to the velocity map m. Therefore, I can define the tomographic operator as the first-order differential of $\mathcal{B}_r$ with respect to m, given a known reflectivity function r,

$$\mathbf{T}_r = \frac{\partial \mathcal{B}_r}{\partial m}. \tag{2.107}$$

$\mathbf{T}_r$ is the Fréchet derivative operator of $\mathcal{B}_r$, which characterizes the linear variations of the Born modeled data as a function of variations in the velocity map m (with a fixed reflectivity function). For notation compactness, I use $\mathbf{T}_{r,m}$ to refer to the tomographic operator evaluated at m. More formally, $\mathbf{T}_{r,m}$ maps a velocity-model perturbation p to a perturbation in the Born-modeled trace d:

$$\begin{aligned} \mathbf{T}_{r,m} : \mathcal{F}(\Omega, \mathbb{R}) &\mapsto \mathcal{F}(\mathcal{T}, \mathbb{R}) \\ p &\mapsto d = \mathbf{T}_{r,m} p. && (2.108) \end{aligned}$$

2.4.2 Forward modeling

To obtain the analytical expression of $\mathbf{T}_{r,m}$, I conduct a similar derivation as for the Born modeling operator. According to equation 2.105, the application of $\mathcal{B}_r$ on a velocity map m is defined as the seismic trace extracted from a field q_m at the receiver position $\mathbf{x}_r$, where q_m is the solution of the following PDE,

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 q_m(\mathbf{x},t)}{\partial t^2} - \nabla^2 q_m(\mathbf{x},t) \quad = \quad r(\mathbf{x})\ddot{u}_m(\mathbf{x},t), \tag{2.109}$$

with the same initial time boundary conditions as in equation 2.2. I now apply a perturbation p to the velocity map m, and the solution $q_{m+p} = q_m + q$ satisfies

$$\frac{1}{\big(m(\mathbf{x})+p(\mathbf{x})\big)^2}\frac{\partial^2 q_{m+p}(\mathbf{x},t)}{\partial t^2} - \nabla^2 q_{m+p}(\mathbf{x},t) \quad = \quad r(\mathbf{x})\ddot{u}_{m+p}(\mathbf{x},t), \tag{2.110}$$

where $u_{m+p} = u_m + u$ corresponds to the wavefield obtained after perturbing the velocity m by p in equation 2.106 :

$$\frac{1}{\big(m(\mathbf{x})+p(\mathbf{x})\big)^2}\frac{\partial^2 u_{m+p}(\mathbf{x},t)}{\partial t^2} - \nabla^2 u_{m+p}(\mathbf{x},t) \quad = \quad s(\mathbf{x},t). \tag{2.111}$$

By combining equations 2.109, 2.110, and 2.111, I can obtain the PDE satisfied by q,

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 q(\mathbf{x},t)}{\partial t^2} - \nabla^2 q(\mathbf{x},t) \quad = \quad r(\mathbf{x})\ddot{u}(\mathbf{x},t) + \frac{2p(\mathbf{x})}{m^3(\mathbf{x})}\ddot{q}_m(\mathbf{x},t) + \epsilon_q(\mathbf{x},t), \tag{2.112}$$

where ϵ_q includes the higher-order terms in p. Therefore, the analytical expression of q is given by

$$q(\mathbf{x},t) = \int\limits_{\mathbf{x}'\in\Omega} g_m(\mathbf{x},t,\mathbf{x}',0) * \Big[r(\mathbf{x}')\ddot{u}(\mathbf{x}',t) + \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')}\ddot{q}_m(\mathbf{x}',t)\Big]\, d\mathbf{x}' + r_{\epsilon_q}(\mathbf{x},t), \tag{2.113}$$

where $r_{\epsilon_q} = o\big(||p||_{\mathcal{P}}\big)$. Furthermore, from equations 2.109 and 2.111, u and q_m can both be expressed as a function of the source wavefield $\ddot{u}_m$:

$$u(\mathbf{x},t) = \int\limits_{\mathbf{x}'\in\Omega} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')}\, g_m(\mathbf{x},t,\mathbf{x}',0) * \ddot{u}_m(\mathbf{x}',t)\, d\mathbf{x}' + r_{\epsilon_u}(\mathbf{x},t) \tag{2.114}$$

$$q_m(\mathbf{x},t) = \int\limits_{\mathbf{x}'\in\Omega} r(\mathbf{x}') g_m(\mathbf{x},t,\mathbf{x}',0) * \ddot{u}_m(\mathbf{x}',t)\, d\mathbf{x}'. \tag{2.115}$$

where $r_{\epsilon_u} = o\big(||p||_{\mathcal{P}}\big)$. I can re-write equation 2.113 as

$$q_{m+p}(\mathbf{x},t) = q_m(\mathbf{x},t) + \int\limits_{\mathbf{x}'\in\Omega} g_m(\mathbf{x},t,\mathbf{x}',0) * \big(\ddot{s}_1(\mathbf{x}',t) + \ddot{s}_2(\mathbf{x}',t)\big)\, d\mathbf{x}' + o\big(||p||_{\mathcal{P}}\big), \tag{2.116}$$

where s_1 and s_2 correspond to two secondary source terms, linear with respect to p, and whose expressions are given by

$$s_1(\mathbf{x},t) = r(\mathbf{x}) \int\limits_{\mathbf{x}'\in\Omega} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')}\, g_m(\mathbf{x},t,\mathbf{x}',0) * \ddot{u}_m(\mathbf{x}',t)\, d\mathbf{x}' \tag{2.117}$$

$$s_2(\mathbf{x},t) = \frac{2p(\mathbf{x})}{m^3(\mathbf{x})} \int\limits_{\mathbf{x}'\in\Omega} r(\mathbf{x}') g_m(\mathbf{x},t,\mathbf{x}',0) * \ddot{u}_m(\mathbf{x}',t)\, d\mathbf{x}'. \tag{2.118}$$

I now extract the wavefields at the receiver location by applying $\mathbf{E}$ to both sides of equation 2.116 and I obtain

$$\mathcal{B}_r(m+p) = \mathcal{B}_r(m) + \mathbf{L}p + o\big(||p||_{\mathcal{P}}\big), \tag{2.119}$$

where $\mathbf{L}$ is linear with respect to p, which implies that $\mathbf{T}_{r,m} = \mathbf{L}$. Therefore, I can now obtain the expression of the tomographic modeling operator for a collection of source/receiver pairs,

$$\boxed{\big(\mathbf{T}_{r,m}p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) = \big(\mathbf{T}_1 p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) + \big(\mathbf{T}_2 p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}),} \tag{2.120}$$

where

$$\boxed{\begin{aligned}\big(\mathbf{T}_1 p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) &= \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_{r^{(j)}}, t, \mathbf{x}, 0) * r(\mathbf{x})\; u^{(1)}_{scat}(\mathbf{x}, t; s^{(i)})\; d\mathbf{x} \\ u^{(1)}_{scat}(\mathbf{x}, t; s^{(i)}) &= \int\limits_{\mathbf{x}'\in\Omega} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')}\; \ddot{g}_m(\mathbf{x}, t, \mathbf{x}', 0) * \ddot{u}_m(\mathbf{x}', t; s^{(i)})\; d\mathbf{x}'\end{aligned}} \tag{2.121}$$

and

$$\boxed{\begin{aligned}\big(\mathbf{T}_2 p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) &= \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_{r^{(j)}}, t, \mathbf{x}, 0) * \frac{2p(\mathbf{x})}{m^3(\mathbf{x})}\; u^{(2)}_{scat}(\mathbf{x}, t)\; d\mathbf{x} \\ u^{(2)}_{scat}(\mathbf{x}, t; s^{(i)}) &= \int\limits_{\mathbf{x}'\in\Omega} r(\mathbf{x}')\; \ddot{g}_m(\mathbf{x}, t, \mathbf{x}', 0) * \ddot{u}_m(\mathbf{x}', t; s^{(i)})\; d\mathbf{x}'.\end{aligned}} \tag{2.122}$$

Equation 2.120 indicates that the application of $\mathbf{T}_{r,m}$ results from the contribution of two analogous terms, $\mathbf{T}_1$ and $\mathbf{T}_2$. In $\mathbf{T}_1$, the source wavefield first interacts with the model perturbation, and then with the reflectivity. Whereas for $\mathbf{T}_2$, the order of scattering is reversed.

Furthermore, this analysis is also valid when an extended reflectivity $\tilde{r}$ is employed during the process. For a subsurface-offset extension, only equations 2.121 and 2.122 need to be modified as follows:

$$\big(\mathbf{T}_1 p\big)(t; s^{(i)}, \mathbf{x}_{r^{(j)}}) = \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_{r^{(j)}}, t, \mathbf{x}, 0) * \int\limits_{\mathbf{h}\in\tilde{\Omega}} r(\mathbf{x}, \mathbf{h})\; u^{(1)}_{scat}(\mathbf{x} - 2\mathbf{h}, t; s^{(i)})\; d\mathbf{h}\; d\mathbf{x} \tag{2.123}$$

$$u^{(2)}_{scat}(\mathbf{x}, t; s^{(i)}) = \int\limits_{\mathbf{x}'\in\Omega} \ddot{g}_m(\mathbf{x}, t, \mathbf{x}', 0) * \int\limits_{\mathbf{h}\in\tilde{\Omega}} r(\mathbf{x}', \mathbf{h})\; \ddot{u}_m(\mathbf{x}' - 2\mathbf{h}, t; s^{(i)})\; d\mathbf{x}'. \tag{2.124}$$

The expression of $\mathbf{T}_1$ and $\mathbf{T}_2$ for the time-lag extension can be obtained in a similar manner and are not shown here.

Geophysical interpretation of the tomographic forward operator

The application of $\mathbf{T}_1$ is illustrated in Figure 2.11, where I embed a reflectivity r (brown dot) and a velocity perturbation p (green dot) in a constant-velocity medium. A source wavefield is propagated into the medium with velocity m. Then, the source wavefield interacts with the model perturbation p (Figure 2.11b), and a first scattered wavefield $u^{(1)}_{scat}$ is generated and propagated. $u^{(1)}_{scat}$ interacts with the reflectivity r (Figure 2.11c), which in turn generates a second scattered wavefield $u^{(1)}_{sec}$. Finally, the second scattered wavefield is propagated and a data perturbation is recorded at the

receiver position (Figure 2.11d).

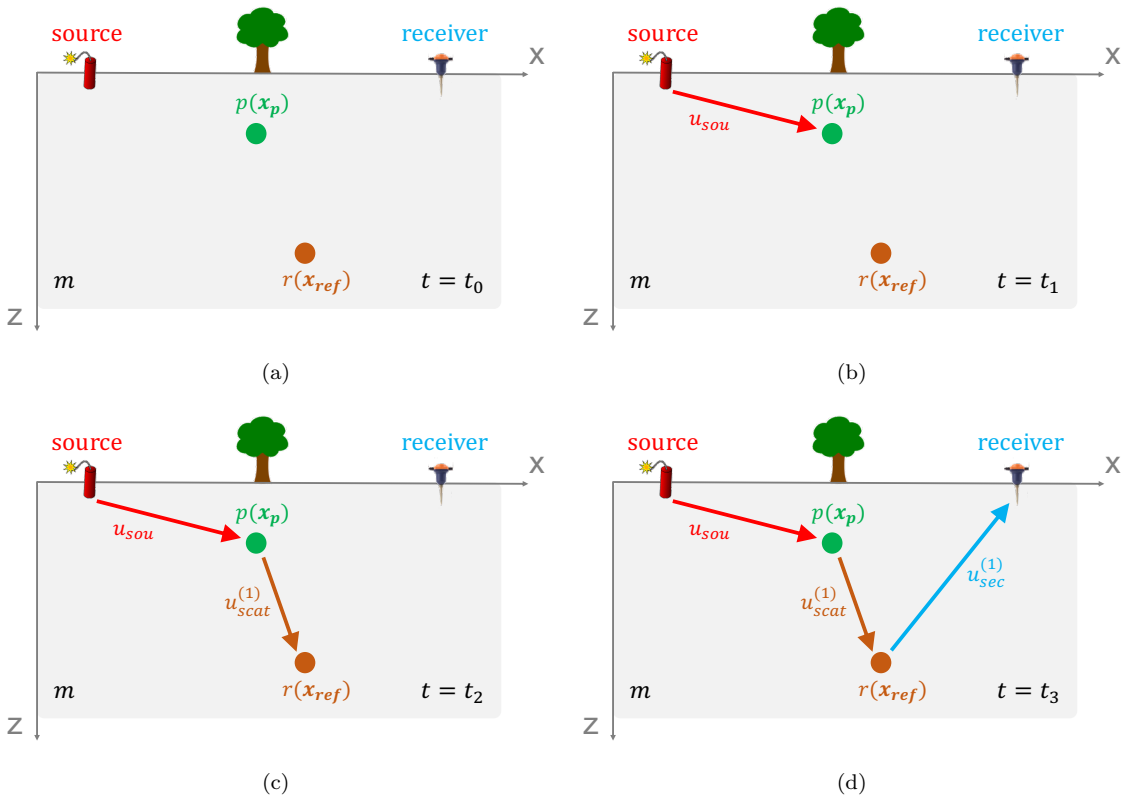


Figure 2.11: Schematic illustrations representing the application of $\mathbf{T}_1$ (equation 2.121) on a model perturbation p (green dot). The reflectivity (brown dot) is assumed to be known and independent of the velocity map m. The red, brown and blue arrows represent the source and scattered wavefields. (a) Initial state. (b) Snapshot of the source wavefield at $t = t_1$ as it interacts with the model perturbation. (c) Snapshot of the first scattered wavefield $u^{(1)}_{scat}$ at $t = t_2$ as it interacts with the reflectivity. (d) Snapshot of the second scattered wavefield $u^{(1)}_{sec}$ at $t = t_3$ as it reaches the receiver position. [**NR**]

A similar analysis can be conducted for $\mathbf{T}_2$ as shown in Figure 2.12. The source wavefield first interacts with the reflectivity r (Figure 2.12b), creates a scattered wavefield $u^{(2)}_{scat}$ which in turn interacts with the model perturbation p (Figure 2.12c). Finally, the second scattered wavefield $u^{(2)}_{sec}$ reaches the receiver location and a data perturbation is recorded (Figure 2.12d).

2.4.3 Adjoint modeling

The expression of the adjoint operator $\mathbf{T}^*_{r,m}$ for a single source/receiver pair is given by

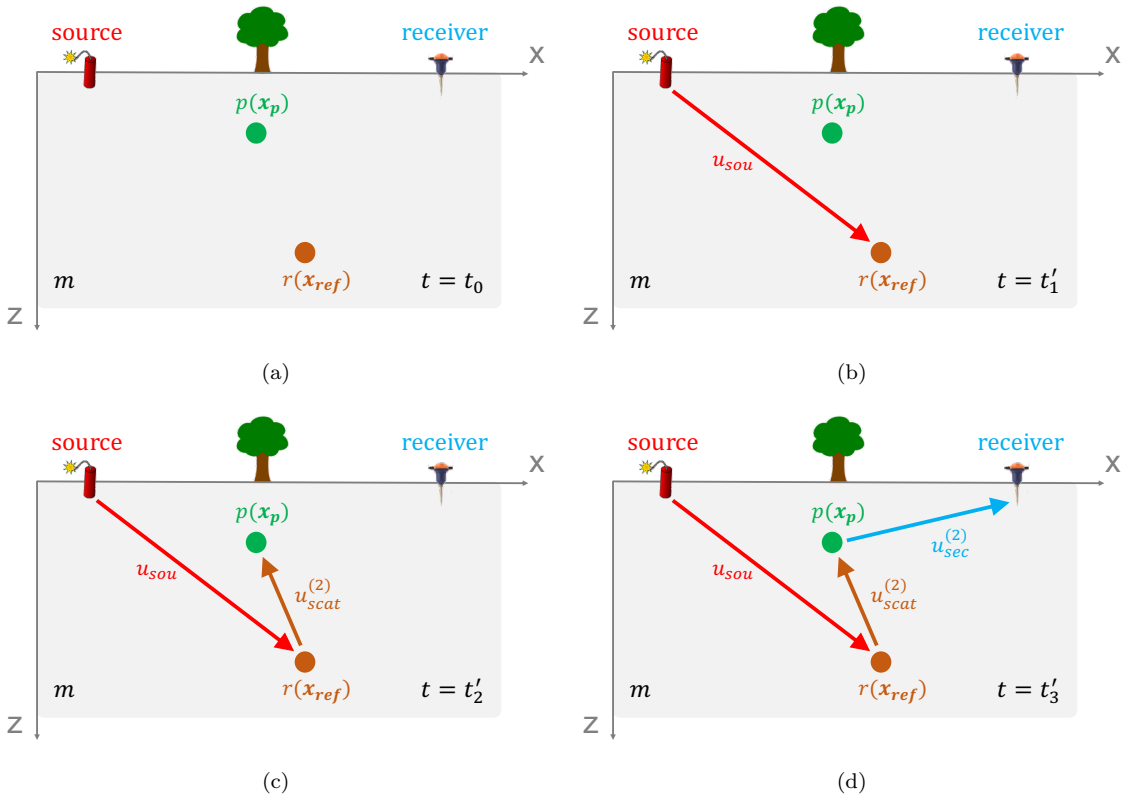


Figure 2.12: Schematic illustrations representing the application of $\mathbf{T}_2$ (equation 2.122) on a model perturbation p (green dot). The reflectivity (brown dot) is assumed to be known and independent of the uniform velocity map m. The red, brown and blue arrows represent the source and scattered wavefields. (a) Initial state. (b) Snapshot of the source wavefield at $t = t_1'$ as it interacts with the reflectivity. (c) Snapshot of the first scattered wavefield $u_{scat}^{(2)}$ at $t = t_2'$ as it interacts with the model perturbation. (d) Snapshot of the second scattered wavefield $u_{sec}^{(2)}$ at $t = t_3'$ as it reaches the receiver position. [**NR**]

$$\big(\mathbf{T}^*_{r,m}d\big)(\mathbf{x}) \quad = \quad \big(\mathbf{T}^*_1 d\big)(\mathbf{x}) + \big(\mathbf{T}^*_2 d\big)(\mathbf{x}) \tag{2.125}$$

I first derive of the analytical expression of $\mathbf{T}^*_1$. By definition, $\mathbf{T}_1$ and $\mathbf{T}^*_1$ must satisfy

$$< \mathbf{T}_1 p, d >_{\mathcal{D}} \quad = \quad < p, \mathbf{T}^*_1 d >_{\mathcal{P}}, \tag{2.126}$$

for all $p \in \mathcal{P}$ and $d \in \mathcal{D}$. The left side of equation 2.126 can be expressed by

$$< \mathbf{T}_1 p, d >_{\mathcal{D}} \quad = \quad \int\limits_{t\in\mathcal{T}} d(t) \int\limits_{\mathbf{x}\in\Omega} g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * r(\mathbf{x})\; u^{(1)}_{scat}(\mathbf{x}, t)\; d\mathbf{x}\; dt \tag{2.127}$$

By replacing $u^{(1)}_{scat}$ by its expression shown in equation 2.121, I can write

$$\begin{aligned}
< \mathbf{T}_1 p, d >_{\mathcal{D}} \quad &= \quad \int\limits_{t,\mathbf{x},\mathbf{x}'} d(t)\; g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * r(\mathbf{x}) \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')}\; \ddot{g}_m(\mathbf{x}, t, \mathbf{x}', 0) * \ddot{u}_m(\mathbf{x}', t)\; d\mathbf{x}'\; d\mathbf{x}\; dt \\
&= \quad \int\limits_{\mathbf{x}'} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')} \int\limits_{\mathbf{x},t} d(t)\; g_m(\mathbf{x}_r, t, \mathbf{x}, 0) * r(\mathbf{x})\; \ddot{g}_m(\mathbf{x}, t, \mathbf{x}', 0) * \ddot{u}_m(\mathbf{x}', t)\; dt\; d\mathbf{x}\; d\mathbf{x}' \\
&= \quad \int\limits_{\mathbf{x}'} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')} \int\limits_{\mathbf{x},t} r(\mathbf{x})\; g_m(\mathbf{x}, t, \mathbf{x}', 0) * \ddot{u}_m(\mathbf{x}', t)\; g_m(\mathbf{x}, 0, \mathbf{x}_r, t) * \ddot{d}(t)\; dt\; d\mathbf{x}\; d\mathbf{x}' \\
&= \quad \int\limits_{\mathbf{x}'} \frac{2p(\mathbf{x}')}{m^3(\mathbf{x}')} \int\limits_{\mathbf{x},t} r(\mathbf{x})\; \ddot{u}_m(\mathbf{x}', t)\; g_m(\mathbf{x}', 0, \mathbf{x}, t) * g_m(\mathbf{x}, 0, \mathbf{x}_r, t) * \ddot{d}(t)\; dt\; d\mathbf{x}\; d\mathbf{x}' \\
&= \quad \int\limits_{\mathbf{x}'} p(\mathbf{x}') \int\limits_{t} \frac{2\ddot{u}_m(\mathbf{x}', t)}{m^3(\mathbf{x}')} \int\limits_{\mathbf{x}} r(\mathbf{x}) g_m(\mathbf{x}', 0, \mathbf{x}, t) * g_m(\mathbf{x}, 0, \mathbf{x}_r, t) * \ddot{d}(t)\; d\mathbf{x}\; dt\; d\mathbf{x}
\end{aligned}$$

The right side of equation 2.126 can be expressed as

$$< p, \mathbf{T}^*_1 d >_{\mathcal{P}} \quad = \quad \int\limits_{\mathbf{x}\in\Omega} p(\mathbf{x}) \big(\mathbf{T}^*_1 d\big)(\mathbf{x})\; d\mathbf{x}. \tag{2.128}$$

Therefore, by using the uniqueness property of the adjoint operator, I obtain the expression of the application of $\mathbf{T}^*_1$ on a dataset composed of a collection of source/receiver pairs,

$$\big(\mathbf{T}_1^* d\big)(\mathbf{x}) \;=\; \sum_{i=1}^{N_s} \int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})} \ddot{u}_m(\mathbf{x},t;s^{(i)})\; u_{adj}^{(1)}(\mathbf{x},t;s^{(i)})\; dt, \tag{2.129}$$

where $u_{adj}^{(1)}$ is defined by:

$$u_{adj}^{(1)}(\mathbf{x},t;s^{(i)}) \;=\; \int\limits_{\mathbf{x}'\in\Omega} r(\mathbf{x}')\; g_m(\mathbf{x},0,\mathbf{x}',t) * \ddot{u}_{rec}(\mathbf{x}',t;s^{(i)})\; d\mathbf{x}', \tag{2.130}$$

and referred to as the adjoint scattered wavefield (u_{rec} is defined in equation 2.74). Therefore, equation 2.129 corresponds to the zero time-lag cross-correlation of the scaled source wavefield with $u_{adj}^{(1)}$. Following a similar derivation, the application of $\mathbf{T}_2^*$ can be written as

$$\big(\mathbf{T}_2^* d\big)(\mathbf{x}) \;=\; \sum_{i=1}^{N_s} \int\limits_{t\in\mathcal{T}} \frac{2}{m^3(\mathbf{x})} u_{scat}^{(2)}(\mathbf{x},t;s^{(i)})\; \ddot{u}_{rec}(\mathbf{x},t;s^{(i)})\; dt, \tag{2.131}$$

where $u_{scat}^{(2)}$ is defined in equations 2.122. In summary,

$$\boxed{\begin{aligned} \big(\mathbf{T}_1^* d\big)(\mathbf{x}) &= \frac{2}{m^3(\mathbf{x})} \sum_{i=1}^{N_s} \ddot{u}_m(\mathbf{x},t;s^{(i)}) \otimes u_{adj}^{(1)}(\mathbf{x},t;s^{(i)}) \\ \big(\mathbf{T}_2^* d\big)(\mathbf{x}) &= \frac{2}{m^3(\mathbf{x})} \sum_{i=1}^{N_s} u_{scat}^{(2)}(\mathbf{x},t;s^{(i)}) \otimes \ddot{u}_{rec}(\mathbf{x},t;s^{(i)}). \end{aligned}} \tag{2.132}$$

Geophysical interpretation of the tomographic adjoint operator

I illustrate the application of $\mathbf{T}_1^*$ in Figure 2.13, and I describe its mechanism in reverse-chronological order. According to equation 2.130, the second-order time-derivative of the receiver wavefield $\ddot{u}_{rec}$ is propagated backward in time, and interacts with the reflectivity at $t = t_2$ (Figure 2.13b). An adjoint scattered wavefield is generated $u_{adj}^{(1)}$, also propagated backward in time. Finally, at $t = t_1$, the source and adjoint scattered wavefields coincide in time and space (Figure 2.13c), which produces a model perturbation p, as shown in Figure 2.13d.

I describe the application of $\mathbf{T}_2^*$ in chronological order, as shown in Figure 2.14. The source wavefield interacts with the reflectivity at $t = t_1'$ (Figure 2.14a), and a scattered wavefield $u_{scat}^{(2)}$ is generated and propagated (Figure 2.14b). At $t = t_2'$, $u_{scat}^{(2)}$ coincides in time and space with the receiver wavefield $\ddot{u}_{rec}$, which produces a model perturbation p (Figure 2.14c). At greater propagation time, $\ddot{u}_{rec}$ reaches the receiver position from where it was generated (Figure 2.14d).

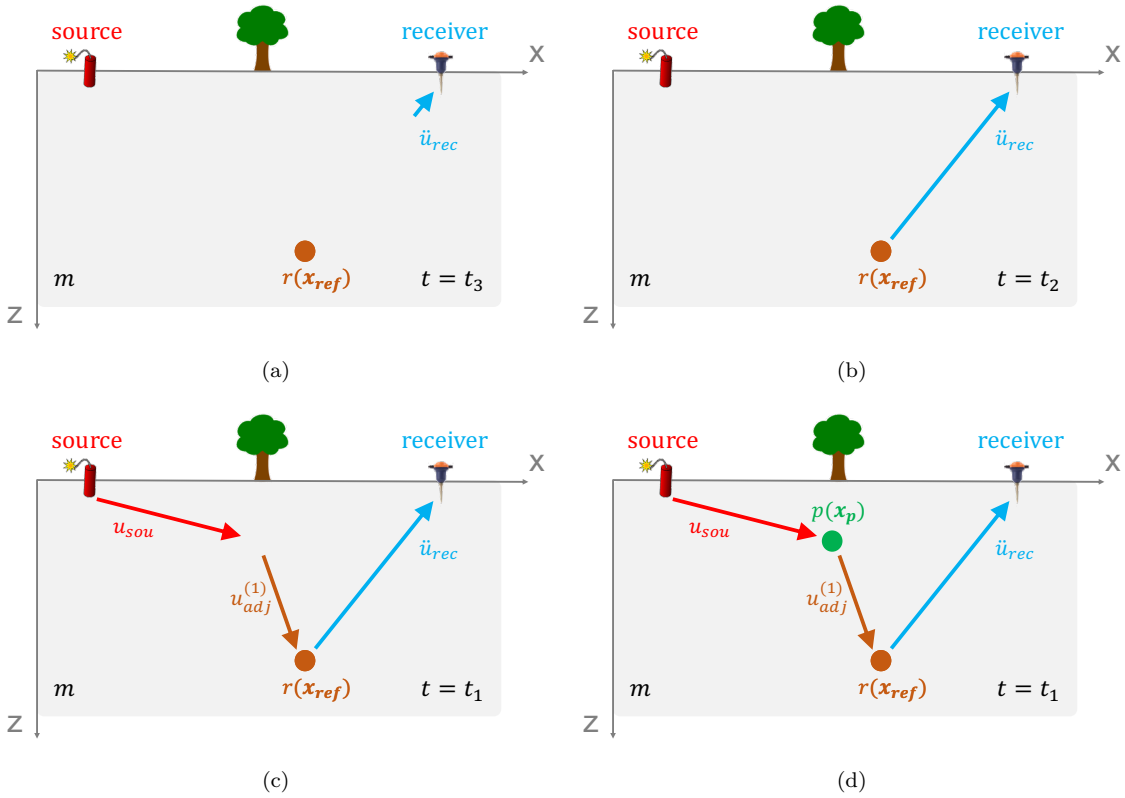


Figure 2.13: Schematic illustrations (in reverse-chronological order) representing the application of $\mathbf{T}_1^*$ (equation 2.129) on a seismic trace d. The reflectivity (brown dot) is assumed to be known and independent of the uniform velocity map m. The red, brown and blue arrows represent the source, scattered, and receiver wavefields, respectively. (a) Snapshot of the receiver wavefield at late propagation time $t = t_3$. (b) Snapshot of the receiver wavefield as it reaches (and interacts) with the reflectivity. (c) Snapshot of the adjoint scattered wavefield $u^{(1)}_{adj}$ at $t = t_1$, which is propagated backward in time. (d) Snapshot of the source and adjoint scattered wavefields as they coincide in time and space, thereby producing a model perturbation p (green dot). [**NR**]

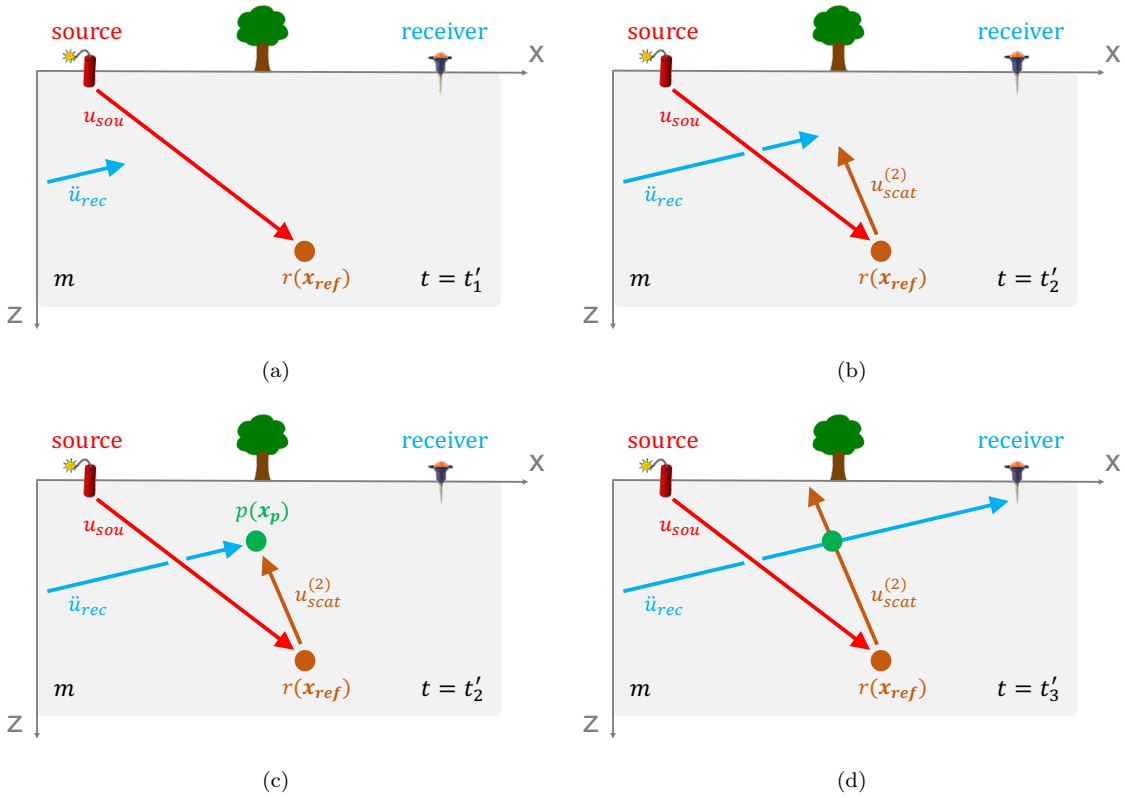


Figure 2.14: Schematic illustrations (in chronological order) representing the application of $\mathbf{T}_2^*$ (equation 2.131) on a seismic trace d. The reflectivity (brown dot) is assumed to be known and independent of the uniform velocity map m. The red, brown and blue arrows represent the source, scattered, and receiver wavefields, respectively. (a) Snapshots of the receiver wavefield and source wavefield as it reaches and interact with the reflectivity at $t = t_1'$. (b) Snapshots of the scattered $u_{scat}^{(2)}$ and receiver wavefields. (c) At $t = t_2'$, both wavefields coincide in time and space, which produces a model perturbation (green dot). (d) Snapshot of the receiver and scattered wavefields at a greater propagation time $t = t_3'$ [**NR**]

2.4.4 Numerical implementation

The discretization and numerical implementation of the tomographic forward and adjoint operators is conducted by expressing them as a composition of linear operators as in equation 2.79. Recall that for a single source/receiver pair, the application of $\mathbf{T_{m,r}}$ on a model perturbation is defined as

$$\mathbf{d} \quad = \quad \mathbf{T_{m,r}}\ \mathbf{p}, \tag{2.133}$$

where $\mathbf{p} \in \mathbb{R}^{N_m}$, $\mathbf{T_{m,r}} \in \mathcal{M}^{N_t \times N_m}$, and $\mathbf{d} \in \mathbb{R}^{N_t}$. Note that $\mathbf{r} \in \mathbb{R}^{N_m}$ in equation 2.133 does not refer a receiver but rather to the discretized reflectivity function.

I first find the expression of the term $\mathbf{T}_1$. According to equation 2.130,

$$\mathbf{u}^{(1)}_{scat} \quad = \quad \mathbf{P_m}\ \mathbf{U}_{sou}\ \mathbf{p}, \tag{2.134}$$

where $\mathbf{u}^{(1)}_{scat} \in \mathbf{R}^{N_w}$, $\mathbf{P_m} \in \mathcal{M}^{N_w \times N_w}$ is the propagation operator (equation 2.40), and $\mathbf{U}_{sou} \in \mathcal{M}^{N_w \times N_m}$ is a scaling operator that multiplies the time slices of the source wavefield by the values of the perturbation $\mathbf{p}$ (as shown in equation 2.80). Futhermore, equation 2.130 implies that

$$\mathbf{d} \quad = \quad \mathbf{F_m}\ \mathbf{R}\ \mathbf{u}^{(1)}_{scat}, \tag{2.135}$$

where $\mathbf{R} \in \mathcal{M}^{N_w \times N_w}$ is given by

$$\mathbf{R} = \begin{bmatrix} \mathbf{r} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{r} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{r} \end{bmatrix}, \tag{2.136}$$

$\mathbf{R}$ is block-diagonal scaling operator that multiplies the time slices of $\mathbf{u}^{(1)}_{scat}$ by the reflectivity $\mathbf{r}$. However, if an extended reflectivity $\tilde{\mathbf{r}}$ is employed, $\tilde{\mathbf{R}} \in \mathcal{M}^{N_w \times N_w}$ is not necessarily diagonal and its application corresponds to performing a convolution (over the extended variable) rather than a simple multiplication. Finally, I can now express $\mathbf{T}_1$ as the following sequence of operators:

$$\mathbf{T}_1 \quad = \quad \mathbf{F_m}\ \mathbf{R}\ \mathbf{P_m}\ \mathbf{U}_{sou}. \tag{2.137}$$

A similar approach can be used for the expression of $\mathbf{T}_2$. Equation 2.122 indicates that

$$\mathbf{d} \quad = \quad \mathbf{F_m}\ \mathbf{U}^{(2)}_{scat}\ \mathbf{p}, \tag{2.138}$$

where $\mathbf{U}^{(2)}_{scat} \in \mathcal{M}^{N_w \times N_m}$ is given by

$$\mathbf{U}^{(2)}_{scat} = \begin{bmatrix} \mathbf{u}^{(2)}_{scat}[1] \\ \vdots \\ \mathbf{u}^{(2)}_{scat}[N_t] \end{bmatrix}, \tag{2.139}$$

$\mathbf{u}^{(2)}_{scat}[i] \in \mathcal{M}^{N_m}$ is a diagonal matrix that multiplies the input by the values of the i^{th} time slice of wavefield $\mathbf{u}^{(2)}_{scat}$. From equation 2.122, I can write

$$\mathbf{u}^{(2)}_{scat} \quad = \quad \mathbf{P_m}\ \mathbf{U}_{sou}\ \mathbf{r}, \tag{2.140}$$

and for an extended reflectivity $\tilde{\mathbf{r}}$, equation 2.140 becomes,

$$\mathbf{u}^{(2)}_{scat} \quad = \quad \mathbf{P_m}\ \tilde{\mathbf{U}}_{sou}\ \tilde{\mathbf{r}}. \tag{2.141}$$

with the same definition of $\tilde{\mathbf{U}}_{sou}$ as in equation 2.102. Finally, $\mathbf{T_{m,r}}$ and its adjoint operator can be expressed as

$$\mathbf{T_{m,r}} \quad = \quad \mathbf{F_m}\ \left[\ \mathbf{R}\ \mathbf{P_m}\ \mathbf{U}_{sou} + \ \mathbf{U}^{(2)}_{scat}\right] \tag{2.142}$$

$$\mathbf{T}^*_{\mathbf{m,r}} \quad = \quad \left[\mathbf{U}^{(2)\,*}_{scat} + \mathbf{U}^*_{sou}\ \mathbf{P}^*_{\mathbf{m}}\ \mathbf{R}^*\right]\ \mathbf{F}^*_{\mathbf{m}} \tag{2.143}$$

The computation of each term $\mathbf{T}_1$ and $\mathbf{T}_2$ requires the storage of two wavefields, (1) the source wavefield $\mathbf{u}_{sou}$ and (2) a scattered wavefield ($\mathbf{u}^{(1)}_{scat}$ for $\mathbf{T}_1$, and $\mathbf{u}^{(2)}_{scat}$ for $\mathbf{T}_2$). For the adjoint process, both terms require the storage of the source and receiver wavefields.

Chapter 3

Theory and design

In this chapter, I present the theoretical framework for FWIME by describing and analyzing the objective function minimized during the optimization process. I show that it is the combination of modified forms of two well-known advanced seismic velocity model building techniques referred to as wave-equation migration velocity analysis (WEMVA) and acoustic full waveform inversion (FWI). By leveraging the robustness of the former with the accuracy and high-resolution nature of the latter, I combine these methods into a consistent and concise mathematical formulation paired with an automatized implementation, which makes the method simple to apply and thus accessible to a broad range of non-expert users. I begin by providing some motivation for my work and I explain the intuition that led me to design FWIME. Then, I present, analyze and dissect the four main building blocks of the FWIME objective function. They include a data-correcting term, an optimal extended perturbation, an annihilating term, and a trade-off parameter (which is the algorithm's unique important adjustable hyper-parameter).

3.1 Motivation

3.1.1 Conventional imaging

The oil and gas industry's standard imaging process is based on a separation of model scales and is typically conducted in three sequential stages (Claerbout, 1985; Biondi and Almomin, 2014). Figure 3.1 shows a schematic diagram representing the spatial resolution of the different features that each step of the conventional imaging workflow is able to recover.

First, tomographic techniques such as migration velocity analysis (MVA) are employed to retrieve low-resolution velocity models of the subsurface (Symes and Kern, 1994; Sava and Biondi, 2004; Yang, 2013; Zhang, 2015; Díaz Pantin, 2016). Such algorithms have heuristically been found to be more immune to initial velocity errors (Stolk and Symes, 2002), as illustrated by the light-blue curve

in Figure 3.1.

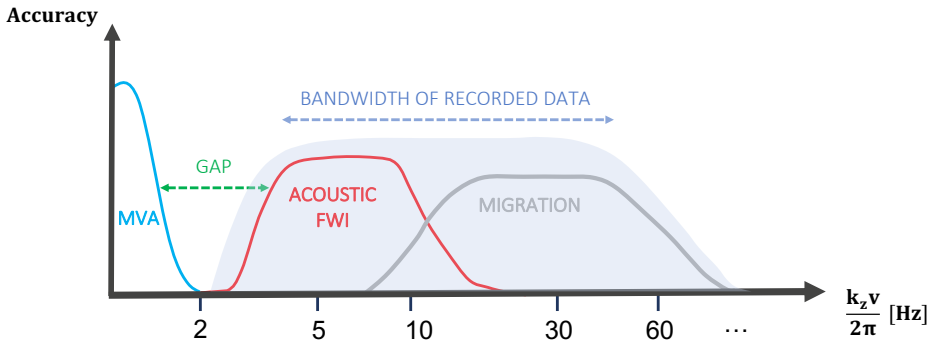


Figure 3.1: 1D schematic diagram showing the accuracy of recovered velocity models for the three conventional imaging steps, plotted as a function of spatial frequency (adapted from Claerbout (1985) and Biondi and Almomin (2014)). The spatial frequency axis is plotted in log scale. [**NR**]

The second step consists in using the output from conventional tomography to recover higher-resolution models by applying acoustic full waveform inversion (FWI), first proposed by Lailly and Bednar (1983); Tarantola (1984) (red curve in Figure 3.1). In the past decade, successful applications of FWI on 3D field data have demonstrated the method's ability at producing accurate and useful solutions (Sirgue et al., 2010; Baeten et al., 2013; Shen et al., 2018). However, this method suffers from two major drawbacks.

On one hand, its success is contingent on already having an accurate initial model in relation with the frequency content of the recorded data, and failing to satisfy this requirement may lead FWI to converge to non-physical solutions, which correspond to spurious local minima present in the loss function that is being minimized (Virieux and Operto, 2009). This problem can potentially be mitigated with the presence of coherent low-frequency long-offset energy within the recorded data. Moreover, gradually increasing the bandwidth of the inverted signal may drive the solution to the correct minimum (Bunks et al., 1995; Brenders and Pratt, 2007; Fichtner, 2010). Unfortunately, such type of data can be challenging and costly to acquire (Dellinger et al., 2016; Brenders et al., 2018; Bate and Mike Perz, 2021), and may therefore not be available, especially in companies' legacy datasets (Warner and Guasch, 2016; Brittan and Jones, 2019).

On the higher end of the spatial frequency spectrum, FWI's resolution is still limited by its computational cost that increases exponentially with the maximum frequency content in the data. Because of the computational cost at high frequencies, the acoustic approximation is commonly made for field-data applications, which may prevent the method from recovering certain elastic properties of the subsurface. In this thesis, in order to simplify notations, I use the FWI acronym to specifically refer to acoustic isotropic constant-density FWI.

In the last step of the imaging sequence, higher-resolution seismic images depicting the various interfaces between rock layers are obtained by applying a seismic migration algorithm (gray curve in Figure 3.1), such as reverse-time migration (RTM) (Baysal et al., 1983), although high-frequency FWI might become an attractive alternative (Lazaratos et al., 2011; Routh et al., 2011). These images can provide useful information about the subsurface geology and can be further processed/interpreted by geophysicists to identify and characterize potential hydrocarbon reservoirs (Buland and Omre, 2003; Grana and Della Rossa, 2010). The quality of such images heavily depends on the of the velocity macro-model provided by the FWI step.

Figure 3.2 shows a successful application of the conventional sequential imaging process on the Marmousi2 model, a well-known challenging 2D synthetic benchmark test for velocity estimation methods (Martin et al., 2006). The initial velocity model (Figure 3.2a) is laterally invariant, and its value is linearly increasing as a function of depth. After applying WEMVA, a low-resolution model is recovered (Figure 3.2b), and then fed into a conventional FWI scheme. The final inverted model shown in Figure 3.2c is very accurate and almost identical to the true model (Figure 3.2d). In this example, the noise-free pressure data are generated and inverted with the same acoustic isotropic two-way wave-equation engine. This scenario does not fully encompass all the challenges encountered with field-data applications, but it illustrates the potential of FWI at recovering excellent solutions.

3.1.2 Limitations of existing techniques

As first pointed out by Claerbout (1985), an information gap between the low and high wavenumbers of the subsurface model is present when dealing with the sequential velocity estimation workflow using surface data (dashed green arrow in Figure 3.1). This phenomenon remains the main bottleneck of the imaging process and results from the fact that conventional tomographic techniques are sometimes not able to produce accurate and high-resolution enough initial models for conventional FWI to recover a satisfactory solution.

Figure 3.3 illustrates this limitation using the same Marmousi2 numerical example as the one in section 3.1.1. However, the initial guess is modified and purposely chosen to be less accurate (Figure 3.3a). After conducting a similar workflow, the final FWI inverted model fails at recovering the main features of the true solution (Figure 3.3b).

The importance of having a good velocity model for the imaging process can be better understood by examining the output of the migration step. Figure 3.4a shows the migrated image computed with the velocity model resulting from the successful FWI scenario (Figure 3.2c). The image accurately maps the interfaces between rocks layers and can be easily interpreted by a team of geophysicists to identify potential hydrocarbon-bearing reservoirs (black and red boxes). On the other hand, a non-physical velocity model such as the one displayed in Figure 3.3b will produce a deteriorated, incoherent, and unreliable map of the subsurface (Figure 3.4b).

To bridge information this gap and create an ultimate inversion technique, multiple methods

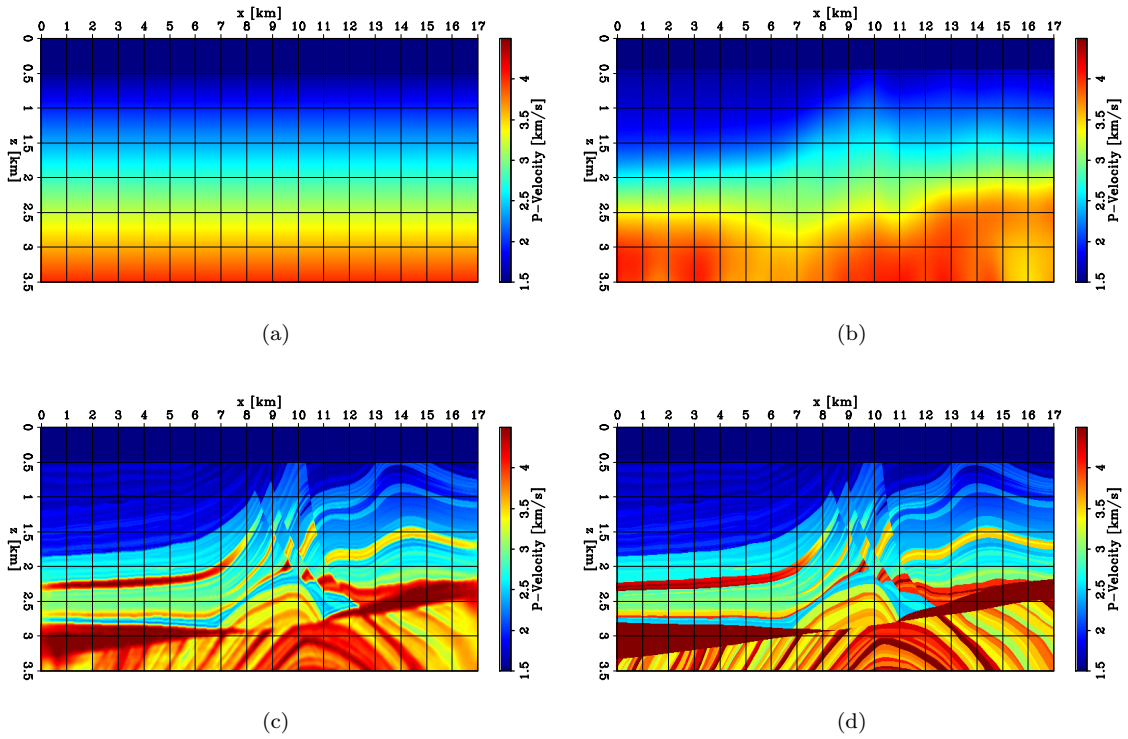


Figure 3.2: Sequence of inverted models throughout the conventional velocity estimation process. (a) Initial model. (b) Inverted model after applying WEMVA. (c) FWI inverted model using (b) as an initial guess. (d) True model. [**CR**]

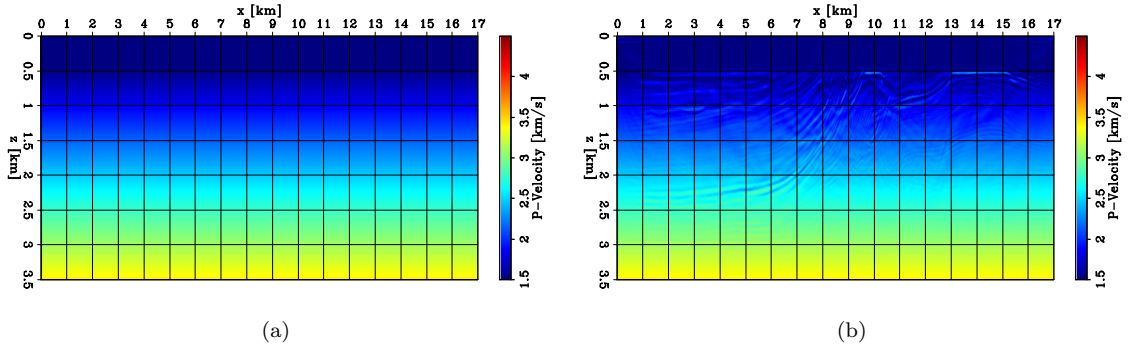
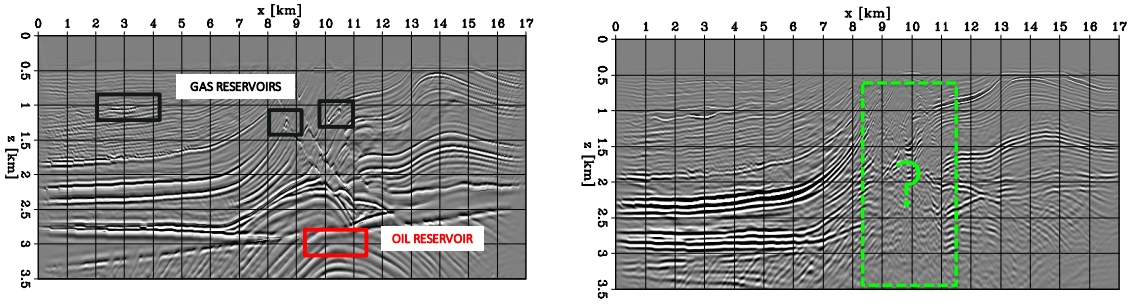


Figure 3.3: 2D panels of velocity models obtained throughout the conventional imaging process. (a) Initial model. (b) Inverted model after applying WEMVA, followed by FWI. [**CR**]



(a) (b)

Figure 3.4: Annotated 2D panels of RTM migrated images computed with (a) the accurate velocity model shown in Figure 3.2c, and (b) the inaccurate velocity model shown in Figure 3.3b. [**NR**]

have been proposed where the conventional FWI problem is modified by either directly extending the unknown model search space (Symes, 2008; Fleury and Perrone, 2012; Biondi and Almomin, 2014; Huang and Symes, 2015; Huang et al., 2017; Barnier and Biondi, 2020; Métivier and Brossier, 2021), by relaxing certain constraints related to the physics of the problem (Van Leeuwen and Herrmann, 2013; van Leeuwen et al., 2014; Warner and Guasch, 2014, 2016; Guasch et al., 2019; Aghamiry et al., 2019), or by measuring the data misfit with a different norm (Métivier et al., 2016, 2018). From an optimization standpoint, the common goal behind all of these approaches is to design alternate and more convex objective functions that share the same global minimum as FWI.

To my knowledge, Symes (2008) was the first author to introduce the idea of combining the robustness of MVA approaches with the accuracy of FWI into one workflow by using a concept referred to as extended modeling. Following this idea, many authors have proposed different methods to efficiently implement this concept. Biondi and Almomin (2014) developed a technique referred to as tomographic full waveform inversion and showed very promising results. However, the proposed algorithm, based on a nested-scheme approach, requires the user to tune many hyper-parameters inherent to their inverse problem formulation. Huang and Symes (2015) showed an interesting solution using the variable projection method (Golub and Pereyra, 1973; Rickett, 2013), but the workflow was still based on an explicit model-scale separation between low- and high-wavenumber components, preventing it from fully bridging the information gap previously described.

3.1.3 Proposed solution

In this thesis, I build upon the work of the aforementioned authors and I propose a new method, referred to as full waveform inversion by model extension (FWIME) that successfully pairs both WEMVA and FWI techniques into one robust and mathematically consistent workflow. The main novelties and contributions of FWIME reside in the design of a new loss function and the optimization

strategy I devise to merge WEMVA and FWI, which results in a more efficient and powerful technique than simply applying them separately and sequentially. In this thesis, I do not mathematically prove that my objective function is more suited for gradient-descent optimization (i.e., free of local minima), but I provide strong numerical evidence to support this claim.

The new objective function contains two components. In the first component, I modify the conventional FWI forward mapping operator (i.e., the acoustic isotropic constant-density finite-difference modeling operator) by adding a data-correcting term coming from the linear mapping of an extended model perturbation into the data space such that the observed data is fully fitted at a given precision. This linear mapping, referred to as extended modeling, is the most important tool of my method: it allows me to linearly predict any event in the data space regardless of the accuracy of the initial velocity model, thereby creating a non-physical forward mapping that always ensures wiggle matching between predicted and observed data.

I add a second component that allows me to progressively remove the contributions of the data-correcting term throughout the inversion process by eliminating all the energy present within the extended model perturbation. This additional component, referred to as the FWIME annihilator, possesses similar beneficial convergence properties as the conventional WEMVA objective function. Moreover, I make use of the variable projection method, which provides three advantages. First, it gives me better control on the phase alignment between predicted and observed data. It also allows me to avoid separating the unknown velocity model parameters into a background (low-wavenumber component) and a reflectivity (high-wavenumber perturbation). Finally, it handles the coupling between the two components of the objective function in an automatic and mathematically consistent fashion, thereby reducing the number of optimization hyper-parameters to only one.

Because of the high computational cost of FWIME, my goal is not to produce models with resolutions as high as the ones obtained with conventional FWI. Instead, I see two practical applications of my algorithm. On one hand, the output of FWIME can be used as an input for conventional acoustic FWI, followed by a standard migration process (Figure 3.5). Alternatively, if a specific hydrocarbon-bearing region within the subsurface has been identified and needs to be further characterized, the FWIME inverted model can be fed directly into a target-oriented elastic full waveform inversion (EFWI) approach, such as the one proposed by Biondi and Barnier (2020) (represented in Figure 3.6). Even though more computationally intensive than its acoustic counterpart, EFWI becomes affordable when applied to smaller target volumes and has the potential to retrieve elastic and petrophysical properties of hydrocarbon reservoirs with more accuracy and robustness than standard ray-based methods (Biondi et al., 2019, 2018; Biondi, 2021).

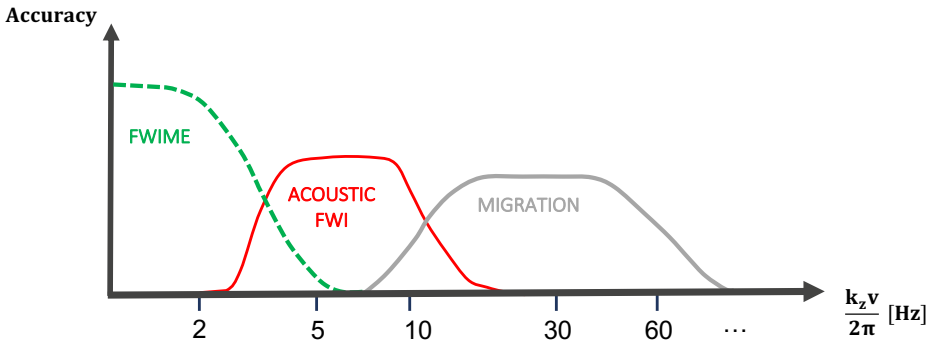


Figure 3.5: 1D schematic diagram showing the accuracy of recovered velocity model as a function of spaial frequency for the first proposed FWIME workflow. FWIME is used to recover an accurate initial model (dashed green curve). A conventional acoustic FWI scheme is then applied to improve the model resolution (red curve). Finally, a standard migration algorithm is conducted to create an accurate map of the subsurface (gray curve). The spatial frequency axis is plotted in log scale. [**NR**]

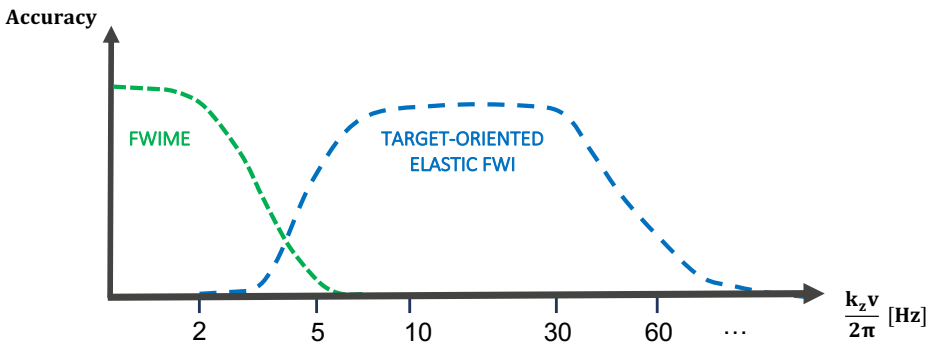


Figure 3.6: 1D schematic diagram showing the accuracy of recovered velocity model as a function of spatial frequency for the second proposed application of FWIME in a production workflow. FWIME is first applied to recover an accurate acoustic model (dashed green curve) which is then directly fed into a target-oriented EFWI (dashed blue curve). The spatial frequency axis is plotted in log scale. [**NR**]

3.2 FWIME formulation

I begin by describing the theoretical framework of FWI and WEMVA, which are the two main algorithms on which FWIME is based. Then, I formally define its objective function and I give some intuition on the design of this method.

3.2.1 Full waveform inversion

In the conventional FWI workflow, the following objective function is minimized

$$\Phi_{\mathrm{FWI}}(\mathbf{m}) = \frac{1}{2} \left\| \mathbf{f}(\mathbf{m}) - \mathbf{d}^{obs} \right\|_2^2 , \tag{3.1}$$

where $\mathbf{m} \in \mathbb{R}^{N_m}$ is the unknown seismic velocity model (discretized and parametrized on a finite-difference grid), $\mathbf{f} : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_d}$ is the discretized acoustic isotropic constant-density forward modeling operator conducted for a collection of fixed source/receiver pairs (defined in subsection 2.1.2), and $\mathbf{d}^{obs} \in \mathbb{R}^{N_d}$ is the seismic data recorded at a set of receivers' locations. $N_m = N_z \times N_x \times N_y$ is the dimension of the model space, where N_x, N_y, and N_z are the number of grid points in each spatial direction. $N_d = N_t \times N_{tr}$ is the dimension of the data space, where N_t is the number of time samples per trace, and N_{tr} is the number of traces in the dataset.

In the following, I assume the existence and uniqueness of a true model $\mathbf{m}_t$ such that $\mathbf{f}(\mathbf{m}_t) = \mathbf{d}^{obs}$, which implies that equation 3.1 possesses a unique global minimum. Due to the large problem scales encountered in 3D field applications, the minimization of equation 3.1 is commonly performed using gradient-based optimization methods. Since $\mathbf{f}$ is nonlinear with respect to $\mathbf{m}$, Φ_{FWI} may bear multiple local minima.

When the initial model $\mathbf{m}_0$ is close enough to the true model $\mathbf{m}_t$, it may belong to the basin of attraction of objective defined in equation 3.1 whose minimum coincides with the global minimum $\mathbf{m}_t$, in which case a gradient-based optimization scheme is sufficient to recover the global minimizer of Φ_{FWI}. Unfortunately, when the initial model is too inaccurate for providing a decent wave propagation simulation (and thus does not lie within the global basin of attraction), a gradient-based optimization scheme may converge to a local minimum corresponding to a non-physical and uninformative representation of the Earth's subsurface, such as the one shown in Figure 3.3b. In the data space, some of the events in the predicted data may initially be shifted by more than half of one cycle from their counterpart in the recorded data, which is referred to as "cycle-skipping". Their cross-correlation contribution to the model update will be zero or in the wrong direction (Virieux and Operto, 2009; Yao et al., 2019). An even worse situation occurs when some predicted data attempt to interpret non-corresponding recorded data leading to an apparent misfit reduction. Such wrong association may lead the iterative process into a meaningless local minimum. In this thesis, I carefully distinguish the concept of "cycle-skipping" (a data-space phenomenon) from "converging to a local minimum" (a model-space pheonmenon).

3.2.2 Wave-equation migration velocity analysis

Wave-equation migration velocity analysis (WEMVA) belongs to a family of techniques aimed at estimating the optimal background velocity model that improves the quality, coherency and focusing of migrated images computed with wave-equation based modeling operators (Biondi and Sava, 1999). The goal of the optimization process is to minimize objective functions of the following form

$$\Phi_{\text{WEMVA}}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{E}\left(\tilde{\mathbf{I}}(\mathbf{m})\right)\right\|_2^2, \tag{3.2}$$

where $\mathbf{m} \in \mathbb{R}^{N_m}$ is the unknown (smooth) seismic velocity model, and $\tilde{\mathbf{I}} \in \mathbb{R}^{N_{\tilde{p}}}$ is an extended migrated image defined by

$$\tilde{\mathbf{I}}(\mathbf{m}) = \tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{d}^{ref}, \tag{3.3}$$

where $\mathbf{d}^{ref} \in \mathbb{R}^{N_d}$ is a subset of the observed data assumed to be primary reflected events. $\tilde{\mathbf{B}}$ denotes the extended Born modeling operator (defined in section 2.3), $*$ symbolizes the adjoint process, and the $\sim$ symbol refers to extended variables and operators. Extensions typically include space lags $\mathbf{h}$, time lags τ, subsurface reflection angles γ, and seismic sources $\mathbf{s}$ (Biondi and Sava, 1999; Biondi and Symes, 2004; Sava and Biondi, 2004; Biondi and Almomin, 2014; Perrone et al., 2014; Zhang, 2015). $N_{\tilde{p}} = N_m \times N_{ext}$ is the dimension of the extended space, where N_{ext} refers to the extension size (i.e., the number of points on the extended axis/axes). $\mathbf{E} : \mathbb{R}^{N_{\tilde{p}}} \mapsto \mathbb{R}^{N_{\tilde{p}}}$ is an operator that typically measures or enhances the defocusing of $\tilde{\mathbf{I}}$, and computes an image residual used to iteratively update the background velocity model $\mathbf{m}$. Thus, minimizing equation 3.2 can be interpreted as finding an optimal model such that the image defocusing is reduced. Various approaches have been developed to design efficient enhancing operators $\mathbf{E}$ by (1) evaluating the curvature of the subsurface angle domain common image gathers (ADCIGs) (Zhang, 2015), by (2) employing the differential semblance optimization (DSO) operator that penalizes the first derivative along the angle axis of ADCIGs, or by (3) measuring the lack of focusing in subsurface-offset extended images (Symes and Carazzone, 1991; Symes and Kern, 1994; Shen and Symes, 2008). While WEMVA methods tend to be less sensitive to the accuracy of the initial model, their output is usually smooth and lacks vertical resolution because only the transmission effects of the velocity are used for surface acquisition geometries (Almomin, 2016).

3.2.3 FWIME objective function

I provide a formal mathematical definition of the FWIME cost function. I begin with the following initial formulation,

$$\Psi_\epsilon(\mathbf{m},\tilde{\mathbf{p}}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \mathbf{d}^{obs}\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}\right\|_2^2, \tag{3.4}$$

where $\mathbf{m} \in \mathbb{R}^{N_m}$ is the unknown discretized seismic velocity model, $\tilde{\mathbf{B}}$ denotes the extended Born modeling operator, and $\tilde{\mathbf{p}} \in \mathbb{R}^{N_{\tilde{p}}}$ is an extended perturbation in either time lags τ or horizontal subsurface offsets $\mathbf{h}$ (defined in subsections 2.3.3 and 2.3.4). In the FWIME framework, the unknown velocity model $\mathbf{m}$ is never extended. The diagonal matrix $\mathbf{D}$ is a modified and invertible form of the differential semblance optimization (DSO) operator that enhances the energy of the extended perturbation $\tilde{\mathbf{p}}$. ϵ is the trade-off parameter between the two components of Ψ_ϵ. Its value is set at the initial step and kept fixed throughout the optimization process.

In equation 3.4, $\tilde{\mathbf{B}}$ is linear with respect to $\tilde{\mathbf{p}}$ but nonlinear with respect to the velocity model $\mathbf{m}$, while $\mathbf{D}$ is linear with respect to $\tilde{\mathbf{p}}$. Therefore, Ψ_ϵ is quadratic with respect to $\tilde{\mathbf{p}}$ (for a fixed $\mathbf{m}$). By employing the variable projection method to minimize equation 3.4 (Golub and Pereyra, 1973; Rickett, 2013; Huang and Symes, 2015), I can explicitly express $\tilde{\mathbf{p}}$ as a function of $\mathbf{m}$ and remove it from the dependencies of Ψ_ϵ on the left side of equation 3.4. The FWIME objective function may be formally defined:

$$\boxed{\Phi_\epsilon(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs}\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\right\|_2^2.} \tag{3.5}$$

This formulation introduces two modifications compared to equation 3.4. The objective function now solely depends on the velocity model $\mathbf{m}$, and the extended perturbation $\tilde{\mathbf{p}}$ has been replaced by $\tilde{\mathbf{p}}_\epsilon^{opt}$, which is defined as the minimizer of the following quadratic objective function (for fixed $\mathbf{m}$ and ϵ values),

$$\Phi_{\epsilon,\mathbf{m}}(\tilde{\mathbf{p}}) = \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right)\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}\right\|_2^2. \tag{3.6}$$

The Hessian of $\Phi_{\epsilon,\mathbf{m}}$ is given by the following expression,

$$\mathbf{H}_{\Phi_{\mathbf{m},\epsilon}} = \tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}, \tag{3.7}$$

For $\epsilon > 0$, $\mathbf{H}_{\Phi_{\mathbf{m},\epsilon}}$ is a positive definite operator as the sum of a positive semi-definite operator $\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m})$ and a positive definite operator, $\epsilon^2\mathbf{D}^*\mathbf{D}$. Hence, $\Phi_{\epsilon,\mathbf{m}}$ has a unique minimizer (and minimum), referred to as the optimal extended perturbation and denoted by $\tilde{\mathbf{p}}_\epsilon^{opt}$, which depends nonlinearly on the velocity model $\mathbf{m}$. Its analytical expression is given by the formal solution of the

normal equations

$$\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) \quad = \quad \tilde{\mathbf{B}}_{\epsilon,\mathbf{D}}^{\dagger}(\mathbf{m}) \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right), \tag{3.8}$$

where $\tilde{\mathbf{B}}_{\epsilon,\mathbf{D}}^{\dagger}$ is the pseudo-inverse matrix of $\tilde{\mathbf{B}}$ in equation 3.6:

$$\tilde{\mathbf{B}}_{\epsilon,\mathbf{D}}^{\dagger}(\mathbf{m}) \quad = \quad \left[\tilde{\mathbf{B}}^{*}(\mathbf{m})\, \tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^{2}\mathbf{D}^{*}\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^{*}(\mathbf{m}). \tag{3.9}$$

In practice, the minimization of equation 3.6 is performed iteratively using a linear conjugate-gradient algorithm (Aster et al., 2018), and is referred to as the variable projection step in FWIME.

The advantage of recasting the inverse problem expressed in equation 3.4 into the problem expressed in equation 3.5 with a linear sub-problem using the variable projection method (equation 3.6) can be further analyzed from an optimization point of view. Equation 3.4 is defined on an increased search space with the use of the additional extended variable $\tilde{\mathbf{p}}$. Since Ψ_{ϵ} is quadratic with respect to $\tilde{\mathbf{p}}$ for a fixed $\mathbf{m}$, the global minimum of Ψ_{ϵ} must be reached for $\mathbf{m}^{opt}$ and for a $\tilde{\mathbf{p}}$ value that minimizes the quadratic cost function $\tilde{\mathbf{p}} \mapsto \Psi_{\epsilon}(\mathbf{m}^{opt}, \tilde{\mathbf{p}})$, which corresponds to equation 3.6. The main advantage of this compact formulation (equation 3.5) is that I now formally invert for a single physical non-extended parameter $\mathbf{m}$ (representing the unknown seismic velocity) while still benefiting from an extended search space. Consequently, I do not need to implement convoluted alternate-direction or gradient-scaling optimization techniques that are usually indispensable for multi-parameter inversions (Operto et al., 2013; Biondi and Almomin, 2014; Le, 2019).

The data-fitting component (first term on the right side of equation 3.5) is a modified FWI objective function where the data-correcting term is used to ensure the phase alignment between modeled and observed data. During the optimization process, the annihilating component gradually forces the L_2-norm of $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})$ to vanish, thereby reducing the contribution of the data-correction term $\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})$.

3.2.4 FWIME $\approx$ FWI + WEMVA

I provide a high-level description of the main intuition that led to the design of FWIME. In equation 3.5, I introduce a data-correcting term $\mathbf{g}_{\epsilon}$ defined by

$$\mathbf{g}_{\epsilon} : \mathbb{R}^{N_m} \quad \mapsto \quad \mathbb{R}^{N_d} \tag{3.10}$$

$$\mathbf{m} \quad \mapsto \quad \mathbf{g}_{\epsilon}(\mathbf{m}) \tag{3.11}$$

in the argument of the FWI objective function (equation 3.1) whose goal is to ensure that

$\mathbf{f}(\mathbf{m})+\mathbf{g}_\epsilon(\mathbf{m})$ fully fits the entire dataset $\mathbf{d}^{obs}$ at any given precision, even for very inaccurate initial velocity models $\mathbf{m}$. With this tool, I first modify the conventional FWI loss function (equation 3.1) as follows,

$$\Phi(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}) + \mathbf{g}_\epsilon(\mathbf{m}) - \mathbf{d}^{obs}\right\|_2^2. \tag{3.12}$$

The right side of equation 3.12 and is referred to as the data-fitting component of FWIME. Assuming that such $\mathbf{g}_\epsilon(\mathbf{m})$ exists and that it can be computed, I now have an enhanced non-physical modeling operator $\mathbf{f}+\mathbf{g}_\epsilon$ that can match the observed data as accurately as needed. With this additional term, the observed data are not "cycle-skipped" anymore but there is no guarantee that Φ is convex nor free of local minima. In fact, the objective function defined in equation 3.12 is constant and null for all models. I use the necessary condition which required that for the target model, the data prediction should eventually be generated by $\mathbf{f}$ only, without the need for a correcting term. I add an annihilating component $A(\mathbf{m}) = \|\mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\|_2^2$ to the objective function that will gradually reduce the contribution of $\mathbf{g}_\epsilon$ during the minimization of Φ, while still ensuring data-matching. Therefore, the FWIME objective function can be symbolically expressed in the following form

$$\Phi(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}) + \mathbf{g}_\epsilon(\mathbf{m}) - \mathbf{d}^{obs}\right\|_2^2 + \frac{\epsilon^2}{2}A(\mathbf{m}), \tag{3.13}$$

where ϵ is the trade-off parameter between the data-fitting and annihilating components. Eventually, if $\mathbf{f}+\mathbf{g}_\epsilon \approx \mathbf{d}^{obs}$ and $\mathbf{g}_\epsilon$ has vanished, the FWIME inverted model coincides with the global minimum of conventional FWI. In this thesis, I do not mathematical prove that the objective function defined in equation 3.13 is more adequate for gradient-descent optimization methods, but I successfully test is on numerical examples.

An intuitive way to understand the advantage of this new formulation is that the data-correcting term $\mathbf{g}_\epsilon(\mathbf{m})$ reduces and controls the relative weight of the data-fitting component with respect to the annihilator in the cost function. The annihilating component gives the freedom to create judicious operators based on WEMVA techniques, which have been heuristically observed to guide the inversion to the global minimum in a more robust manner (Symes and Carazzone, 1991; Symes and Kern, 1994; Stolk and Symes, 2002). The rate at which the annihilator reduces the contribution of the data-correcting term is controlled by the value of ϵ and is key to ensure smooth convergence (Fu and Symes, 2017). Finally, equation 3.13 can be interpreted as the sum of two terms: a data-fitting component which is a modified FWI problem where the forward modeling combines physical wave propagation with an unphysical additional term, and an annihilating component that possesses

similar features as a WEMVA objective function:

$$\boxed{\text{FWIME} \approx \text{FWI} + \text{WEMVA.}} \tag{3.14}$$

3.2.5 "Phase-only" inversion

For field data, I apply a trace-by-trace normalization operator to the modeled and observed data in order to ensure that both have approximately similar relative amplitudes, which allows the inversion scheme to focus more on a phase-comparison approach (Routh et al., 2011; Shen, 2014). The new FWIME objective function is modified as follows,

$$\Phi^n_\epsilon(\mathbf{m}) \;=\; \frac{1}{2}\left\|\mathbf{n}(\mathbf{f}(\mathbf{m})) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}_\epsilon(\mathbf{m}) - \mathbf{n}(\mathbf{d}^{obs})\right\|^2_2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}^{opt}_\epsilon(\mathbf{m})\right\|^2_2. \tag{3.15}$$

where $\mathbf{n} : \mathbb{R}^{N_d} \mapsto \mathbb{R}^{N_d}$ is a trace-by-trace normalization operator defined by

$$\mathbf{n}(\mathbf{f}_{s,r}(\mathbf{m})) \;=\; \frac{\mathbf{f}_{s,r}(\mathbf{m})}{\|\mathbf{f}_{s,r}(\mathbf{m})\|_2 + c} \tag{3.16}$$

$$\mathbf{n}(\mathbf{d}^{obs}_{s,r}) \;=\; \frac{\mathbf{d}^{obs}_{s,r}}{\|\mathbf{d}^{obs}_{s,r}\|_2 + c} \tag{3.17}$$

where $\mathbf{f}_{s,r}$ and $\mathbf{d}^{obs}_{s,r}$ are the predicted and observed traces from receiver r generated by source s, and $c > 0$ is a user-defined constant that ensures numerical stability. Equation 3.6 is modified accordingly,

$$\Phi^n_{\epsilon,\mathbf{m}}(\tilde{\mathbf{p}}) = \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \left[\mathbf{n}(\mathbf{d}^{obs}) - \mathbf{n}(\mathbf{f}(\mathbf{m}))\right]\right\|^2_2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}\right\|^2_2. \tag{3.18}$$

In equation 3.15, the trace normalization operator $\mathbf{n}$ is purposely not applied to the data-correcting term $\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}_\epsilon(\mathbf{m})$ to ensure that $\Phi^n_{\epsilon,\mathbf{m}}$ remains quadratic with respect to $\tilde{\mathbf{p}}$, which is a necessary condition for the application of the variable projection method.

3.3 Dissection of the FWIME objective function

In this section, I provide a thorough analysis of the four constitutive blocks of FWIME: (1) the data-correcting term, (2) the extended optimal perturbation, (3) the annihilator, and (4) the trade-off parameter. I illustrate how they are effectively computed on a 2D numerical example based on the Marmousi2 model. Even though my analysis is conducted for the conventional FWIME objective

function (equation 3.5), it is straight forward to extend it to the phase-only formulation proposed in equation 3.15.

3.3.1 The data-correcting term

The data-correcting is a tool used in FWIME to fit all the events in the observed data at any given precision. When the current velocity model estimate $\mathbf{m}$ is too inaccurate, some events in the recorded data $\mathbf{d}^{obs}$ may be incorrectly modeled (or may even be missing) by the conventional forward mapping $\mathbf{f}(\mathbf{m})$. The goal of the data-correcting term is to predict these missing events contained in the residual $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$ by combining the concept of extended Born modeling with the variable projection method. It has been numerically observed that an extended Born modeling operator $\tilde{\mathbf{B}}$ (with an adequate extension) combined with a well-chosen extended perturbation $\tilde{\mathbf{p}}^*$ can linearly predict any data misfit regardless of the accuracy of the velocity model (Symes, 2008; Huang and Symes, 2015; Barnier et al., 2018). This statement relies on the existence of a perturbation $\tilde{\mathbf{p}}^* \in \mathbb{R}^{N_{\tilde{p}}}$ that satisfies

$$\forall \mathbf{m},\ \forall \mathbf{d}^{obs},\ \forall \delta > 0,\ \exists \tilde{\mathbf{p}}^*\ s.t.\ \left\| \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^* - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\big) \right\|_2^2 < \delta. \tag{3.19}$$

In order to avoid such formalism, conditions such as the one shown in inequality 3.19 will be replaced by

$$\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^* \quad \approx \quad \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}). \tag{3.20}$$

The computation of $\tilde{\mathbf{p}}^*$ can be conducted iteratively by casting equation 3.20 into a least-squares optimization problem and by finding the minimizer of the following quadratic function (for a fixed $\mathbf{m}$):

$$\chi_{\mathbf{m}}(\tilde{\mathbf{p}}) \quad = \quad \frac{1}{2} \left\| \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\big) \right\|_2^2. \tag{3.21}$$

The minimization of $\chi_{\mathbf{m}}$ is mathematically and computationally equivalent to conducting an extended linearized waveform inversion. It is also a special case of the FWIME variable projection step (equation 3.6) where $\epsilon = 0$, and thus $\chi_{\mathbf{m}} = \Phi_{0,\mathbf{m}}$. Consequently, $\tilde{\mathbf{p}}^*(\mathbf{m}) = \tilde{\mathbf{p}}_0^{opt}(\mathbf{m})$ is the FWIME optimal extended perturbation and its expression is given by

$$\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}) \quad = \quad \tilde{\mathbf{B}}_0^{\dagger}(\mathbf{m})\big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\big), \tag{3.22}$$

where $\tilde{\mathbf{B}}_0^{\dagger}(\mathbf{m}) = [\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m})]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})$ is the pseudo-inverse of $\tilde{\mathbf{B}}(\mathbf{m})$ in equation 3.21. Since $\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m})$ is a semi-definite positive matrix, $\tilde{\mathbf{p}}_0^{opt}(\mathbf{m})$ may not be unique. Furthermore, if $\tilde{\mathbf{p}}_0^{opt}(\mathbf{m})$ is such that $\Phi_{0,\mathbf{m}}(\tilde{\mathbf{p}}_0^{opt}(\mathbf{m})) = 0$, equation 3.20 is satisfied at any given precision, and

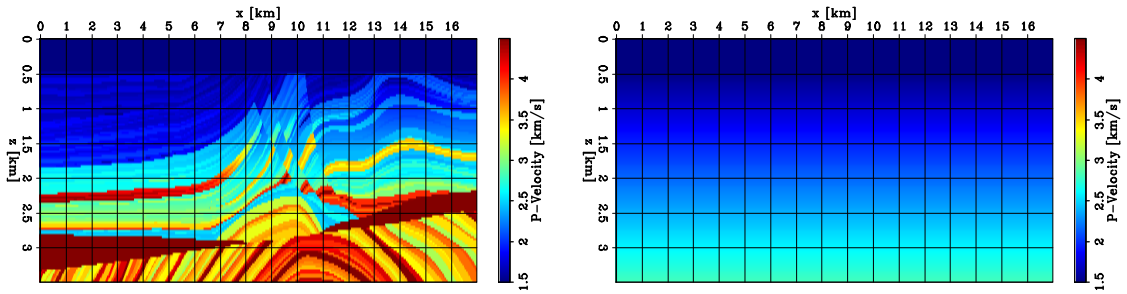
$$\begin{aligned} \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}) &= \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{B}}_0^{\dagger}(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right) \\ &\approx \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}), \end{aligned} \tag{3.23}$$

where $\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_0^{opt}(\mathbf{m})$ is precisely the FWIME correcting term for $\epsilon = 0$. In such case, the extension is said to be approximately linearly invertible on the data space and $\tilde{\mathbf{B}}(\mathbf{m})$ is a surjective linearized operator (Symes, 2008). The use of extended modeling is crucial because when the velocity model generates data with substantial kinematic errors, the minimization of equation 3.21 with a non-extended Born modeling operator fails to reduce its misfit to zero. Therefore, this powerful tool is fundamental for the FWIME workflow: by combining extended modeling with the variable projection method (and by setting $\epsilon = 0$), the data-correcting term can match the data misfit $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$ as accurately as numerical precision allows. I do not claim to prove the existence of $\tilde{\mathbf{p}}_0^{opt}$ such that $\Phi_{0,\mathbf{m}}(\tilde{\mathbf{p}}_0^{opt}) = 0$ for any $\mathbf{m}$. However, the numerical tests conducted in this thesis using either time-lag or subsurface offset extensions indicate that such minimizer exists, even for very inaccurate $\mathbf{m}$.

I illustrate the potential of extended modeling by computing the data-correcting term on a numerical example based on the Marmousi2 model shown in Figure 3.7a. I generate noise-free pressure data with a two-way acoustic finite-difference propagator using a grid-spacing of 30 m in both directions. At the surface, I place 140 sources every 120 m, and 567 receivers every 30 m. Data are modeled with a wavelet containing energy restricted to the 4-13 Hz frequency range, and are recorded for 7 s. The initial model $\mathbf{m}_0$ (Figure 3.7b) is purposely chosen to be inaccurate: it is laterally invariant and linearly increasing with depth (a 500 m thick water layer is added on top). Figure 3.8 shows two velocity profiles extracted at $x = 5$ km and $x = 10$ km, respectively.

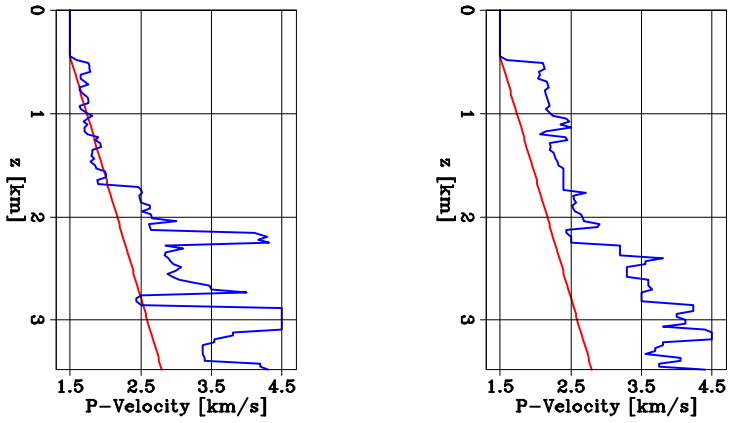
Figure 3.9a shows a representative shot gather of the observed data, $\mathbf{d}^{obs}$, generated by a source placed at $x = 1.2$ km. The recorded data contain both reflected and refracted energy. Figure 3.9b displays the analogous shot gather computed with the initial model, $\mathbf{f}(\mathbf{m}_0)$, which fails to predict most of the events in the observed data, and Figure 3.9c is the initial data residual, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$.

I set $\epsilon = 0$ and I minimize equation 3.6 with 100 iterations of linear conjugate gradient using three different forms of $\tilde{\mathbf{B}}$: (1) non-extended (conventional Born), (2) extended with time lags τ, and (3) extended with horizontal subsurface offsets h_x. For both time lags and subsurface offsets, I use 141 points of extension which allows τ to range from -1.12 s to 1.12 s, and h_x to range from -2.1 km to 2.1 km. The convergence curves for these optimizations are shown in Figure 3.10. Both time-lag (red curve) and horizontal subsurface offset (pink curve) extensions manage to reduce the objective function value by more than 99.5%, and a more accurate matching can be obtained by conducting



(a) (b)

Figure 3.7: 2D panels of velocity models. (a) Marmousi2 model. (b) Initial model $\mathbf{m}_0$. [**ER**]



(a) (b)

Figure 3.8: Depth velocity profiles extracted at (a) $x = 5$ km and (b) $x = 10$ km. The blue and red curves represent the true and initial models, respectively. [**ER**]

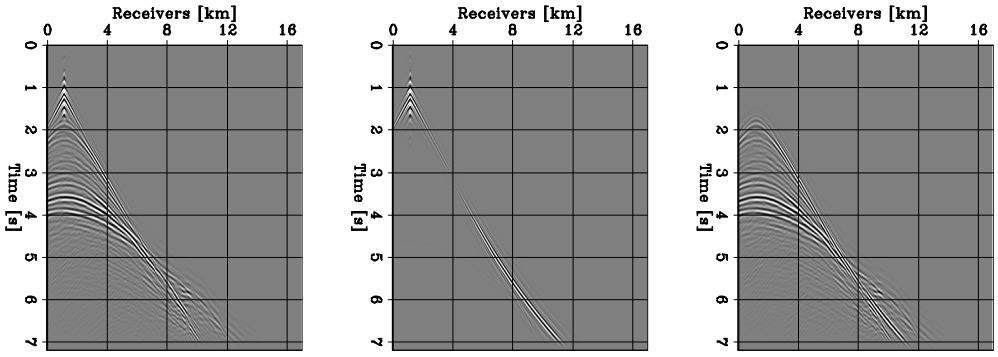


Figure 3.9: Shot gathers generated by a source located at $x = 1.2$ km. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data with the initial model, $\mathbf{f}(\mathbf{m}_0)$. (c) Initial data residual, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**ER**]

more iterations. Not surprisingly, the conventional (non-extended) Born operator is unable to match the data misfit as well as its extended counterparts (blue curve). For this example, setting ϵ to zero is only done to illustrate the ability of extended to match any type of data. For practical applications (when conducting a FWIME inversion scheme), ϵ should indeed be set to a strictly positive value.

The effectiveness of extended modeling can also be appreciated by examining the data-space residuals. Figure 3.11a shows a shot gather extracted from the initial data residual, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. Figure 3.11b is the FWIME data-correcting term computed after minimization of equation 3.6 with a horizontal subsurface-offset extension. Clearly, the data-correcting term $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0)$ is able to capture all the events (with the correct amplitudes) that were missed by the inaccurate nonlinear prediction (Figure 3.11a). Figure 3.11c displays the difference between the data-correcting term and the initial data residual, $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$, which is numerically close to zero. An important remark is that this level of data matching is achieved for all shots within the dataset, and not just for only one specific shot gather. Figure 3.12a shows analogous panels for the non-extended optimization. As expected, the data-correcting term stemming from the non-extended Born operator fails to match most of the deeper reflections and the refracted energy at larger offsets (Figure 3.12c).

3.3.2 The optimal extended perturbation

In FWIME, the optimal extended perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$ is the output of a linear mapping of the conventional FWI data residual into an extended (non-physical) model space (equation 3.8). Even though computing $\tilde{\mathbf{p}}_\epsilon^{opt}$ is mechanically and computationally equivalent to conducting an extended least-squares reverse-time migration (ELSRTM), its purpose is fundamentally different. ELSRTM is driven by physics, and only a subset of the recorded data, the primary reflected events $\mathbf{d}^{ref}$, is

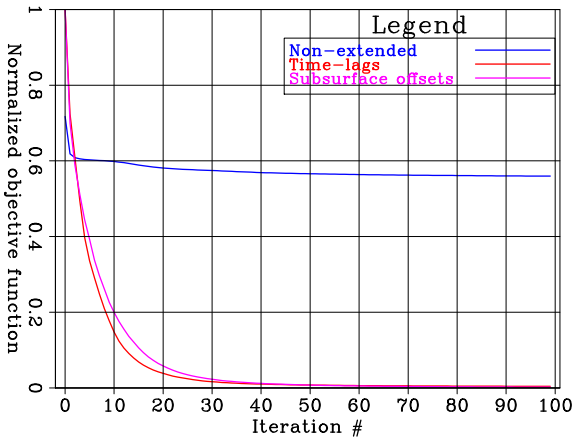
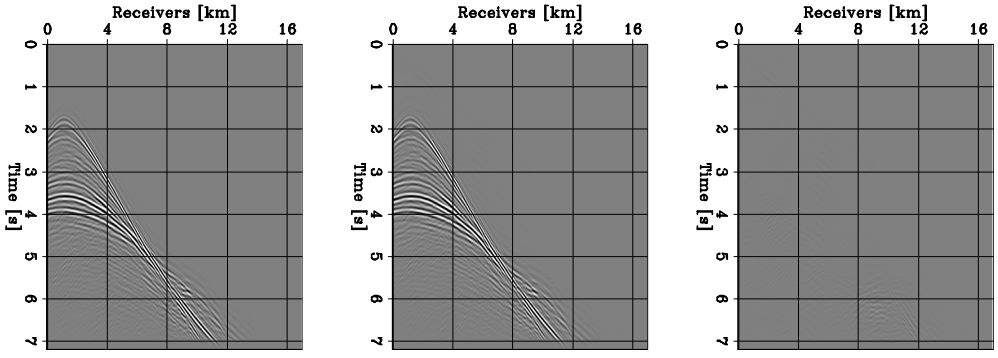
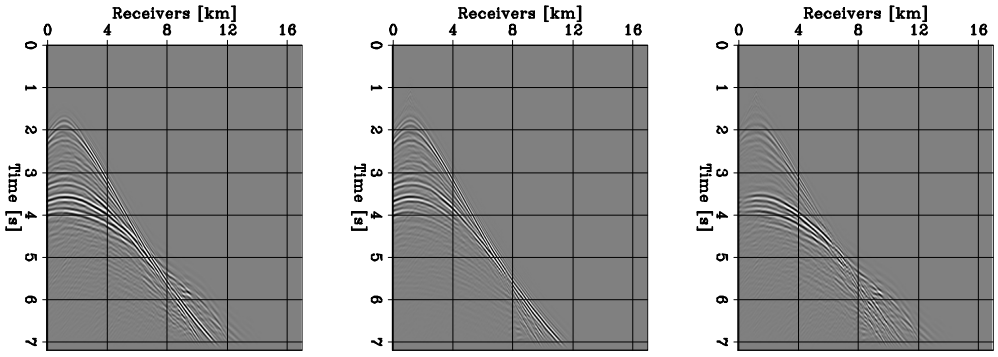


Figure 3.10: Normalized objective functions corresponding to the minimization of equation 3.6 with three different Born operators. Non-extended Born (blue curve), time-lag extension (red curve), and horizontal subsurface offset extension (pink curve). All three curves are normalized with the same value. [**CR**]



(a) (b) (c)

Figure 3.11: Shot gathers generated by a source located at $x = 1.2$ km. (a) Initial data misfit, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. (b) FWIME data-correcting term computed with $\epsilon = 0$ and using a horizontal subsurface-offset extension, $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0)$. (c) Difference between the data-correcting term and the initial data-residual, $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$. All panels are displayed with the same grayscale. [**CR**]



(a) (b) (c)

Figure 3.12: Shot gathers generated by a source located at $x = 1.2$ km. (a) Initial data misfit, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. (b) FWIME data-correcting term computed using a non-extended Born operator, $\mathbf{B}(\mathbf{m}_0)\mathbf{p}_0^{opt}(\mathbf{m}_0)$. (c) Difference between the data-correcting term and the initial data-residual, $\mathbf{B}(\mathbf{m}_0)\mathbf{p}_0^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$. All panels are displayed with the same grayscale. [**CR**]

inverted. The aim is to find a coherent, focused and geologically consistent extended image $\tilde{\mathbf{I}}^{opt}$ such that its demigration produces a set of reflections that match the ones present in the recorded data (Leader, 2015). This process is done by minimizing the following quadratic function,

$$\varphi_{\mathbf{m}}(\tilde{\mathbf{I}}) = \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{I}} - \mathbf{d}^{ref}\right\|_2^2, \tag{3.24}$$

where $\mathbf{m} \in \mathbb{R}^{N_m}$ is the (fixed) background velocity model, and $\tilde{\mathbf{I}} \in \mathbb{R}^{N_{\tilde{p}}}$ is an extended image expected to only contain short-wavelength components (e.g., seismic reflectors mapping interfaces between rocks layers). Low-wavenumber features may arise with the use of two-way wave-equation propagators, but are usually considered noise and removed with various techniques (Youn and Zhou, 2001; Fletcher et al., 2005; Guitton et al., 2007; Fei et al., 2010; Liu et al., 2011). The optimization problem defined in equation 3.24 can also be regularized to incorporate a priori geological subsurface information and mitigate the effects of uneven illumination patterns in the migrated image (Prucha, 2004).

In FWIME, $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ is computed with the same mathematical mechanism as in equation 3.24, and may therefore share similar features with a conventional extended image (Sava and Fomel, 2006). However, there are two major differences. First, during the FWIME variable projection step, the full data mismatch $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$ is inverted (equation 3.6). This term may include all types of waves such as transmissions, refractions, and (but not only) reflections, as shown in the previous numerical example (Figure 3.9c). The mapping of this data mismatch from the data space into $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ will potentially introduce low- and high-wavenumber events with certain characteristics in the extended

space of $\tilde{\mathbf{p}}_\epsilon^{opt}$ that can provide quantitative information about the errors in the current velocity model estimate $\mathbf{m}$. Thus no filtering or restriction on $\tilde{\mathbf{p}}_\epsilon^{opt}$'s wavenumber content should be applied at any stage of the FWIME workflow (it is important to keep all the information within $\mathbf{p}_\epsilon^{opt}$). Intuitively, $\mathbf{p}_\epsilon^{opt}$ possesses the same dimensions as an extended image, but serves as a metric in the extended-model space to assess the errors between the physical prediction $\mathbf{f}(\mathbf{m})$ and the observed data $\mathbf{d}^{obs}$: complex overlapping events present in the data space can be mapped and more easily untangled into the extended space of $\mathbf{p}_\epsilon^{opt}$.

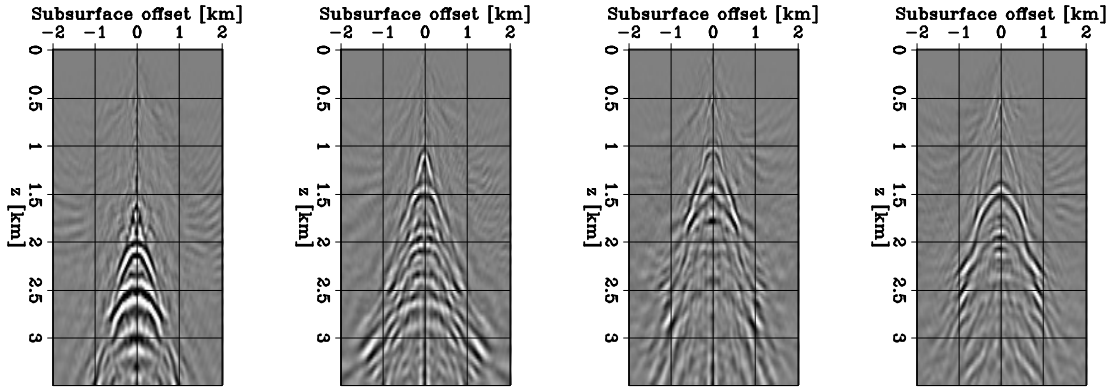
The second major difference is that in a noise-free environment and assuming the FWIME optimization scheme can converge to the unique global minimum $\mathbf{m}_t$, the data-prediction error $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_t)$ should eventually vanish, which implies that $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_t)$ must also vanish (according to equation 3.8). In contrast, current techniques aim at improving the coherency and focusing of images/perturbations, which should be representative of the Earth's subsurface (Biondi and Symes, 2004; Sava and Biondi, 2004; Biondi and Almomin, 2014). In FWIME, I do not assign any physical meaning to $\tilde{\mathbf{p}}_\epsilon^{opt}$: it is only a tool used to predict the events in the recorded data that were missed by the physical modeling $\mathbf{f}(\mathbf{m})$, and its target value is the null vector.

I examine the features of $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed with $\epsilon = 0$ on the Marmousi2 numerical example described in the previous section. Figures 3.13 and 3.14 show CIGs extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ at four horizontal positions computed with a space-lag (horizontal subsurface-offset) and time-lag extension, respectively. Clearly, $\tilde{\mathbf{p}}_\epsilon^{opt}$ possesses similar features as conventional extended images: for both space and time lags, some coherent clusters of energy (corresponding to the mapping of reflections observed in the data-space) are mapped away from the physical plane, especially towards greater depths where the velocity error in $\mathbf{m}$ is the largest. Such events can easily be interpreted by examining SOCIGs (Figure 3.13), where the frowning moveout indicates that the velocity used for propagation is too low, a behavior commonly observed in conventional extended images (Biondi, 2006). The compounding effect of the kinematic errors over depth can also be detected by analyzing extended depth planes extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ for both types of extensions, as shown in Figures 3.15 and 3.16.

3.3.3 The annihilating component

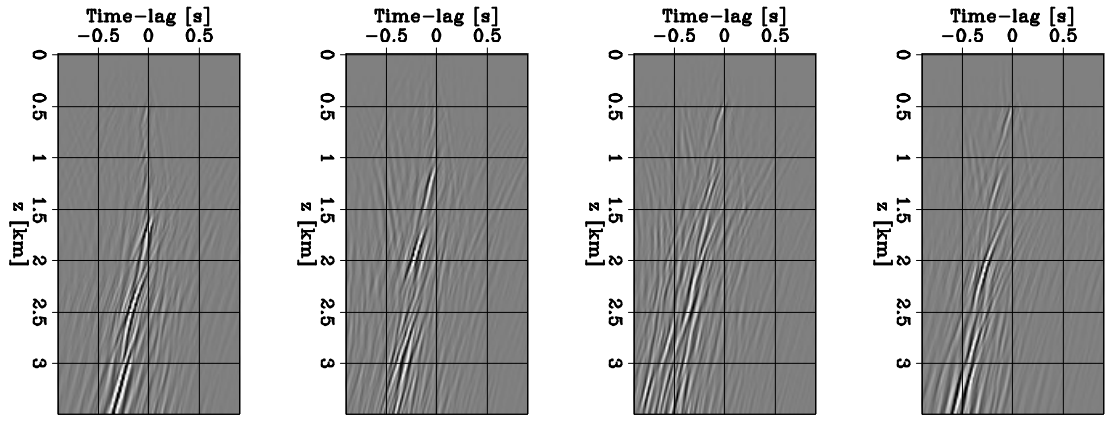
By adding a data-correcting term to the data-fitting component, I effectively create a new non-physical modeling operator, $\mathbf{f} + \mathbf{g}_\epsilon$. I showed that by setting $\epsilon = 0$, I could satisfy $\mathbf{f} + \mathbf{g}_0 \approx \mathbf{d}^{obs}$. With this new non-physical modeling operator, the predicted data are not cycle-skipped but there is no guarantee that the FWIME objective function possesses a wider basin of attraction about the global minimum. In order to guide the optimization towards the global solution, the contribution of the data-correcting term must be gradually reduced during the inversion process, which can be achieved by forcing the L_2-norm of $\mathbf{p}_\epsilon^{opt}$ to vanish, while still ensuring that equation 3.20 remains satisfied.

I add an annihilating term (second component on the right side of equation 3.5) which employs a



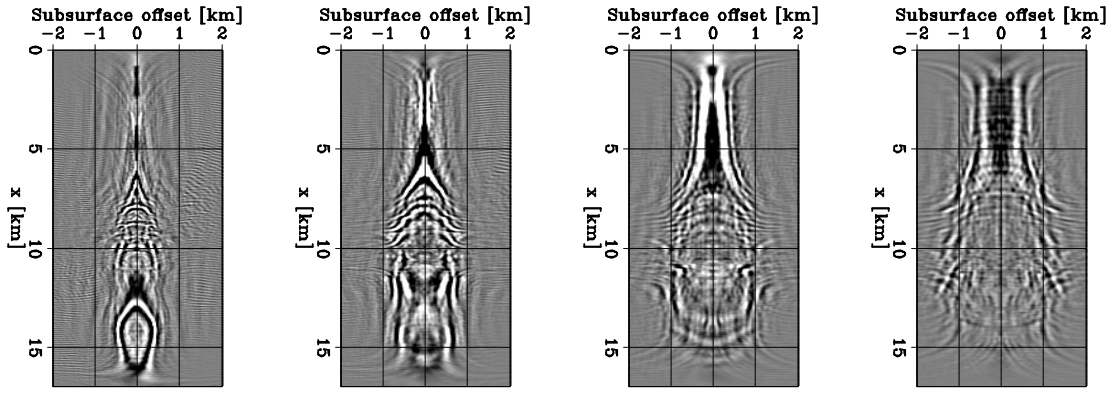
(a) (b) (c) (d)

Figure 3.13: SOCIGs extracted at four horizontal positions from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$. (a) $x = 5.0$ km, (b) $x = 8$ km, (c) $x = 10$ km, and (d) $x = 13$ km. All panels are displayed with the same grayscale. [**CR**]



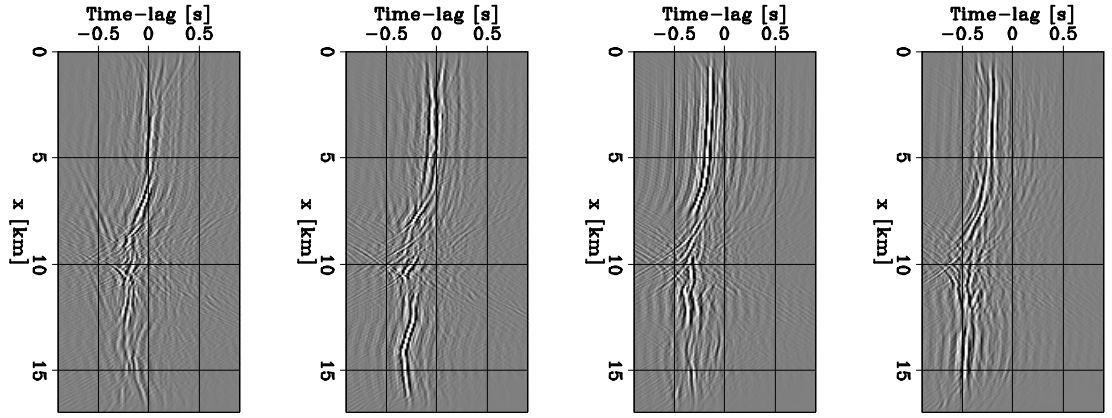
(a) (b) (c) (d)

Figure 3.14: TLCIGs extracted at four horizontal positions from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$. (a) $x = 5.0$ km, (b) $x = 8$ km, (c) $x = 10$ km, and (d) $x = 13$ km. All panels are displayed with the same grayscale. [**CR**]



(a) (b) (c) (d)

Figure 3.15: Extended depth planes extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed with a horizontal subsurface-offset extension h_x at four depths. (a) $z = 1.5$ km, (b) $z = 2$ km, (c) $z = 2.5$ km, and (d) $z = 3$ km. All panels are displayed with the same grayscale. [**CR**]



(a) (b) (c) (d)

Figure 3.16: Extended depth planes extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed with a time-lag extension τ at four depths. (a) $z = 1.5$ km, (b) $z = 2$ km, (c) $z = 2.5$ km, and (d) $z = 3$ km. All panels are displayed with the same grayscale. [**CR**]

modified form of the DSO operator $\mathbf{D}$, first proposed by Symes and Kern (1994). This operator has been extensively and successfully used for computing optimal image residuals in MVA algorithms (Sava and Biondi, 2004; Symes, 2008). It enhances features in the extended migrated images created by the presence of errors in the velocity model. For horizontal subsurface offset and time-lag extended images, it is a diagonal operator that multiplies each point of the extended image by a value proportional to its distance to the physical plane. Embedded into a MVA workflow, it rewards images having most of their energy focused in the vicinity of the zero-subsurface and zero time-lag planes. Moreover, Stolk and Symes (2002); Symes (2008) have carefully studied the behavior of this operator and demonstrated its ability to "yield optima that are robust against large errors in the initial model estimates."

I take advantage of the beneficial properties of the DSO operator to extract kinematic information from $\mathbf{p}_\epsilon^{opt}$ and guide the inversion towards the optimal solution. However, since the goal is not to obtain a well-focused image (but to make the perturbation vanish), I modify the DSO operator by also penalizing energy located on the physical plane of $\mathbf{p}_\epsilon^{opt}$, but with a smaller penalty weight. Figure 3.17 shows the absolute value of the coefficients in the conventional (blue curve) and modified (red curve) DSO operators for a fixed point $M(z_M, x_M, h_x)$ within $\mathbf{p}_\epsilon^{opt}$ as function of its distance $|h_x|$ to the physical plane. An analogous penalty function is employed for the time-lag extension.

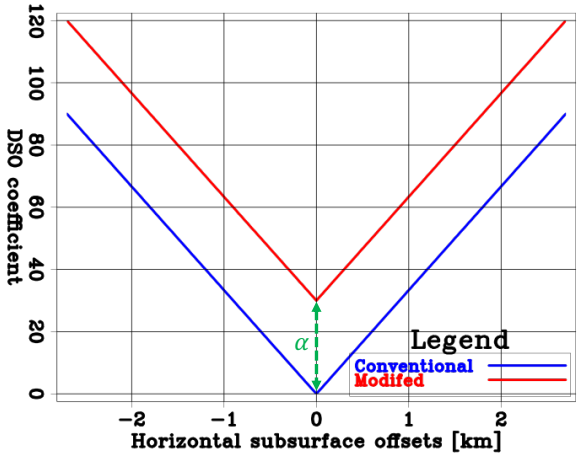


Figure 3.17: Absolute value of the penalty function for the conventional (blue curve) and modified (red curve) DSO operators applied to a fixed point $M(z_M, x_M, h_x)$ where h_x is the horizontal subsurface-offset value. [**NR**]

From a optimization standpoint, the modification of operator $\mathbf{D}$ is important because it ensures, under certain assumptions, that the FWIME and the conventional FWI objective functions share the same unique global minimum $\mathbf{m}_t$ (Appendix A).

3.3.4 The trade-off parameter ϵ

In FWIME, ϵ is an important hyper-parameter to select because its value affects the shape of the objective function. This hyper-parameter is set at the initial step and is fixed throughout the optimization process. It allows me to control the level of data fitting (with the data-correcting term) by adjusting the penalty applied to $\tilde{\mathbf{p}}_\epsilon^{opt}$. During the minimization of equations 3.5 and 3.6, low ϵ-values will impose less constraint on the annihilating component, thereby allowing more energy to be mapped into $\tilde{\mathbf{p}}_\epsilon^{opt}$, even far from the physical plane. This allows the data-correcting term to match the data misfit $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$ (i.e., satisfy equation 3.20) with more accuracy. In fact, by setting $\epsilon = 0$, the FWIME objective function is constant and numerically close to zero for all numerically reasonable velocity models $\mathbf{m}$. Therefore, very low ϵ-values are not optimal and may slow down the convergence. Conversely, high ϵ-values will reward minimizing the annihilating component rather than the data-fitting component, thereby not mitigating the cycle-skipping effect, which may lead FWIME to converge to a local minimum. Finally, when ϵ tends to infinity, the L_2-norm of $\mathbf{p}_\epsilon^{opt}$ converges to zero and FWIME is mathematically equivalent to FWI (Appendix B). In this thesis, I do not develop a mathematical method to automatically select an optimal ϵ-value, but I use a trial-and-error approach based on examining a subset of the CIGs extracted from the optimal extended perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed at the initial step. Fortunately, for the 2D numerical tests and 3D field applications, I observe that the sensitivity of FWIME with respect to ϵ is quite limited as long as the proper order of magnitude is determined.

Furthermore, Fu and Symes (2017) successfully show that adjusting the trade-off parameter throughout the inversion process can potentially increase its efficiency. However, in FWIME, I purposely choose to keep ϵ fixed as an effort to reduce the need for human input. As a result, FWIME is formulated in a compact and mathematically consistent manner that only requires the tuning of one single hyper-parameter at the initial step. This feature makes my approach easy to apply.

Chapter 4

Optimization: A model-space multi-scale approach

In this chapter, I describe the optimization process of FWIME with gradient-descent (i.e., local optimization) numerical schemes. I begin by analyzing and describing the structure of the FWIME gradient. I show that for inaccurate initial models, its tomographic component is responsible for recovering the missing low wavenumbers during the initial stages of the inversion process, even when refracted and/or low-frequency energy is absent from the recorded data. Then, I devise and present a model-space multi-scale optimization workflow which is crucial for the success of my method. In this approach, I parametrize the unknown velocity model on spline grids and I simultaneously invert the full data-bandwidth while gradually refining the grid spacing with iterations. The inverted model for a given grid is then used as the initial guess for the following inversion performed with a finer grid. The grid refinement rate allows me to better control and slowly increase the wavenumber content (i.e., the spatial resolution) of the model updates, which constitutes the multi-scale aspect of my technique. In addition, this approach alleviates the need for tedious data filtering and event selection typically required for conventional FWI (Shen, 2014) because all wave modes are inverted at once with the same procedure. Finally, I successfully invert cycle-skipped data on three synthetic tests where conventional FWI converges to local minima. Each test is designed to generate simple datasets containing only one wave mode (transmission, reflection or refraction) in order to demonstrate that FWIME can invert any type of seismic data using the same algorithm and without the need for intensive hyper-parameter tuning.

4.1 Conventional optimization

I begin by deriving and describing the structure of the FWIME gradient with respect to the unknown velocity model $\mathbf{m}$. I then examine more closely the tomographic component of the gradient and I illustrate on a simple numerical experiment its ability to capture low-wavenumber features from reflected data. Finally, I summarize the full inversion workflow.

4.1.1 FWIME gradient

The gradient of the FWIME objective function (equation 3.5) with respect to the velocity model $\mathbf{m}$ is given by

$$\nabla\Phi_\epsilon(\mathbf{m}) \quad = \quad \mathbf{M}\Big[\mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}))\Big]\Big[\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs}\Big], \tag{4.1}$$

where $\mathbf{B}^*$ is the adjoint of the conventional (non-extended) Born modeling operator, $\mathbf{T}^*$ is the adjoint of the data-space tomographic operator first proposed by Biondi and Almomin (2014) (defined in section 2.4), and $\mathbf{M}$ is a masking operator that may be used to prevent the gradient from updating certain regions of the model (e.g., the water layer). Equation 4.1 (whose derivation is shown in Appendix C.1) can be expressed as the sum of two terms,

$$\nabla\Phi_\epsilon(\mathbf{m}) \quad = \quad \mathbf{M}\left(\nabla\Phi_\epsilon^B + \nabla\Phi_\epsilon^T\right), \tag{4.2}$$

with

$$\nabla\Phi_\epsilon^B(\mathbf{m}) \quad = \quad \mathbf{B}^*(\mathbf{m})\mathbf{r}_d^\epsilon(\mathbf{m}), \tag{4.3}$$

$$\nabla\Phi_\epsilon^T(\mathbf{m}) \quad = \quad \mathbf{T}^*(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}))\mathbf{r}_d^\epsilon(\mathbf{m}), \tag{4.4}$$

where the adjoint source $\mathbf{r}_d^\epsilon(\mathbf{m})$ is the argument of the FWIME data-fitting component. Its expression is given by

$$\mathbf{r}_d^\epsilon(\mathbf{m}) = \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs}. \tag{4.5}$$

$\nabla\Phi_\epsilon^B$ is referred to as the "Born" gradient of FWIME and shares kinematic similarities with the conventional FWI gradient (they employ the same operator but use different adjoint sources). The second component, $\nabla\Phi_\epsilon^T$, is referred to as the "tomographic" (or "WEMVA") gradient. This term

arises from the differentiation of the data-correcting term (with respect to the velocity model $\mathbf{m}$) and is essential for the FWIME workflow to recover the missing low-wavenumber components of the velocity model at early stages of the optimization process (Barnier et al., 2018). The FWIME gradient can be decomposed into two terms,

$$\boxed{\nabla^{\text{FWIME}} \quad = \quad \nabla^{\text{FWI}} + \nabla^{\text{WEMVA}}.} \tag{4.6}$$

After conducting many numerical tests, I observe that this structure produces three regimes throughout the inversion workflow. The first stage can been seen as a tomographic or WEMVA regime, where the low-wavenumber components of the velocity model are recovered. If the initial velocity model is very inaccurate, the tomographic gradient tends to dominate (in terms of amplitude), and the FWIME scheme mainly relies on the contribution of this term to avoid converging to local minima. As the inversion progresses, FWIME enters an intermediate regime where both gradient components share similar amplitudes and contribute equally to the total search direction. Finally, when the inverted model is close to the true solution, FWIME enters what I refer to as the "linear regime": the search direction is mainly guided by the Born component which primarily updates the high-wavenumber features of the velocity. At this point, conventional FWI is able to converge towards the global solution without any extension strategy.

One of the main advantages of FWIME is its ability to automatically manage the transitions between the three different regimes without the need to apply any scale mixing or manual enhancement of the gradient components, in contrast with the method proposed in Biondi and Almomin (2014). This is achieved by the use of two ingredients in the inversion workflow. First, the variable projection method allows $\tilde{\mathbf{p}}_\epsilon^{opt}$ to handle the coupling between both tomographic and Born gradients. Second, a model-space multi-scale strategy is applied to gradually increase the wavenumber content of the model updates, which I describe in the next section.

4.1.2 FWIME gradient for phase-only formulation

The FWIME gradient for the phase-only formulation (equation 3.15) can be obtained by modifying equation 4.1 as follows,

$$\nabla\Phi_\epsilon^n(\mathbf{m}) = \mathbf{M}\Big[\mathbf{B}^*(\mathbf{m})\mathbf{N}^*(\mathbf{f}(\mathbf{m})) + \mathbf{T}^*(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}))\Big]\left(\mathbf{n}(\mathbf{f}(\mathbf{m})) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{n}(\mathbf{d}^{obs})\right), \tag{4.7}$$

where $\mathbf{n}$ is the trace normalization operator, and $\mathbf{N}$ is the self-adjoint Jacobian operator of $\mathbf{n}$, defined by

$$\begin{aligned} \mathbf{N} : \mathbb{R}^{N_d} &\mapsto \mathbb{R}^{N_d} \\ \mathbf{x} &\mapsto \frac{\partial \mathbf{n}(\mathbf{x})}{\partial \mathbf{x}}. \end{aligned} \tag{4.8}$$

The derivation of equation 4.7 and the expression of $\mathbf{N}$ are described in Appendices C.2 and C.3, respectively.

4.1.3 A closer look at the tomographic operator

The tomographic component of the gradient is crucial for the success of FWIME. It is computed by applying the adjoint of the data-space tomographic operator $\mathbf{T}$ to the FWIME data-residuals $\mathbf{r}_d^\epsilon$ (equation 4.4). In Chapter 2, I provide a formal derivation and description of $\mathbf{T}$. In this section, I further analyze it from a physical standpoint. I apply its forward and adjoint mappings on a numerical example, and I illustrate its potential at accurately recovering low-wavenumber features of the subsurface from reflection data.

Forward mapping

Mathematically, $\mathbf{T} : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_d}$ is the Jacobian operator of the data-correcting term with respect to the velocity model $\mathbf{m}$, given a fixed optimal extended perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$. That is,

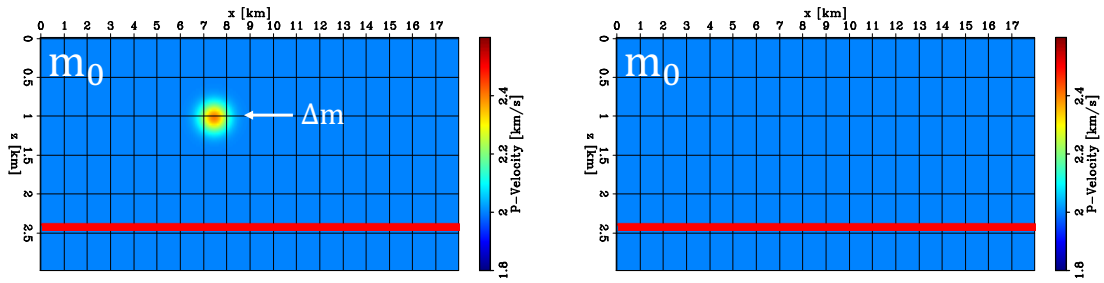
$$\mathbf{T}(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}) = \frac{\partial\big(\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}\big)}{\partial \mathbf{m}}\Big|_{\tilde{\mathbf{p}}_\epsilon^{opt}}. \tag{4.9}$$

For a small velocity perturbation $\Delta\mathbf{m}$, $\mathbf{T}$ is a linear operator that relates $\Delta\mathbf{m}$ to changes in the Born-modeled data, $\Delta\mathbf{d}^{Born}$, assuming a fixed extended reflectivity $\tilde{\mathbf{p}}_\epsilon^{opt}$.

I compute the output of $\mathbf{T}$ by applying it to a small velocity perturbation embedded in a homogeneous background $\mathbf{m}_0$ containing one horizontal reflector, which plays the role of $\tilde{\mathbf{p}}_\epsilon^{opt}$ (non-extended for the sake of simplicity) (Figure 4.1a). In this example, $\tilde{\mathbf{p}}_\epsilon^{opt}$ is fixed, the background model $\mathbf{m}_0$ is set to 2 km/s, and the horizontal reflector is located at a depth of $z = 2.4$ km (Figure 4.1b). The Gaussian positive velocity perturbation $\Delta\mathbf{m}$ reaches a maximum value of 0.5 km/s. I place 150 sources and 600 receivers at the surface every 120 m and 30 m, and I compute $\Delta\mathbf{d}^{tomo} = \mathbf{T}(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})\Delta\mathbf{m}$. Figure 4.2 shows two representative shot gathers extracted from $\Delta\mathbf{d}^{tomo}$, generated with sources located at $x = 1$ km and $x = 12$ km, respectively.

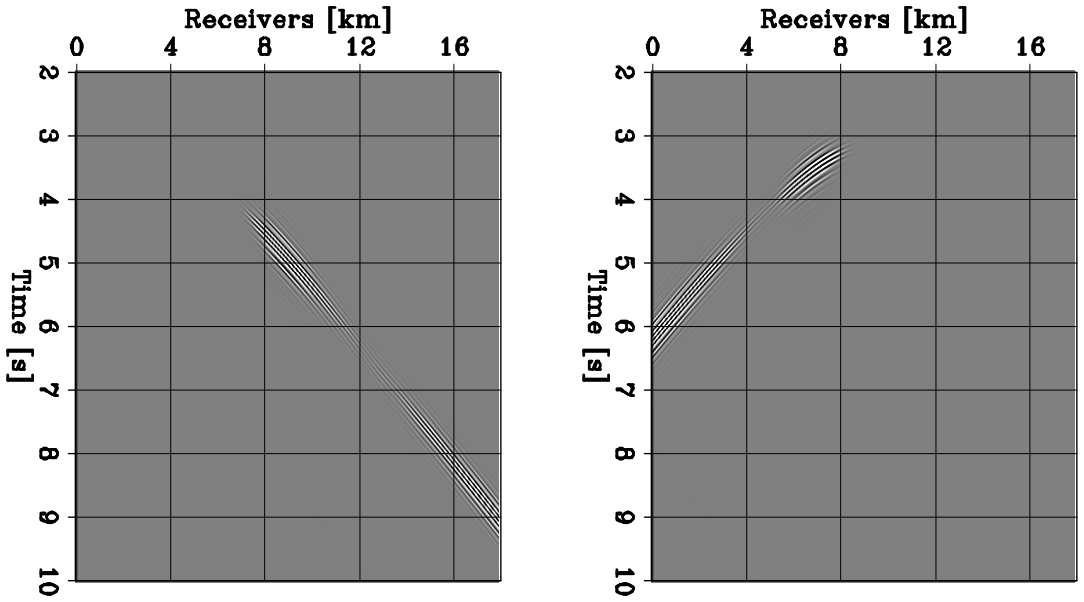
Adjoint mapping

Examining the output of the forward mapping of $\mathbf{T}$ is challenging to interpret, but the adjoint mapping $\mathbf{T}^* : \mathbb{R}^{N_d} \mapsto \mathbb{R}^{N_m}$ provides better physical insight on the properties of the tomographic



(a) (b)

Figure 4.1: 2D panels of velocity models. The red horizontal line represents the reflector playing the role of $\tilde{\mathbf{p}}_\epsilon^{opt}$ (fixed and non-extended for this specific example). (a) True model $\mathbf{m}_0 + \Delta\mathbf{m}$. (b) Background model $\mathbf{m}_0$. [**NR**]



(a) (b)

Figure 4.2: Representative shot gathers resulting from the application of the data-space tomographic operator $\mathbf{T}(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})$ to the velocity perturbation $\Delta\mathbf{m}$, for sources located at (a) $x = 1$ km, and (b) $x = 12$ km. [**ER**]

gradient as it linearly relates data-space perturbations to velocity-model perturbations. I now apply $\mathbf{T}^*$ to $\Delta\mathbf{d}^{tomo}$, and I obtained the following velocity perturbation $\Delta\mathbf{m}_{tomo}$,

$$\Delta\mathbf{m}_{tomo} = \mathbf{T}^*(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})\ \mathbf{T}(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})\ \Delta\mathbf{m}. \tag{4.10}$$

Figures 4.3a and 4.3b show $\Delta\mathbf{m}_{tomo}$ computed according to equation 4.10 for the same sources placed at $x = 1$ km and $x = 12$ km, respectively, which result in low-wavenumber (smooth) changes in the velocity model, even with the use of pure reflection data. The model perturbations $\Delta\mathbf{m}_{tomo}$ shown in Figures 4.3a and 4.3b display a pattern commonly known as the "rabbit ears" updates in conventional FWI, and demonstrate the ability of the FWIME tomographic gradient to reconstruct the long-wavelength features of deeper targets that may not be illuminated by refracted energy, such as diving waves. Figure 4.3c shows the analogous map computed for the entire collection of available source/receiver pairs (equation 4.10). By comparing it to the true perturbation (Figure 4.3d), the velocity update lacks vertical resolution but the recovered anomaly seems to be accurately positioned.

As I later show in this thesis with more complex numerical examples, the FWIME tomographic gradient possesses similar features as the ones obtained with reflection full waveform inversion (RFWI) (Xu et al., 2012; Brossier et al., 2015; Zhou et al., 2015), but one of the main differences between the two approaches is the fact that FWIME employs an extended reflectivity. Even though this model extension increases the computational cost of my method, the additional degrees of freedom it provides allow the tomographic gradient to produce satisfactory model updates even if the velocity error falls outside of the linearization approximation, thereby making FWIME more robust.

4.1.4 FWIME inversion workflow

I summarize the main steps of the FWIME workflow (minimization of equation 3.5) in algorithm 1. The FWIME scheme begins by the selection of an extension type. As first noticed by Biondi and Almomin (2014), extending $\tilde{\mathbf{p}}_\epsilon^{opt}$ with time lags seem to allow the data-correcting term to efficiently capture large time shifts for both reflected and refracted waves. Additionally, for 3D field applications, the time-lag extension requires a single additional axis (compared to two axes for space lags), thereby improving the computational efficiency of the method.

Note that in algorithm 1, the full data bandwidth is simultaneously inverted from the start, and all wave modes are employed (including reflected energy) to potentially produce long-wavelength updates (as shown in the previous section by the analysis of operator $\mathbf{T}^*$). Hence, data-space multi-scale approaches used in conventional FWI such as the one proposed by Bunks et al. (1995) are not suited for my method. In the next section, I present an alternate multi-scale strategy that enables the simultaneous inversion of the full data bandwidth and to gradually increase the resolution of the

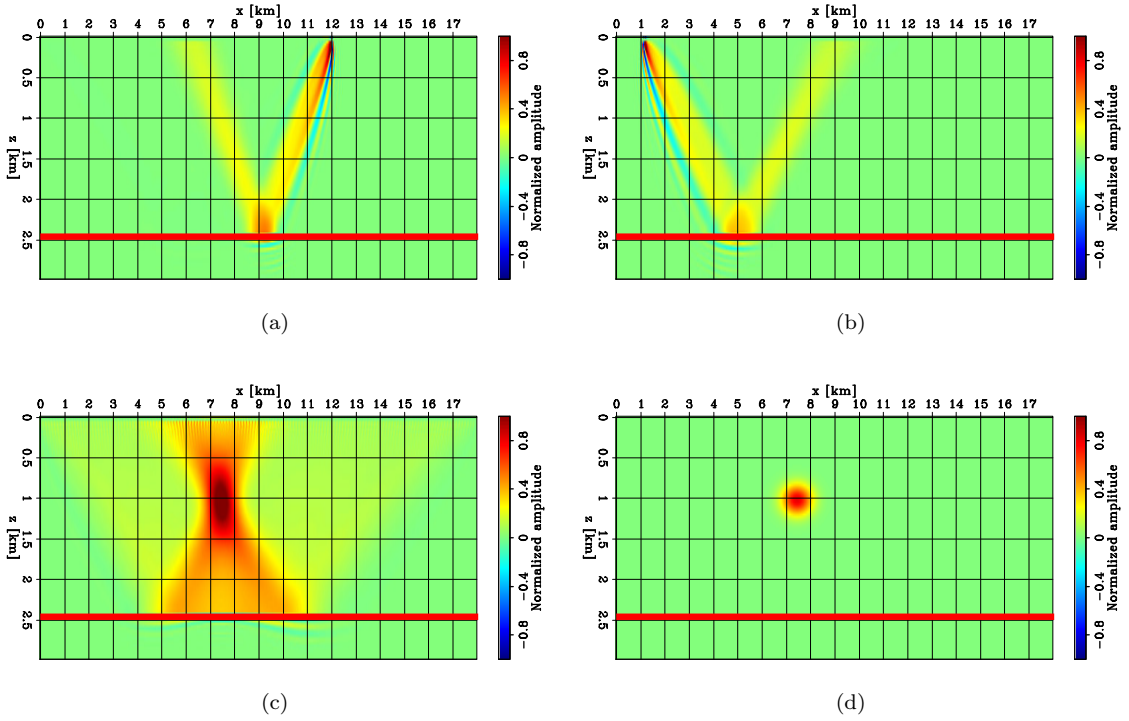


Figure 4.3: Panels showing velocity perturbations computed by the application of $\mathbf{T}^*(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})\mathbf{T}(\mathbf{m}_0, \tilde{\mathbf{p}}_\epsilon^{opt})$ to panel (d). (a) Velocity perturbation computed with a single source placed at $x = 1$ km. (b) Velocity perturbation computed with a single source placed at $x = 12$ km. (c) Velocity perturbation computed with all available sources. (d) True velocity perturbation. Panels (a)-(c) are normalized with the same value. [**NR**]

Algorithm 1: FWIME

- Select the initial model $\mathbf{m}_0$
- Select the extension type and the length of the extended axis
- Select the hyper-parameter ϵ (fixed throughout the optimization process)
- For $i = 0, \ldots, n_{iter} - 1$
  1. Compute $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_i)$
  2. Compute $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i) = \underset{\tilde{\mathbf{p}}}{\operatorname{argmin}} \frac{1}{2} \left\| \tilde{\mathbf{B}}(\mathbf{m}_i)\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_i)\right) \right\|_2^2 + \frac{\epsilon^2}{2} \left\| \mathbf{D}\tilde{\mathbf{p}} \right\|_2^2$
  3. Compute objective function value $\Phi_\epsilon(\mathbf{m}_i)$
  4. Set $\mathbf{r}_d^\epsilon(\mathbf{m}_i) = \mathbf{f}(\mathbf{m}_i) + \tilde{\mathbf{B}}(\mathbf{m}_i)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i) - \mathbf{d}^{obs}$
  5. Compute FWIME gradient $\nabla_{\mathbf{m}}\Phi_\epsilon(\mathbf{m}_i) = \mathbf{M}\left[\mathbf{B}^*(\mathbf{m}_i) + \mathbf{T}^*(\mathbf{m}_i)\right]\mathbf{r}_d^\epsilon(\mathbf{m}_i)$
  6. Compute search direction $\mathbf{s}_i$
  7. Compute step length γ_i
  8. Update model $\mathbf{m}_{i+1} = \mathbf{m}_i + \gamma_i \mathbf{s}_i$

model updates while maintaining robust convergence properties.

The computation of $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i)$ (step 2 of algorithm 1) corresponds to the variable projection step of FWIME, and is conducted iteratively by minimizing equation 3.6 with a linear conjugate gradient algorithm. The number of linear iterations needed for convergence is problem-dependent but for 3D field applications, we observe that approximately 50 iterations are sufficient. This step is equivalent to performing an extended linearized waveform inversion, and accounts for 95% to 99% of the total computational cost of FWIME. In addition, $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i)$ must be re-evaluated each time the velocity model $\mathbf{m}_i$ is modified and the objective function estimation (step 3 of algorithm 1) is thus the main computational bottleneck of the workflow. In contrast, the most computationally intensive step for conventional FWI methods corresponds to the gradient evaluation. In order to reduce the number objective function evaluations during the step-length estimation process (step 7 of algorithm 1), I employ the method proposed by Moré and Thuente (1994) (which only requires one objective function evaluation and one gradient computation to find an appropriate step length), rather than more conventional step-length estimation strategies such as parabolic line search (Press et al., 1989; Nocedal and Wright, 2006). One remaining challenge in FWIME is to improve the efficiency of step 2 of algorithm 1. Hou and Symes (2017) successfully developed a computationally cost-effective method based on an approximate inverse of the space-lag extended Born modeling operator to minimize objective functions such as the one shown in equation 3.21,

$$\chi_{\mathbf{m}}(\tilde{\mathbf{p}}) \quad = \quad \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right)\right\|_2^2. \tag{4.11}$$

However, there are three issues preventing me from directly applying the authors' method to the variable projection step in FWIME. First, their workflow was developed for a space-lag extended Born modeling operator and needs to be adapted to time-lag extension. In addition, the approximate inverse formulation is only valid when the annihilating term in equation 3.6 is null (i.e., for $\epsilon = 0$). Finally, the data residual $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$ must only contain reflected energy. In this thesis, I do not propose a solution to these issues and I leave such investigation for future work.

4.2 A model-space multi-scale approach

In this section, I provide some motivation for developing a model-space multi-scale strategy, I describe my proposed method, and I show how I incorporate it into the FWIME workflow.

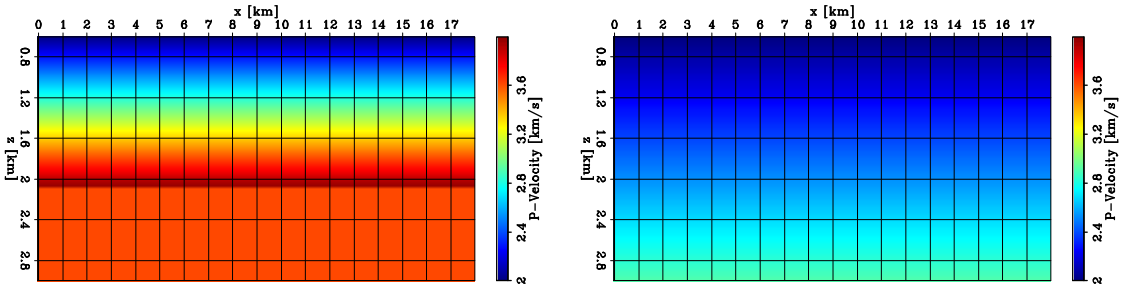
4.2.1 Motivation

For any waveform inversion, it seems crucial to accurately recover the missing long spatial-wavelength components at early stages, and then gradually increase the resolution of the model updates. In fact, it has been observed that the extent of the basin of attraction of FWI about the global minimum increases when the low-frequency component of the data is inverted in a data-space multi-scale manner (Bunks et al., 1995; Fichtner, 2010). Additionally, Mora (1989) shows the connection between the propagation direction of the source and receiver wavefields and the wavenumber updates introduced by their cross-correlation (i.e., the model scale that is updated at each iteration). Finally, Sirgue and Pratt (2004) extend this discussion and describe the connection between the data frequency content and the model updates and propose a method to select the frequency band to be inverted.

Unlike conventional methods, FWIME accepts the presence (and the simultaneous inversion) of the full data-bandwidth, which may include all wave types and all available frequencies from the start. I develop a new workflow where the data-space multi-scale strategy is substituted by a model-space multi-scale one, and the resolution of the model updates is controlled and gradually increased throughout the optimization procedure (Barnier et al., 2019). This process is achieved by considering a velocity re-parametrization on spatially adjustable non-uniform grids with the use of basic-spline (B-spline) basis functions (De Boor, 1986; Shene, 2011), instead of a finite-difference grid. However, all wavefield modeling and propagations are still conducted on conventional finite-difference grids. The FWIME workflow (algorithm 1) begins by recovering the low-wavenumber components by using a coarse-grid model representation, thereby limiting the resolution of the updates. As the inversion progresses, the grid sampling is gradually refined and the inverted model on a given grid is then used

as the initial guess for the following inversion performed on a finer/denser grid until an accurate solution is successfully recovered. The benefit of this approach is that the spline parameterization and its refinement rate provides the ability to control and gradually increase the resolution of the model-updates with iterations.

I illustrate the need to incorporate this new multi-scale approach on a numerical example where I compute the initial FWIME search direction. I design a 18 km-wide and 3 km-deep laterally invariant velocity model $\mathbf{m}_{true}$ composed of a shallow homogeneous layer, a linear $v(z)$ velocity gradient, and a sharp horizontal reflector at a depth of 2.1 km (Figure 4.4a). The initial velocity model $\mathbf{m}_0$ (Figure 4.4b) is also laterally invariant and composed of the same shallow homogeneous layer, but the velocity gradient in the deeper region is chosen to be inaccurate enough for conventional FWI to converge to a non-physical solution (the inversion results are not shown here). In addition, $\mathbf{m}_0$ does not contain any reflector. Figure 4.5 shows the velocity profiles of the true (blue curve) and initial (red curve) models.



(a) (b)

Figure 4.4: 2D panels of (a) the true and (b) initial velocity models. [**ER**]

At the surface, I place 150 source every 120 m and 600 receivers every 30 m. I generate noise-free pressure data with a two-way acoustic modeling operator and a source containing energy restricted to the 9-18 Hz range. For this numerical example, I propose to solely invert reflected energy and I apply a data-muting mask $\mathbf{M}_d$ on all shot records to mute events occurring at offsets greater than 3.0 km. Figure 4.6a shows a representative shot gather of the raw observed data $\mathbf{d}^{obs}$ (left panel), and the muted data $\mathbf{M}_d\mathbf{d}^{obs}$ (middle panel) for a shot located at $x = 8$ km. Figure 4.6c shows the muted initial data difference $\mathbf{M}_d\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\right)$.

I conduct the variable projection step of FWIME (step 2 of algorithm 1) by minimizing objective function 3.6 with 60 iterations of linear conjugate gradient and $\epsilon = 2.5 \times 10^{-7}$. I use a time-lag extension τ for $\tilde{\mathbf{p}}_\epsilon^{opt}$ with a total of 91 points sampled at 16 ms, allowing τ to range from -0.72 s to 0.72 s. Figure 4.7 shows a TLCIG extracted at $x = 9$ km from $\tilde{\mathbf{p}}_\epsilon^{opt}$. The event with strong energy located at negative time lags (white arrow) corresponds to the mapping of the reflection from the sharp horizontal interface (white arrow in Figure 4.6c) into the extended space of $\tilde{\mathbf{p}}_\epsilon^{opt}$. As

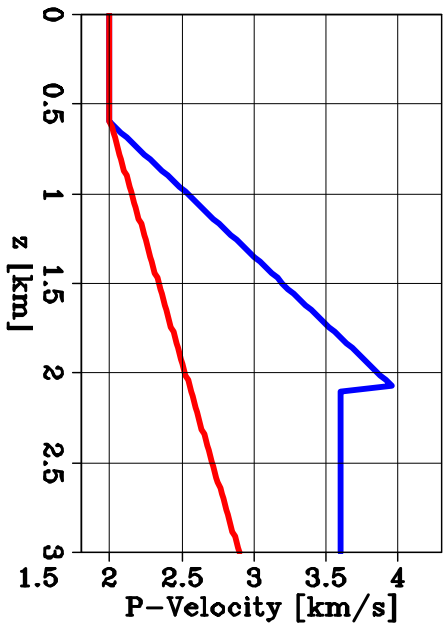
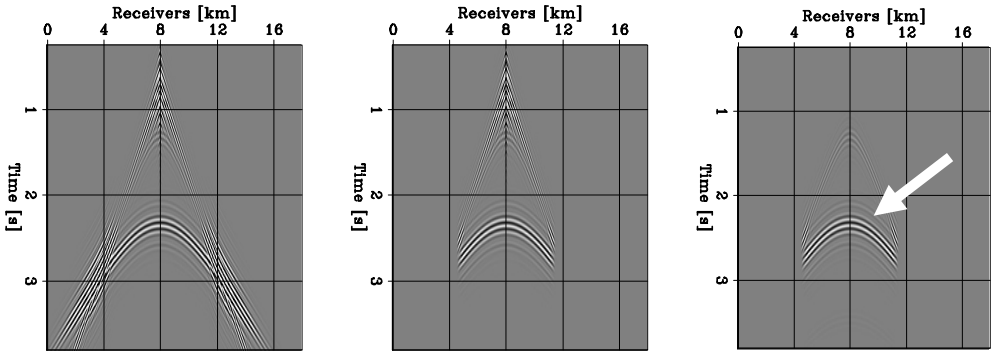


Figure 4.5: 1D profiles of the true model (blue curve), and the initial model (red curve). [**ER**]



(a) (b) (c)

Figure 4.6: Representative shot gathers generated by a source located at $x = 8$ km. (a) Observed data with no muting applied, $\mathbf{d}^{obs}$. (b) Observed data after muting, $\mathbf{M}_d\mathbf{d}^{obs}$. (c) Initial data difference after muting, $\mathbf{M}_d(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0))$. All panels are displayed with the same grayscale. [**ER**]

expected, the position of its maximum energy is shifted away from the physical axis (where $\tau = 0$ s), which in this example indicates that the initial velocity is lower than the true velocity. Additionally, Figure 4.7 illustrates how $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ contains crucial kinematic information and shows the importance of using an extended perturbation with a large-enough extension (otherwise the information would be lost).

Figure 4.8a shows the Born component of the initial FWIME search direction. As anticipated, it is similar to the initial FWI search direction (Figure 4.8d). In both cases, the position of the sharp interface is too shallow (due to the velocity error within the initial model). Figure 4.8b displays the initial tomographic search direction, which seems promising by comparing it to the true search direction $\mathbf{s}_{true}$ shown in Figure 4.8e. Its amplitude, however, is much smaller than the Born component (Figures 4.8a and 4.8b are normalized by different values for display purposes). Finally, Figure 4.8c represents the total FWIME search direction $\mathbf{s}_{total}$, obtained by summing the two panels on the first row. Even though the tomographic component manages to accurately recover the missing low wavenumbers of the velocity, its amplitude is overwhelmed by the Born update which will likely lead the optimization scheme to a local minimum.

One potential solution would be to manually control the relative amplitudes of the two components by assigning more weight to the tomographic gradient at early stages, and gradually adjusting the relative weights throughout the inversion process. However, this approach would be very user intensive and challenging to automatize. Alternatively, one could apply a spatial smoothing filter to manually limit the spatial resolution of the velocity updates (Biondi and Almomin, 2014). While this rather standard filtering approach may be valid in many practical situations, it is not consistent in the context of optimization. We prefer to consider a re-parametrization of the velocity model with B-spline basis functions which naturally filters high-wavenumber effects and allows a natural and more spatially flexible refinement, thanks to the subdivision property of B-splines.

4.2.2 Velocity parametrization using B-splines

B-spline basis functions are commonly used in computer-aided design and graphic to draw smooth curves and surfaces passing in the vicinity of a set of control points, also referred to as "spline nodes" (Shene, 2011). I employ these functions to represent seismic velocity models on a coarse and potentially spatially non-uniform grid (referred to as the spline grid) and I create a linear operator (referred to as the spline operator) that maps them onto a finer grid (the finite-difference propagation grid). This mapping does not require to fit the control points exactly, and is therefore technically not an interpolation method. However, releasing this constrain provides great flexibility for the spline grid positioning and provides the ability to take into account prior geological knowledge of the Earth's subsurface. Additionally, unlike radial basis functions (RBF), B-spline functions have very limited support (i.e., they are non-zero only locally), which makes them computationally efficient for 3D applications.

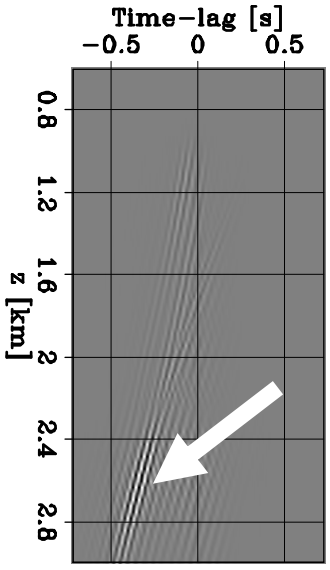


Figure 4.7: TLCIG extracted at $x = 9$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_0)$ computed at the initial stage of the FWIME workflow. [**CR**]

Mathematical framework

I follow the theory described in Shene (2011) and I modify it for my application. I define the spline operator of order p as the following linear mapping $\mathbf{S}^{(p)}$:

$$\begin{aligned} \mathbf{S}^{(p)} : \mathbb{R}^{N_{m_c}} &\mapsto \mathbb{R}^{N_{m_f}} \\ \mathbf{m}^c &\mapsto \mathbf{m}^f = \mathbf{S}^{(p)}\, \mathbf{m}^c. \end{aligned} \tag{4.12}$$

$\mathbf{m}^f \in \mathbb{R}^{N_{m_f}}$ represent the seismic velocity model parametrized on a "fine" uniform finite-difference grid, where $N_{m_f} = N_x \times N_y \times N_z$ represents the total number of finite-difference grid points. $\mathbf{m}^c \in \mathbb{R}^{N_{m_c}}$ is a representation of the velocity model on a predefined coarse and (potentially) non-regularly spaced grid, and N_{m_c} is the number of points on the coarse grid (i.e., the number of spline nodes or control points). The adjoint of the spline operator is therefore a mapping from the finite-difference grid into the coarse grid,

$$\begin{aligned} \left(\mathbf{S}^{(p)}\right)^* : \mathbb{R}^{N_{m_f}} &\mapsto \mathbb{R}^{N_{m_c}} \\ \mathbf{m}^f &\mapsto \mathbf{m}^c = \left(\mathbf{S}^{(p)}\right)^*\, \mathbf{m}^f. \end{aligned} \tag{4.13}$$

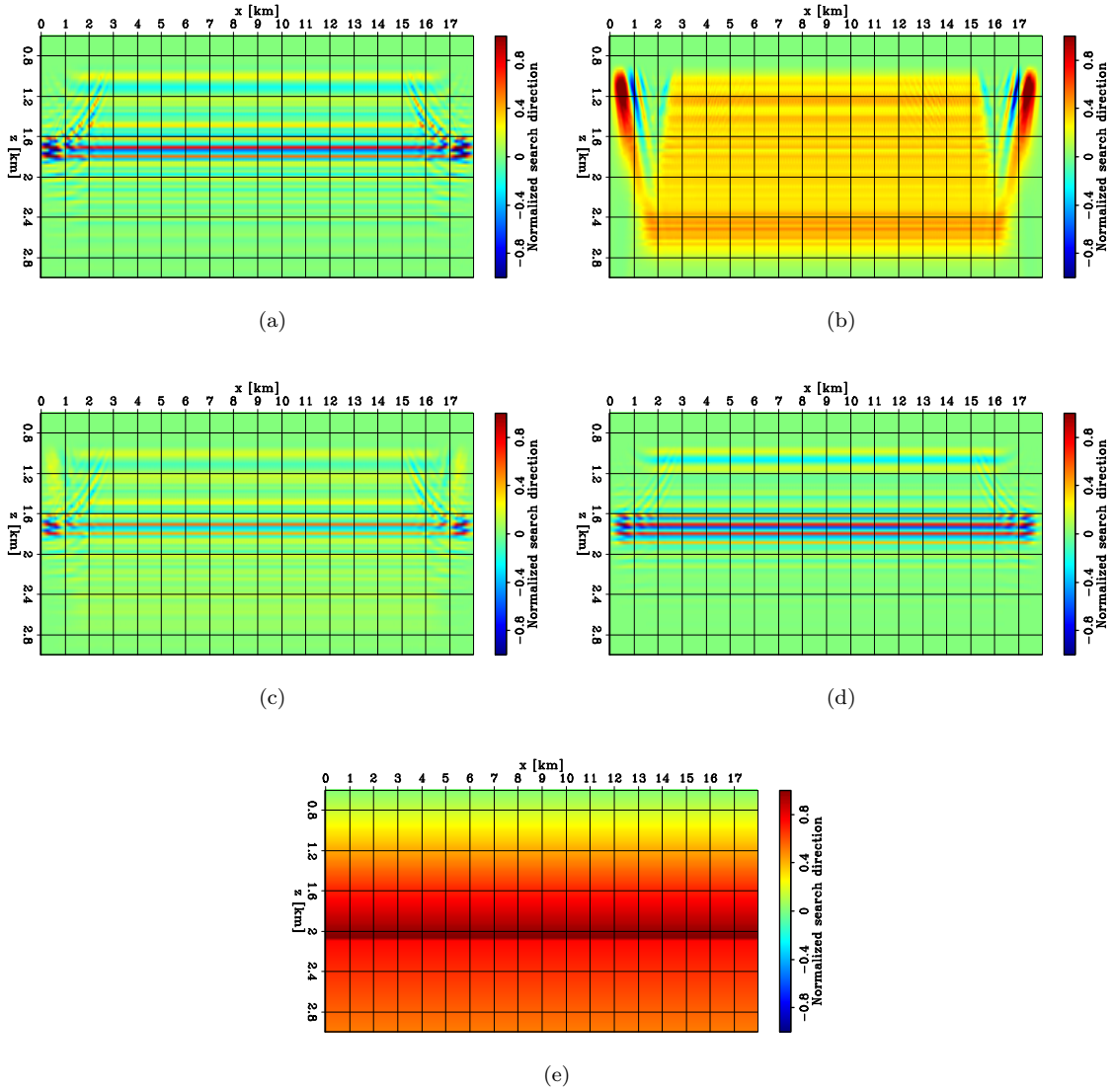


Figure 4.8: 2D panels of initial search directions. (a) FWIME Born component. (b) FWIME tomographic component. (c) FWIME total search direction (obtained by summing (a) and (b)), $\mathbf{s}_{total}$. (d) FWI search direction. (e) True search direction, $\mathbf{s}_{true}$. For display purposes, panel (b) is normalized with a smaller value than in panel (a), and the amplitude of (b) is much smaller than the one of (a). [**CR**]

I denote by $\mathcal{G}_c \subset \mathbb{R}^{N_{m_c}}$ and $\mathcal{G}_f \subset \mathbb{R}^{N_{m_f}}$ the set of points composing the spline and finite-difference grids, respectively. The entries of the spline operator $\mathbf{S}^{(p)} \in \mathbb{R}^{N_{m_f} \times N_{m_c}}$ are computed using B-spline basis functions of order p whose expressions are given by the Cox-de Boor recursion formula (De Boor, 1986). They depend both on the interpolation technique (i.e., the type of basis functions employed) and on the way the coarse grid is arranged. In the following, I set $p = 3$ to ensure that the reconstructed functions in the output space are $\mathcal{C}^2$ within the area of interest (i.e., where wavefields are modeled). To simplify notations, and I do not explicitly write the dependency of the spline operators on p.

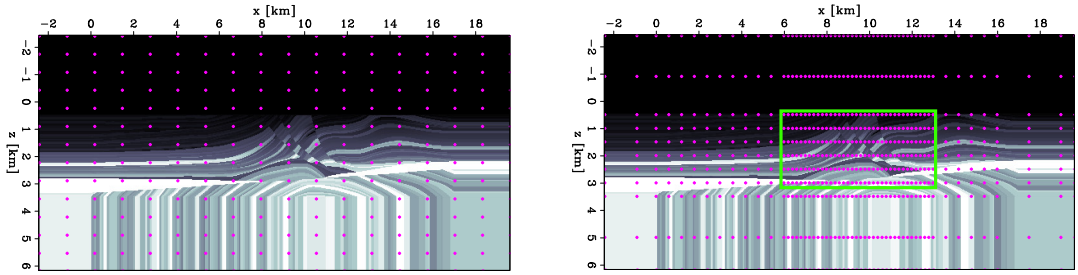
In the inversion scheme, $\mathbf{m}^c$ is now the unknown set of parameters I wish to recover, and its entries should be seen as weights rather than seismic velocity values. Hence, their actual magnitudes/units are not informative nor directly interpretable. However, once $\mathbf{m}^c$ is known, the corresponding velocity field can be inferred by simply applying the forward spline operator $\mathbf{S}$ to map the inverted model onto the finite-difference grid. To gain better insight on this mapping, I express the velocity value at the i^{th} point on the finite-difference grid as a function of the model values at the spline nodes. For that, I examine the i^{th} row of equation 4.12, which is given by

$$\mathbf{m}_i^f = \sum_{k=1}^{N_{m_c}} S_{ik}\, \mathbf{m}_k^c. \tag{4.14}$$

Equation 4.14 simply indicates that the velocity value at the i^{th} point on the finite-difference grid can be expressed as a linear combination of the weights at each spline node. For example, S_{ik} is the contribution of spline node k to the velocity value computed at the i^{th} point on the finite-difference grid. In addition, B-spline functions possess a compact support which makes the corresponding mapping operator $\mathbf{S}$ very sparse. For a given point on the finite-difference grid, only a maximum of six (in 2D) and nine (in 3D) coefficients are non-zero while still ensuring that the reconstructed velocity function (on the finite-difference grid) is $\mathcal{C}^2$ (when $p = 3$). This implies that at most nine terms from the sum in the right-side of equation 4.14 will contribute to the computation of $\mathbf{m}_i^f$. More stringent conditions on the level of smoothness and continuity (i.e., for $p > 3$) will increase the number of non-zero coefficients. In comparison, RBFs are more computationally intensive but offer more flexibility on the coarse grid disposition, which could be useful when representing models containing sharp velocity interfaces/contrasts embedded in a smooth (low-wavenumber) surrounding (Dahlke, 2019). I plan to investigate their use in future work.

I illustrate the application and the properties of my proposed spline operator $\mathbf{S}$ by parametrizing a 2D velocity field based on the Marmousi2 benchmark model on two different spline grids. The underlying assumption for such model representation is that denser spline grids will be able to better represent higher-wavenumber features from the velocity field (i.e., reconstruct higher-resolution models), while coarser spline grids will tend to spatially smooth the velocity model (i.e., behave as a

low-pass filter in the wavenumber domain). Figure 4.9 shows two spline grid dispositions (the pink dots correspond to the spline nodes) overlaid on the Marmousi2 velocity model displayed with the absorbing boundaries used for the finite-difference propagation. If no prior geological information is known, I can simply choose to represent the velocity model on a regularly sampled spline grid, as shown in Figure 4.9a. For this regular mesh, the distance between two consecutive nodes is set to 0.7 km and 1.2 km in the vertical and horizontal directions, respectively. I can also use the fact that there will likely be very little model updates in the absorbing boundaries and therefore reduce the spline grid density in this region of the model. Furthermore, I can leverage prior geological information if the velocity model is likely to contain high-wavenumber features within a specific region, and adapt the spline mesh by increasing the node density within the zone of interest, as shown by the green box in Figure 4.9b. As previously mentioned, the spline nodes cannot be arbitrarily placed and must be arranged in a net disposition. However, each direction can have its own irregular sampling, which gives plenty of flexibility.



(a) (b)

Figure 4.9: 2D panels of spline meshes overlaid on the Marmousi2 velocity model. (a) Regular spline mesh. (b) Irregular spline mesh. The pink dots indicate the position of spline nodes. [**NR**]

Figure 4.10 shows the spatial smoothing effect resulting from sequentially applying operator $\mathbf{S}$ and then $\mathbf{S}^*$ to the Marmousi2 velocity model $\mathbf{m}_{true}$ shown in Figure 4.10a (displayed with the absorbing boundary layers used for the finite-difference propagation). Figures 4.10b and 4.10c show the application of $\mathbf{S}^*$ on $\mathbf{m}_{true}$ for the regular mesh (Figure 4.9a) and irregular mesh (Figure 4.9b), respectively. The vertical and horizontal labels correspond to the spline node indices, and the actual value at each grid point can not be interpreted as a velocity field. Figures 4.10d and 4.10e show the application of the forward mapping $\mathbf{S}$ on the panels shown in Figures 4.10b and 4.10c, respectively. As expected, the effect of mapping the velocity model onto the "regular" spline grid and then back to the finite-difference grid introduces a strong spatially-homogeneous smoothing effect (Figure 4.10d). Moreover, increasing the grid density in certain regions of the model allows the mapping to preserve high-resolution features, as shown in Figure 4.10e (central region of the model). Therefore, one of the main advantages of the B-spline parametrization (compared to more conventional smoothing

methods based on wavenumber-domain filtering) is its ability to easily and efficiently apply non-uniform spatial smoothing for different regions of the velocity model.

4.2.3 An alternate multi-scale workflow

Another advantage of this general framework is that it may also be easily and elegantly implemented in many types of waveform inversion techniques (Barnier et al., 2019), as well as for different parametrization methods (e.g., RBFs). Here, I incorporate the proposed B-spline parametrization strategy into the FWIME workflow. In the following, I assume I have already constructed a coarse spline grid $\mathcal{G}_c$, a finite-difference grid $\mathcal{G}_f$, and the corresponding spline operator $\mathbf{S}: \ \mathcal{G}_c \mapsto \mathcal{G}_f$ as defined in section 4.2.2.

A modified FWIME objective function

Recall that the FWIME objective function defined on $\mathcal{G}_f$ (equation 3.5) is given by

$$\Phi_\epsilon(\mathbf{m}^f) \quad = \quad \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}^f) + \tilde{\mathbf{B}}(\mathbf{m}^f)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}^f) - \mathbf{d}^{obs}\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}^f)\right\|_2^2, \tag{4.15}$$

where $\mathbf{m}^f$ is the velocity model represented on $\mathcal{G}_f$. Therefore, $\Phi_\epsilon : \mathcal{G}_f \mapsto \mathbb{R}^+$. I now modify equation 4.15 by re-parametrizing Φ_ϵ on the coarse spline grid $\mathcal{G}_c$, I introduce the spline operator $\mathbf{S}$, and I substitute $\mathbf{m}^f$ by $\mathbf{S}\mathbf{m}^c$. The new FWIME objective function is now given by

$$\tilde{\Phi}_\epsilon(\mathbf{m}^c) \quad = \quad \frac{1}{2}\left\|\mathbf{f}(\mathbf{S}\mathbf{m}^c) + \tilde{\mathbf{B}}(\mathbf{S}\mathbf{m}^c)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{S}\mathbf{m}^c) - \mathbf{d}^{obs}\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{S}\mathbf{m}^c)\right\|_2^2, \tag{4.16}$$

where $\tilde{\Phi}_\epsilon : \mathcal{G}_c \mapsto \mathbb{R}^+$, and

$$\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{S}\mathbf{m}^c) = \underset{\tilde{\mathbf{p}}}{\operatorname{argmin}} \ \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{S}\mathbf{m}^c)\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{S}\mathbf{m}^c)\right)\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}\right\|_2^2. \tag{4.17}$$

In equation 4.16, the dimension of the search space (i.e., model space) has been reduced from $\mathbb{R}^{N_{m_f}}$ to $\mathbb{R}^{N_{m_c}}$ (N_{m_c} is usually much smaller than N_{m_f}). It is important to note that in equations 4.16 and 4.17, $\tilde{\mathbf{p}}_\epsilon^{opt}$ is never parametrized on the spline grid, but rather on the (finer) finite-difference grid because it may contain all wavenumber components at any stage of the inversion process, as explained in section 3.3.2. In fact, $\tilde{\mathbf{p}}_\epsilon^{opt}$ is the mapping into the extended space of all the events present in the observed data $\mathbf{d}^{obs}$ that our modeling $\mathbf{f}(\mathbf{S}\mathbf{m}^c)$ failed to predict. These events can (and will likely) include all types of waves, such as refractions and reflections.

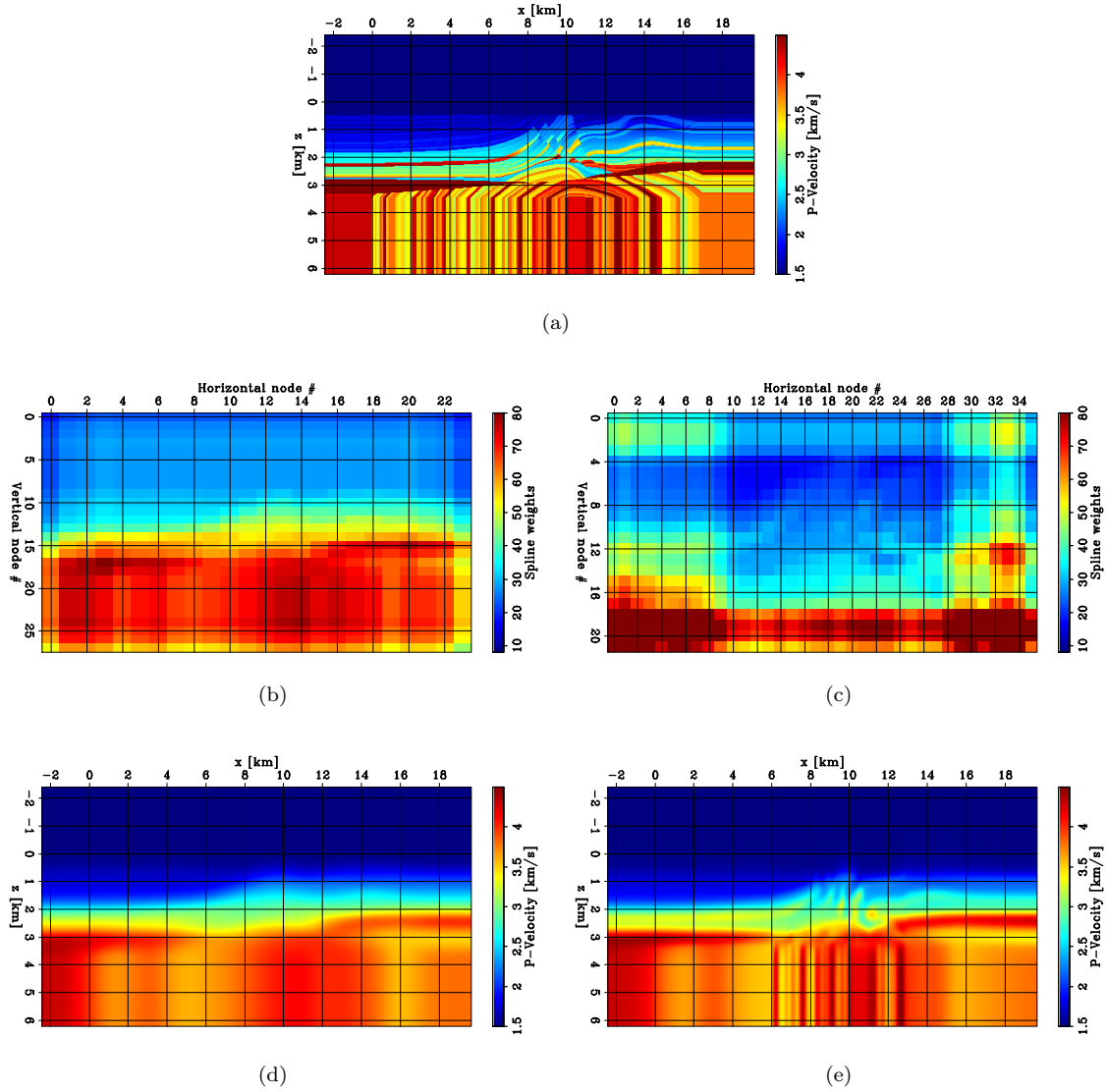


Figure 4.10: 2D panels of the Marmousi2 velocity model after sequential applications of the spline operator $\mathbf{S}$ and its adjoint, $\mathbf{S}^*$. (a) True velocity model $\mathbf{m}_{true}$. (b) Application of $\mathbf{S}^*$ on $\mathbf{m}_{true}$ for the spline mesh shown in Figure 4.9a. (c) Application of $\mathbf{S}^*$ on $\mathbf{m}_{true}$ for the spline mesh shown in Figure 4.9b. (d) Application of $\mathbf{S}$ on panel (b). (e) Application of $\mathbf{S}$ on panel (c). [**ER**]

Gradient computation

The gradient of the modified FWIME objective function is obtained by applying the chain rule, which results in mapping the conventional FWIME gradient from the finite-difference grid onto the spline grid,

$$\nabla_{\mathbf{m^c}} \tilde{\Phi}_\epsilon(\mathbf{m^c}) \quad = \quad \mathbf{S}^* \, \nabla_{\mathbf{m}^f} \Phi(\mathbf{m}^f), \tag{4.18}$$

and the phase-only formulation follows an analogous modification,

$$\nabla_{\mathbf{m^c}} \tilde{\Phi}^n_\epsilon(\mathbf{m^c}) \quad = \quad \mathbf{S}^* \, \nabla_{\mathbf{m}^f} \Phi^n(\mathbf{m}^f). \tag{4.19}$$

To illustrate the usefulness of such strategy on the FWIME gradient/search direction, I re-visit the numerical example proposed in section 4.2.1. I generate a spline mesh regularly sampled at 0.3 km in both directions (the finite-difference grid is sampled at 0.03 km in both directions), and I compute the new FWIME search direction (equation 4.18). Figures 4.11a-c show the new Born, tomographic, and total FWIME search directions displayed on the finite-difference grid, which are obtained by applying operator $\mathbf{SS}^*$ to the panels shown in Figures 4.8a-c, respectively. The FWIME search direction is now much more promising: it is solely guided by the tomographic component and accurately captures the missing low wavenumbers. In this numerical example, the spline parametrization behaves in a similar fashion as a low-pass filter (in the spatial frequency domain) and removes the undesired high-wavenumber features introduced by the Born update (Figure 4.8a).

Initial model computation

In order to minimize equation 4.16, I first construct an initial velocity model defined on the coarse grid, $\mathbf{m}^c_0 \in \mathcal{G}_c$. Naturally, $\mathbf{m}^f_0$ is designed on $\mathcal{G}_f$, and then mapped onto $\mathcal{G}_c$. This is achieved by finding the unique minimizer $\mathbf{m}^c_0 \in \mathcal{G}_c$ that satisfies the following equation:

$$\mathbf{m}^c_0 \quad = \quad \underset{\mathbf{m}^c \in \mathcal{G}_c}{\operatorname{argmin}} \, \frac{1}{2} \left\| \mathbf{S}\mathbf{m}^c - \mathbf{m}^f_0 \right\|_2^2. \tag{4.20}$$

For a well-chosen grid pair $(\mathcal{G}_c, \mathcal{G}_f)$, it can be shown that operator $\mathbf{S}^*\mathbf{S}$ is invertible, which is usually the case when the finite-difference grid is much more densely sampled than the spline grid. Therefore, this property ensures the existence (and uniqueness) of $\mathbf{m}^c_0$.

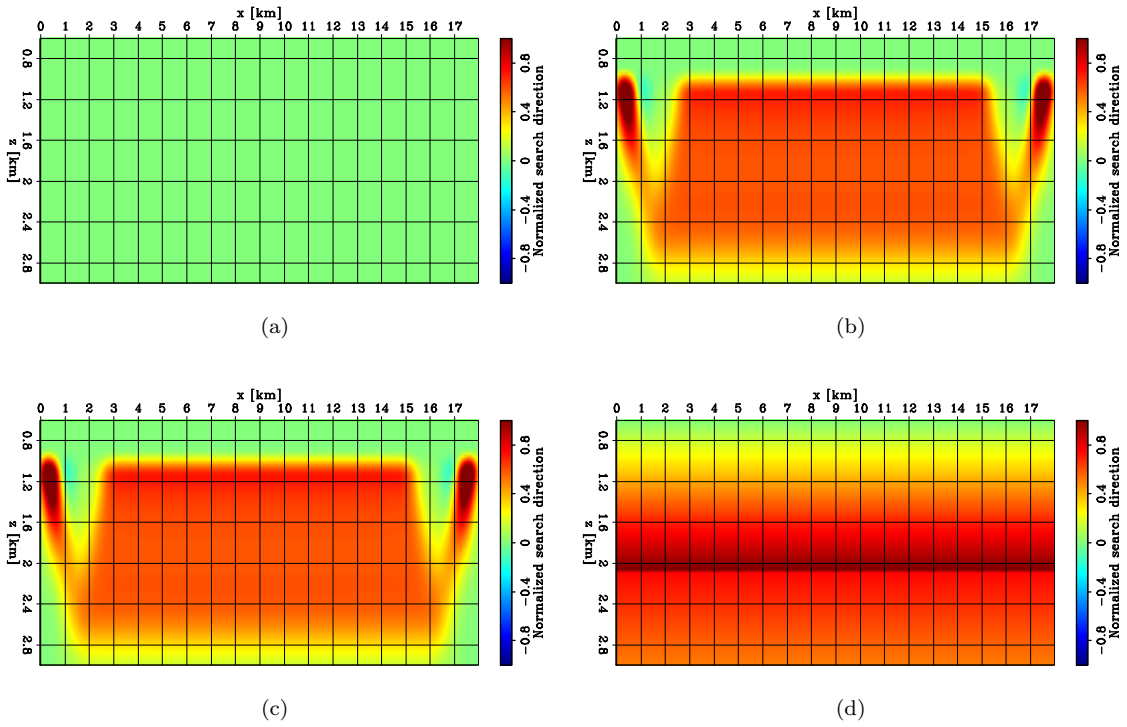


Figure 4.11: 2D panels of initial search directions. (a) FWIME Born search direction obtained by applying $\mathbf{SS}^*$ to the panel shown in Figure 4.8a. (b) FWIME tomographic search direction obtained by applying $\mathbf{SS}^*$ to the panel shown in Figure 4.8b. (c) FWIME search direction obtained by summing panels (a) and (b). (d) True search direction. Panels (a), (b) and (c) are normalized by the same value and displayed with the same color scale. [**CR**]

Inversion workflow

The workflow I employ to minimize equation 4.16 is summarized in algorithm 2 (assuming a fixed spline grid). Note that the computations in steps (f) through (i) are conducted on the spline grid. When $\tilde{\Phi}$ has been successfully optimized, the inverted model $\mathbf{m}^c_{opt}$ can be mapped onto the finite-difference grid for better interpretation/visualization by applying operator $\mathbf{S}$,

$$\mathbf{m}^f_{opt} = \mathbf{S}\mathbf{m}^c_{opt}. \tag{4.21}$$

Algorithm 2: FWIME with B-spline parametrization

1. Select a finite-difference grid $\mathcal{G}_f$
2. Construct a coarse grid $\mathcal{G}_c$ and its mapping operator $\mathbf{S}$
3. Select the extension type and the length of the extended axis
4. Select the hyper-parameter ϵ (fixed throughout the optimization process)
5. Design an initial model on the finite-difference grid, $\mathbf{m}^f_0$
6. Compute the initial model on the coarse grid, $\mathbf{m}^c_0 = \underset{\mathbf{m}^c \in \mathcal{G}_c}{\operatorname{argmin}} \frac{1}{2} \left\| \mathbf{S}\mathbf{m}^c - \mathbf{m}^f_0 \right\|_2^2$
7. For $i = 0, ..., n_{iter} - 1$
   (a) Map current model estimate onto the finite-difference grid, $\mathbf{m}^f_i = \mathbf{S}\,\mathbf{m}^c_i$
   (b) Compute $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}^f_i)$
   (c) Compute $\tilde{\mathbf{p}}^{opt}_\epsilon(\mathbf{m}^f_i) = \underset{\tilde{\mathbf{p}}}{\operatorname{argmin}} \frac{1}{2} \left\| \tilde{\mathbf{B}}(\mathbf{m}^f_i)\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}^f_i)\right) \right\|_2^2 + \frac{\epsilon^2}{2} \left\| \mathbf{D}\tilde{\mathbf{p}} \right\|_2^2$
   (d) Set $\mathbf{r}^\epsilon_d(\mathbf{m}^f_i) = \mathbf{f}(\mathbf{m}^f_i) + \tilde{\mathbf{B}}(\mathbf{m}^f_i)\tilde{\mathbf{p}}^{opt}_\epsilon(\mathbf{m}^f_i) - \mathbf{d}^{obs}$
   (e) Compute objective function value $\Phi_\epsilon(\mathbf{m}^f_i)$
   (f) Compute conventional FWIME gradient with respect to the model parametrized on the spline grid
   $$\nabla_{\mathbf{m}^c}\Phi_\epsilon(\mathbf{m}^c_i) = \mathbf{S}^*\mathbf{M}\left[\mathbf{B}^*(\mathbf{m}^f_i) + \mathbf{T}^*(\mathbf{m}^f_i)\right]\mathbf{r}^\epsilon_d(\mathbf{m}^f_i)$$
   (g) Compute search direction $\mathbf{s}^c_i \in \mathcal{G}_c$
   (h) Compute step length γ_i
   (i) Update model on $\mathcal{G}_c$, $\mathbf{m}^c_{i+1} = \mathbf{m}^c_i + \gamma_i \mathbf{s}^c_i$

As described in section 4.2.1, the inversion shown in algorithm 2 can be incorporated into a model-space multi-scale approach where the spline grid is gradually refined throughout the optimization

process (the finite-difference grid remains fixed). I start the FWIME workflow with a coarse spline grid $\mathcal{G}_{c_0}$ (along with its corresponding spline operator $\mathbf{S}_0$). I minimize the FWIME objective function for that particular spline grid (algorithm 2), and the inverted model $\mathbf{m}_{opt}^{c_0}$ is then used as initial guess for the following inversion performed on the next denser grid $\mathcal{G}_{c_1}$. This multi-scale process is repeated n_g times (where n_g is the number of coarse grids) until the inverted model is satisfactory, or when the coarse grid $\mathcal{G}_{c_{n_g}}$ coincides with the finite-difference grid. The hyper-parameter ϵ stays fixed, even throughout the spline grid refinements. I summarize this multi-scale process in algorithm 3.

Algorithm 3: FWIME with a model-space multi-scale approach

---

1. Select a finite-difference grid $\mathcal{G}_f$
2. Construct a collection of n_g spline grids $\{\mathcal{G}_{c_i}\}_{0\leq i<n_g}$ and their spline operators $\{\mathbf{S}_i\}_{0\leq i<n_g}$
3. Design an initial model on the finite-difference grid, $\mathbf{m}_0^f$
4. Compute the initial model on the initial coarse grid $\mathcal{G}_{c_0}$,

$$\mathbf{m}_0^{c_0} = \underset{\mathbf{m}^{c_0}\in\mathcal{G}_{c_0}}{\operatorname{argmin}} \frac{1}{2}\left\|\mathbf{S}_0\mathbf{m}^{c_0} - \mathbf{m}_0^f\right\|_2^2$$

5. Select the extension type and the length of the extended axis
6. Select the hyper-parameter ϵ
7. For $i = 0, ..., n_g - 1$
   (a) Minimize the FWIME objective function on $\mathcal{G}_i$ using $\mathbf{m}_0^{c_i}$ as initial guess and by applying step 7 of algorithm 2, and obtain $\mathbf{m}_{opt}^{c_i}$
   (b) Map $\mathbf{m}_{opt}^{c_i}$ onto the finite-difference grid, $\mathbf{m}_{opt,i}^f = \mathbf{S}_i\,\mathbf{m}_{opt}^{c_i}$
   (c) Convert the FWIME inverted model on spline grid $\mathcal{G}_i$ into a model parametrized on the new spline grid, $\mathcal{G}_{i+1}$:

$$\mathbf{m}_0^{c_{i+1}} = \underset{\mathbf{m}^{c_{i+1}}\in\mathcal{G}_{c_{i+1}}}{\operatorname{argmin}} \frac{1}{2}\left\|\mathbf{S}_{i+1}\mathbf{m}^{c_{i+1}} - \mathbf{m}_{opt,i}^f\right\|_2^2$$

   (d) Use $\mathbf{m}_0^{c_{i+1}}$ as initial guess for the inversion on $\mathcal{G}_{i+1}$

4.3 Numerical examples

I design three 2D synthetic examples where I illustrate FWIME's ability to accurately and automatically invert simple cycle-skipped datasets composed of one specific type of wave. My goal is to carefully analyze and show the reader how each wave mode is inverted with the exact same

algorithm, without the need to filter/select any specific event from the dataset. In each case, conventional data-space multi-scale FWI converges to unsatisfactory solutions. In the first example, I simulate a borehole experiment and the dataset solely contains transmitted waves. Then, I re-visit a similar experiment as the one proposed by Mora (1989) where reflection data containing complex waveforms are generated. Finally, I invert a dataset only composed of diving waves. For all three experiments, I generate and invert noise-free pressure data with the same two-way acoustic isotropic constant-density finite-difference propagator.

4.3.1 Inversion of transmitted data

I conduct a transmission experiment where I place 20 sources every 50 m inside a 1 km-deep vertical borehole, and 100 receivers every 10 m in a second identical borehole. The distance between the two boreholes is 1 km, and the true velocity model is uniform and set to 2.5 km/s. Figure 4.12 shows a schematic representation of the experiment setup.

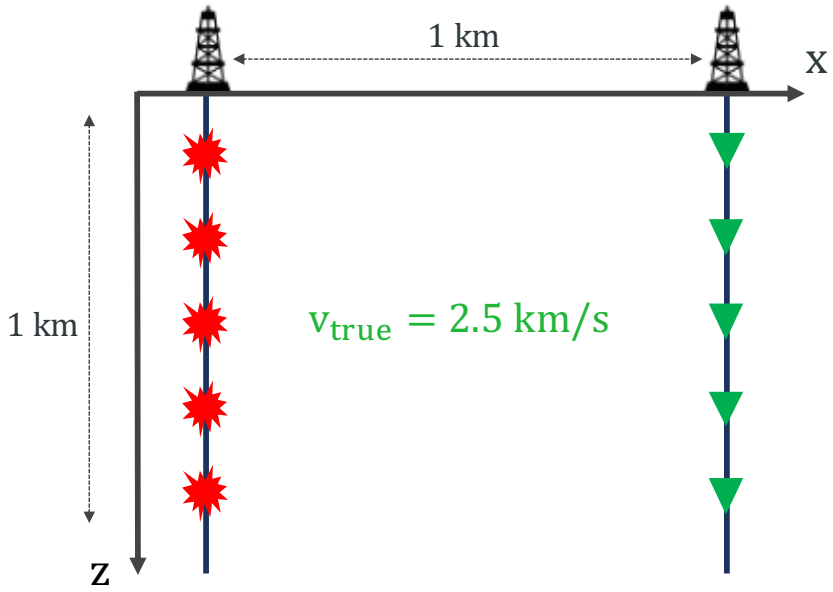
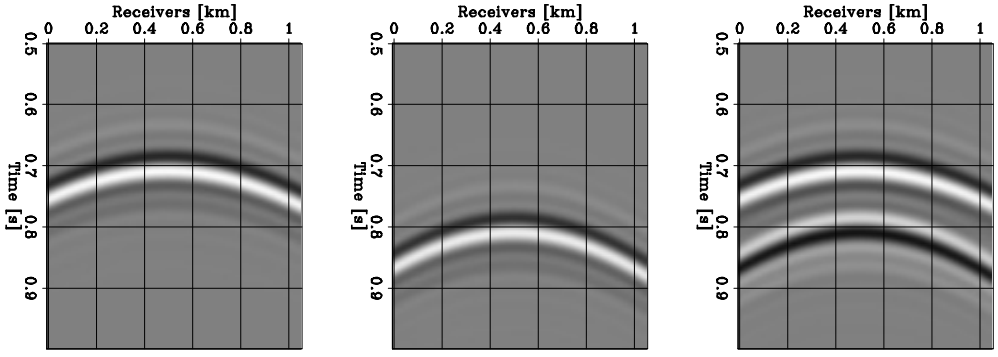


Figure 4.12: Schematic diagram illustrating the experiment setup for my proposed numerical example. [**NR**]

I generate the dataset with a finite-difference grid spacing of 10 m in both directions. The frequency spectrum of the source is strictly limited to the 9-35 Hz range. The initial velocity model $\mathbf{m}_0$ is uniform and set to 2.0 km/s. Figures 4.13 shows the observed data $\mathbf{d}^{obs}$, the initial data prediction $\mathbf{f}(\mathbf{m}_0)$, and the initial data difference $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ for a shot gather generated by a source placed at $z = 0.5$ km in the left borehole. As expected, conventional data-space

multi-scale FWI converges to a local minimum, and the final FWI data-residuals are cycle-skipped (Figures 4.14b and c).

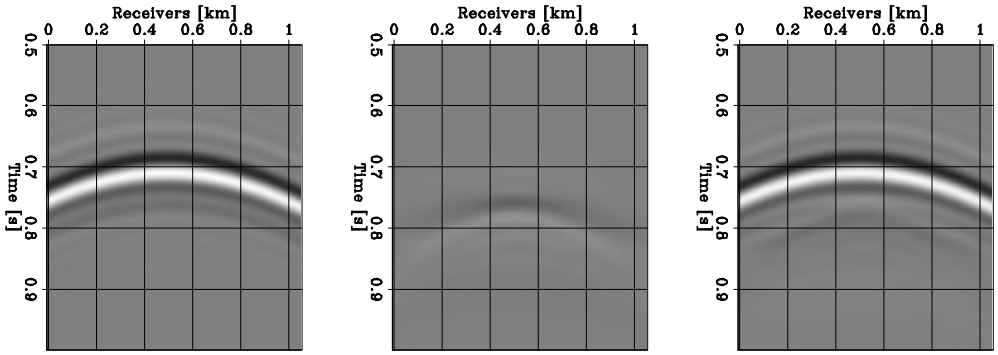


(a) (b) (c)

Figure 4.13: Representative shot gathers generated by a source placed at $z = 0.5$ km in the left borehole. (a) Observed data, $\mathbf{d}^{obs}$. (b) Initial prediction, $\mathbf{f}(\mathbf{m}_0)$. (c) Initial data difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**ER**]

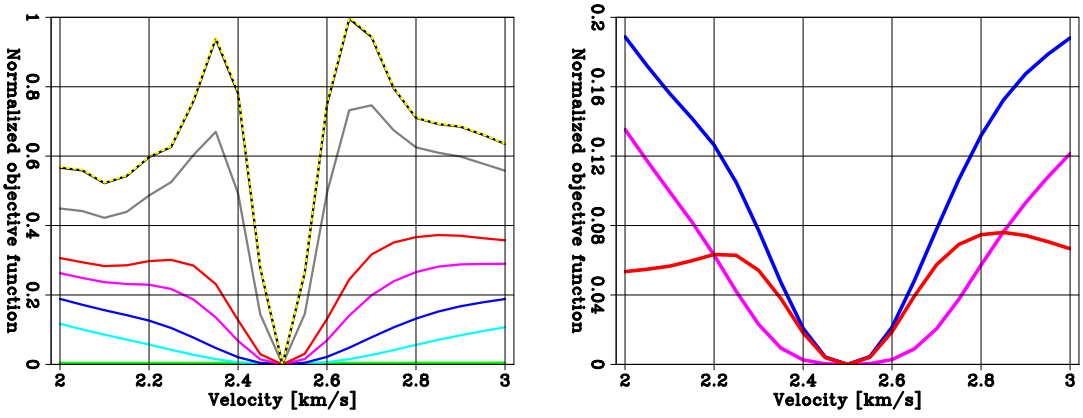
I sample both FWI and FWIME objective functions (using the full data bandwidth in both cases) for uniform velocity models ranging from 2.0 km/s to 3.0 km/s by increments of 0.05 km/s, and for seven ϵ-values ranging from $\epsilon_1 = 0$ to $\epsilon_7 = 1.0\times10^{-4}$ (Figure 4.15a). As expected, the FWI objective function presents local minima (yellow dashed curve in Figure 4.15a), but for certain ϵ-values, the FWIME objective function is monotonically decreasing toward the global solution (dark- and light-blue curves in Figure 4.15a). For these ϵ-values, the FWIME formulation managed to remove all local minima (for this range of models) and guarantees global convergence for gradient-based methods when inverting a scalar parameter (i.e., by assuming the true model is homogeneous). Figure 4.15b displays the three components of the FWIME objective function computed with $\epsilon = 1.5 \times 10^{-6}$, which corresponds to the dark-blue curve in Figure 4.15a. The data-fitting component has been convexified (pink curve), but local minima are now present within the annihilating component (red curve). However, the total objective function is free of local minima. As expected, when $\epsilon = 0$, the FWIME objective function is approximately constant and equal to zero (green curve in Figure 4.15a): the data-correcting term satisfies equation 3.23, and the FWIME data-fitting term vanishes for all velocity models $\mathbf{m}$. Conversely, as the ϵ-value increases, the FWIME objective function converges pointwise to the FWI objective function, which numerically illustrates the property demonstrated in Appendix B. In fact, for $\epsilon_7 = 1.0 \times 10^{-4}$, the FWI and FWIME objective functions are nearly identical (solid black curve and yellow dashed curve in Figure 4.15a).

I set $\epsilon = 1.5 \times 10^{-6}$ and I conduct 100 iterations of the FWIME workflow by simultaneously inverting the full data bandwidth, starting with the same uniform velocity model $\mathbf{m}_0$ set to 2.0



(a) (b) (c)

Figure 4.14: Representative shot gathers generated by a source placed at $z = 0.5$ km in the left borehole. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data computed with the final FWI inverted model, $\mathbf{f}(\mathbf{m}_{FWI})$. (c) Final FWI data residual, $\Delta\mathbf{d}(\mathbf{m}_{FWI}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWI})$. All panels are displayed with the same grayscale. [**CR**]



(a) (b)

Figure 4.15: Normalized objective functions computed for homogeneous models ranging from 2.0 km/s to 3.0 km/s. (a) FWIME objective functions computed with increasing ϵ-values: $\epsilon_1 = 0.0$ (green curve), $\epsilon_2 = 2.0 \times 10^{-7}$ (light-blue curve), $\epsilon_3 = 1.5 \times 10^{-6}$ (dark-blue curve), $\epsilon_4 = 7.5 \times 10^{-6}$ (pink curve), $\epsilon_5 = 1.0 \times 10^{-5}$ (red curve), $\epsilon_6 = 5.0 \times 10^{-5}$ (grey curve), and $\epsilon_7 = 1.0 \times 10^{-4}$ (solid black curve). The yellow dashed curve shows the conventional FWI objective function. (b) Components of the FWIME objective function (equation 4) computed with $\epsilon = \epsilon_3$ (dark blue curve in panel (a)). The blue curve shows the total FWIME objective function, the pink curve is the data-fitting component, and the red curve is the annihilating component scaled by $\frac{\epsilon^2}{2}$. [**CR**]

km/s. For this specific numerical example, the FWIME workflow successfully managed to retrieve an accurate solution without the need to employ the model-space multi-scale strategy (the unknown velocity model is thus parametrized directly on the finite-difference grid from the start). Additionally, I do not assume spatial uniformity of the velocity model and I invert for all $N_m = 10^4$ unknown model parameters. I use a time-lag extended axis with 81 points sampled at $\Delta\tau = 4$ ms. The variable projection step (step 2 of algorithm 1) is performed with 30 iterations of linear conjugate gradient. Figure 4.16 shows the resulting FWIME convergence curves (solid lines). All three components of the objective function converge to zero, which indicates that the scheme has successfully recovered the global solution. On the same plot, I superimpose the FWI objective function evaluated at each iteration of FWIME (red dashed line). This curve is not the result of an inversion process, but simply the values that the FWI objective function would have taken for this sequence of FWIME inverted models. These observations show that in this case, the FWIME optimization path is insensitive to the local minima present in the conventional FWI objective function. This analysis is also confirmed by the average velocity of the inverted models from the two optimization schemes (Figure 4.17). I choose this model metric because of the inherent uncertainty in the conventional traveltime tomography problem (Squires et al., 1994).

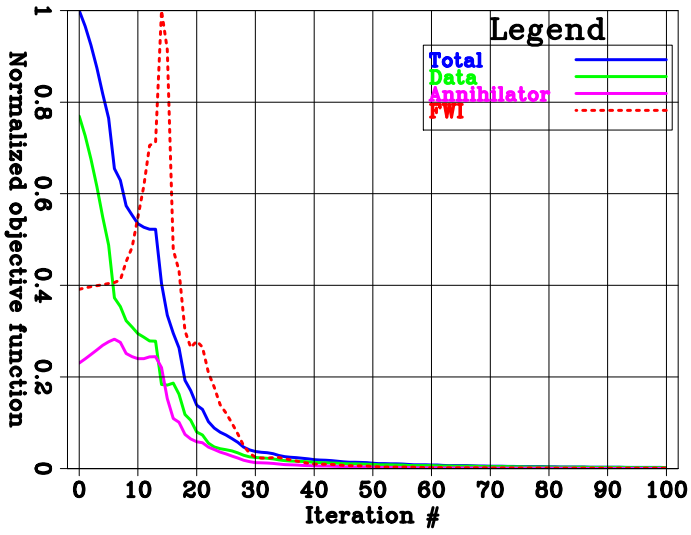


Figure 4.16: Normalized convergence curves as a function of iterations. The solid curves correspond to the three components of the FWIME objective function obtained by simultaneously inverting the full bandwidth of the transmission dataset with $\epsilon = 1.5 \times 10^{-6}$. The blue curve corresponds to the total FWIME objective function, the red curve is the data-fitting component, and the pink curve is the annihilating component. The red dashed curve shows the value of the conventional FWI objective function evaluated at each inverted FWIME model (it is not the output of an optimization process). [**CR**]

Figure 4.18 shows the difference between observed and predicted data, $\Delta\mathbf{d}(\mathbf{m}_i) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_i)$,

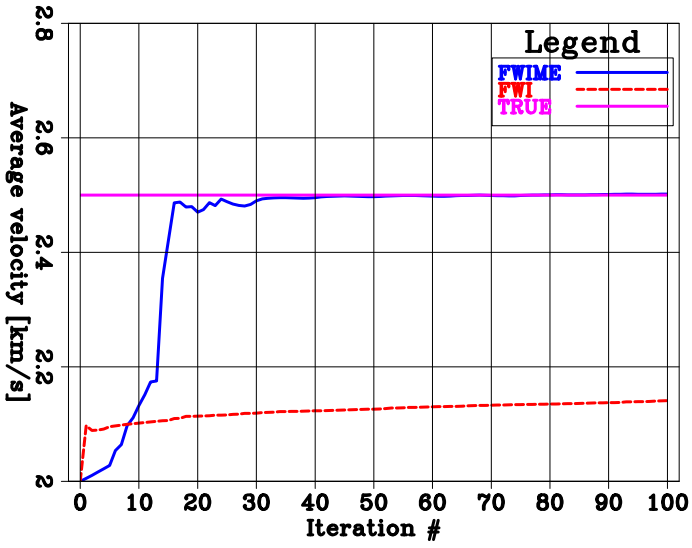
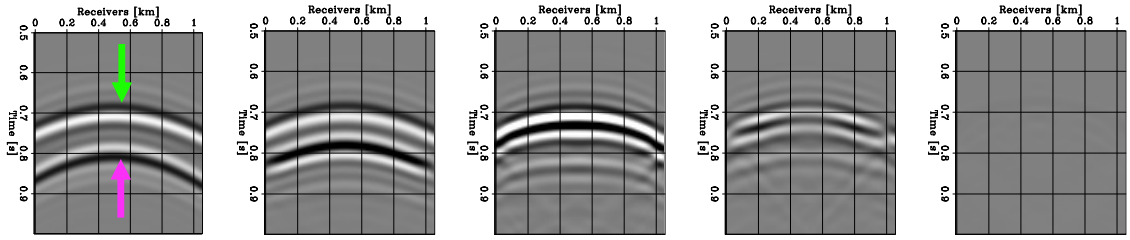


Figure 4.17: Inverted average velocity as a function of iterations for FWIME (solid blue curve) and conventional FWI (red dashed curve). The pink curve shows the true velocity value, v_{true} = 2.5 km/s. [**CR**]

computed with the FWIME inverted models at iterations 0, 10, 15, 20, and 100. Figure 4.18a indicates that the initial model largely underestimates the true velocity value (i.e., $\mathbf{m}_0 < \mathbf{m}_{true}$) and the data are cycle-skipped. The green and pink arrows correspond to the observed data $\mathbf{d}^{obs}$ and predicted data $\mathbf{f}(\mathbf{m}_0)$, respectively. At iteration 10, the velocity model has been updated in the correct direction and the time-shift between predicted and observed data has shrunk (Figure 4.18b). At iteration 15, the two events begin to overlap with a misaligned phase, which corresponds to the increase in the FWI objective function in Figure 4.15a (yellow dashed curve) and Figure 4.16 (red dashed curve). This effect begins to disappear at iteration 20 as the phases of the predicted and observed data start to align (Figure 4.18d). At the last iteration, the predicted and observed data are almost identical (Figure 4.18e), which proves that the FWIME workflow has converged to the global solution.

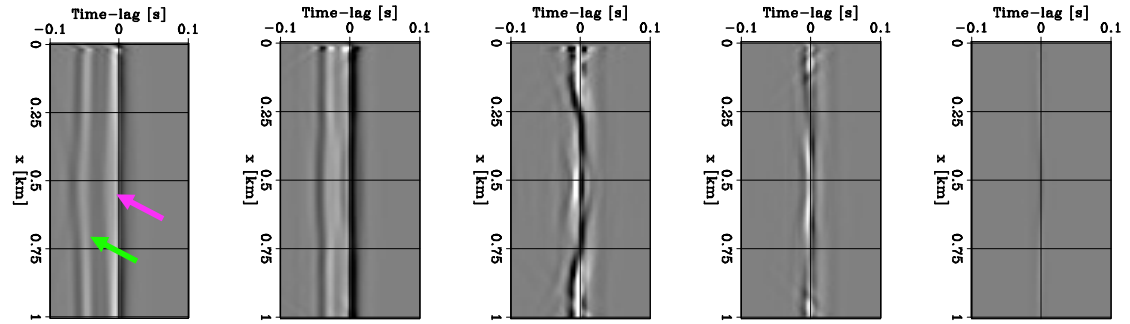
The simplicity of this dataset makes it easy to identify each event and provides insight on the connection between data and extended space. I conduct of an analogous step-by-step analysis of $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i)$. Figure 4.19 shows the evolution of a TLCIG extracted at z = 0.5 km from $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_i)$ computed at iterations 0, 10, 15, 20, and 100 of the FWIME inversion process. At the initial step (Figure 4.19a), I observe the presence of two separate vertical clusters of energy, which correspond to the mapping (by minimizing equation 3.6) of the two events observable from the data space (green and pink arrows in Figure 4.18a) into $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$. First, the event corresponding to $\mathbf{f}(\mathbf{m}_0)$ (pink arrow in Figure 4.18a) is mapped into $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$ at τ = 0 s: no extension in $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$ is needed to generate such an event because all modeled wavefields propagate with velocity $\mathbf{m}_0$. The second



(a) (b) (c) (d) (e)

Figure 4.18: Shot gathers generated by a source placed at $z = 0.5$ km in the left borehole. Each panel corresponds to the difference between observed data and predicted data computed with the inverted FWIME model $\mathbf{m}_i$ at five stages of the optimization workflow, $\Delta\mathbf{d}(\mathbf{m}_i) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_i)$. (a) Initial step, (b) iteration 10, (c) iteration 15, (d) iteration 20, and (e) iteration 100 (last iteration). All panels are displayed with the same grayscale. [**CR**]

vertical cluster of energy is located away from the physical axis (green arrow in Figure 4.19a) and corresponds the mapping of the observed data $\mathbf{d}^{obs}$ (green arrow in Figure 4.18a) into $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_0)$. In this case, an extension is required to generate an event with an apparent propagation velocity $\mathbf{m}_{true}$. Moreover, the fact that the energy focuses at negative values of τ confirms that the initial velocity model $\mathbf{m}_0$ is too slow. Note that all 100 shot gathers such as the one displayed in Figure 4.18a are employed to compute $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_0)$ when minimizing equation 3.6. As the optimization progresses and the velocity model becomes more accurate, the energy within $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_i)$ begins to diminish (starting from time lags with larger magnitude) and to gradually focus toward the physical axis, as shown in Figures 4.19b-d. Finally, at the end of the FWIME workflow (iteration 100), the energy within $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_{100})$ completely vanishes (Figure 4.19e).



(a) (b) (c) (d) (e)

Figure 4.19: TLCIGs extracted at $z = 0.5$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_i)$ at five stages of the FWIME process. (a) Initial step, (b) iteration 10, (c) iteration 15, and (d) iteration 20, and (e) iteration 100 (last iteration). All panels are displayed with the same grayscale. [**CR**]

4.3.2 Inversion of reflection data

I test FWIME on a reflection-dominated dataset generated by a model similar to the one proposed in Mora (1989). More specifically, I use this numerical example to show the need of combining the conventional FWIME workflow with the model-space multi-scale approach. The true model is 4 km wide and is composed of two homogeneous horizontal layers with velocity values of 2.7 km/s and 2.25 km/s, respectively. The interface between the two horizontal layers is located at a depth of z = 1.1 km. In the top layer, I embed a circular-shaped low-velocity anomaly with sharp contours and a velocity value of 2.2 km/s, which is 19% lower than the top-layer velocity value (Figure 4.20d). The initial model $\mathbf{m}_0$ is homogeneous and set to 2.7 km/s (Figure 4.20a).

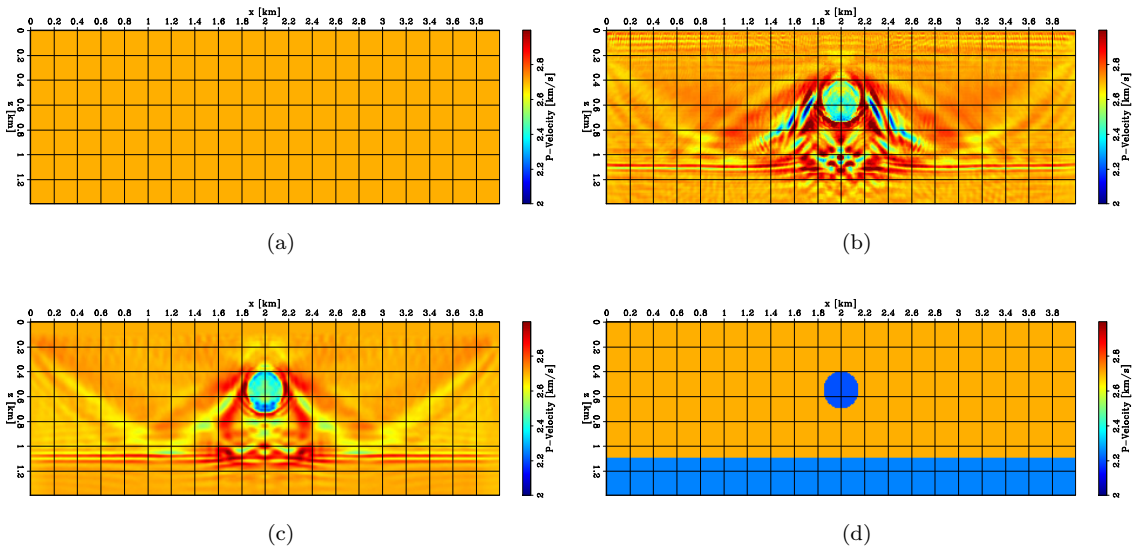


Figure 4.20: 2D panels of velocity models, based on Mora (1989). (a) Initial model. (b) Inverted model obtained by conducting conventional data-space multi-scale FWI (3 frequency bands). (c) FWIME inverted model without any multi-scale strategy. (d) True model. [**CR**]

The noise-free pressure data are generated using a finite-difference grid spacing of $\Delta z = \Delta x = 10$ m, and with a source containing energy strictly restricted to the 20-50 Hz frequency range. I choose this unrealistic frequency range to ensure that conventional multi-scale FWI fails to retrieve a physical solution. I set 40 sources and 400 receivers at the surface with a spacing of 100 m and 10 m, respectively. Figure 4.21a shows the initial data difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$, for a source placed $x = 2$ km. Besides reflected energy, triplications stemming from the presence of the low-velocity anomaly can also be observed in the recorded events.

I first conduct a conventional data-space multi-scale FWI workflow using four frequency bands spanning the 20-50 Hz frequency range. The inverted model, shown in Figure 4.20b, indicates that FWI has converged to a local minimum. Figure 4.21b displays the data-residual computed at the

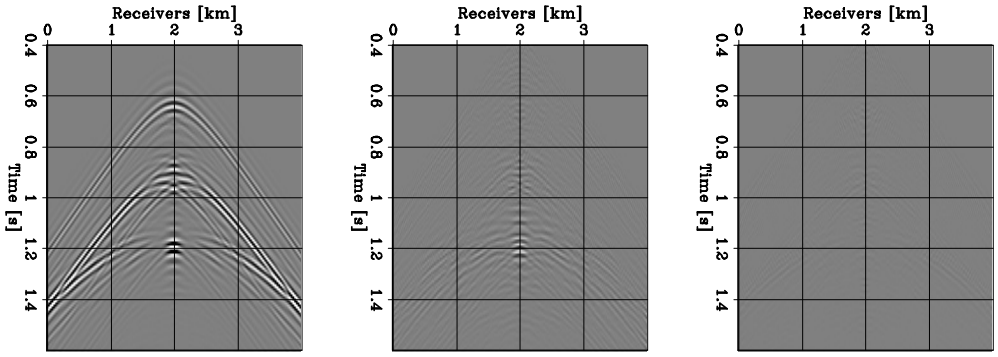


Figure 4.21: Shot gathers generated with a source located at $x = 2$ km showing the difference between observed and predicted data, $\Delta\mathbf{d}(\mathbf{m}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})$, computed with different velocity models. (a) Initial model (Figure 4.20a). (b) FWI inverted model (Figure 4.20b). (c) FWIME inverted model with a model-space multi-scale approach (Figure 4.25c). [**CR**]

last iteration of FWI, which confirms that the inverted model is unable to accurately predict the complex waveform (i.e., the triplications in the wavefield) generated by the presence of the low-velocity anomaly.

For the FWIME workflow, the full 20-50 Hz data bandwidth is inverted at once. $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ is extended in time-lags with 101 points sampled at $\Delta\tau = 8$ ms, and the variable projection step is performed with 50 iterations of linear conjugate gradient. Figure 4.22 shows the different components of the initial FWIME search direction computed on the finite-difference grid (without any spline re-parametrization). As expected, the Born component (Figure 4.22a) is similar to the conventional initial FWI search direction: the reflections from the dataset are mapped as high-wavenumber migration isochromes into the model space (Zhou et al., 2015). This is confirmed by examining the amplitude spectra of the spatial Fourier transforms of the initial FWI search direction (Figure 4.23a), and the FWIME's Born component (Figure 4.23b). Both update directions are missing the low-wavenumber information present in the ideal search direction (Figure 4.23e). Moreover, since the initial background velocity model is inaccurate (absence of the low-velocity circular anomaly), these migration isochromes are initially misplaced and will likely guide the inversion to a non-physical solution, especially in the zone between the bottom of the anomaly and the horizontal interface. The tomographic update is more promising and recovers regions of the spectrum that were not captured by neither the FWI nor the Born component (Figures 4.22b). Nevertheless, the total search direction (Figures 4.22c and 4.23d) are contaminated by the migration isochromes from the Born component. Therefore, if no multi-scale strategy is employed, FWIME converges to a local minimum, as shown in Figure 4.20c.

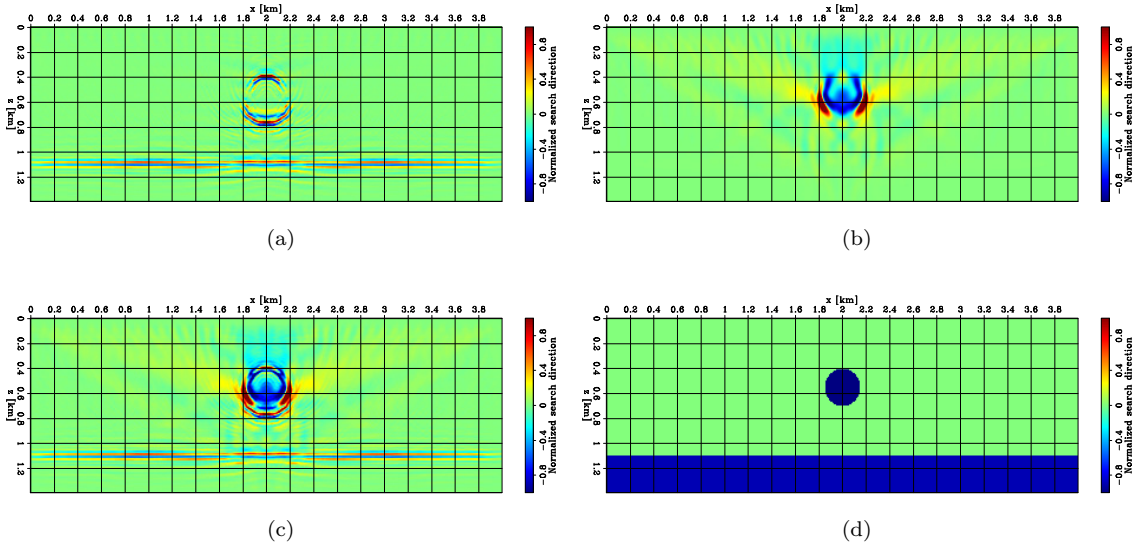


Figure 4.22: 2D panels of normalized initial search directions. (a) FWIME Born search direction. (b) FWIME tomographic search direction. (c) FWIME total search direction. (d) Ideal search direction. Panels (a), (b), and (c) are normalized by the same value and displayed on the same color scale. [**CR**]

To overcome this issue, I use a sequence of three spatially-uniform spline grids sampled at 50 m, 20 m, and 10 m, respectively (the third grid coincides with the finite-difference propagation grid). Figures 4.24a-c show the initial FWIME search directions after applying operator $\mathbf{SS}^*$ to Figures 4.22a-c ($\mathbf{S}$ is the spline mapping operator). The amplitude of the Born component is now much smaller than the amplitude of the tomographic update, the spurious high-wavenumber features have been removed, and the search direction is improved (compared to Figure 4.22c).

I can now successfully apply my multi-scale FWIME workflow. I use a fixed ϵ-value of 1.7×10^{-7} throughout the entire process. Each spline grid refinement is automatically triggered when the numerical solver is unable to find a step length that decreases the objective function. Figures 4.25a-c show the sequence of inverted models at the end of each spline grid. The final recovered model is excellent and manages to accurately reconstruct the velocity values in the shadow zone located between the bottom of the anomaly and the horizontal interface. The sharpness of the anomaly is also well captured, as shown by the vertical (Figure 4.26a) and horizontal (Figure 4.26b) velocity profiles extracted at $x = 2$ km and $z = 0.6$ km, respectively (the oscillating behavior of the model is due to the limited frequency range available in the dataset). In addition, the difference between the observed and predicted data computed with the final FWIME model is shown in Figure 4.21c and confirms the quality of the inversion result. In this experiment, the sensitivity of the inverted model with respect to the trade-off parameter ϵ was very limited. Similar results as the one shown

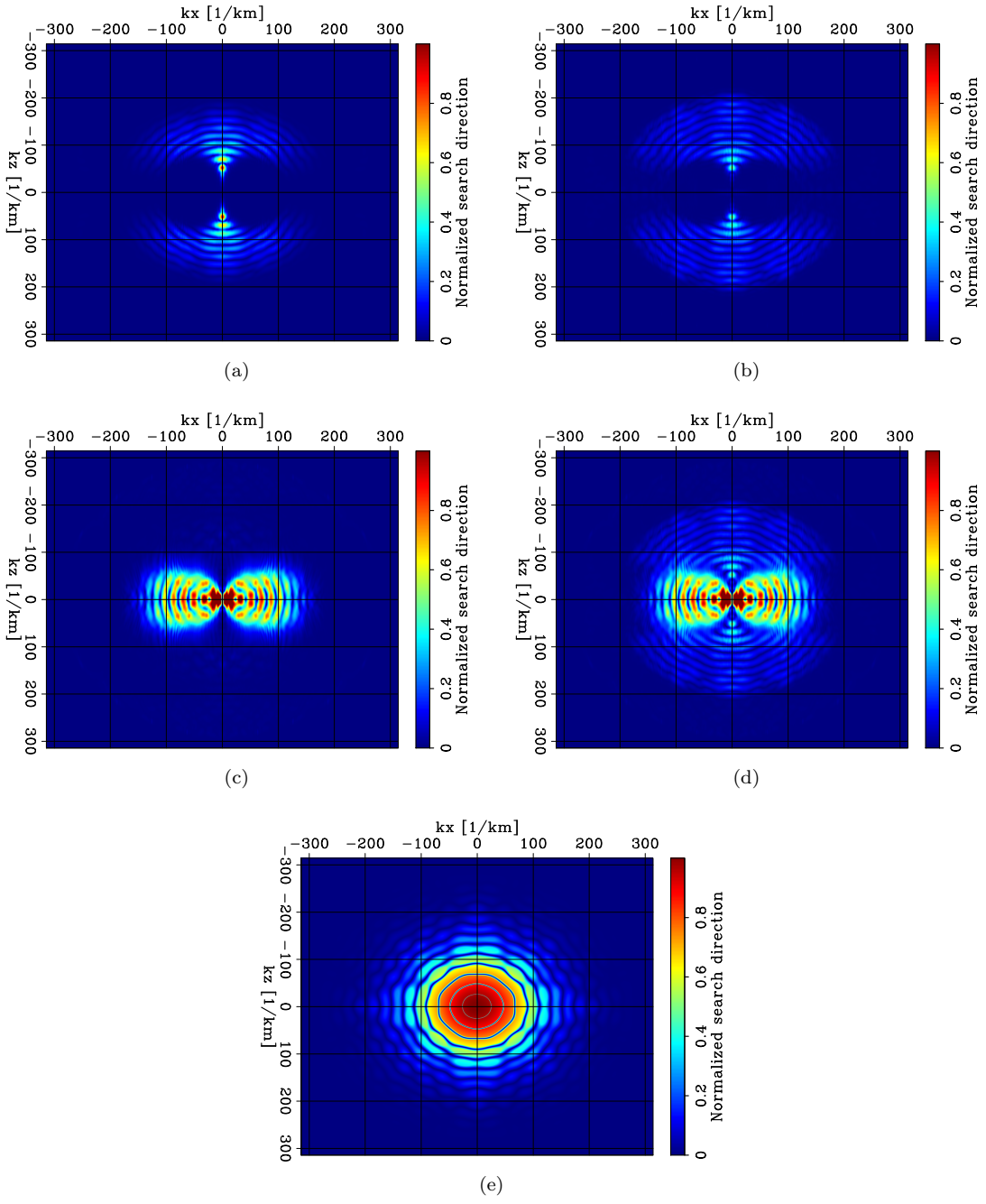


Figure 4.23: Amplitude spectra of the spatial Fourier transforms of initial search directions. (a) Conventional FWI. (b) FWIME's Born component. (c) FWIME's tomographic component. (d) FWIME total initial search direction (sum of panels (b) and (c)). (e) Ideal search direction. Panels (b), (c) and (d) are normalized by the same value and displayed on the same color scale. [**CR**]

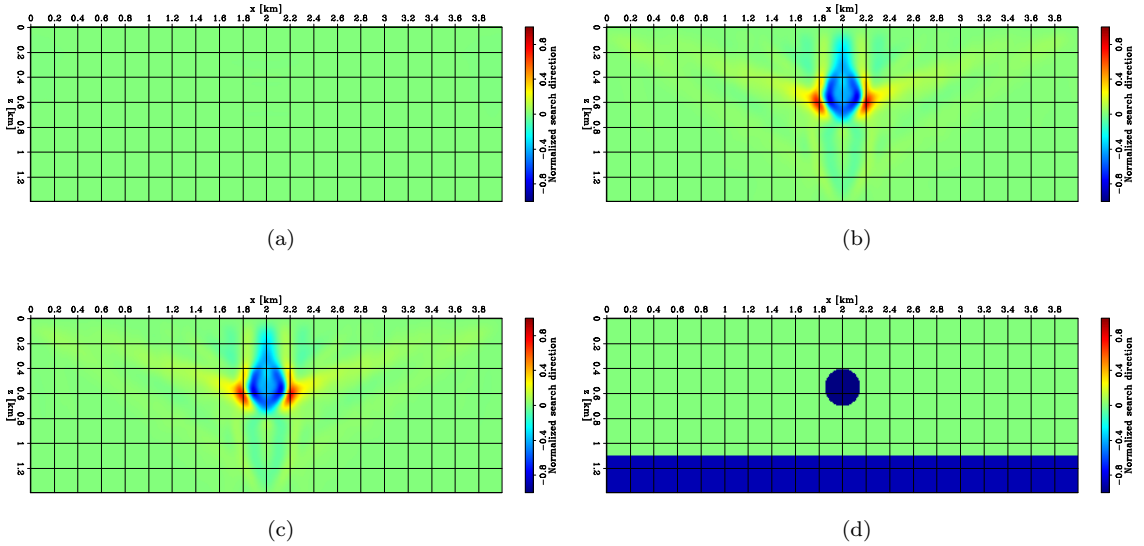
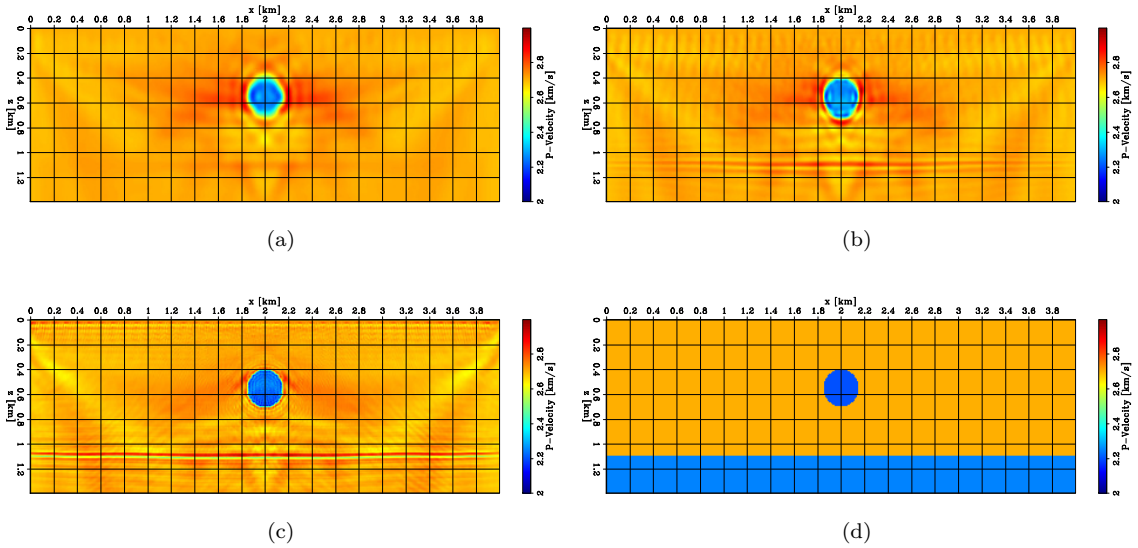


Figure 4.24: 2D panels of normalized initial search directions after their mapping onto the first spline grid (i.e., after applying operator $\mathbf{SS}^*$ to the analogous panels in Figure 4.22). (a) FWIME Born search direction. (b) FWIME tomographic search direction. (c) FWIME total search direction. (d) Ideal search direction (no spline mapping was applied to this panel). Panels (a), (b), and (c) are normalized by the same value and displayed on the same color scale. [**CR**]

in Figure 4.25c were obtained for ϵ-values ranging within one order of magnitude.

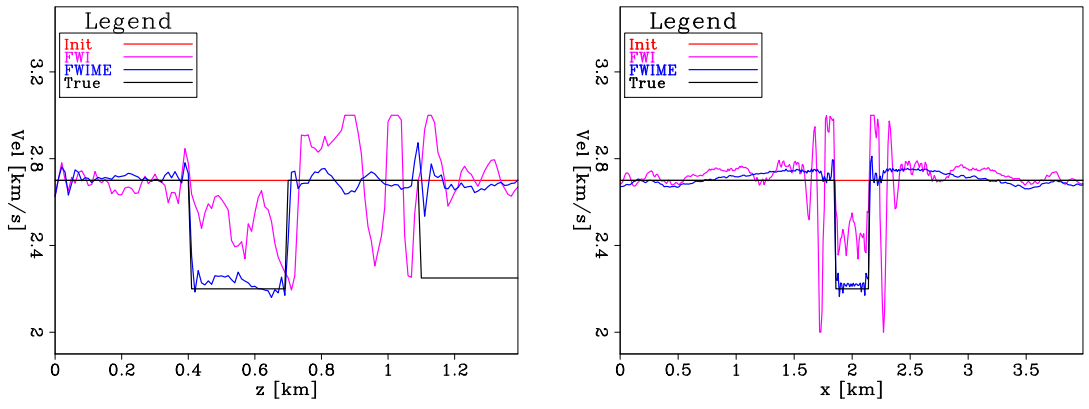
4.3.3 Inversion of diving waves

In this numerical experiment, I invert a dataset solely composed of diving waves where the inaccurate initial velocity model produces very large kinematic errors in the predicted data. The dataset is generated with a source wavelet containing energy strictly limited to the 3-12 Hz frequency range, which prevents conventional FWI from leveraging the low-frequency signal (below 3 Hz) to overcome the cycle-skipping phenomenon. The true model is 16 km wide by 2.8 km deep, and is discretized with a finite-difference grid spacing of 30 m in both directions. It is composed of a 0.4 km-thick homogeneous layer placed on top of second horizontally-invariant layer whose values linearly increase with depth, as shown in Figures 4.27d and 4.28 (black curve). The initial model $\mathbf{m}_0$ is chosen to be unrealistically inaccurate (Figure 4.27a). It is homogeneous and set to 2.0 km/s (dark-blue curve in Figure 4.28). I place 137 sources and 550 receivers at the surface, spaced every 120 m and 30 m, respectively. Figures 4.29a-c show a representative shot gather corresponding to the observed data $\mathbf{d}^{obs}$, the initial prediction $\mathbf{f}(\mathbf{m}_0)$, and initial data difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ computed for a source placed at $x = 1.2$ km.



(a) (b) (c) (d)

Figure 4.25: 2D panels of inverted velocity models obtained at different stages of the FWIME model-space multi-scale workflow. (a) Inverted model after 32 iterations on first spline grid. (b) Inverted model after 20 iterations on second spline grid. (c) Final inverted model after 30 iterations on the finite-difference grid. (d) True model. A total of 82 L-BFGS iterations were used to obtain the result in panel (c). [**CR**]



(a) (b)

Figure 4.26: Velocity profiles of the true (black curve), initial (red curve), conventional FWI (pink curve), and FWIME (blue curve) inverted models. (a) Depth velocity profiles extracted at $x = 2$ km. (b) Horizontal velocity profiles extracted at $z = 0.6$ km. [**CR**]

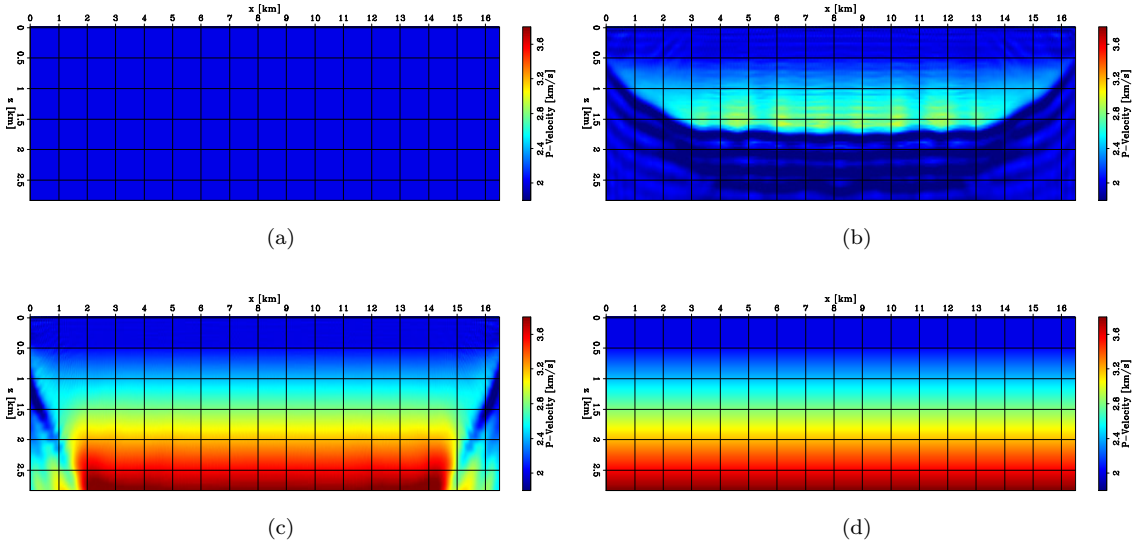


Figure 4.27: 2D panels of velocity models. (a) Initial model. (b) Inverted model with a conventional data-space multi-scale FWI (using three frequency bands). (c) Final FWIME inverted model. (d) True model. Panel (c) was obtained using a model-space multi-scale FWIME approach with three different spline grids and a total of 40 iterations. [**CR**]

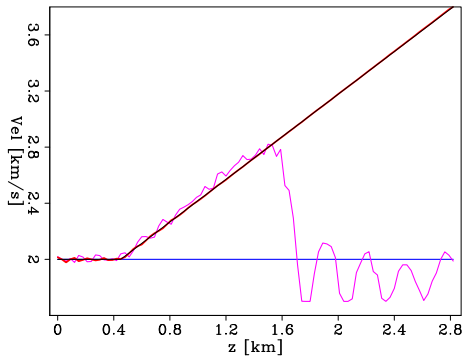
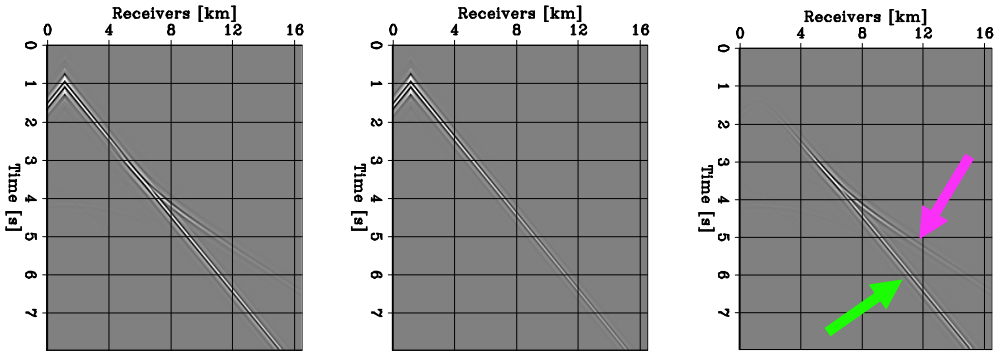


Figure 4.28: Depth velocity profiles extracted at $x = 8$ km from the initial model (blue curve), the FWI model (pink curve), the FWIME model (red curve), and the true model (black curve). The black and red curves are very similar and overlap each other. [**CR**]



(a) (b) (c)

Figure 4.29: Representative shot gathers generated with a source located at $x = 1.2$ km. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data with the initial model, $\mathbf{f}(\mathbf{m}_0)$. (c) Initial data difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**ER**]

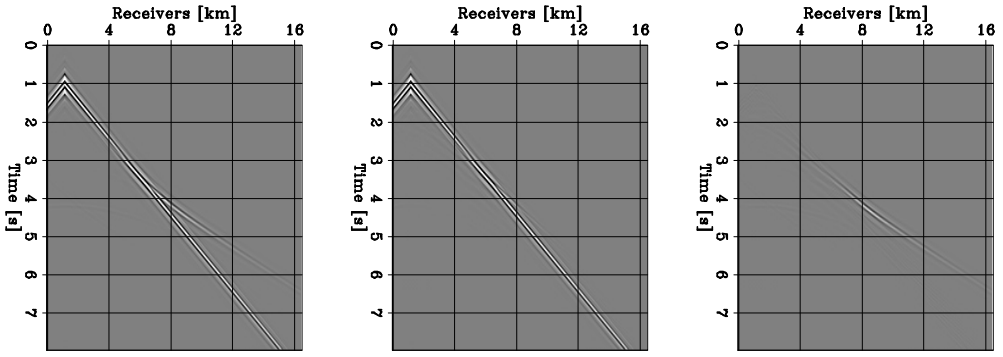
Conventional FWI

I apply a conventional multi-scale FWI workflow using three frequency bands spanning the available 3-12 Hz bandwidth, starting with the uniform model $\mathbf{m}_0$. For each frequency band, I conduct 500 iterations of L-BFGS. FWI fails to recover the correct velocity model for depths greater than 1.6 km, as shown in Figures 4.27b and 4.28 (pink curve). In addition, Figures 4.30b and c display the predicted data and data residual computed with the final FWI model and show that the recovered model is unable to accurately predict refracted events (i.e., diving waves) for offsets greater than 7 km.

FWIME initial step

For the computation of $\tilde{\mathbf{p}}_\epsilon^{opt}$, I use a time-lag extension, which is known to be better suited than horizontal subsurface offsets to model large time shifts between oberved and predicted data. The extended axis is composed of 101 points sampled at $\Delta\tau = 32$ ms, which allows τ to range from -1.6 s to 1.6 s.

The full potential of extended modeling (and the ability of the data-correcting term $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$ to match any data misfit even for inaccurate background velocity models) can be better appreciated by closely examining the first variable projection step in the FWIME workflow. The initial data difference $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ contains two events (Figure 4.29c). The first event possesses a linear moveout (green arrow) and corresponds to the phase mismatch between the direct arrivals from the true and initial models. The second event (pink arrow) is the diving wave present in the observed data that is not modeled by our initial prediction (due to the constant velocity value).



(a) (b) (c)

Figure 4.30: Representative shot gathers generated with a source located at $x = 1.2$ km. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data with the final FWI inverted model, $\mathbf{f}(\mathbf{m}_{FWI})$. (c) Data difference, $\Delta\mathbf{d}(\mathbf{m}_{FWI}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWI})$. All panels are displayed with the same grayscale. [**CR**]

On one hand, a data-correcting term $\mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0)$ (computed by minimzing equation 3.6 with a non-extended linear Born modeling operator, the initial velocity model $\mathbf{m}_0$, and $\epsilon = 0$) would have no chance to generate diving waves that would fit the initial data difference. Mathematically, this means that

$$\mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0) \neq \Delta\mathbf{d}(\mathbf{m}_0). \tag{4.22}$$

Figure 4.31b shows the data-correcting term computed with a non-extended Born operator, $\mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0)$. As expected, it fails to predict the refracted energy. This observation is confirmed by examining the prediction error, $\mathbf{r}_d^\epsilon(\mathbf{m}_0) = \mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0) - \Delta\mathbf{d}(\mathbf{m}_0)$ in Figure 4.31c. On the other hand, the use of an extended Born operator allows the data-correcting term to accurately match the initial data misfit (with a linear operator) and a constant velocity model $\mathbf{m}_0$ as shown in Figure 4.32b and 4.32c. That is,

$$\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0) \approx \Delta\mathbf{d}(\mathbf{m}_0), \tag{4.23}$$

Figure 4.33a shows a TLCIG extracted at $x = 8$ km from $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$ computed at the initial step with $\epsilon = 7.5 \times 10^{-5}$. In the shallow part of the model where $\mathbf{m}_0$ is accurate (for $z \leq 1.6$ km), the energy within $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$ is focused in the vicinity of the zero-lag axis (region within the green oval). This energy cluster corresponds to the mapping of the direct arrival (green arrow in Figure 4.29c) into $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$, which could in fact be modeled without any extension. In the deeper region, the velocity error increases and the coherent energy is gradually positioned at large-amplitude negative

time lags. This energy cluster (pink oval), which corresponds to the mapping of the diving waves (pink arrow in Figure 4.29c), can only be captured with the use of an extended perturbation and a modeling operator.

In this particular example, the initial model $\mathbf{m}_0$ was extremely inaccurate and our extended axis τ had to span an unusually wide range of time lags to satisfy equation 3.20. By comparing TLCIGs for diving waves (Figure 4.33a) with the ones for transmitted waves (Figure 4.19a), I observe that different wave types are mapped into energy clusters that possess various moveout characteristics in the extended space of $\tilde{\mathbf{p}}_\epsilon^{opt}$. As I show in this thesis, one of the main advantages of FWIME is its capacity to invert any wave type with the same mechanism, regardless of the type of moveout within $\tilde{\mathbf{p}}_\epsilon^{opt}$.

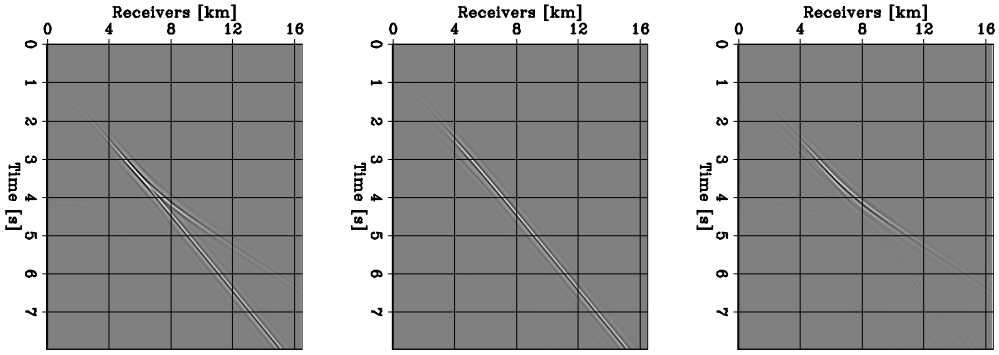
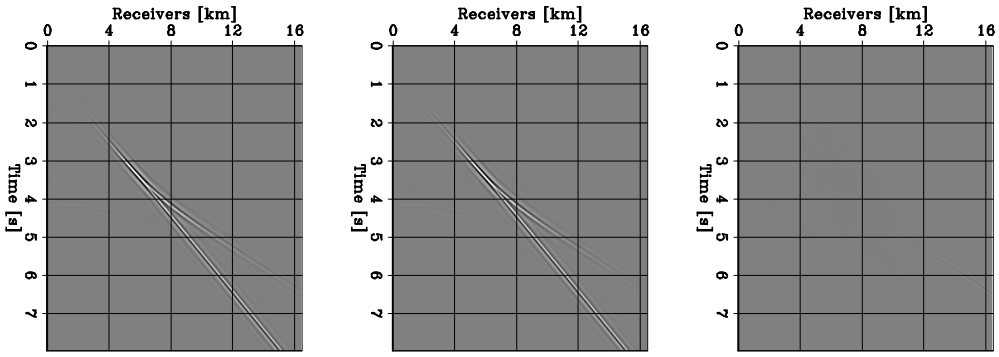


Figure 4.31: Representative shot gathers generated with a source located at $x = 1.2$ km. (a) Initial data difference: $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. (b) Data-correcting term computed with a non-extended Born operator (and $\epsilon = 0$): $\mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0)$. (c) Prediction error between the data-correcting term and the initial data residual: $\mathbf{r}_d^\epsilon(\mathbf{m}_0) = \mathbf{B}(\mathbf{m}_0)\mathbf{p}_\epsilon^{opt}(\mathbf{m}_0) - \Delta\mathbf{d}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**CR**]

FWIME workflow

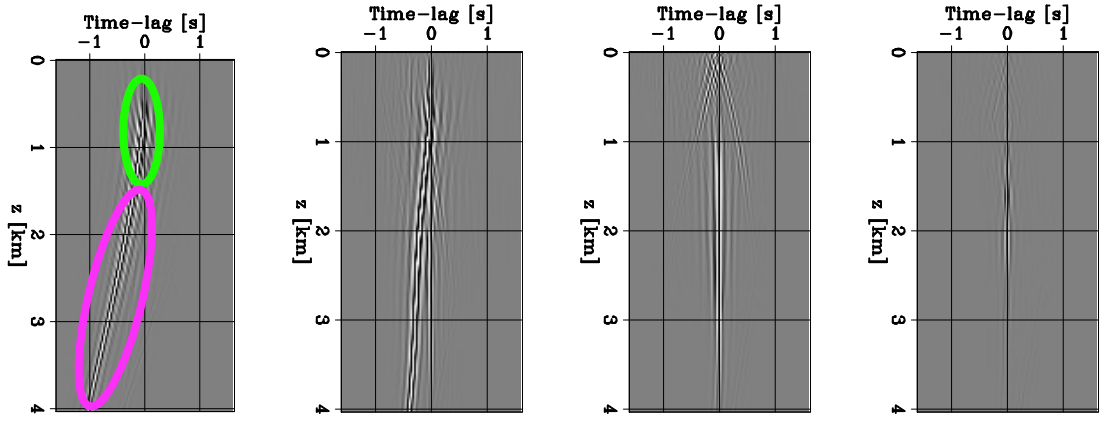
Figures 4.34a-c show the Born, tomographic, and total first search directions for the FWIME workflow (computed with $\epsilon = 7.5 \times 10^5$) on the finite-difference grid (no spline mapping is applied yet). The total search direction seems promising and is dominated by the tomographic component. In order to impose lateral smoothness on the inverted model, I begin the multi-scale FWIME scheme by designing a coarse and spatially uniform initial spline grid, where $\Delta z = 1$ km, and $\Delta x = 1.8$ km. The search directions obtained after re-parametrization on the initial spline grid are shown in Figures 4.35a-c, and are much more geologically consistent.

I simultaneously invert the full available bandwidth from the data, I fix the ϵ-value to 7.5×10^5,



(a) (b) (c)

Figure 4.32: Representative shot gathers generated with a source located at $x = 1.2$ km. (a) Initial data difference: $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. (b) Data correcting term computed with a time-lag extended Born operator: $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0)$. (c) Prediction error between the data-correcting term and the initial data residual: $\mathbf{r}_d^\epsilon(\mathbf{m}_0) = \tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_0) - \Delta\mathbf{d}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**CR**]



(a) (b) (c) (d)

Figure 4.33: TLCIGs extracted at $x = 8$ km from $\tilde{\mathbf{p}}_\epsilon^{opt}$ at four stages of FWIME. (a) Initial FWIME step, (b) iteration 3, (c) iteration 10, (d) iteration 25. All panels are displayed with the same grayscale. [**CR**]

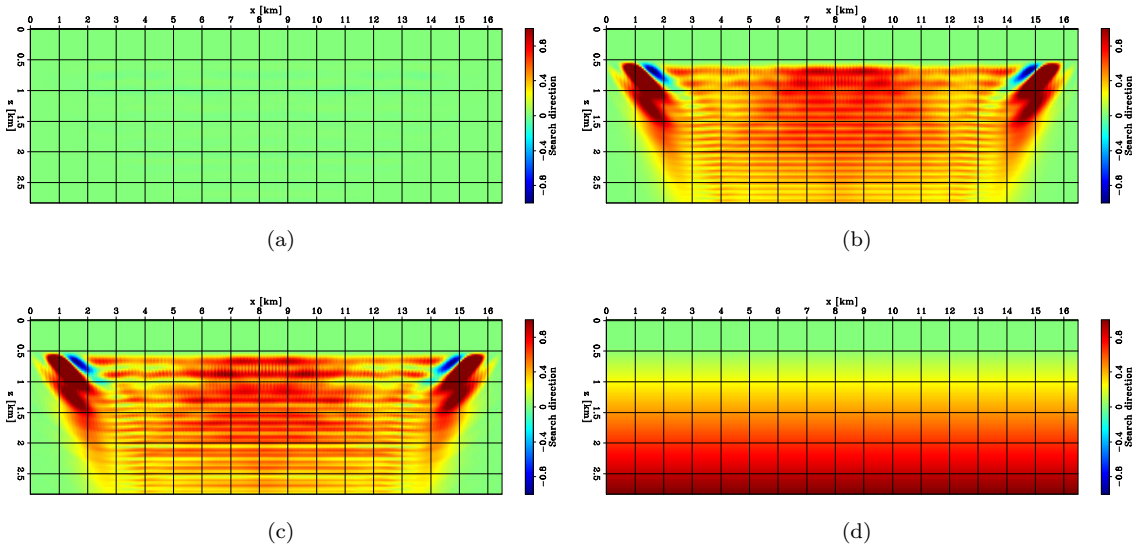


Figure 4.34: 2D panels of normalized initial search directions. (a) FWIME Born search direction. (b) FWIME tomographic search direction. (c) FWIME total search direction. (d) Ideal search direction. Panels (a), (b), and (c) are normalized by the same value and displayed on the same color scale. [**CR**]

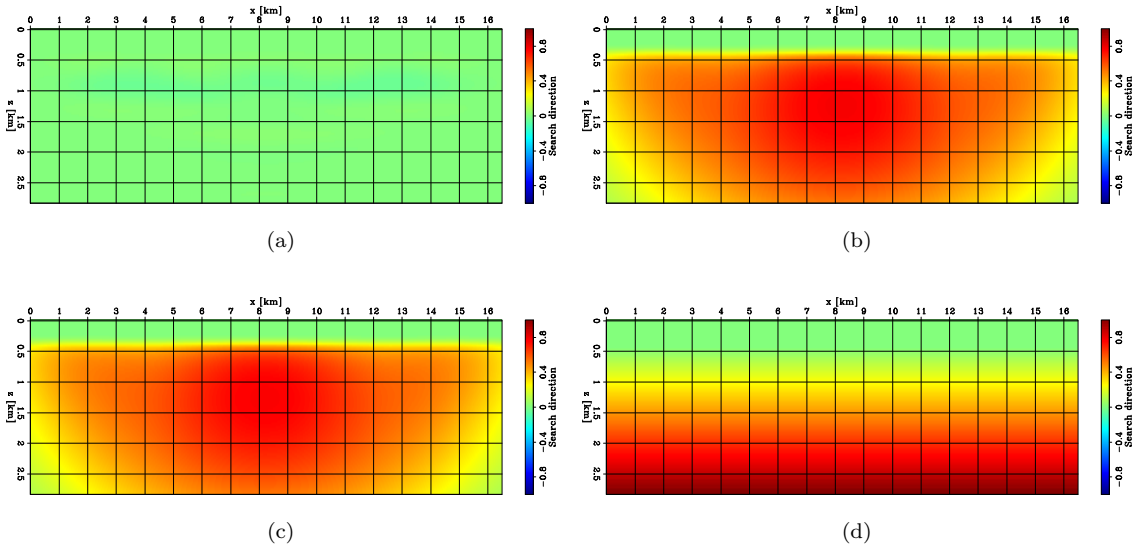


Figure 4.35: 2D panels of normalized initial search directions after applying operator $\mathbf{SS}^*$ (where $\mathbf{S}$ is the spline operator for the first grid). (a) FWIME Born search direction. (b) FWIME tomographic search direction. (c) FWIME total search direction. (d) Ideal search direction. Panels (a), (b), and (c) are normalized by the same value and displayed on the same color scale. [**CR**]

and I use a sequence of three spline grids (the last grid coincides with the finite-difference grid). The final inverted model obtained after 40 L-BFGS iterations of FWIME is shown in Figure 4.27c. While it has recovered an accurate solution down to a depth of approximately 2.8 km (red curve in Figure 4.28), it suffers from small edge artifacts inherent to the limited acquisition aperture. Figure 4.33 shows the evolution of a TLCIG extracted at $x = 8$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed at iterations 0, 3, 10, and 25. Most of the kinematic errors have been corrected after 10 iterations as most of the energy within $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ is focused in the vicinity of the zero time-lag axis, at which point the inverted model could be used as input for the less computationally costly FWI.

Finally, in the FWIME scheme, I embed prior information by using an extremely coarse initial spline grid. For a fair comparison, I also attempt to improve the FWI result by combining the same spline parametrization sequence as for FWIME (Barnier et al., 2019) with the more conventional data-space multi-scale approach. However (and as expected), I do not observe any uplift in the quality of the inverted model.

Chapter 5

Applying FWIME to realistic 2D synthetic examples

In this chapter, I apply FWIME on more realistic and challenging 2D synthetic examples to show its ability at successfully and simultaneously inverting datasets composed on all types of waves. The first test is conducted on a model proposed by the Seiscope consortium, which is designed to assess the robustness of FWIME against the ill-posedness coming from wrong association of predicted and observed waveforms. In the second example, I invert cycle-skipped data dominated by reflected events on the Marmousi2 model (Martin et al., 2006). The third example focuses on the use of refracted energy (diving waves) to simultaneously recover high- and low-resolution features of the North Sea region of the 2004 BP model (Billette and Brandsberg-Dahl, 2005). In the fourth example, I show that FWIME can employ transmitted energy to successfully delineate a salt body without the use of low-frequency energy and with an initial model containing no salt. In each case, conventional FWI fails to recover a useful solution due to the lack of low-frequency energy combined with inaccurate initial models, whereas FWIME manages to recover an excellent solution. The noise-free pressure data are generated and inverted with the same acoustic isotropic constant-density two-way wave-equation engine.

5.1 Seiscope's syncline model

I apply FWIME on a synthetic example designed and proposed by the Seiscope consortium. The true velocity model is shown in Figure 5.1d. The model is approximately 12 km wide, 3.5 km deep, and it is composed of two horizontal layers with a low-velocity synclinal inclusion (basin) embedded in the second (deeper) layer. The initial velocity (Figure 5.1a) is identical to the true model but does not contain the synclinal basin. The velocity values are set to $v_{top} = 2.8$ km/s and $v_{bottom} = 4.0$

km/s in the top and bottom layers, respectively. The difference between the two velocity models is shown in Figure 5.9d.

I generate a dataset with a finite-difference numerical scheme using a grid spacing of 50 m in both directions and a band-passed Ricker wavelet containing energy only within the 1.5-6.5 Hz frequency range (Figure 5.2). I place 48 shots spaced every 250 m and 255 receivers every 50 m. All acquisition devices are placed at a constant depth of 50 m. Figure 5.3 shows two representative shot gathers for sources located at $x = 0.3$ km and $x = 6.3$ km containing energy only within the 1.5-6.5 Hz frequency range.

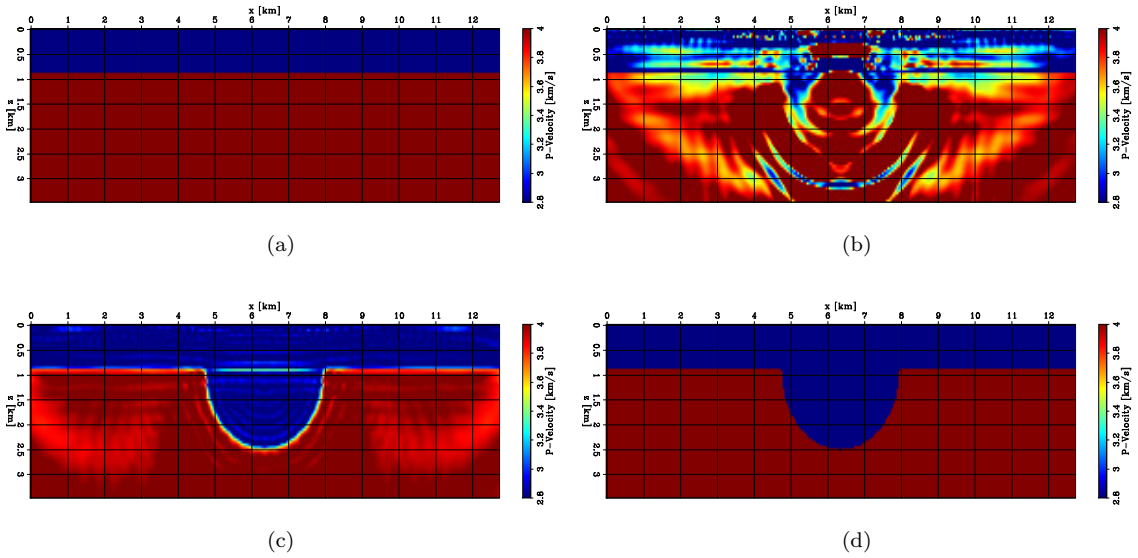


Figure 5.1: 2D panels of velocity models. (a) Initial model. (b) Inverted model after conventional data-space multi-scale FWI. (c) Final FWIME inverted model. (d) True model. [**CR**]

I begin by describing the motivation that led the Seiscope consortium to design this test, and I analyze the complexity of the data by examining a sequence of snapshots of a representative wavefield. Finally, I show that FWIME can recover an accurate solution whereas conventional FWI converges to a local minimum.

This synthetic problem was designed and developed by a team of researchers led by Weiguang He at the Seiscope consortium, and kindly shared to the Stanford Exploration Project (SEP) in May 2019. The goal of this experiment is to test the robustness of various data-driven waveform inversion methods against the ill-posedness coming from wrong association of predicted and observed waveforms. This ill-posedness is related to phase identification where different kind of synthetic and observed waveforms are tentatively compared, which could be quite frequent when the current model estimate predicts more or less waveforms than the velocity model corresponding to the observed waveforms (J. Virieux, personal communication, May 2019). In order to generate this type of

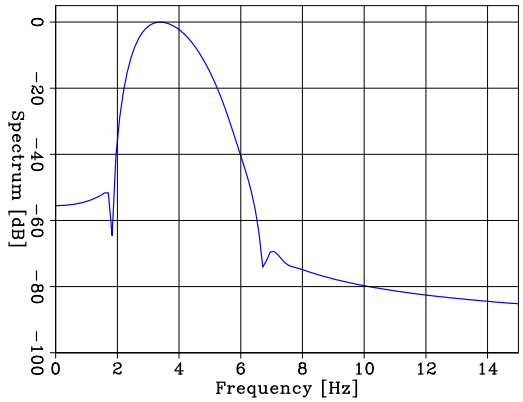
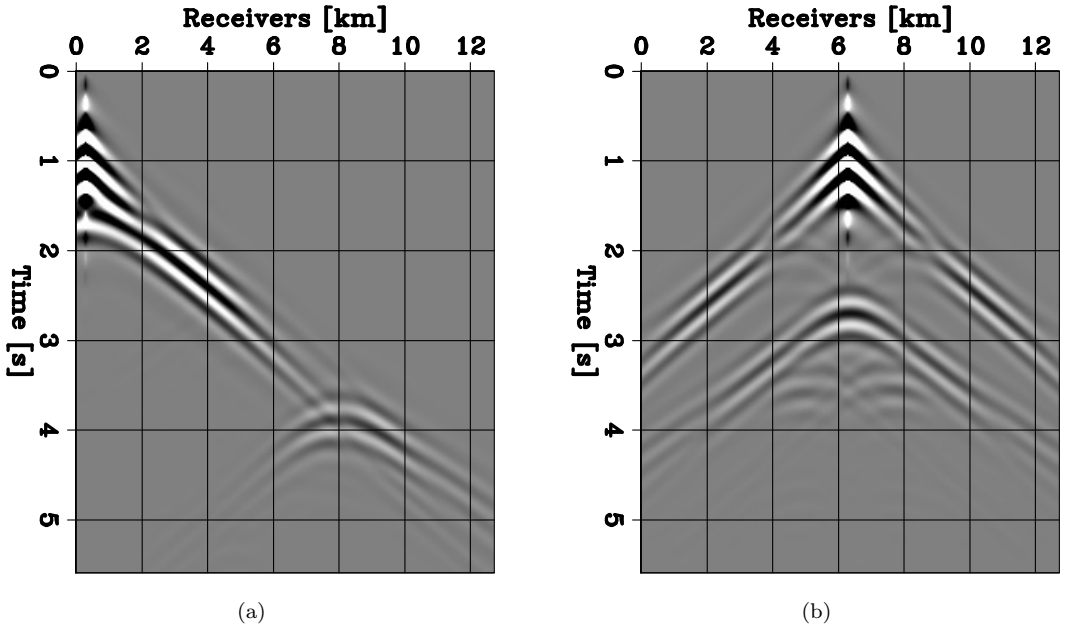


Figure 5.2: Amplitude spectrum of the seismic source used to generate the dataset for this numerical example. [**ER**]



(a) (b)

Figure 5.3: Representative shot gathers generated by a seismic source containing energy restricted to the 1.5-6 Hz frequency range. (a) Source located at $x = 0.3$ km. (b) Source located at $x = 6.3$ km. All panels are displayed with the same grayscale. [**ER**]

scenario with a simple numerical experiment, the authors embed a synclinal basin into the bottom layer of a horizontal two-layer model. The basin curvature is such that the reflected events from the bottom of the syncline generate wavefield triplications, which are recorded in the data. These events overlap with the reflections from the shallow interface between the two layers. The complexity of the recorded waveforms can be observed in the shot gathers displayed in Figure 5.3. Moreover, this numerical experiment poses another difficult challenge because its initial model contains a vast region (relative to the dimensions of the area of interest) with mispositioned sharp boundaries and incorrect strong velocity-contrasts with its surroundings.

Both difficulties (phase identification and large areas of kinematic errors coupled with mispositioned sharp interfaces) provide a similar - though simpler - scenario to the one encountered when exploring in regions with the presence of complex overburdens, such as salt bodies. This similarity is due to the complicated waveforms generated and recorded at the surface (e.g., prismatic reflections from rugose top-salt interface) as well as the difficulty to accurately delineate these complex overburdens. Therefore, testing FWIME on such an example can help gain better insight on its potential ability to perform well in these cases.

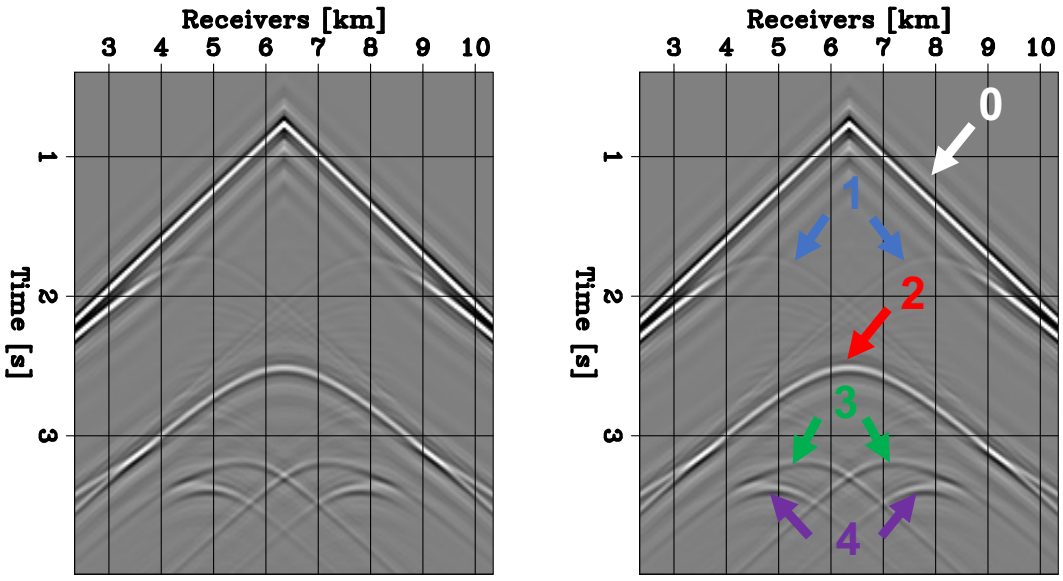
In order to analyze the origin and the nature of the complex waveforms present in the recorded data, I generate one shot gather using a higher-frequency seismic source placed at a horizontal position of $x = 6.3$ km (Figure 5.4a). Figure 5.4b shows the same shot gather where five different events have been identified and labeled with colored arrows, which can be further interpreted by examining a sequence of snapshots of the wavefield that gave rise to this recorded shot gather (Figure 5.5):

- The first event (event 0, white arrow in Figure 5.4b) displays a linear moveout and corresponds to the direct arrival of the incident wavefield (which does not interact with the interface between the two layers).

- Event 1 (blue arrows in Figures 5.4b and 5.5b) corresponds to the back-scattering of the incident wavefield as it interacts with the sharp corners on the edges of the synclinal basin. The corners act as diffracting points and the recorded event shows a hyperbolic moveout observable in the shot gather.

- Event 2 is generated when the incident wavefield reflects from the bottom of the basin. The wavefield later refocuses at the focal point of the syncline (Figure 5.5b), which creates a wavefield similar to the one that would have been generated by a virtual point source located at the focal point (red arrow in Figure 5.4b), directly observable by examining the wavefield snapshots (Figures 5.5d-g).

- Event 3 (green arrows in Figure 5.4b) is generated by the overlapping of a reflection from the bottom of the synclinal basin with the "tails" of the incident scattered wavefield after its interaction with the sharp corners (event 1) (at this point, I cannot fully explain the mechanism



(a) (b)

Figure 5.4: (a) Shot gather for a source located at x = 6.35 km containing energy within the 2-30 Hz frequency range. (b) Identical shot gather as (a) with labeled seismic events. All panels are displayed with the same grayscale. [**NR**]

that gives rise to this event). Nevertheless, this generates two up-going wavefields (event 3) observable in Figures 5.5f-k (green arrows).

- Event 4 (purple arrows in Figure 5.4b) stems from the diffraction of both up-going wavefields (i.e., event 3) by the corners of the synclinal basin. Their respective wave path can be observed in Figures 5.5j-l.

Finally, Figures 5.6 shows the observed, predicted and data difference computed with the initial velocity model for a shot located at $x = 6.3$ km. Figures 5.6c clearly illustrates the overlapping of true and wrongly predicted events (i.e., the reflection of the incident background wavefield from the misplaced horizontal interface present in the initial model).

I first conduct a conventional FWI approach to illustrate its inability at recovering the true model. Figures 5.1b shows the inverted model after 500 iterations of L-BFGS, which confirms that the FWI objective function has converged to a local minimum. Figures 5.7b and c show the predicted data $\mathbf{f}(\mathbf{m}_{FWI})$ and data difference $\Delta\mathbf{d}(\mathbf{m}_{FWI}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWI})$ computed with the FWI inverted model, which is unable to accurately predict the triplications generated by the synclinal feature.

I apply FWIME by simultaneously inverting all data available within the 1.5-6 Hz frequency range. I use two spline grids throughout the inversion scheme. The first spline grid has a sampling of 500 m in both directions. However, in the vicinity of the interface between the two layers, I use a

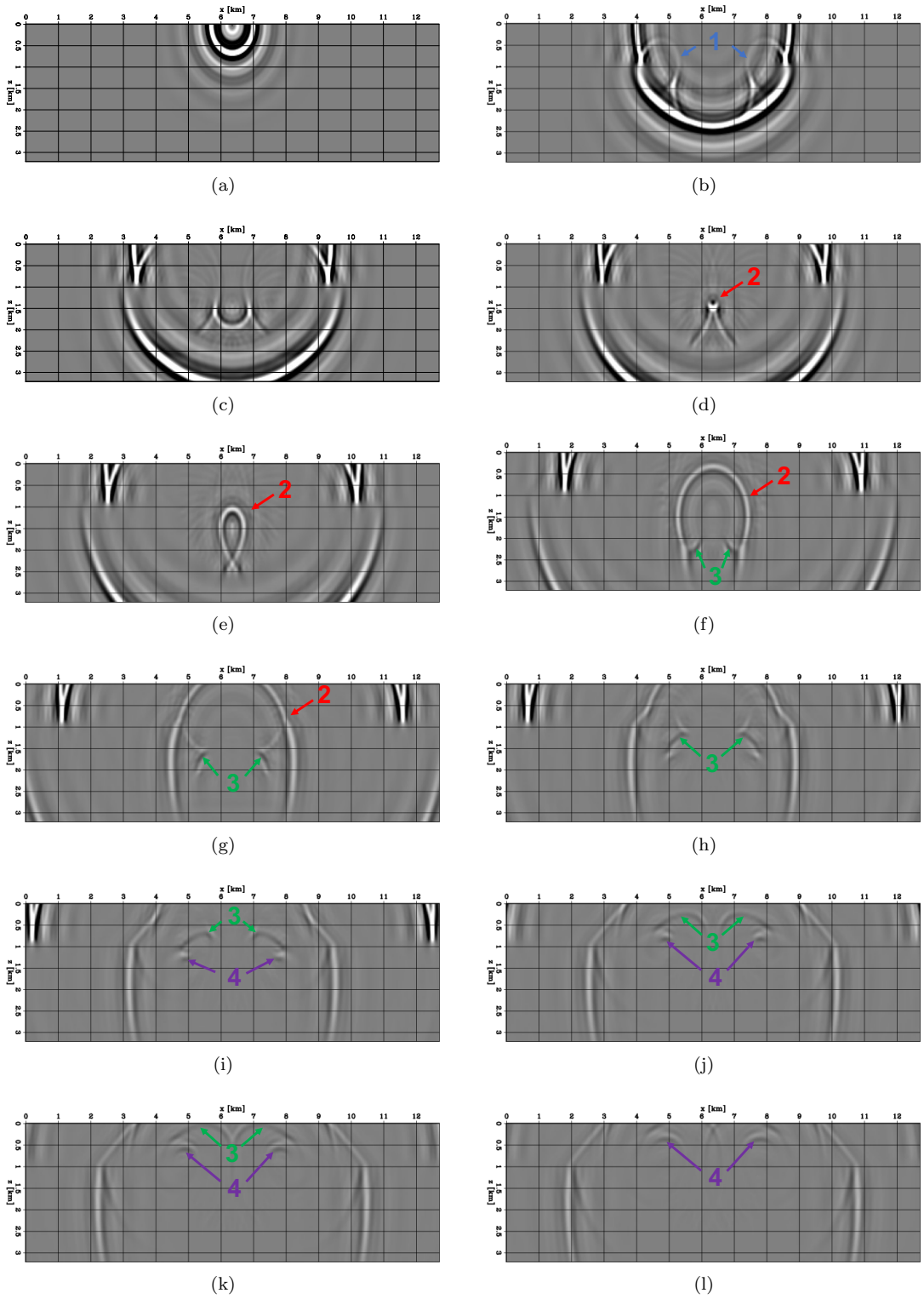


Figure 5.5: Wavefield snapshots generated by a source located at $x = 6.35$ km and containing energy within the 2-30 Hz frequency range, giving rise to the shot record shown in Figure 5.4. Each panel shows the wavefield at a given time step. All panels are displayed with the same grayscale. [**NR**]

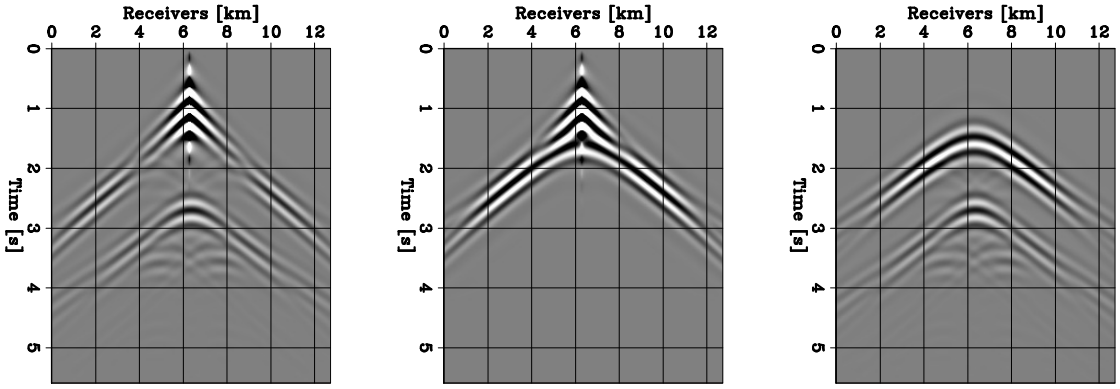


Figure 5.6: Shot gather for a source located at $x = 6.3$ km containing energy within the 1.5-6 Hz frequency range. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data with the initial velocity model, $\mathbf{f}(\mathbf{m}_0)$. (c) Initial data difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$. All panels are displayed with the same grayscale. [**ER**]

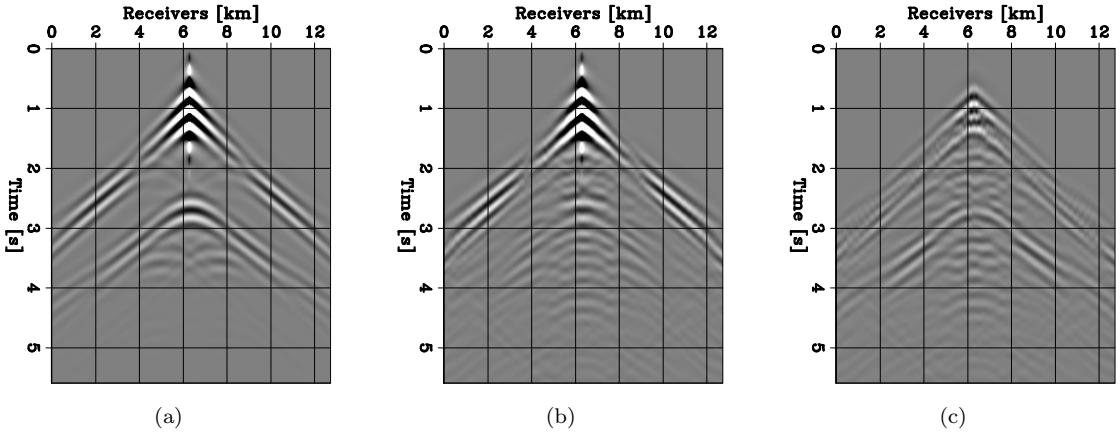
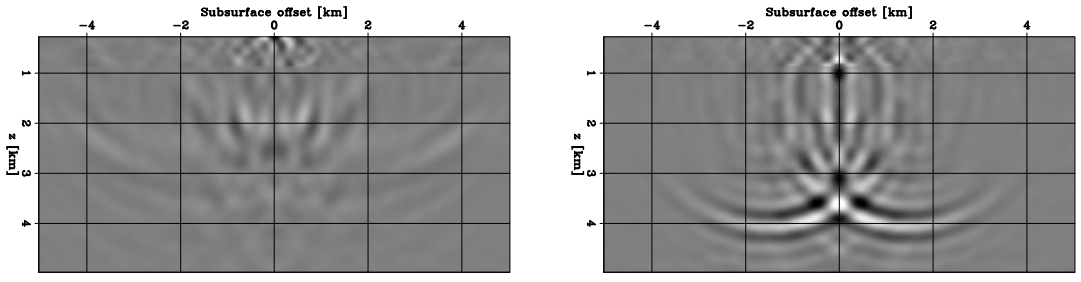


Figure 5.7: Representative shot gather for a source placed at x = 6.3 km. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data with the inverted model using conventional FWI, $\mathbf{f}(\mathbf{m}_{FWI})$. (c) Data difference, $\Delta\mathbf{d}(\mathbf{m}_{FWI}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWI})$. All panels are displayed with the same grayscale. [**CR**]

spacing of 50 m in the z-direction in order to preserve the sharpness present in the initial model. The second spline grid coincides with finite-difference grid. I use a horizontal subsurface-offset extension h_x for $\tilde{\mathbf{p}}_\epsilon^{opt}$ which is adequate for a dataset dominated by reflection events (Biondi and Almomin, 2014). Moreover, due to the large kinematic errors in the initial data-prediction, I use 201 points of extension, which allows h_x to range from -5 km to 5 km. Each variable projection step is conducted by minimizing objective function shown in 3.6 with 60 iterations of linear conjugate gradient.

Figure 5.8 shows two horizontal SODCIGs extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ at $x = 4.0$ km and $x = 6.0$ km, respectively. The SODCIG located at $x = 4$ km contains weaker energy/events due to the fact that at that location, the initial model is already accurate (the initial data-prediction $\mathbf{f}(\mathbf{m}_0)$ matches the observed data $\mathbf{d}^{obs}$ quite well for shots positions near $x = 4$ km). For SODCIGs closer to the basin, more energy from the data-difference is being mapped back into $\tilde{\mathbf{p}}_\epsilon^{opt}$. The initial kinematic error is so large that a maximum subsurface offset range of 8 km is needed to fully capture the energy (Figures 5.8b).

Figure 5.9a shows the scaled Born component of the first search direction and Figure 5.9b displays the scaled tomographic component. Note that the panels in Figures 5.9a and 5.9b are not scaled by the same value. The Born component is approximately two orders of magnitude weaker than the tomographic component, even without the use of a spline parametrization. Indeed, this is expected (and desired) at early stages of the FWIME workflow when the model updates should primarily be guided by the tomographic term. Figure 5.9c shows the total search direction (i.e., the sum of the two panels in Figures 5.9a and 5.9b), which is approximately identical to the tomographic component. Even though this search direction may seem promising (by comparing it to the ideal update in Figure 5.9d), its mapping onto the first spline grid removes most of the high-wavenumber artifacts and provides a much more favorable search direction (Figure 5.10a).



(a) (b)

Figure 5.8: Horizontal SODCIGs of the extended optimal perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed at the initial step and extracted at two horizontal positions. (a) $x = 4.0$ km. (b) $x = 6.0$ km. Both figures are displayed with the same grayscale. [**CR**]

I conduct 28 L-BFGS iterations of FWIME on the first spline grid. The spline grid refinement is triggered when the solver is unable to find a proper step length. The inverted model after the

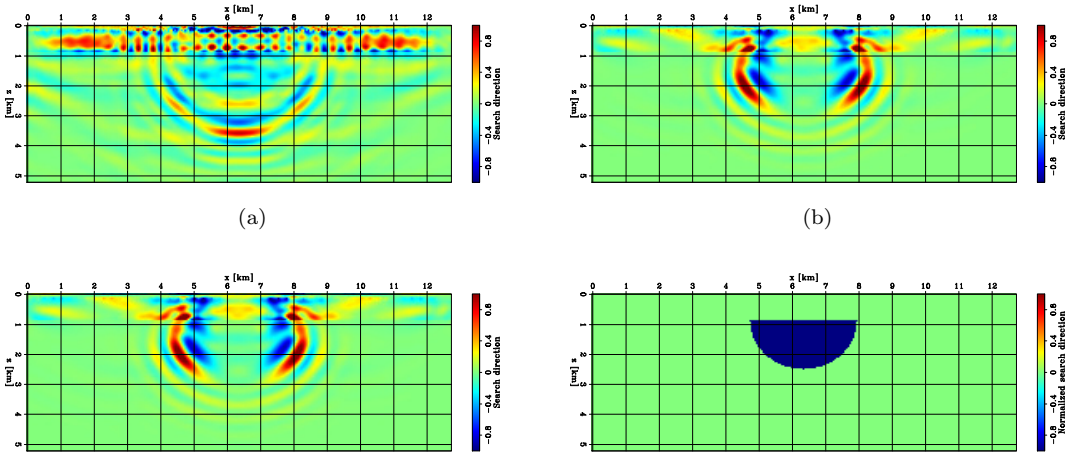


Figure 5.9: Scaled FWIME initial search directions on the finite-difference grid (without spline re-parametrization). (a) Born search direction. (b) Tomographic search direction. (c) Total search direction (sum of panels (a) and (b)). (d) True search direction. Note that (a) is scaled by a different value than (b) and (c) for clarity purpose. The amplitude of panel (a) is approximately two orders of magnitude smaller than the one of panels (b) and (c). [**CR**]

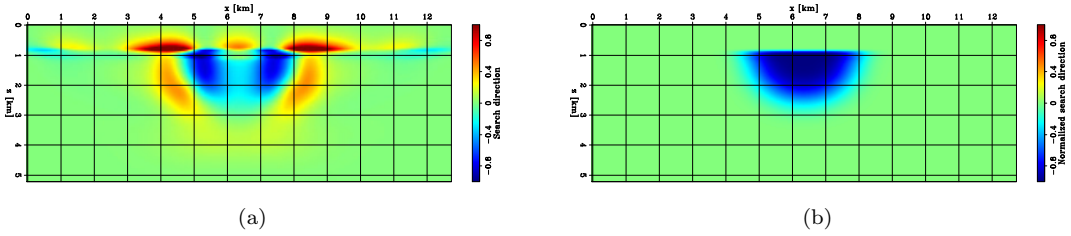


Figure 5.10: Scaled search directions after mapping on the first spline grid. (a) Total FWIME initial search direction. (b) True search direction. [**CR**]

first spline grid is shown in Figure 5.11a. I use the inverted model on spline 1 as the initial model for spline 2 (which coincides with the finite-difference grid), on which I conduct an additional 87 L-BFGS iterations of FWIME. The inverted model after inversion on spline 2 is shown in Figures 5.11b. Finally, I use the FWIME inverted model on spline grid 2 as an initial model for conventional FWI. The inverted FWIME model is shown in Figure 5.11c. Figures 5.12a and b show vertical and horizontal velocity profiles of the initial, true and FWIME inverted models at various positions. The final result is quite accurate but still contains a thin reflector located at the initially wrongly-positioned horizontal interface (at an approximate depth of $z = 0.8$ km). However, this artifact will likely disappear if the frequency content of the data were to be increased. Even though the FWIME results shown in Figure 5.11 are computed for a fixed ϵ-value of 1.0×10^{-5}, I also experimented different values (ranging from $\epsilon = 0.5 \times 10^{-5}$ to $\epsilon = 5.0 \times 10^{-5}$), which led to similar success.

Figure 5.13 shows the evolution of the predicted data $\mathbf{f}(\mathbf{m}_i)$ (middle column) and the data-difference $\Delta\mathbf{d}(\mathbf{m}_i) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_i)$ (right column) at various stages of the FWIME workflow (for a source placed at $x = 6.3$ km). The inversion begins by recovering a model that is able to generate some of the triplications in the wavefield (Figure 5.13e) by first removing a portion of the incorrect high-velocity zone at the bottom of the basin (Figure 5.11a). After the inversion on spline 2 (third row of Figure 5.13), most of the basin has been recovered, and the reflection from the initially mis-positioned horizontal interface (at $z \approx 0.8$ km) has been removed from the predicted data (Figure 5.13h). At the final stage (fourth row of Figure 5.13), the prediction is even more accurate, and the data-residuals have completely vanished (Figure 5.13k-l)

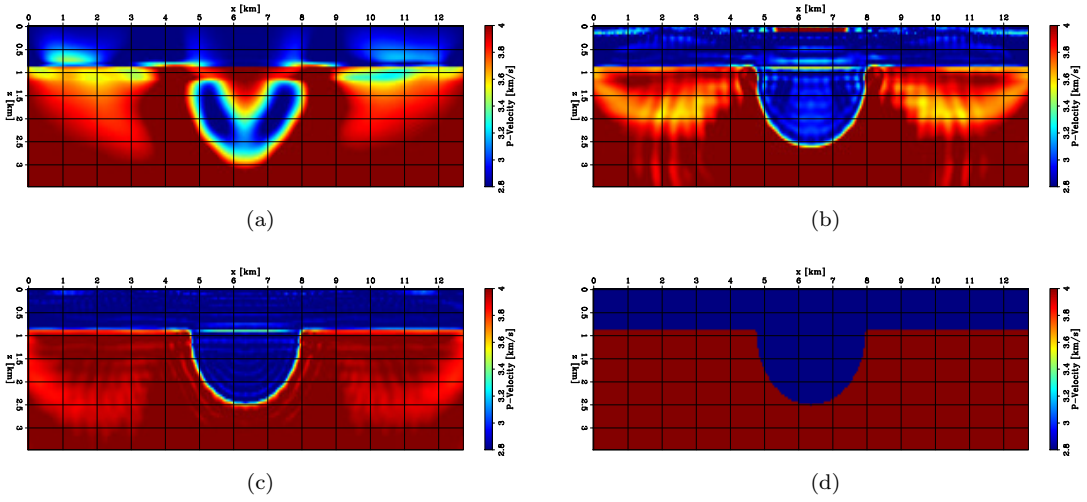


Figure 5.11: Inverted models at various stages of the FWIME workflow. (a) Inverted model after 28 iterations of FWIME on spline 1. (b) Inverted model after 87 iterations of FWIME on spline 2. (c) Inverted model after applying 200 iterations conventional data-space multi-scale FWI using (b) as an initial model. (d) True model. [**CR**]

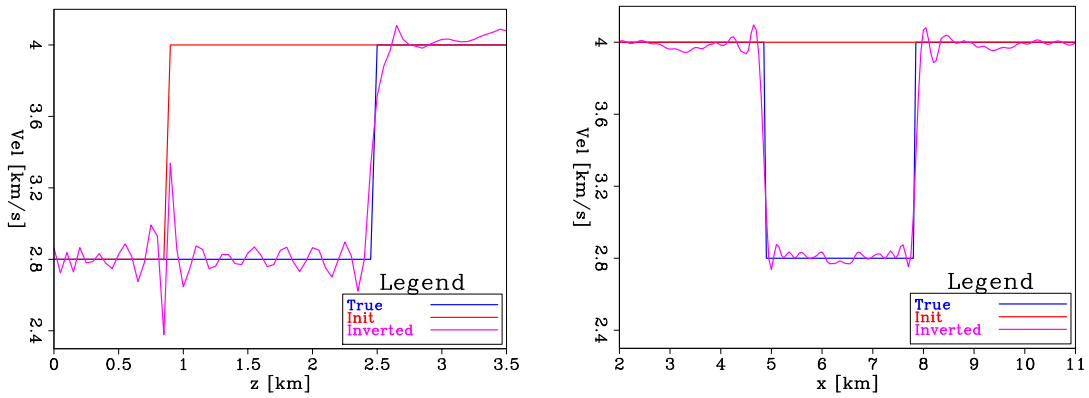


Figure 5.12: Velocity profiles of the initial model (red curve), true model (blue curve), and the final FWIME inverted model. (a) Vertical profile extracted at $x = 6$ km. (b) Horizontal profile extracted at $z = 1.5$ km. [**CR**]

For quality-control purposes, I migrate the observed data (to which I subtract the direct arrival) using the initial velocity model (Figure 5.14a) and the final velocity model (Figure 5.14b). The second panel shows a clear improvement, which is confirmed by the angle-domain common images gathers (ADCIGs) extracted at various positions in Figure 5.15. In the third and fourth columns of Figure 5.15, it is clear that the bottom of the basin (initially wrongly positioned) has been imaged quite accurately with a large improvement in the flatness of the ADCIGs. By examining the first row of Figure 5.15, it is interesting to notice that FWIME seems to be able to leverage any type of coherent moveout from the extended space of $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ to accurately update the velocity model. Even tough not tested here, I anticipate that more conventional techniques based on interpreting the curvature of the the moveouts within the ADCIGs (Zhang, 2015) would have not performed as well.

5.2 Marmousi2 model

The Marmousi2 synthetic model simulates a geologically complex subsurface structure and is well-suited for calibrating velocity-model building algorithms (Martin et al., 2006). I use this benchmark test to show that FWIME can successfully invert reflection-dominated datasets and recover accurate solutions without the need for low-frequency signal or good initial guesses. Compared to the original Marmousi (Versteeg and Grau, 1990), this model is wider and more representative of current long-offset offshore acquisition geometries. The sediments are placed under a thicker layer of water of approximately 475 m, which reduces the presence of refracted energy recorded in the data. Finally, additional hydrocarbon reservoirs and stratigraphic features have been added to increase the overall geological complexity of the model (Martin et al., 2002).

The true model $\mathbf{m}_{true}$ is 17 km wide and 3.5 km deep (Figure 5.16d). The initial model $\mathbf{m}_0$

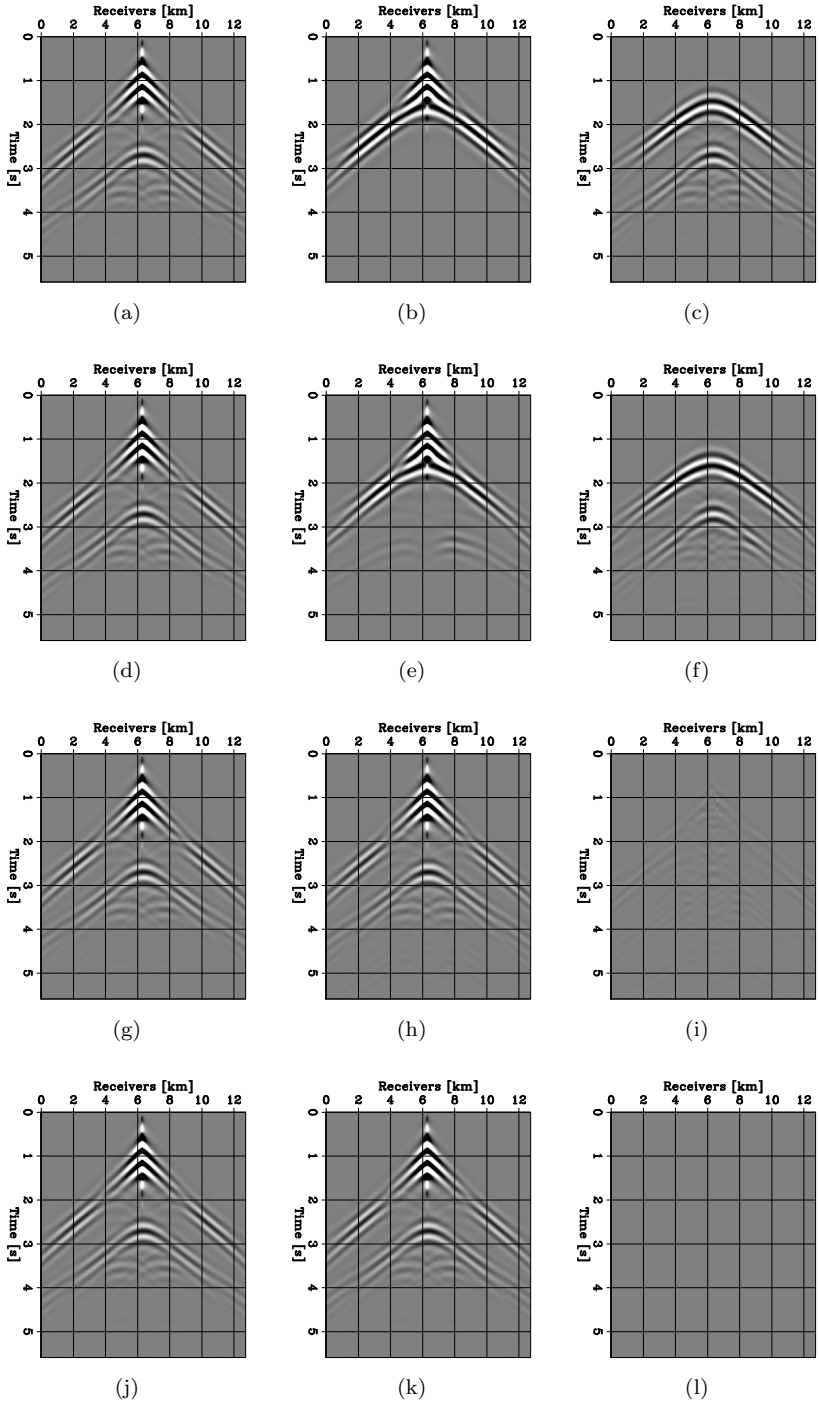


Figure 5.13: Observed data (left column), predicted data (middle column), and data difference (right column) for a shot located at $x = 6.3$ km computed with the FWIME inverted models at various stages of the FWIME workflow. Initial model (first row), inverted model on spline 1 (second row), inverted model spline 2 (third row), and final inverted model (last row). All panels are displayed on the same grayscale. [**CR**]

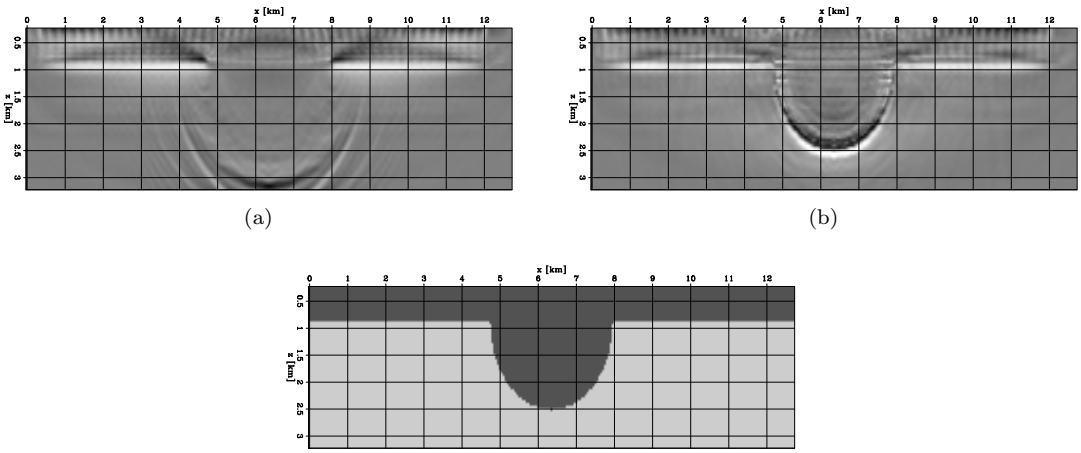


Figure 5.14: Zero-offset migrated images computed with different velocity models. (a) Initial velocity model. (b) Final FWIME inverted model. (c) True velocity model (for reference). Panels (a) and (b) are displayed with the same grayscale. [**CR**]

is designed to be inaccurate (Figure 5.16a). Below the water layer (the bathymetry is assumed to be known for this problem), the velocity is laterally invariant and linearly increasing with depth. Velocity profiles of the true and initial models extracted at four horizontal positions are displayed in Figure 5.17 (red and and black curves). At the surface, I place 140 sources every 120 m, and 567 fixed receivers every 30 m. I generate noise-free pressure data with a source wavelet containing energy restricted to the 4-13 Hz range (Figure 5.18a), and a finite-difference grid spacing of 30 m in both directions. Figures 5.19a and 5.19d show two representative shot gathers from the observed data for sources placed at $x = 1.2$ km and $x = 8.4$ km. Even though some refracted energy is recorded at offsets larger than 7 km, the dataset is strongly dominated by reflected events. In addition, the inaccuracy of the initial data can be appreciated in Figures 5.19b and 5.19e, which show that the initial prediction $\mathbf{f}(\mathbf{m}_0)$ fails to generate any reflection. The inaccuracy of $\mathbf{m}_0$ coupled with the lack of low-frequency and refracted energy (i.e., diving waves) make this test quite challenging for conventional velocity-model building techniques. To illustrate this claim, I apply data-space multi-scale FWI using five frequency bands (Figure 5.18b), which fails to retrieve a useful solution, especially in the deeper section of the model (Figure 5.16b).

I describe the initial step of FWIME which includes a hyper-parameter tuning process and the computation of the initial search directions. Then, I present and carefully analyze my inversion results.

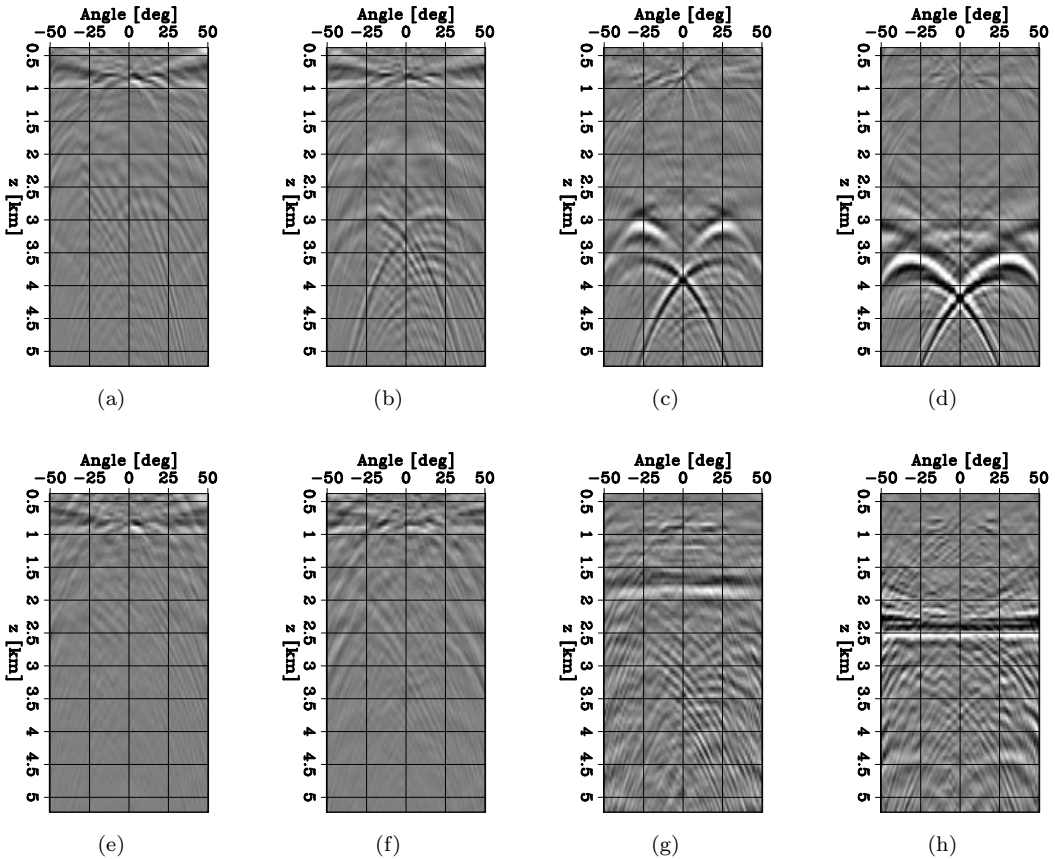


Figure 5.15: Angle domain common image gathers (ADCIGs) computed with the initial model (top row) and the final FWIME inverted model (bottom row) at four different horizontal positions. First column is at $x = 3$ km, second column is at $x = 4$ km, third column is at $x = 5$ km, and fourth column is at $x = 6$ km. All panels are displayed with the same grayscale. [**CR**]

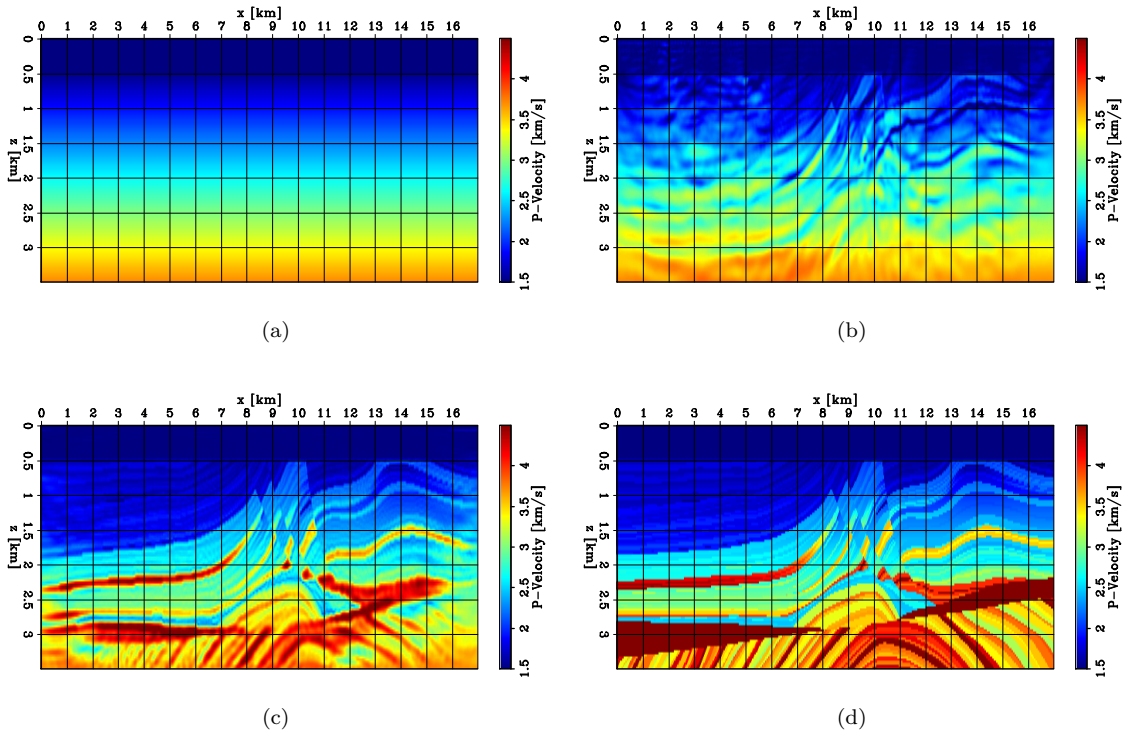


Figure 5.16: 2D panels of velocity models. (a) Initial model. (b) Inverted model after conventional data-space multi-scale FWI using five frequency bands. (c) Final FWIME inverted model. (d) True model. [**CR**]

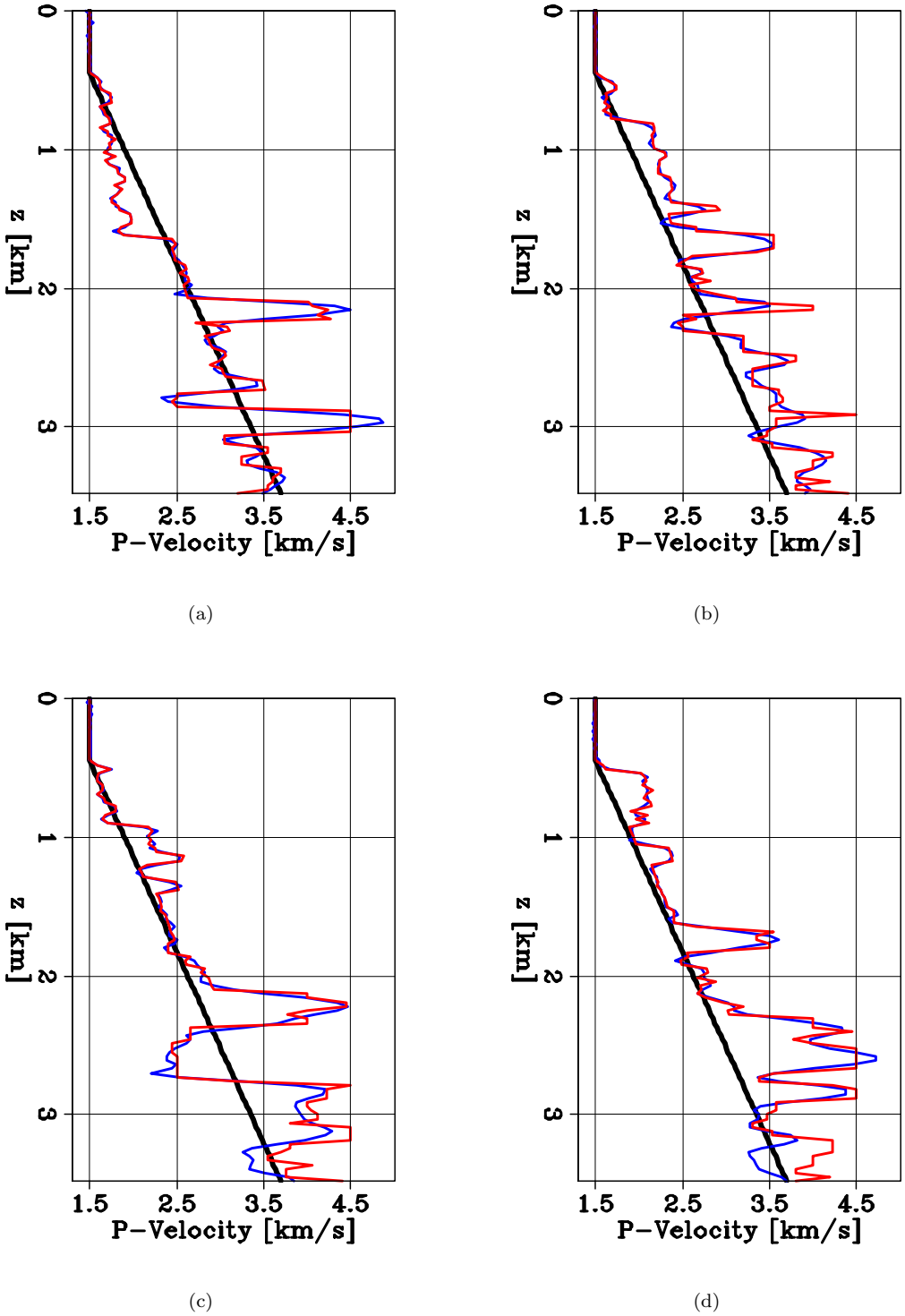


Figure 5.17: Depth velocity profiles extracted at (a) $x = 6$ km, (b) $x = 9$ km, and (c) $x = 11$ km, and (d) $x = 13$ km. The black curve represents the initial model, the red curve is the true model, and the blue curve is the inverted FWIME model. [**CR**]

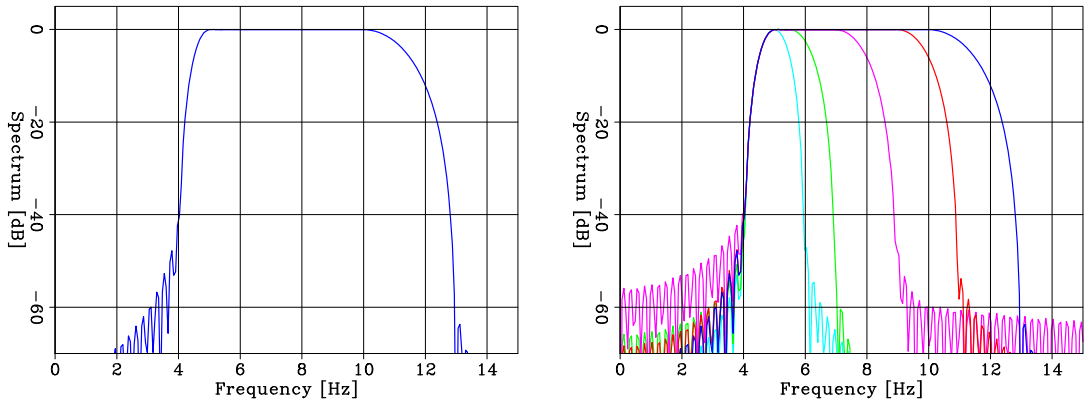


Figure 5.18: Amplitude spectra of the seismic sources used to generate the various datasets for this numerical example. (a) Source used for the FWIME workflow. (b) Sequence of sources used for the data-space multi-scale FWI workflow. [**ER**]

Selection of the extension

The first step consists in selecting the optimal extension type and the length of the extended axis for $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ such that the initial data-correcting term $\tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0)$ is able to match the initial data difference $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ (i.e., satisfy equation 3.23). As explained in section 3.3.1, this condition ensures that without penalizing defocused energy within $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ (i.e., when $\epsilon = 0$), the initial prediction error can be fully explained by the mapping of $\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0)$ into the data space. Indeed, failing to select an adequate extension length may result in cycle-skipping, which in turn may lead FWIME to converge to a local minimum. For computational efficiency, I wish to find the optimal extension type that satisfies equation 3.23 with the smallest number of points on the extended axis. For that, I set $\mathbf{m} = \mathbf{m}_0$ and $\epsilon = 0$, and I minimize objective function 3.6 for (1) time lags and (2) horizontal subsurface offset, by conducting two separate inversions. For both inversions, I use an extended axis of 101 points. Hence, τ ranges from -0.8 s to 0.8 s, whereas h_x ranges from -1.5 km to 1.5 km. Figure 5.20a shows the normalized convergence curves for the minimization of objective function 3.6 (initial variable projection step) using time lags (blue curve), horizontal subsurface offsets (red curve), and with no extension (pink curve). As expected, the non-extended inversion fails to reduce the misfit to zero. In addition, the time-lag extension seems to perform better than subsurface offsets as it manages to decrease the objective function value by approximately 99.4% after 60 iterations of linear conjugate gradient, and by 99.9% after 100 iterations (compared to 96% and 97% for subsurface offsets), thereby satisfying the condition expressed in equation 3.23 more efficiently. Indeed, a more accurate matching (up to numerical precision) would require more iterations. For this numerical example, I choose a time-lag extension with a length of 101 points (a very conservative number) sampled at 16 ms, and I limit the number of linear conjugate gradient

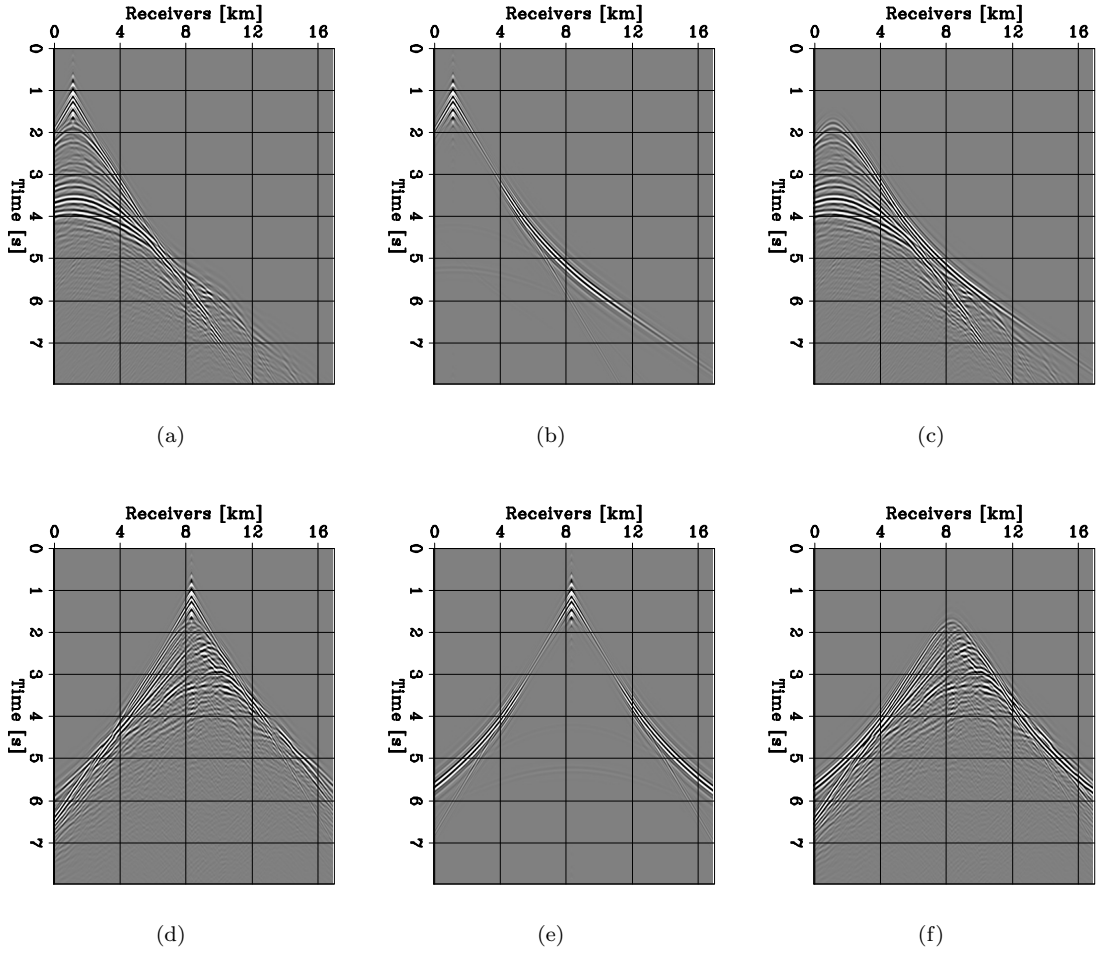
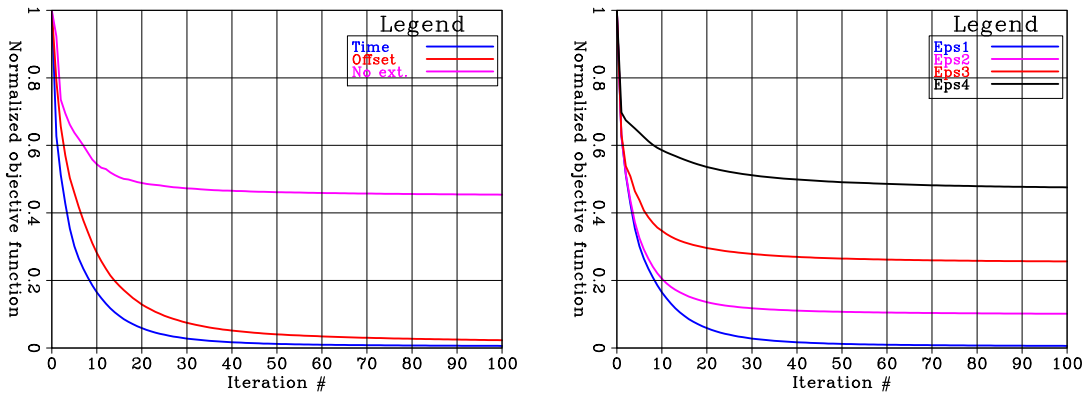


Figure 5.19: Representative shot gathers for sources placed at $x = 1.2$ km (first row) and $x = 8.4$ km (second row). Observed data, $\mathbf{d}^{obs}$ (first column), Predicted data with the initial model, $\mathbf{f}(\mathbf{m}_0)$. (second column), and data-difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ (third column). All panels are displayed with the same grayscale. [**ER**]

iterations to 60 for the variable projection step. To further reduce the computational cost for 3D field applications, a smaller number of points on the extended axis could likely be chosen (while still satisfying equation 3.23).

Selection of the trade-off parameter ϵ

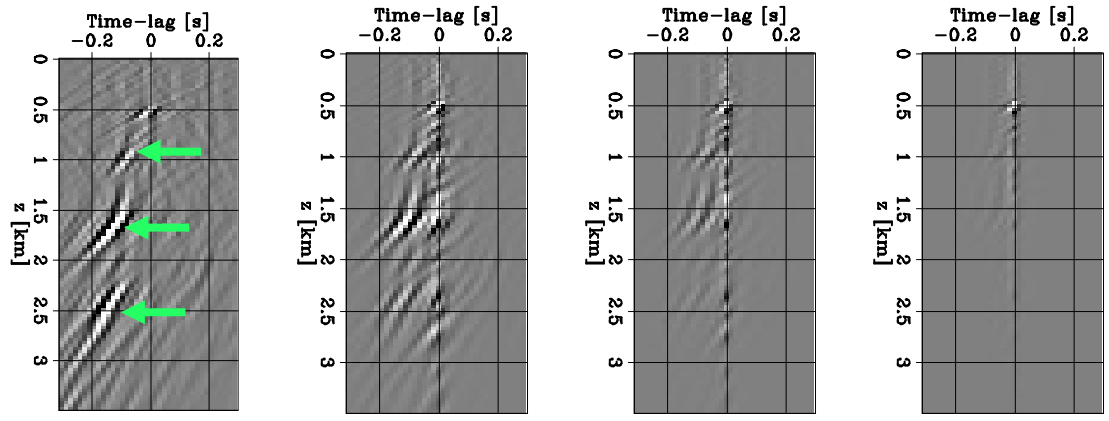
The second step consists in selecting the optimal ϵ-value. As explained in section 3.3.4, ϵ serves as a trade-off parameter that controls the level of data-fitting by penalizing the presence of energy within $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ during the minimization of objective functions 3.5 and 3.6. To illustrate its effect on this numerical example, I compute $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ (using the initial velocity model $\mathbf{m}_0$) for four ϵ-values by minimizing objective function 3.6. The corresponding convergence curves are shown in Figure 5.20b. As the ϵ-value increases, the ability of the data-correcting term to predict the initial data misfit $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ is reduced. Figure 5.21 shows a TLCIG extracted at $x = 14$ km from the four inverted $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ (corresponding to the four convergence curves shown in Figure 5.20b), while Figure 5.22 represents the difference between the corresponding data-correcting term and the initial data-residual (i.e., the adjoint source defined in equation 4.5): $\mathbf{r}_d^{\epsilon}(\mathbf{m}_0) = \tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$.



(a) (b)

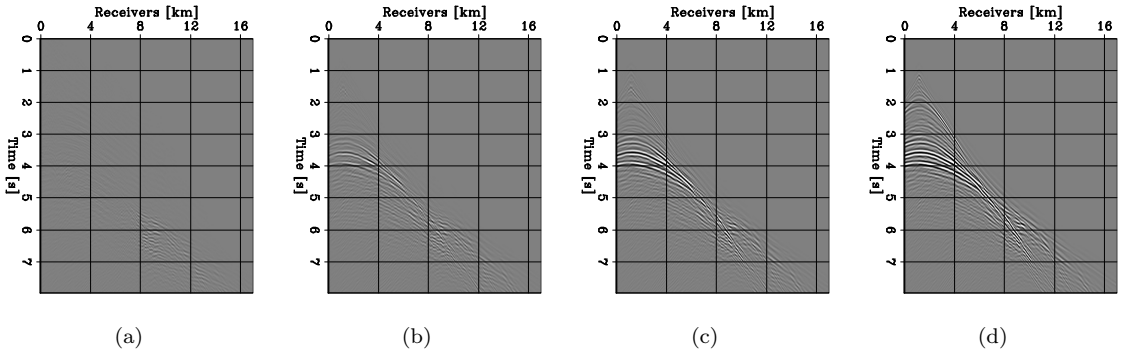
Figure 5.20: Normalized objective functions corresponding to the minimization of equation 3.6 (initial variable projection step) with different extensions and ϵ-values. (a) Time-lag extension (blue curve), horizontal subsurface-offset extension (red curve), and non-extended (pink curve). For (a), I set $\epsilon = 0$. (b) Time-lag extension with different ϵ-values: $\epsilon = 0$ (blue curve), $\epsilon = 1.5 \times 10^{-5}$ (red curve), $\epsilon = 5.0 \times 10^{-5}$ (pink curve), and $\epsilon = 5.0 \times 10^{-4}$ (black curve). In both panels, all curves are normalized by the same value. [**CR**]

For $\epsilon = 0$, no constraint is applied on $\tilde{\mathbf{p}}_{\epsilon}^{opt}$, and therefore a considerable amount of energy is mapped away from the physical plane (Figure 5.21a). Consequently, the data-correcting term is able to match the initial data residual with high accuracy, and almost no coherent signal is present within the adjoint source (Figure 5.22a). However, increasing the ϵ-value penalizes defocused events within $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ and reduces the amount of energy spread away from the physical plane (Figures 5.21b and c).



(a) (b) (c) (d)

Figure 5.21: TLCIGs extracted at $x = 14$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}_0)$ computed with four different ϵ-values. (a) $\epsilon = 0$, (b) $\epsilon = 1.5 \times 10^{-5}$, (c) $\epsilon = 5.0 \times 10^{-5}$, and (d) $\epsilon = 5.0 \times 10^{-4}$. All panels are plotted for $\tau \in [-0.3 \text{ s}, \ 0.3 \text{ s}]$ for display purposes. However, all the computations are conducted for the full time-lag range with $\tau \in [-0.8 \text{ s}, \ 0.8 \text{ s}]$. All panels are displayed with the same grayscale. [**NR**]



(a) (b) (c) (d)

Figure 5.22: Representative shot gathers generated by a source located at $x = 1.2$ km displaying the FWIME adjoint source, $\mathbf{r}_d^{\epsilon}(\mathbf{m}_0) = \tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$, computed for four different ϵ-values ($\mathbf{m}_0$ is fixed). (a) $\epsilon = 0$, (b) $\epsilon = 1.5 \times 10^{-5}$, (c) $\epsilon = 5.0 \times 10^{-5}$, and (d) $\epsilon = 5.0 \times 10^{-4}$. All panels are displayed with the same grayscale. [**CR**]

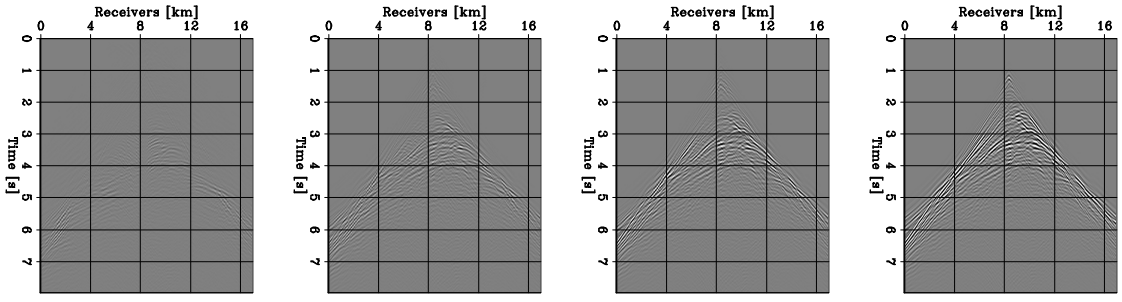


Figure 5.23: Representative shot gathers generated by a source located at $x = 8.4$ km displaying the FWIME adjoint source, $\mathbf{r}_d^{\epsilon}(\mathbf{m}_0) = \tilde{\mathbf{B}}(\mathbf{m}_0)\tilde{\mathbf{p}}_0^{opt}(\mathbf{m}_0) - \big(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)\big)$, computed for four different ϵ-values ($\mathbf{m}_0$ is fixed). (a) $\epsilon = 0$, (b) $\epsilon = 1.5 \times 10^{-5}$, (c) $\epsilon = 5.0 \times 10^{-5}$, and (d) $\epsilon = 5.0 \times 10^{-4}$. All panels are displayed with the same grayscale. [**CR**]

The data-correcting term is then unable to accurately match the initial data-misfit (Figures 5.22b and c). Eventually, for an extremely high ϵ-value, both $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ and the data-correcting term vanish (Figures 5.21d and 5.22d), and $\mathbf{r}_d^{\epsilon}(\mathbf{m}_0) \approx \mathbf{f}(\mathbf{m}_0) - \mathbf{d}^{obs}$. Mathematically, minimizing the FWIME cost function (equation 3.5) with this large ϵ-value would correspond to conducting a non-extended FWIME, or equivalently, conventional FWI.

During the minimization of objective function 3.6, the background velocity model $\mathbf{m}$ is constant, and thus the ϵ-value only affects the amplitude of the events within $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ rather than the kinematics (the reflectors' positions are not affected by ϵ). In Figure 5.21a (i.e., for $\epsilon = 0$), there are three distinct events located in the extended space at negative time-lag values (green arrows), which provide valuable information on velocity errors present in the initial model. In this particular case, they indicate that the initial velocity is too low. On one hand, selecting a too small ϵ-value is sub-optimal because the amplitude of the corresponding adjoint source $\mathbf{r}_d^{\epsilon}(\mathbf{m}_0)$ (employed in the gradient computation) would not contain any coherent information (its amplitude would be numerically close to zero), as shown in Figure 5.22a. On the other hand, a too-high ϵ-value would force the amplitude of the events in the extended space to vanish, and the information they carry about velocity errors would be lost (Figure 5.21d). In addition, the data-correcting term would not be able to match the initial data residual Figures 5.22d, resulting in cycle-skipping. For this numerical example, I observe that for $\epsilon \in [1.5 \times 10^{-5}, 5.0 \times 10^{-5}]$, FWIME converges to similar accurate solutions.

In this thesis, I propose to select the ϵ-value by examining a subset of the TCLIGs extracted from the initial $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed with a few ϵ-values. While I acknowledge the user-intensive nature of this approach, I decide to leave the investigation and development of a more efficient strategy for future work.

Figure 5.24 shows the Born, tomographic, and total FWIME search directions on the finite-difference grid computed with $\epsilon = 1.5 \times 10^{-5}$. The total search direction is mainly guided by the tomographic component and seems to accurately capture some of the low-wavenumber features present in the ideal search direction in the shallow region (Figures 5.24c and d). However, it is overwhelmed by high-wavenumber noise. To mitigate this effect, I use an initial coarse spline grid with a sampling of 0.2 km and 1 km in the vertical and horizontal directions, respectively. Figure 5.25 shows the analogous panels from Figure 5.24 after their mapping onto the first spline grid by applying operator $\mathbf{SS}^*$ (equations 4.12 and 4.13). As expected, the total FWIME search direction is improved.

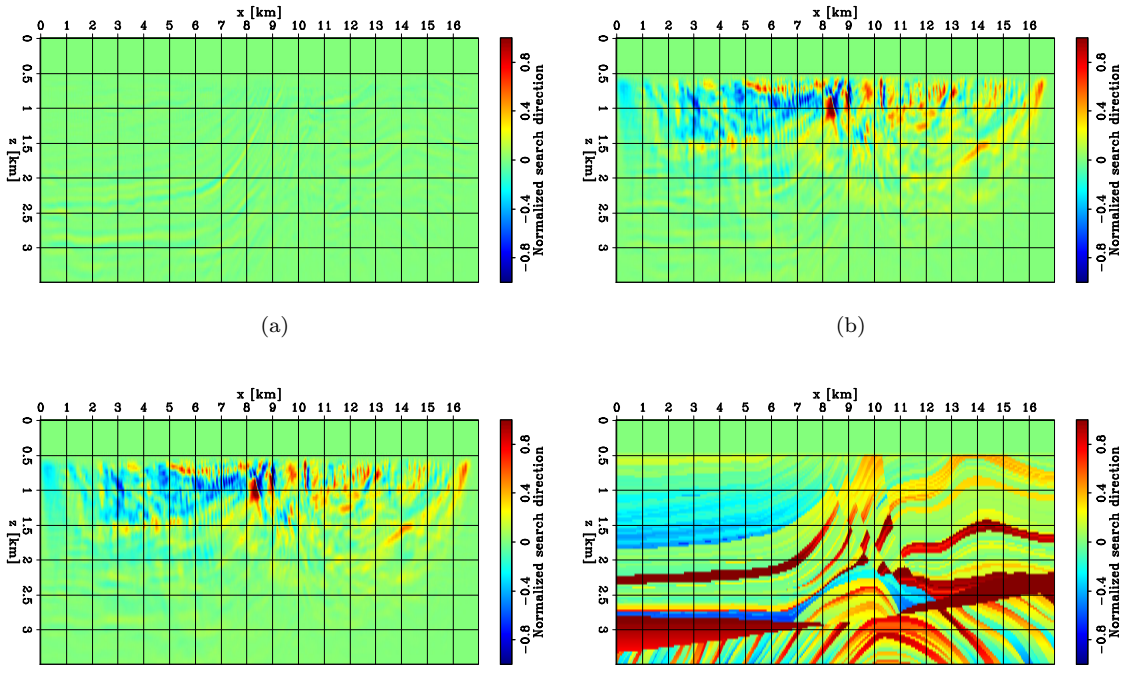


Figure 5.24: Initial FWIME search directions on the finite-difference grid (before spline re-parametrization), computed for $\epsilon = 1.5 \times 10^{-5}$. (a) Born component, (b) tomographic component, (c) total search direction, and (d) true search direction. Panels (a)-(c) are normalized with the same value. [**CR**]

For the FWIME scheme, I use a sequence of four spline grids and I keep the same ϵ-value throughout the entire process ($\epsilon = 1.5 \times 10^{-5}$). Each spline grid refinement is triggered when the stepper is unable to find an appropriate step length for that particular grid. The spacing in the second and third grids are obtained by halving the spacing from the previous ones. The final spline grid coincides with the finite-difference grid ($\Delta z = \Delta x = 30$ m). These parameters are summarized in Table 5.1.

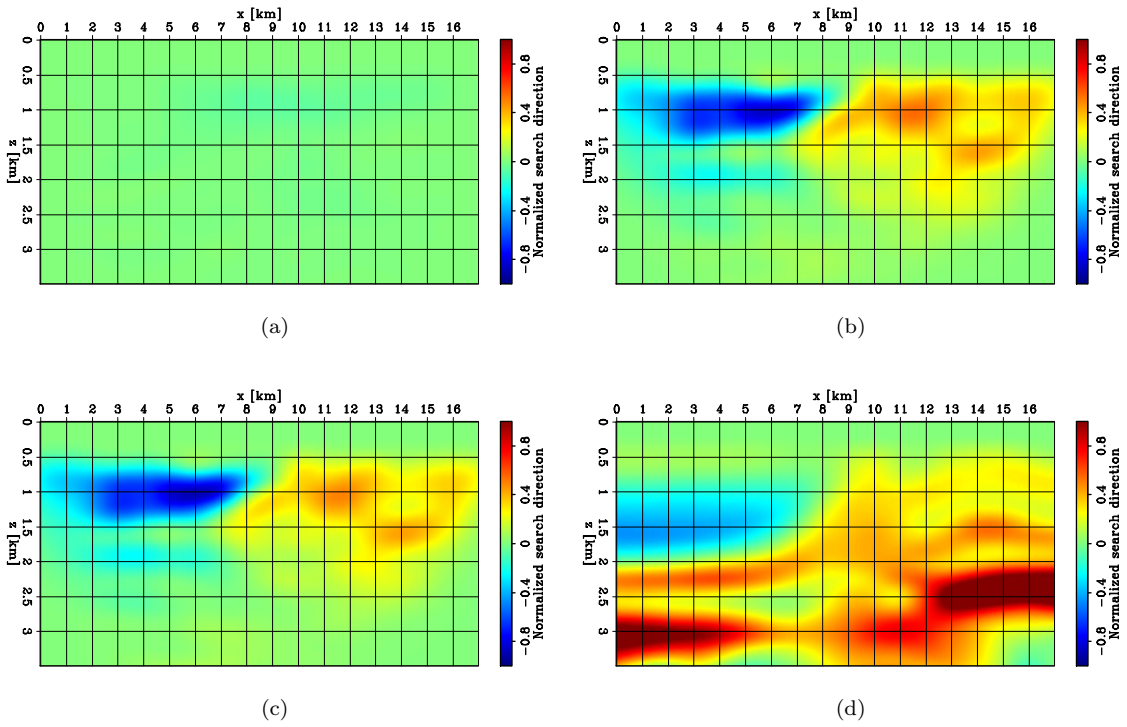


Figure 5.25: Initial FWIME search directions after their mapping on the initial spline grid, computed for $\epsilon = 1.5 \times 10^{-5}$. (a) Born component, (b) tomographic component, (c) total search direction, and (d) true search direction. Panels (a)-(c) are normalized with the same value. [**CR**]

Grid number	Δz [km]	Δx [km]
1	0.2	0.9
2	0.1	0.45
3	0.05	0.21
4	0.03	0.03

Table 5.1: Parameters of the sequence of spline grids used for the model-space multi-scale FWIME scheme. Spline 4 coincides with the finite-difference grid.

Figure 5.26b-d show the FWIME inverted model after spline 1, 2, and 3, respectively. The final FWIME inverted model after 240 iterations of L-BFGS is shown in Figure 5.26e. Even though it suffers from minor edge effects due to the limited acquisition aperture, it is very accurate, as confirmed by the velocity profiles shown in Figure 5.17 (blue curves).

Figure 5.27a shows the value of the normalized FWIME objective functions components as a function of iterations throughout the four different stages of the inversion process (i.e., on the four spline grids). The blue curve corresponds to the total FWIME objective function, the red curve corresponds to the FWIME data-fitting component, and the pink curve displays the FWIME annihilating component. The three abrupt changes in convexity occurring at iterations 20, 80, and 180 correspond to a spline grid refinement.

Figure 5.27b displays the normalized total FWIME objective function (blue curve) along with the value of the conventional FWI objective function evaluated at each inverted model during the FWIME sequence (red curve). The red curve is not the result of an inversion process, but simply an evaluation of the conventional FWI objective function at each FWIME inverted model. It can also be interpreted as a measure of how well the FWIME estimated model explains all the events in the recorded data. Moreover, it is not monotonically decreasing, which illustrates that the FWIME algorithm has successfully created an alternate descent-path towards the global minimum that could not have been taken by conventional FWI. Eventually, both curves in Figure 5.27b converge to zero (up to numerical precision) which means that FWIME has managed to find an inverted model that matches all the events on the observed data without the need for the additional data-correcting term.

I analyze $\tilde{\mathbf{p}}_\epsilon^{opt}$ by examining the evolution of the TLCIG at $x = 14$ km (Figure 5.29) and the zero-lag cross section of $\tilde{\mathbf{p}}_\epsilon^{opt}$ (Figure 5.28) extracted at four stages of the inversion process. Initially, the energy is clustered away from the physical plane (Figure 5.29a) and the zero-lag cross-section lacks coherency (Figure 5.28a). As the velocity model becomes more accurate, the energy gradually focuses towards the physical plane (Figures 5.29b-d), and the coherency of the zero-lag section is simultaneously enhanced (Figures 5.28b-d). As expected, when algorithm converges, $\tilde{\mathbf{p}}_\epsilon^{opt}$ vanishes (Figures 5.29e and 5.28e).

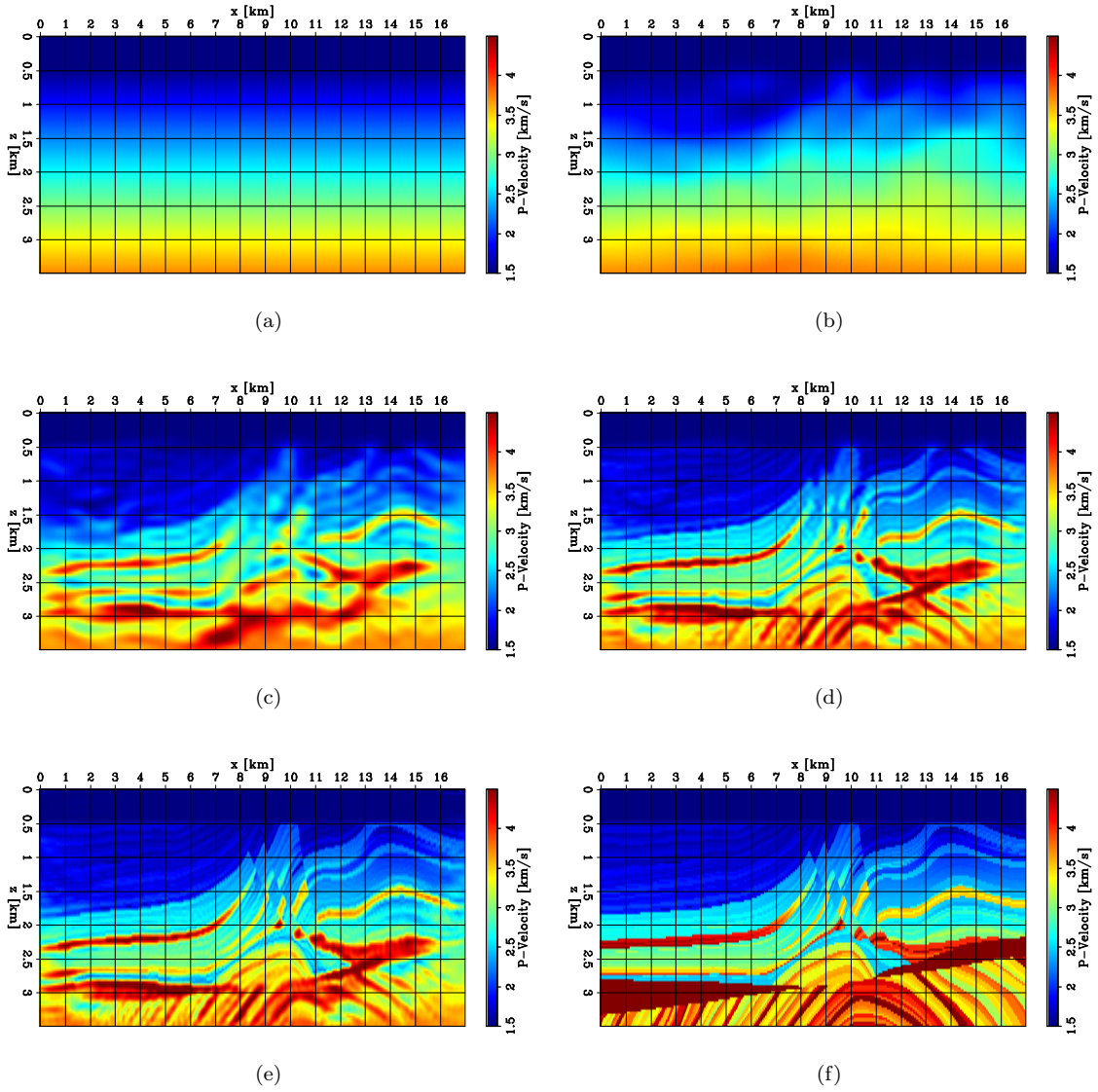


Figure 5.26: Inverted models at different stages of the model-space multi-scale FWIME workflow with $\epsilon = 1.5 \times 10^{-5}$. (a) Initial model. (b) Inverted model after 25 iterations on the first spline grid. (c) Inverted model after 55 iterations on the second spline grid. (d) Inverted model after 110 iterations on the third spline grid. (e) Inverted model after 50 iterations on the fourth (finite-difference) grid. (f) True model. [**CR**]

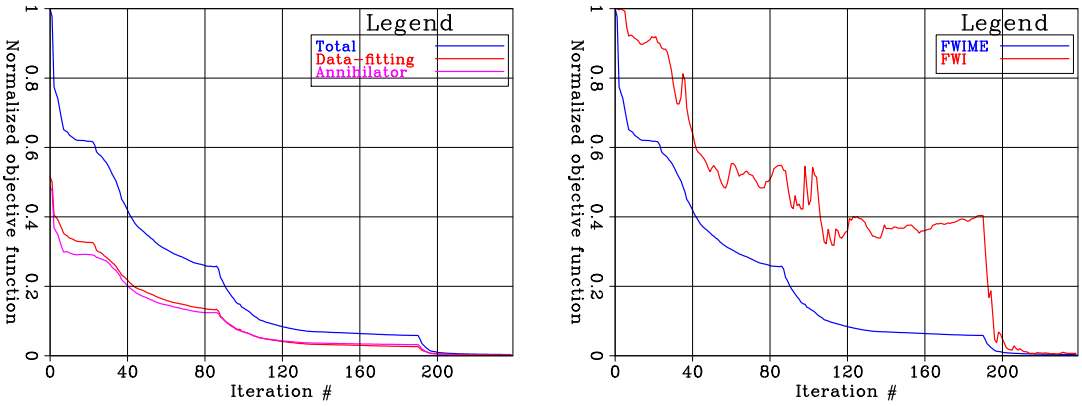


Figure 5.27: Normalized objective functions. (a) Total FWIME objective function (blue curve), FWIME data-fitting component (red curve), and FWIME annihilating component (pink curve). (b) Total FWIME objective function (blue curve), and FWI objective function evaluated at each FWIME inverted model (red curve). [**CR**]

As I explain in section 3.1.3, the goal of my algorithm is to recover a starting model for conventional FWI to converge to a useful solution. In this numerical example, I conduct FWIME until completion in order to illustrate the potential of the method. However, for 3D field applications, the optimization would have been stopped at earlier stages. Figure 5.30a shows the result of applying FWI using the FWIME inverted model on the first spline grid, and indicates that 20 iterations of FWIME would have been sufficient to converge to an accurate solution. Beyond that point, FWI and FWIME perform equally well (Figure 5.30b-d).

I apply an imaging quality-control step in order to assess the accuracy of the inverted velocity model. Figures 5.31a and b show the zero-offset sections of the initial and final migrated images, respectively. By comparing them to the true model (Figure 5.31c), I can clearly see an improvement in the image, especially in the vicinity of the reservoir (pink box) where none of the features were initially visible. This observation is also confirmed by noticing the flattening of the subsurface angle domain common image gathers (ADCIG) shown in Figures 5.32 extracted at four horizontal positions ranging between $x = 9$ km and $x = 12$ km. The top and bottom rows of Figure 5.32 shows the ADCIGs computed with the initial and final models, respectively. As expected, the final velocity improves the image of the different interfaces at the reservoir level.

Reducing FWIME's computational cost

The most computationally intensive component of FWIME is the variable projection step, which consists in iteratively minimizing the quadratic objective function defined in equation 3.6 with a linear conjugate-gradient scheme. From a mathematical standpoint, this equation must be accurately solved for the expression of the gradient in equation 4.1 to be valid, as I demonstrate in Appendix C.

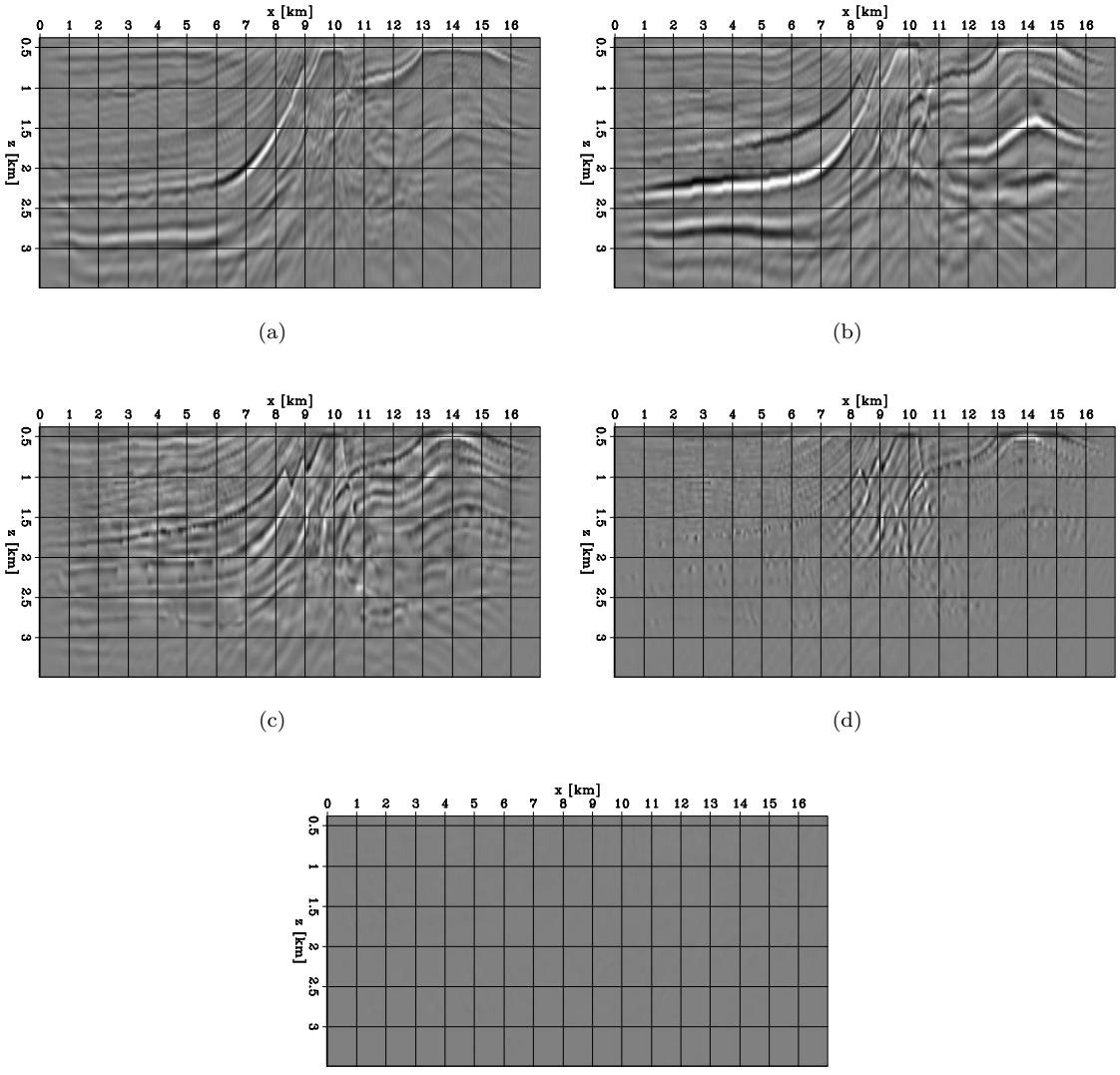


Figure 5.28: Zero time-lag sections of $\mathbf{p}_{\epsilon}^{opt}$ computed at four stages of the FWIME workflow. (a) Initial step. (b) After inversion on spline 1. (c) After inversion on spline 2. (d) After inversion on spline 3. (e) Final step. All panels are displayed with the same grayscale. [**CR**]

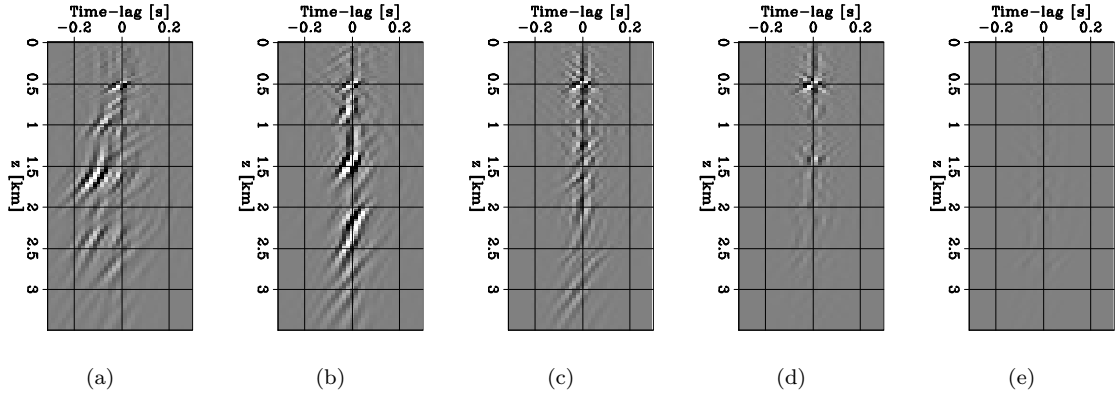


Figure 5.29: Time-lag common image gathers (TLCIG) extracted at $x = 14$ km from $\mathbf{p}_\epsilon^{opt}$ computed at four stages of the FWIME workflow. (a) Initial step. (b) After inversion on spline 1. (c) After inversion on spline 2. (d) After inversion on spline 3. (e) Final step. All panels are plotted for $\tau \in [-0.3 \text{ s}, \ 0.3 \text{ s}]$ for display purposes. However, all the computations are conducted for the full time-lag range with $\tau \in [-0.8 \text{ s}, \ 0.8 \text{ s}]$. All panels are displayed with the same grayscale. [**CR**]

From a numerical aspect, this condition implies that equation 3.6 should be minimized until full convergence. That is, by conducting "enough" linear conjugate-gradient iterations. In order to reduce the computational cost of my method, I assess how solving this quadratic problem less accurately (i.e., with less linear conjugate-gradient iterations) impacts the quality of the FWIME solution.

The inverted model shown in Figure 5.26e is obtained by solving the variable projection problem with 60 iterations, which accounts for 97% of the total FWIME cost. The resulting convergence curve computed at the initial step is shown in Figure 5.34. I now conduct analogous FWIME schemes with the same hyper-parameter selection but by reducing the number of linear iterations during the variable projection step. Figures 5.33a-5.33d show the FWIME inverted models obtained by minimizing equation 3.6 with 7, 10, 15, and 20 linear iterations, respectively. Moreover, for a fair cost comparison, the inverted models in Figures 5.26e and 5.33 were obtained with a total of 240 nonlinear iterations of L-BFGS. The inverted results become less accurate for 10 linear iterations or less (Figures 5.33b and 5.33a), but seem unaffected when the number of iterations is set to 15 or higher (Figures 5.33c and 5.33d). In this numerical test, reducing the number of linear iteration to 15 corresponds to a computational cost decrease of 73%. Even though this behavior may be case dependent, it shows potential value in trying to reduce the number of linear iterations for the variable projection step in FWIME, especially for 3D applications.

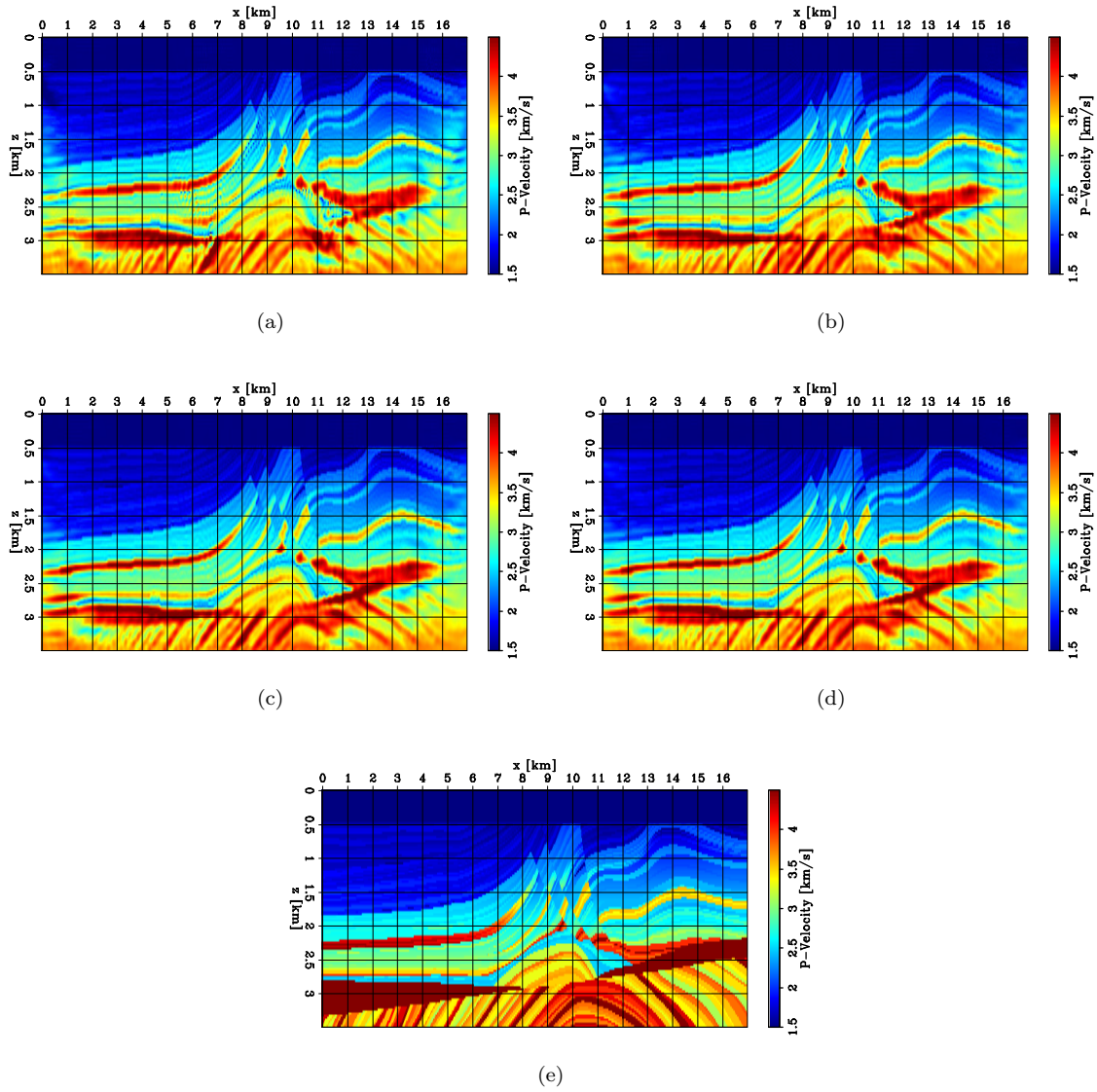


Figure 5.30: 2D panels showing the result of applying FWI using various FWIME inverted models as initial guesses. (a) FWIME inverted model on spline 1 (Figure 5.26b). (b) FWIME inverted model on spline 2 (Figure 5.26c). (c) FWIME inverted model on spline 3 (Figure 5.26d). (d) Final FWIME inverted model (Figure 5.26e). (e) True model. [**CR**]

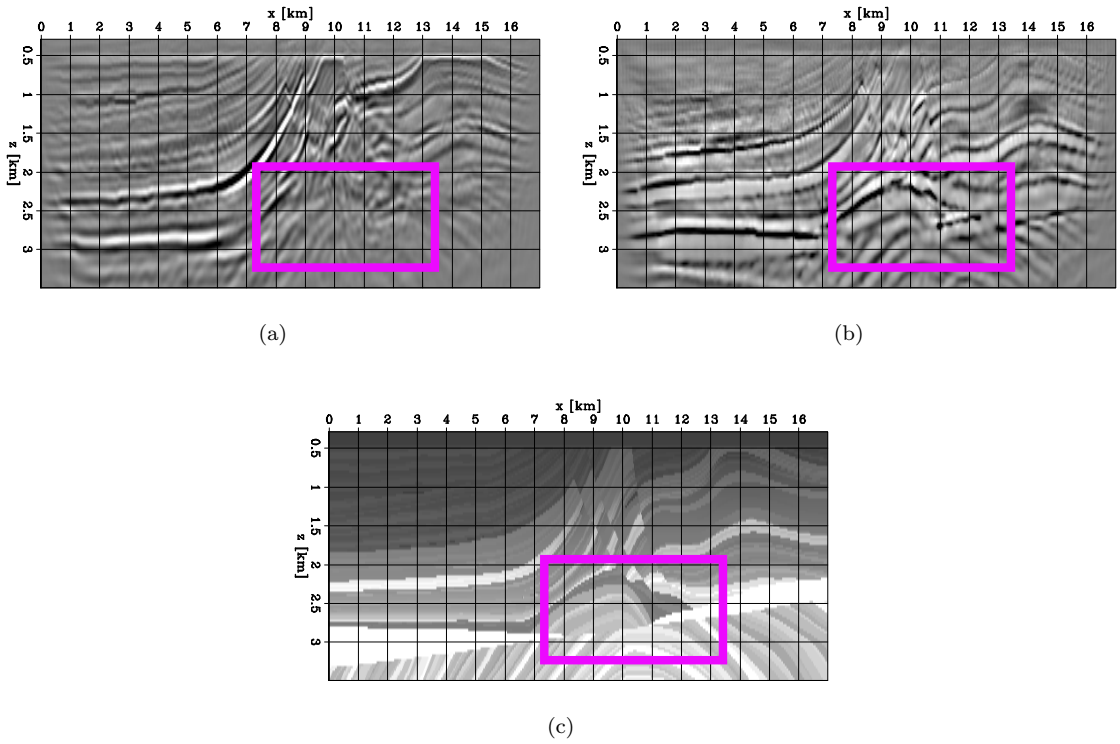


Figure 5.31: Zero-offset sections of migrated images computed with different velocity models. (a) Initial model. (b) FWIME inverted model. (c) True model. [**NR**]

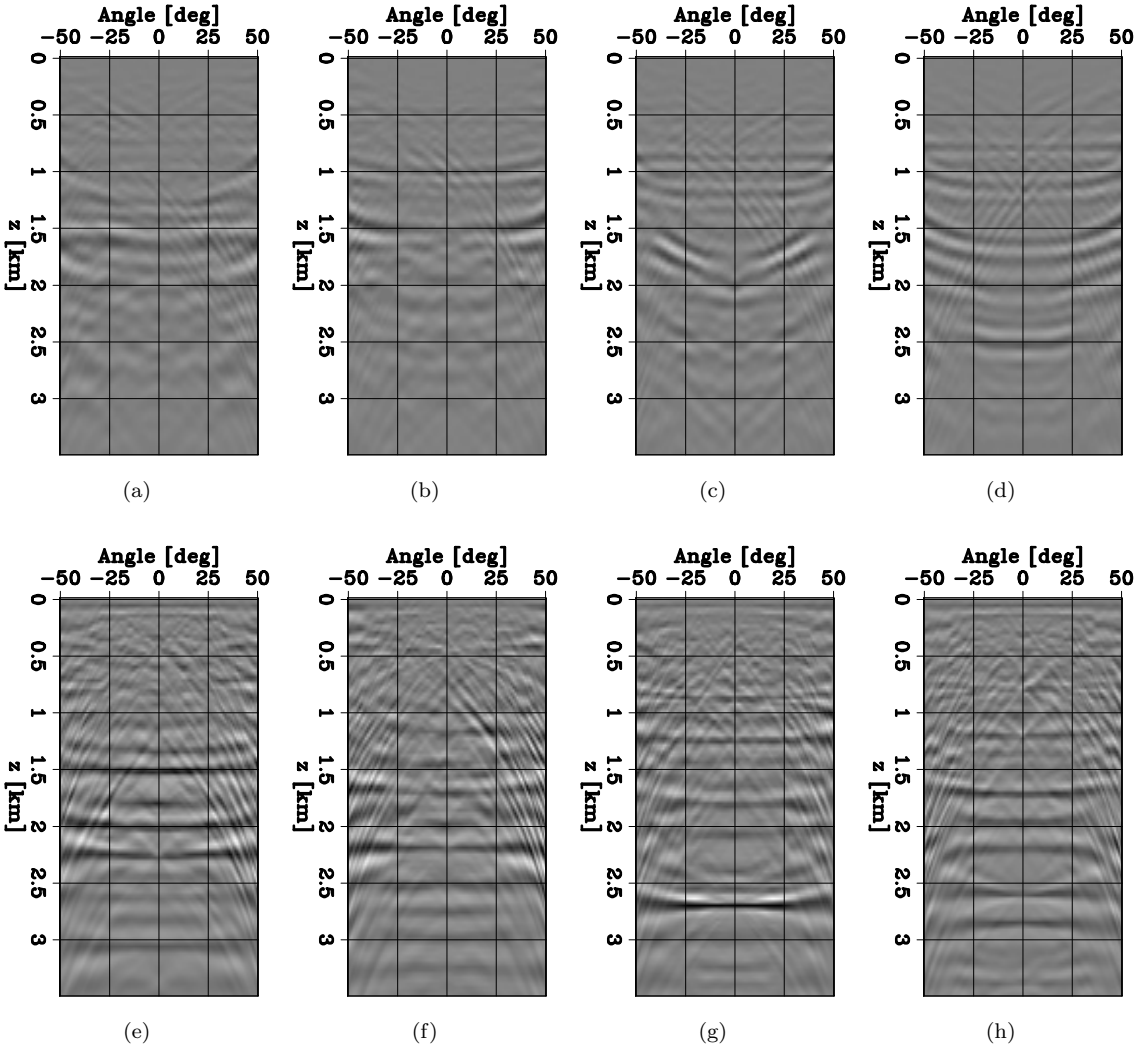


Figure 5.32: Angle domain common image gathers (ADCIGs) computed with the initial model (top row) and the final FWIME inverted model (bottom row). First column is for $x = 9$ km, second column is for $x = 10$ km, third column is for $x = 11$ km, and fourth column is for $x = 12$ km. All panels are displayed on the same grayscale. [**CR**]

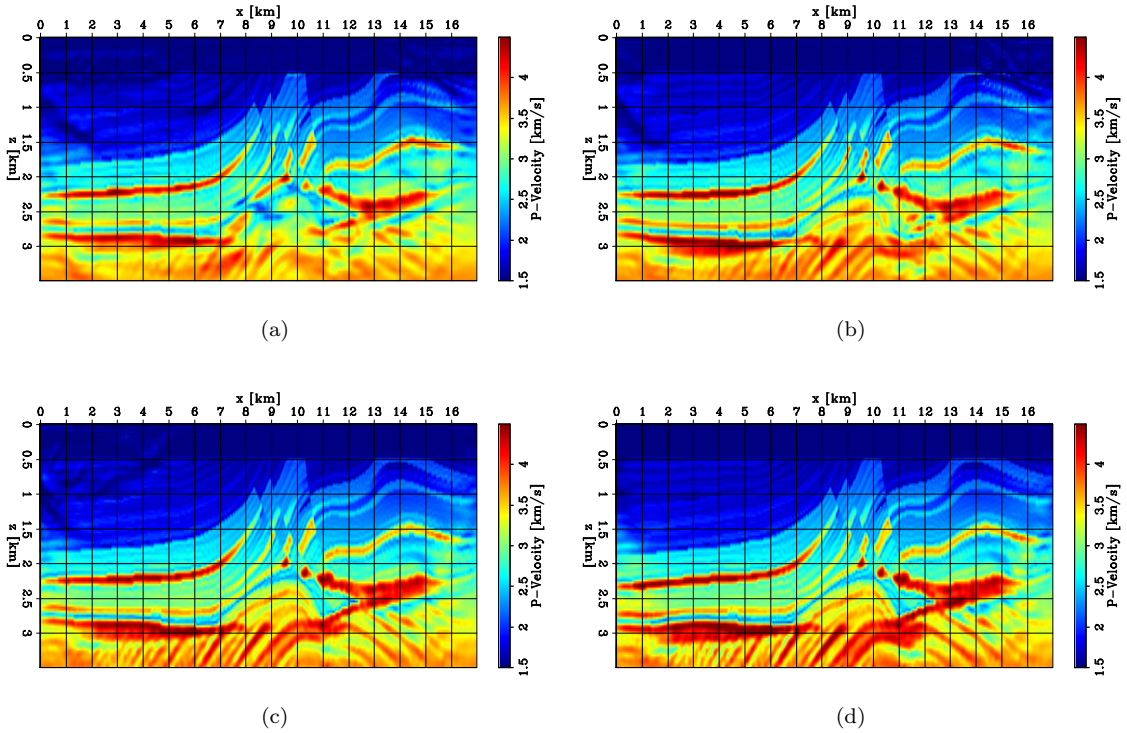


Figure 5.33: FWIME inverted models computed with the same hyper-parameter selection as for the result shown in Figure 5.26e but with less linear conjugate-gradient iterations for the variable projection step. (a) 7 iterations, (b) 10 iterations, (c) 15 iterations, and (d) 20 iterations. [**CR**]

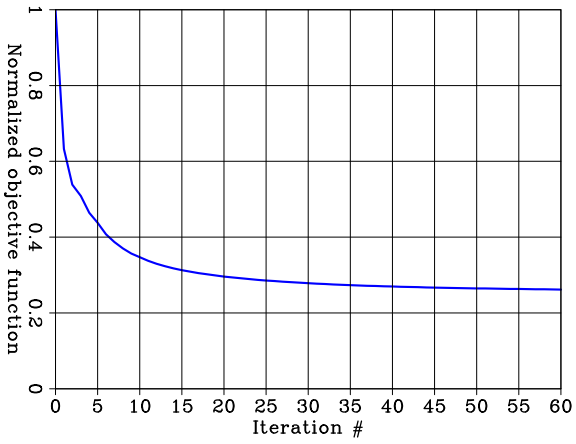


Figure 5.34: Convergence curve corresponding to the minimization of equation 3.6 conducted at the initial step of the inversion with $\epsilon = 1.5 \times 10^{-5}$. [**CR**]

Effect of free-surface multiples

I investigate how the presence of free-surface multiples recorded with marine acquisition geometries affects the performance of FWIME. I generate a new dataset with a free-surface boundary condition at the water surface, and absorbing-boundary conditions in all other directions (Robertsson, 1996; Biondi, 2021). Both sources and receivers are placed at a depth of 30 m below the surface. All other acquisition parameters are identical to the ones used for our previous analysis. Figure 5.35 shows two shot gathers generated by sources positioned at $x = 1.2$ km (first column) and $x = 14.4$ km (second column) using an absorbing-boundary condition (first row) and a free-surface boundary condition at the water/air interface (second row). The free-surface multiples are clearly observable (white arrows) and their amplitude is strong due to the relatively small thickness of the water layer. In the central part of the section (corresponding to the original Marmousi model), the multiples overlap with the various reflected events generated by the complex geological structures, as shown in Figure 5.35d (Guo et al., 2020).

We conduct data-space multi-scale FWI on this new dataset, which converges to an unsatisfactory solution (Figure 5.36a). We then apply FWIME in a similar fashion as for our previous analysis.

Figure 5.37 shows TLCIGs extracted at four horizontal positions (ranging from $x = 11$ km to $x = 14.5$ km) from $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_{init})$ computed at the intial step (by minimizing equation 3.6) using the original dataset with absorbing boundaries (top row), and then using the new dataset with the free-surface boundary condition (bottom row). In both rows of Figures 5.37, we observe clusters of energy located at negative time lags indicating that the initial velocity values are too low for this region of the model (as we saw in the previous analysis). However, for the free-surface case (bottom row), additional clusters of coherent energy are present at positive time lags, which correspond to the mapping of the free-surface multiples into the extended space of $\tilde{\mathbf{p}}_\epsilon^{opt}$ (indicated by the green arrows in Figures 5.37e-h). Since free-surface multiples generally propagate with a lower velocity than primary reflections recorded with the same traveltime, their positions on the extended axis wrongfully indicate that the initial velocity $\mathbf{m}_{init}$ is too high, thereby misleading the FWIME search direction for that region of the model.

Figures 5.38a and b show the initial FWIME search directions computed on the first spline grid (displayed on the finite-difference grid) with the original dataset and with the new dataset, respectively. By comparing these panels to the true search direction (Figure 5.38c), we can see that the free-surface multiples seem to guide the inversion in the wrong direction, especially in the right side of the model. The final FWIME result obtained after a total of 120 iterations of L-BFGS (using the same sequence of spline grids as for the original test) is shown in Figure 5.36b. The inversion manages to accurately reconstruct the sharp horizontal reflectors in the left part of the model, but the fails to recover the shallow complex region of the model located between $x = 9$ km and $x = 15$ km. For this acquisition geometry, the presence of free-surface multiples seem to harm the quality of the inverted solution (compared to the original result shown in Figure 5.36c). To mitigate this

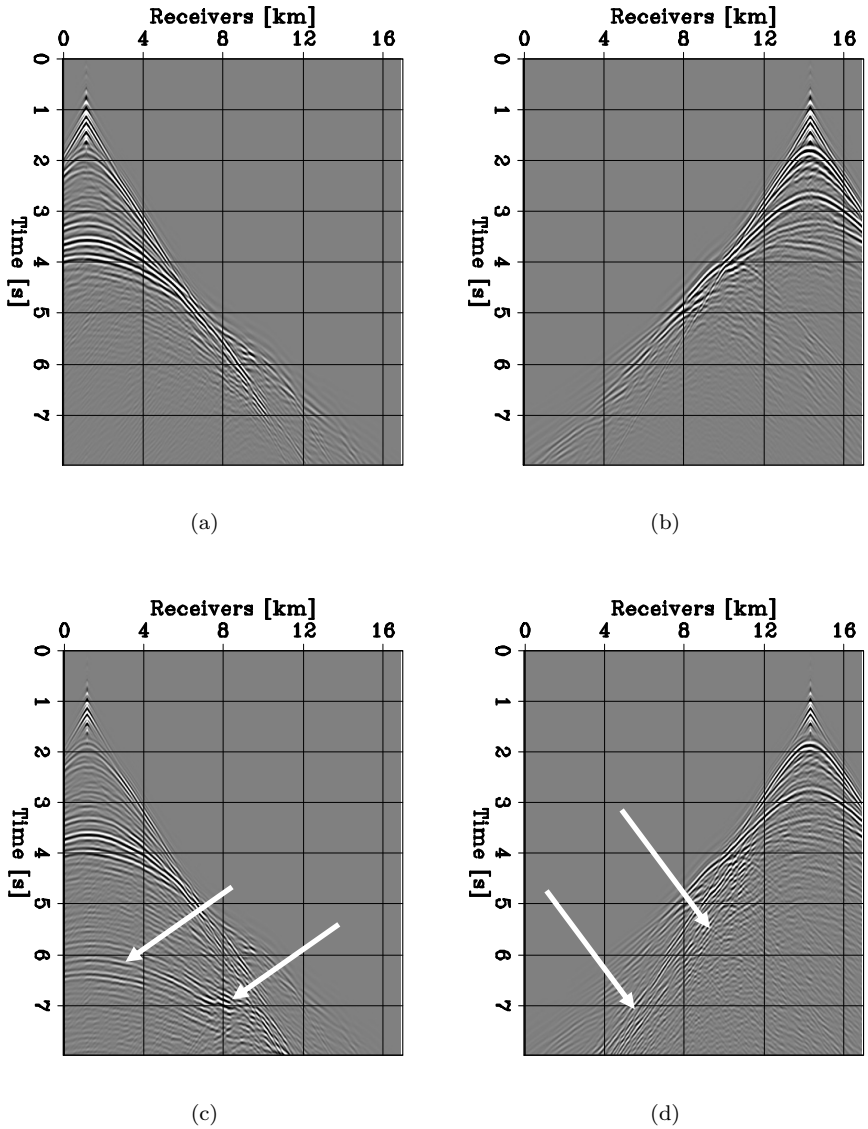


Figure 5.35: Representative shot gathers for sources placed at $x = 1.2$ km (first column) and $x = 14.4$ km (second column). The top row shows the data modeled with absorbing-boundary conditions in all directions. The bottom row shows the data modeled with a free-surface boundary condition at the water surface. All panels are displayed with the same grayscale. [**ER**]

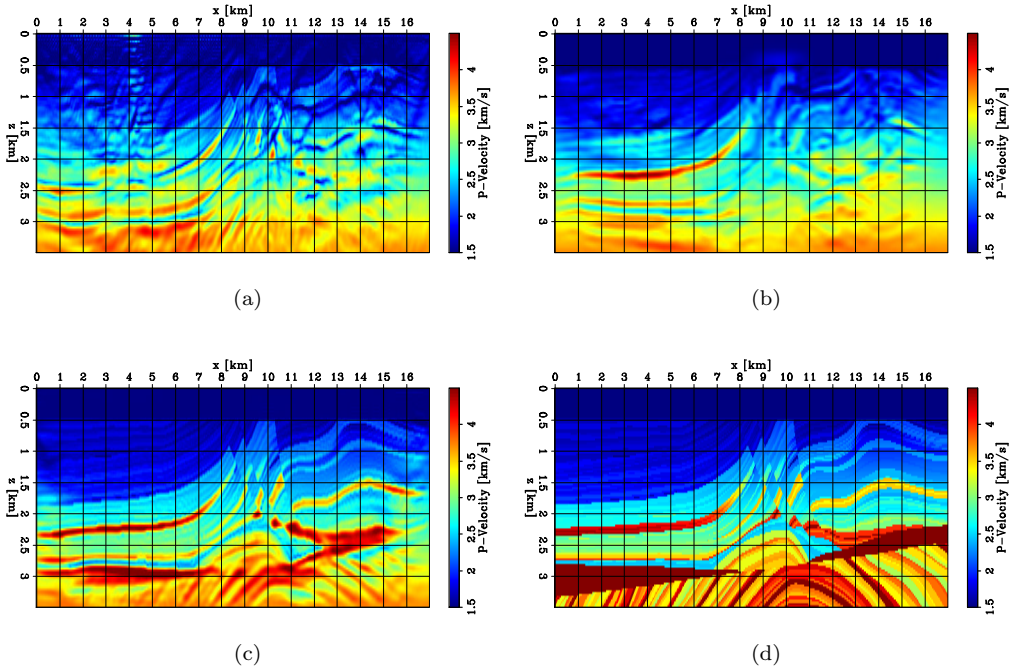


Figure 5.36: 2D panels of velocity models. (a) FWI model inverted using the dataset modeled with a free-surface. (b) Final FWIME model inverted using the dataset modeled with a free-surface. (c) FWIME model inverted with the original dataset. (d) True model. [**CR**]

effect, we propose to use surface-related multiple elimination techniques when applying FWIME to offshore data acquired in a similar scenario (Verschuur et al., 1992; Baumstein and Hadidi, 2006; Dragoset et al., 2010; Siahkoohi et al., 2019).

5.3 BP 2004 model: North Sea region

I invert data generated from the North Sea region of the BP 2004 benchmark model. The true model $\mathbf{m}_{true}$ is 29 km-wide and 5.5-km deep, and is shown in Figure 5.39d. The initial velocity model $\mathbf{m}_0$ (Figure 5.39a) is horizontally invariant and linearly increasing with depth, and contains substantial errors as shown by the 1D velocity profiles in Figures 5.40 and 5.41. At the surface, I place 182 sources every 160 m, and 728 fixed receivers every 40 m. The data are generated with a source containing energy strictly limited to the 3-9 Hz frequency range (Figure 5.42a), and are recorded for 13 s. As shown in Figures 5.43a and d, the observed data are dominated by diving waves and the initial prediction indicates the presence of multiple cycle-skipped events (Figure 5.43c and 5.43f).

This example tests the ability of FWIME at inverting a dataset containing mostly refracted energy (rather than using mainly reflected energy as I illustrated with the numerical example in

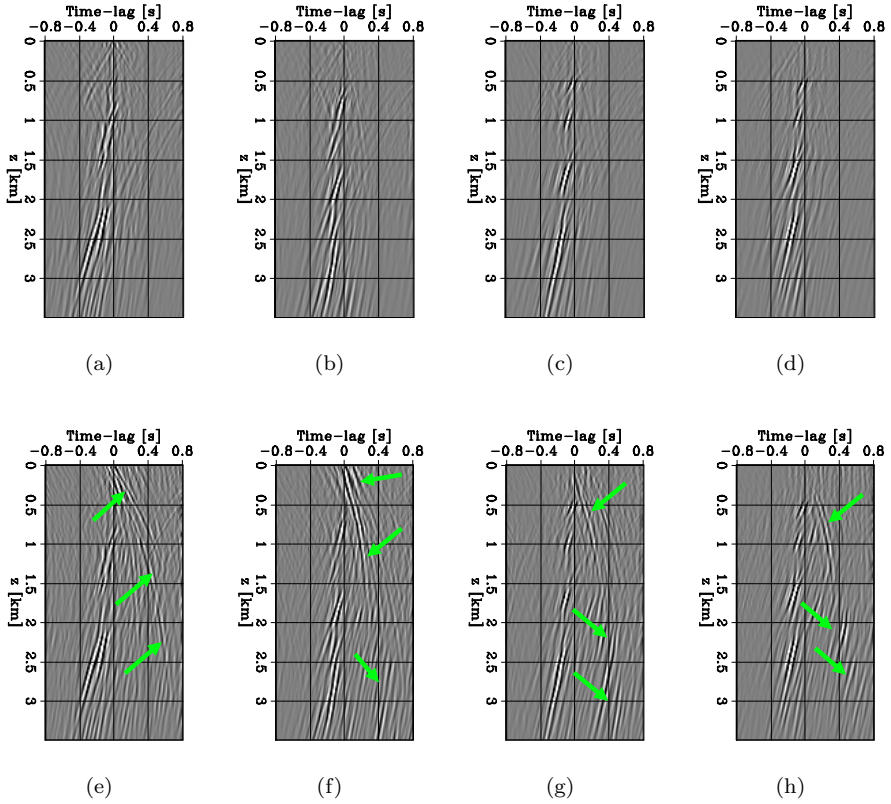


Figure 5.37: TLCIGs extracted at $x = 11$ km, $x = 12.5$ km, $x = 13.5$ km, and $x = 14.5$ km from $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}_{init})$ computed with the dataset using absorbing boundaries (top row), and with a free-surface boundary condition (bottom row). All panels are displayed with the same grayscale. [**CR**]

section 5.2) to simultaneously recover model features of various scales. The first difficulty is to obtain the correct velocity trend (low-resolution component) in the deeper regions of the model. The second challenge consists in accurately delineating the high-resolution features, which include the low-velocity zones in the shallow region and two high-velocity anomalies underneath the mud volcano. As expected, conventional multi-scale FWI (using a sequence of four frequency bands shown in Figure 5.42b) fails to retrieve the optimal solution (Figure 5.39b).

For the FWIME process, I follow a similar hyperparameter-tuning analysis as for the numerical example described in section 5.2, and I set $\epsilon = 1.75 \times 10^{-5}$. In order to account for the large kinematic errors in the initial data prediction, I use a time-lag extension spanning the $[-1.2\text{ s},\ 1.2\text{ s}]$ interval with 101 points sampled at $\Delta\tau = 24$ ms. I use a sequence of 4 spline grids, and each spline grid refinement is triggered when the stepper is unable to find a step length that decreases the objective function value. The initial spline is chosen to be very coarse with $\Delta z = 1.0$ km and $\Delta x = 2.4$ km. The spacing in the second and third grids are obtained by halving the spacing from

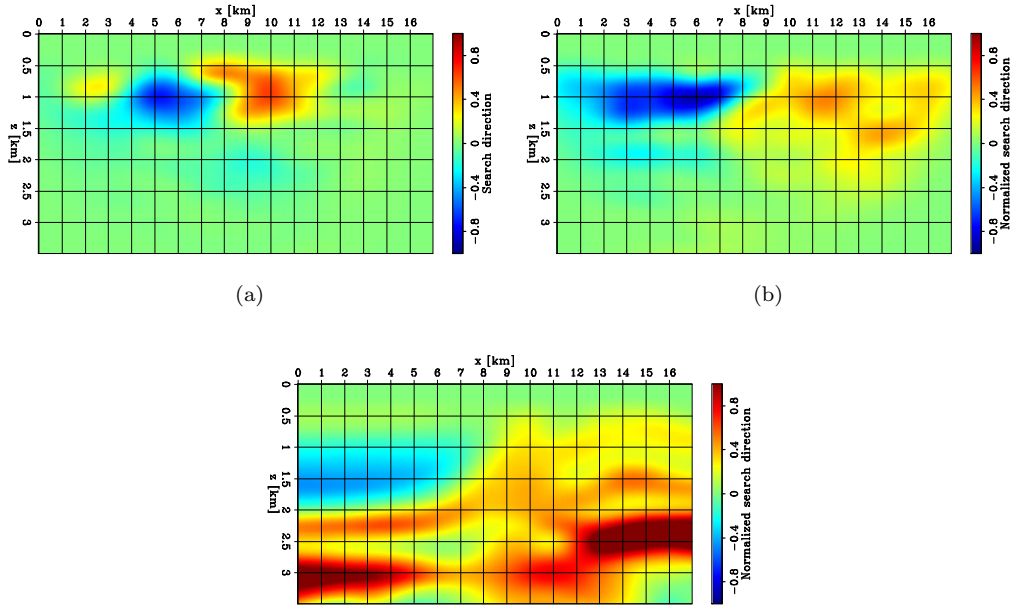


Figure 5.38: Normalized initial search directions computed on the first spline grid (displayed on the finite-difference grid). (a) FWIME search direction computed with the new dataset containing free-surface multiples. (b) FWIME search direction computed with the original dataset. (c) True search direction. Panels (a) and (b) are normalized with the same value. [**CR**]

the previous ones. The final spline grid coincides with the finite-difference grid ($\Delta z = \Delta x = 40$ m). I assume the bathymetry is known and for the first, second and third grid, I use a finer spatial sampling of 100 m in the vicinity of the water bottom to account for this sharp interface. Table 5.2 summarizes the spline grid parameters used in this numerical example.

Grid number	Δz [km]	Δx [km]
1	0.9	2.4
2	0.6	1.2
3	0.2	0.28
4	0.04	0.04

Table 5.2: Parameters of the sequence of spline grids used for the model-space multi-scale FWIME scheme. Spline 4 coincides with the finite-difference grid.

The benefit of the spline parametrization can be observed by examining the initial search direction (Figure 5.44). Before its mapping onto the initial spline grid, the FWIME first search direction $\mathbf{s}_{init}$ (Figure 5.44a) contains spurious high-resolution artifacts, which are not present in the true search

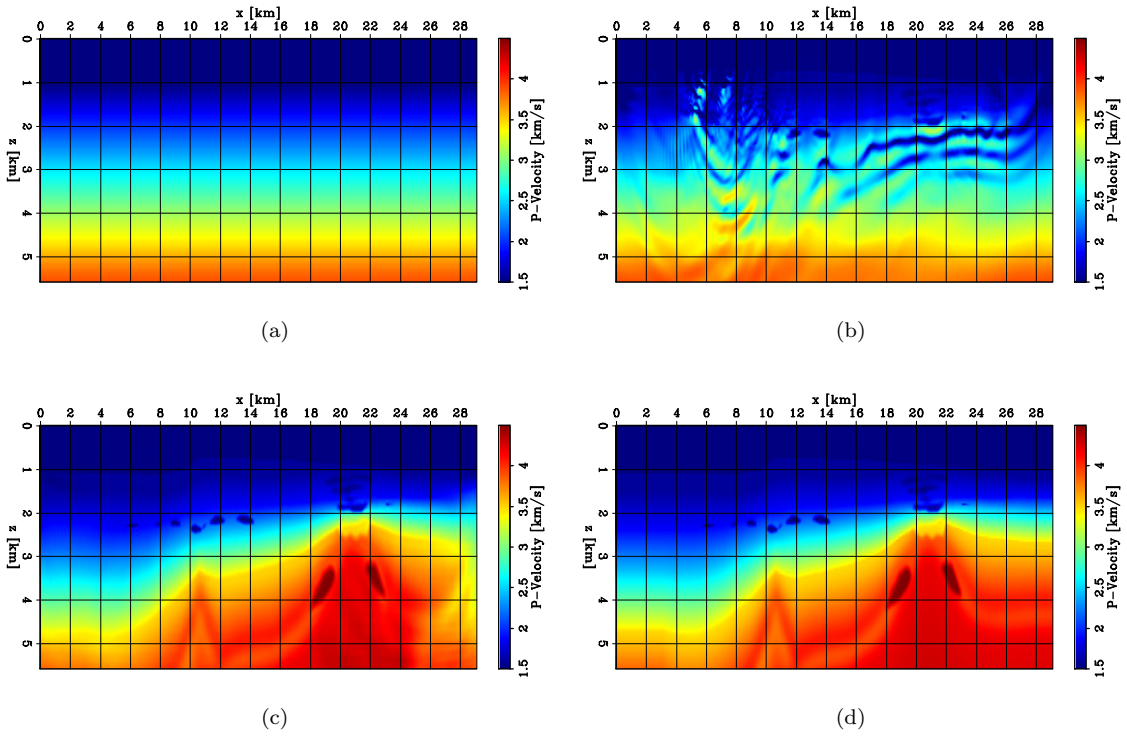


Figure 5.39: 2D panels of velocity models. (a) Initial model. (b) Inverted model after conventional data-space multi-scale FWI using four frequency bands. (c) Final FWIME inverted model. (d) True model. [**CR**]

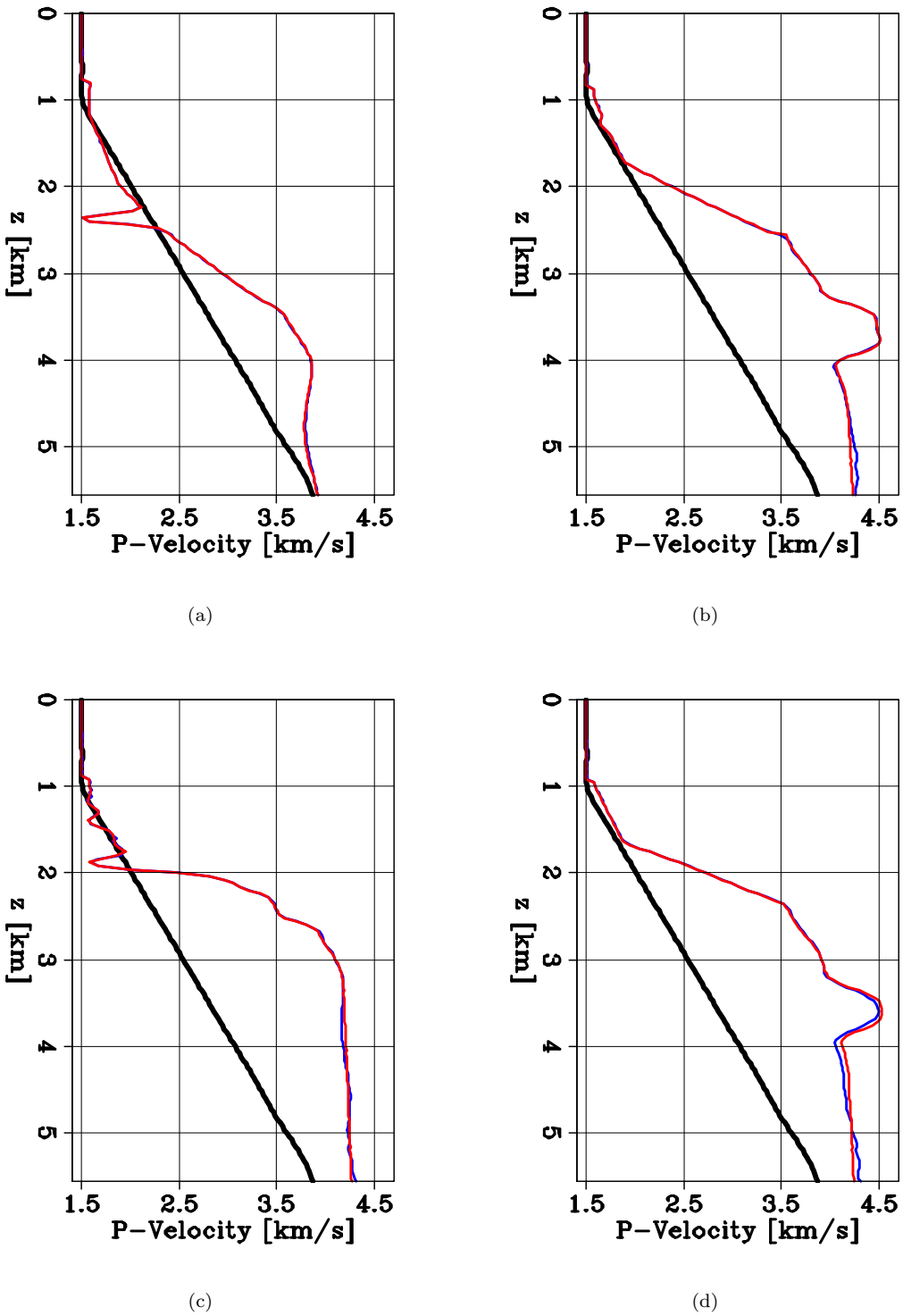


Figure 5.40: Depth velocity profiles extracted at (a) $x = 6$ km, (b) $x = 9$ km, and (c) $x = 11$ km, and (d) $x = 13$ km. The black curve represents the initial model, the red curve is the true model, and the blue curve is the FWIME inverted model. [**CR**]

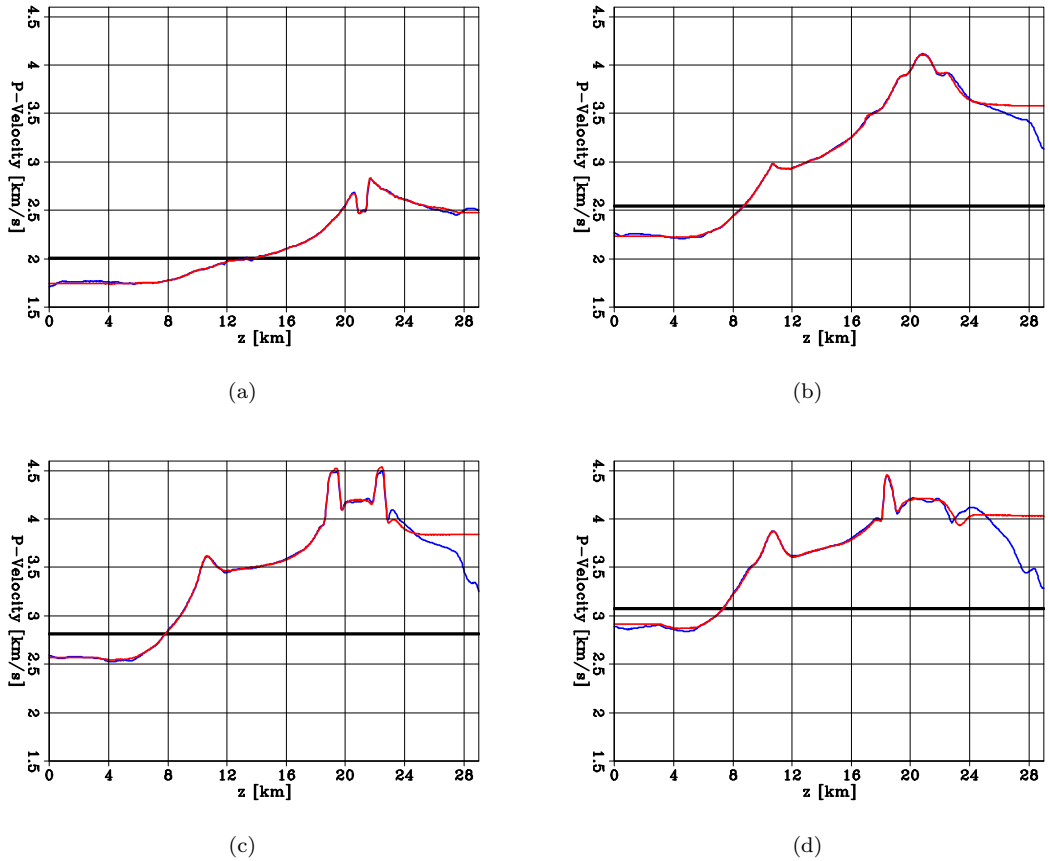
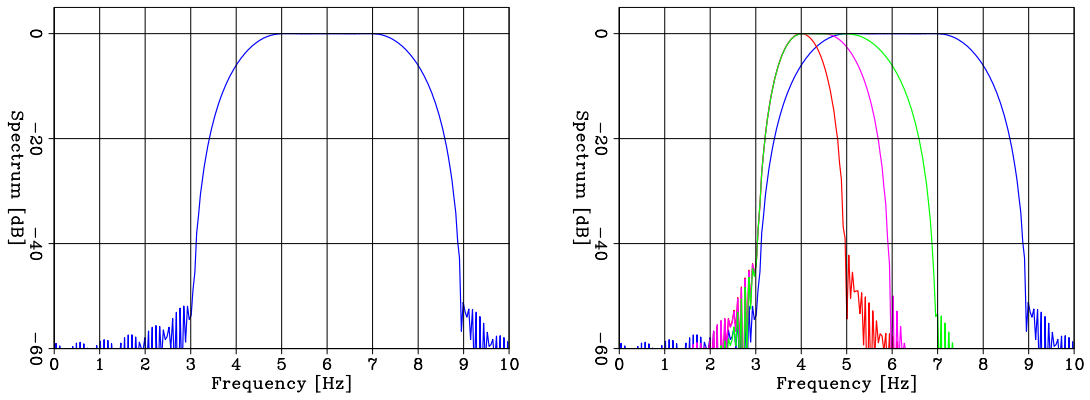
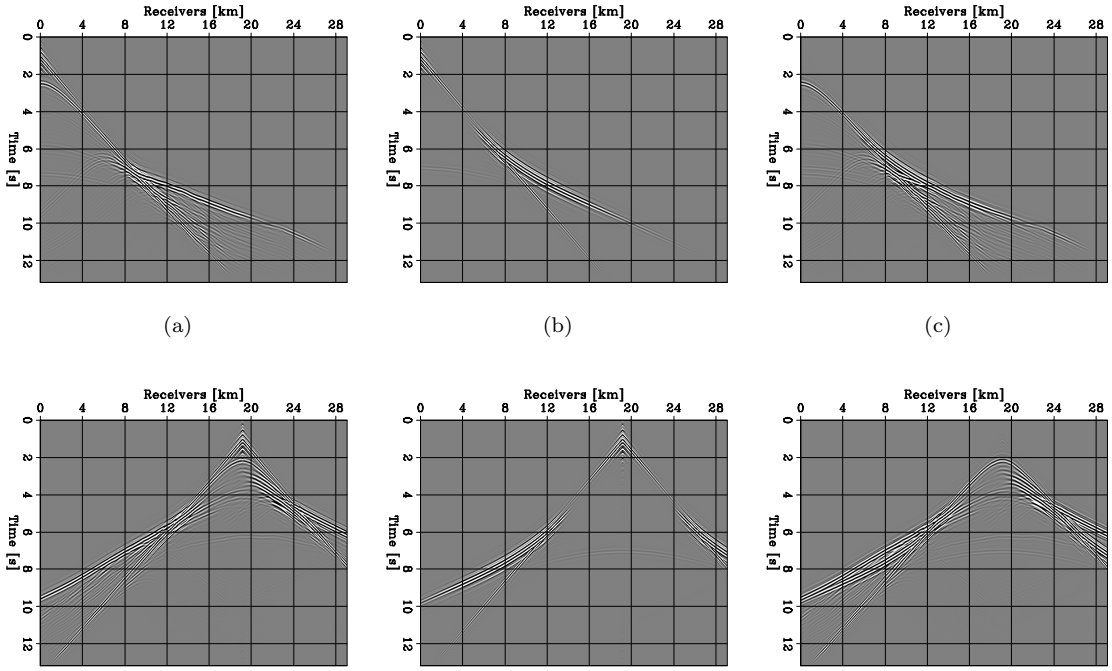


Figure 5.41: Horizontal velocity profiles extracted at (a) $z = 2$ km, (b) $z = 3$ km, (c) $x = 3.5$ km, and (d) $x = 4$ km. The black curve represents the initial model, the red curve is the true model, and the blue curve is the FWIME inverted model. [**CR**]



(a) (b)

Figure 5.42: Amplitude spectra of the seismic sources employed in this numerical example. (a) Source used for the FWIME workflow. (b) Sequence of sources used for the data-space multi-scale FWI workflow. [**ER**]



(a) (b) (c)

(d) (e) (f)

Figure 5.43: Representative shot gathers for sources placed at $x = 0$ km (first row) and $x = 19$ km (second row). Observed data, $\mathbf{d}^{obs}$ (first column), predicted data with the initial model, $\mathbf{f}(\mathbf{m}_0)$ (second column), and initial data-difference, $\Delta\mathbf{d}(\mathbf{m}_0) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_0)$ (third column). All panels are displayed with the same grayscale. [**ER**]

direction $\mathbf{s}_{true} = \mathbf{m}_{true} - \mathbf{m}_0$ (Figure 5.44b). Figures 5.44c and d show the FWIME and ideal search directions after their mapping onto the first spline grid (i.e., after applying $\mathbf{SS}^*$ to $\mathbf{s}_{init}$ and to $\mathbf{s}_{true}$), respectively. The high-wavenumber artifacts have been removed and the FWIME search direction is improved as it shows stronger similarity with the ideal search direction.

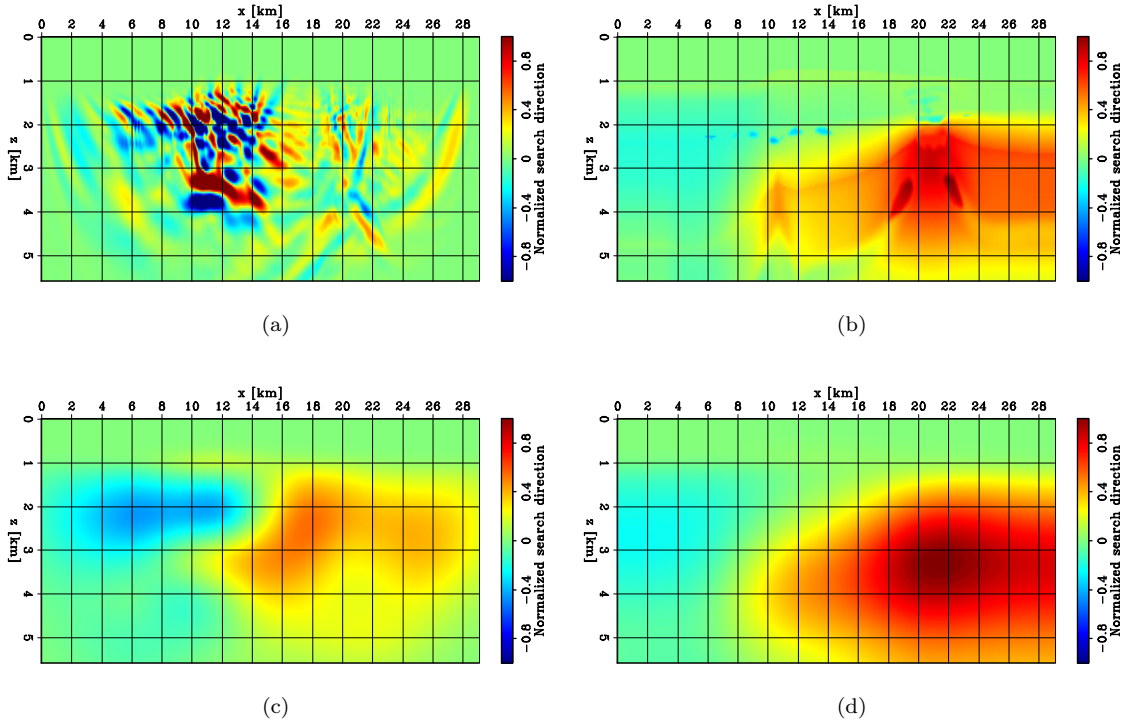


Figure 5.44: Normalized initial search directions. (a) FWIME initial search direction $\mathbf{s}_{init}$ before applying any spline parametrization, computed with $\epsilon = 1.75 \times 10^{-5}$. (b) True search direction, $\mathbf{s}_{true} = \mathbf{m}_{true} - \mathbf{m}_0$. (c) FWIME initial search direction after its mapping on the initial spline grid (displayed on the finite-difference grid), $\mathbf{s}_{init}^{spline} = \mathbf{SS}^*\mathbf{s}_{init}$. (d) True search direction after its mapping on the initial spline grid (displayed on the finite-difference grid), $\mathbf{s}_{true}^{spline} = \mathbf{SS}^*\mathbf{s}_{true}$. [**CR**]

The inverted models after the first and second grids (Figures 5.45b and c) show an accurate recovery of the very long-wavelength anomalies (i.e., the velocity trend) missing in the initial model. In a test not shown here, I verify that the inverted model after the second grid is accurate enough for conventional FWI to retrieve a solution very similar to the final FWIME inverted model. The quality of the final FWIME inverted model (obtained after a total of 217 iterations of L-BFGS) is excellent (Figure 5.45e): it manages to recover both the long- and short-wavelength anomalies. This result is also confirmed by observing the vertical and horizontal velocity profiles extracted at various positions in Figures 5.40 and 5.41.

For quality-control purpose, I examine the zero-lag cross-section of $\mathbf{p}_\epsilon^{opt}$ and one TLCIG extracted

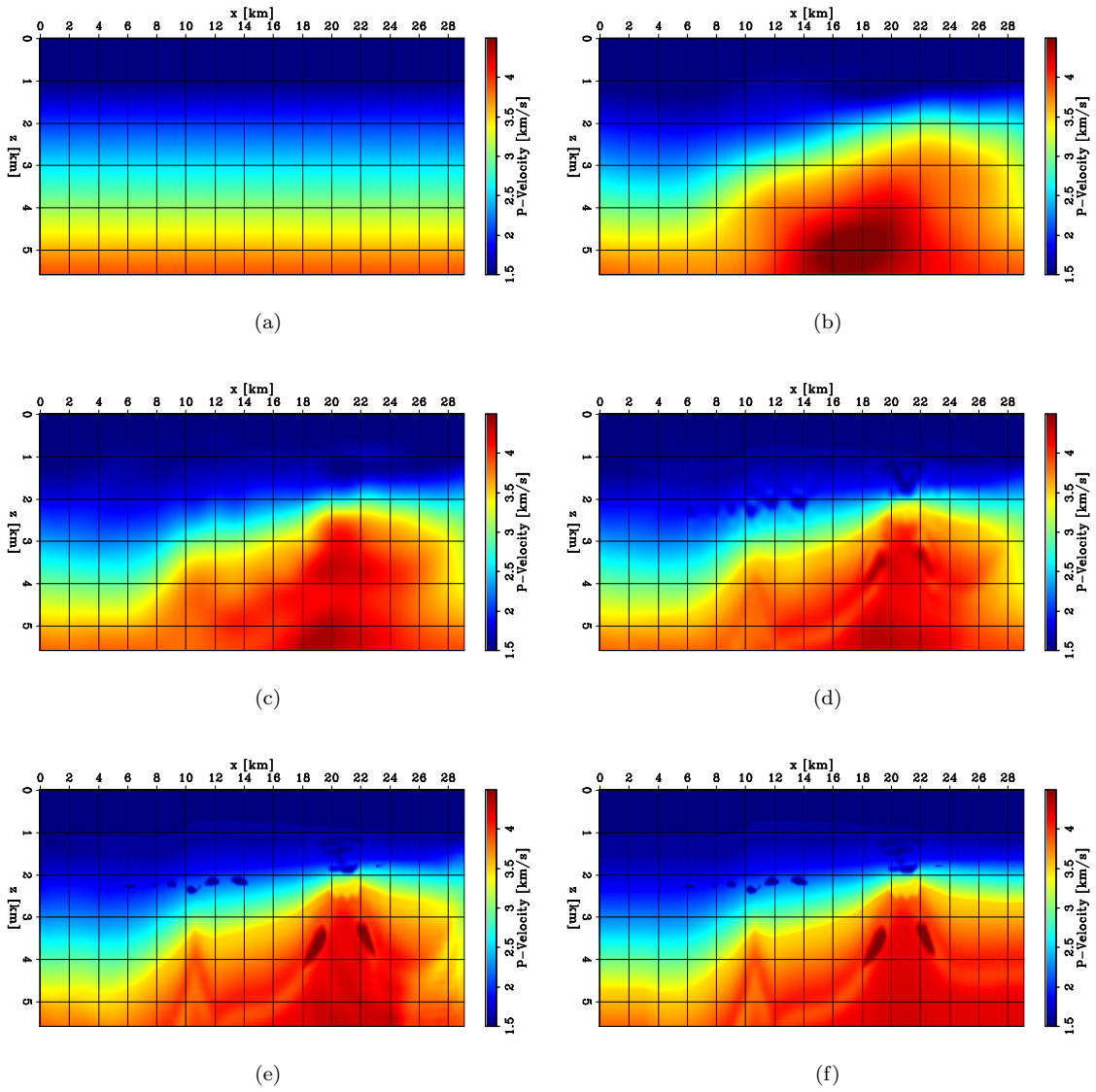


Figure 5.45: Inverted models at different stages of the model-space multi-scale FWIME workflow with $\epsilon = 1.75 \times 10^{-5}$. (a) Initial model. (b) Inverted model after 77 iterations on the first spline grid. (c) Inverted model after 60 iterations on the second spline grid. (d) Inverted model after 33 iterations on the third spline grid. (e) Inverted model after 124 iterations on the fourth (finite-difference) grid. (f) True model. The total number of L-BFGS iterations used to obtained panel 5.45e is 217. [**CR**]

at $x = 22$ km at five stages of the inversion. In a similar fashion as for the Marmousi2 example in section 5.2, a high amount of energy is initially mapped away from the physical plane, confirming that the starting model velocity values are too low (Figure 5.47a). Simultaneously, the zero-lag cross-section lacks coherency (Figure 5.46a). As the inversion progresses, the FWIME inverted models become more accurate, and the energy focuses in the vicinity of the physical plane of $\mathbf{p}_{\epsilon}^{opt}$ Figures 5.47b-d. As expected, the energy within $\mathbf{p}_{\epsilon}^{opt}$ vanishes completely (Figures 5.46e and 5.47e) once the algorithm has converged.

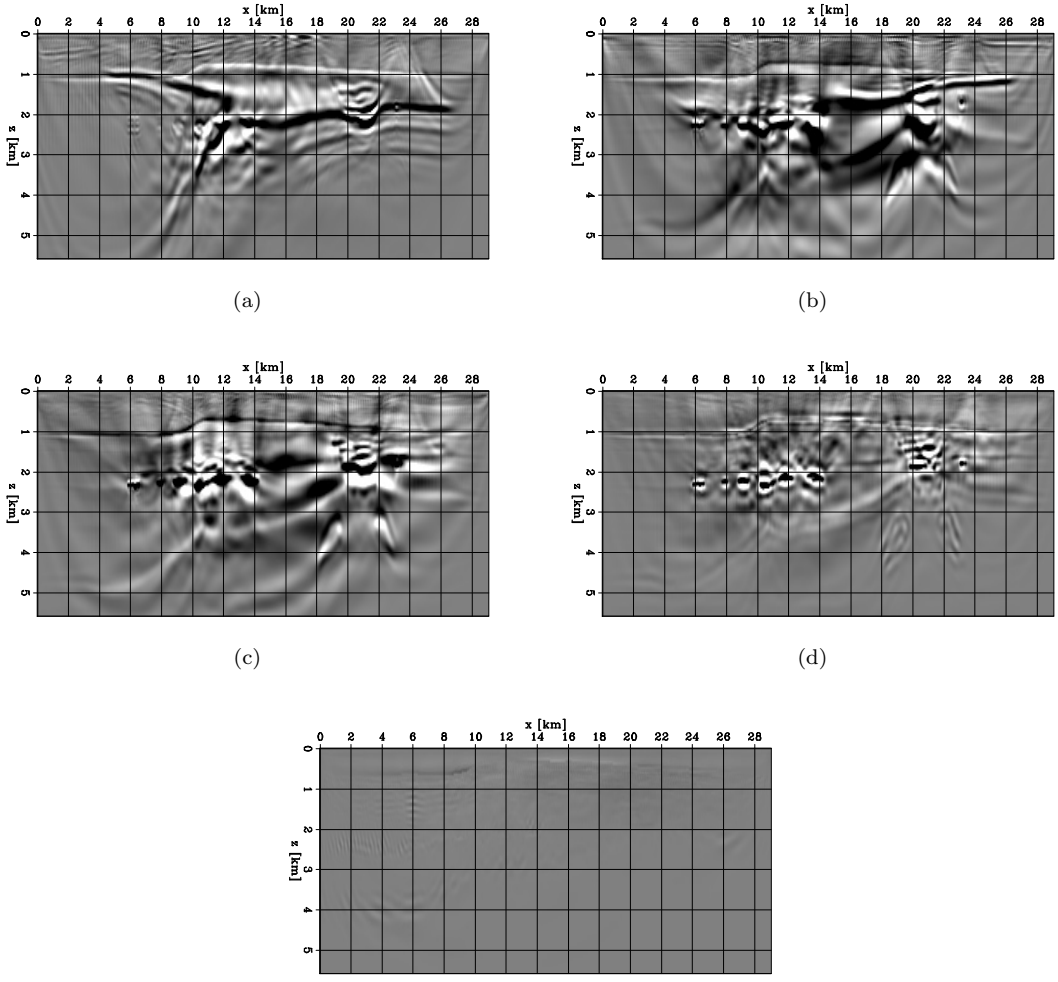
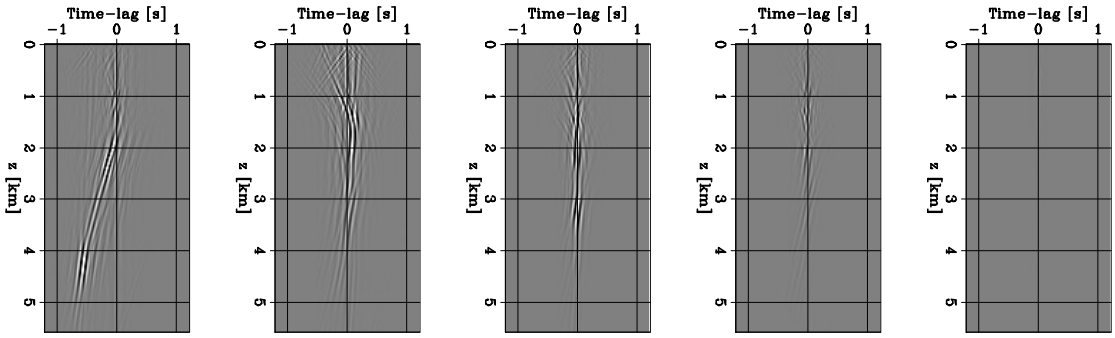


Figure 5.46: Zero time-lag sections of $\mathbf{p}_{\epsilon}^{opt}$ computed at four stages of the FWIME workflow. (a) Initial step. (b) After inversion on spline 1. (c) After inversion on spline 2. (d) After inversion on spline 3. (e) Final step. All panels are displayed with the same grayscale. [**CR**]



(a) (b) (c) (d) (e)

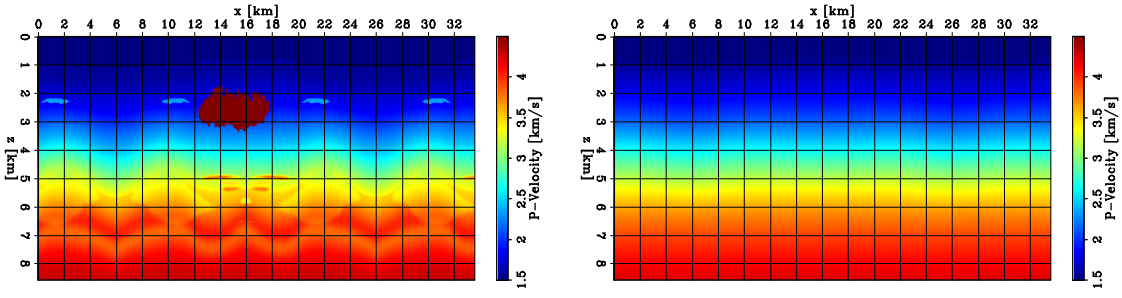
Figure 5.47: Time-lag common image gathers (TLCIG) extracted at $x = 22$ km from $\mathbf{p}_{\epsilon}^{opt}$ computed at four stages of the FWIME workflow. (a) Initial step. (b) After inversion on spline 1. (c) After inversion on spline 2. (d) After inversion on spline 3. (e) Final step. All panels are displayed with the same grayscale. [**CR**]

5.4 Salt model

Following the recent deployment of ultra-long offset full-azimuth sparse-node acquisition surveys (Bate and Mike Perz, 2021), I design a test to show that FWIME can leverage refracted energy for velocity model building in complex subsalt regions. I show its potential at providing significant imaging improvements without the need for low-frequency energy nor accurate initial model. The true model (Figure 5.48a) is 33 km wide and 8.5 km deep. It is composed of a sediment background (modified from Billette and Brandsberg-Dahl (2005)) in which a 4 km-wide and 1 km-thick salt body is embedded. The ultimate goal is to accurately image a set of four thin high-velocity layers and one low-velocity anomaly located underneath the salt. The initial model is obtained by strongly smoothing the sediment background and does not contain any information about the presence of salt (Figure 5.48b and black curves in Figures 5.49).

To simulate an ultra-long offset acquisition survey, I place 840 fixed receivers every 40 m at the surface and I generate noise-free pressure data with 139 sources. The distance between two consecutive sources is set to 240 m, and the maximum recorded offset for this test is 33 km. For the FWIME workflow, the source wavelet contains energy limited to the 3-9 Hz range, and the data are recorded for 15 s. This example proposes an ideal acquisition geometry (the salt body is well illuminated by diving waves propagating through it) and is conducted to show the usefulness of combining FWIME with novel long-offset node acquisition surveys to build accurate velocity models for FWI.

I first conduct a multi-scale FWI using the initial model shown in Figure 5.48b assuming the presence of unrealistic low-frequency energy in the recorded data, as shown by the five wavelet



(a) (b)

Figure 5.48: 2D panels of velocity models. (a) True model. (b) Initial model. [**NR**]

spectra in Figure 5.50a. The final inverted model is extremely accurate (Figure 5.51b) and provides an estimate of the best recoverable solution using waveform-inversion schemes with this particular acquisition geometry. I apply a second FWI workflow (with the same initial model) assuming that no energy below 1.8 Hz is present (which is still quite optimistic for field data). The sequence of wavelet spectra employed for this scheme is shown in Figure 5.50b. In this case, FWI fails to converge to a physical solution, which highlights the difficulty of retrieving the salt body when starting with a pure sediment model (Figure 5.51c).

For the FWIME scheme, I invert data generated with a source wavelet whose spectrum is shown in Figure 5.50c. The lowest useful frequency is set to 3 Hz, which is higher compared to the spectrum of the lowest band for conventional FWI (blue curve in Figure 5.50b). I use a time-lag extension spanning the $[-1.2\text{ s},\ 1.2\text{ s}]$ interval with 101 points sampled at $\Delta t = 24$ ms and I set $\epsilon = 2.5 \times 10^{-4}$. The parameters of the five spline grids employed are shown in Table 5.3. In addition, I assume the bathymetry is known and for the first, second, third, and fourth grids, I use a finer spatial sampling of 100 m in the vicinity of the water bottom to allow the recovered velocity models to correctly predict the strong reflection generated from this interface (even at early iterations). The sequence of FWIME inverted models on each grid is shown in Figure 5.52. The inversion on the first grid is crucial as it starts building a low-resolution salt body (i.e., a geobody with smooth contours) (Figure 5.52b). Then, the contours of the salt are gradually sharpened as the spline grid is refined (Figure 5.52c-f). Once the salt body is accurately captured, the thin subsalt reflectors are then recovered on the finite-difference grid, and the final inverted model is accurate (Figure 5.52g).

Figure 5.53a displays the normalized FWIME objective function (blue curve) along with the value of the conventional FWI objective function computed at each inverted model during the FWIME sequence (red curve). As expected, the red curve is not monotonically decreasing but seems to behave slightly differently than the analogous curve obtained in the example based on the Marmousi2 model from section 5.2 (Figure 5.27b). Figure 5.53b shows a closeup of Figure 5.53a

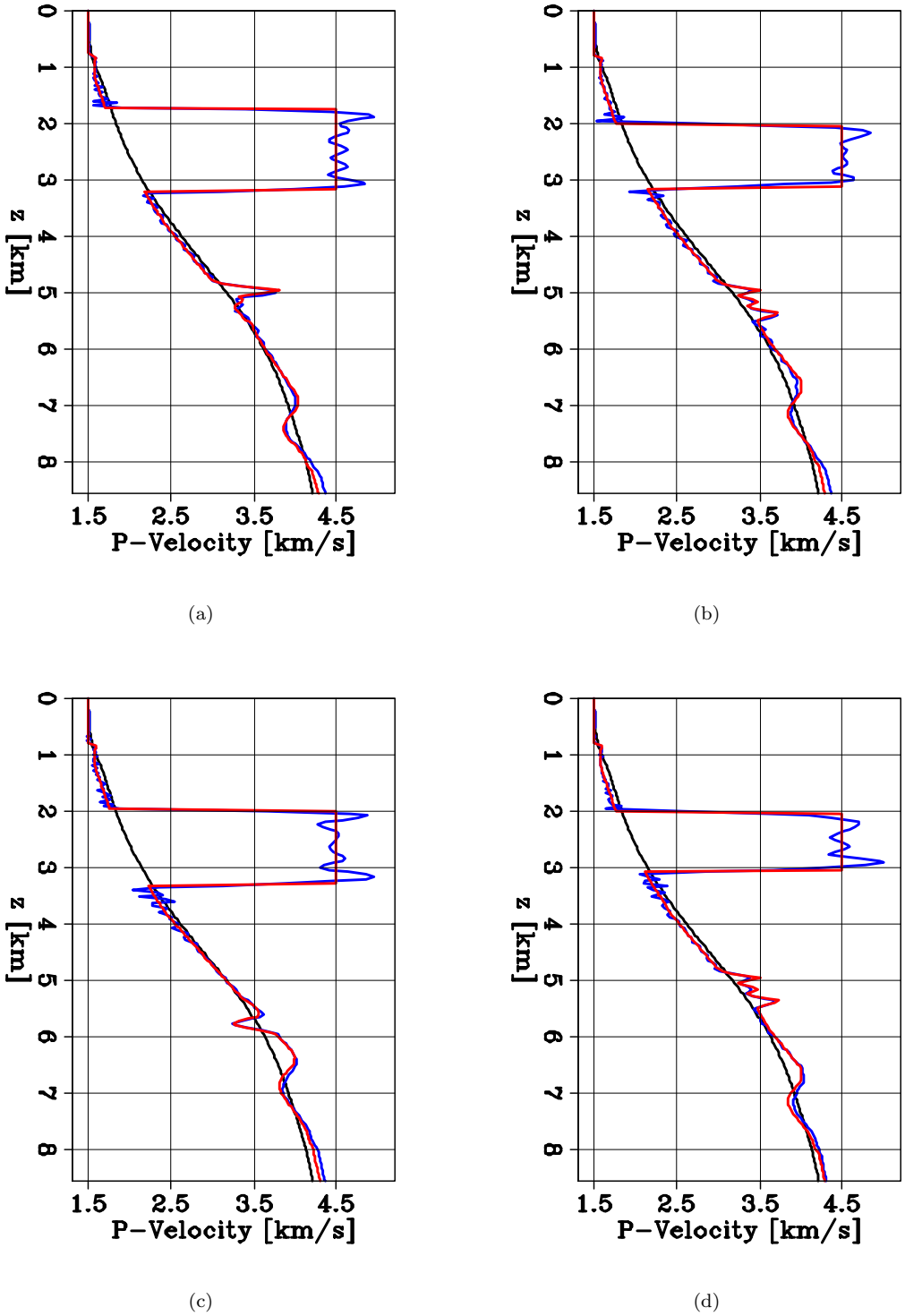


Figure 5.49: Depth velocity profiles extracted at (a) $x = 14$ km, (b) $x = 15$ km, and (c) $x = 16$ km, and (d) $x = 17$ km. The black curve represents the initial model, the red curve is the true model, and the blue curve is the final FWIME inverted model. [**CR**]

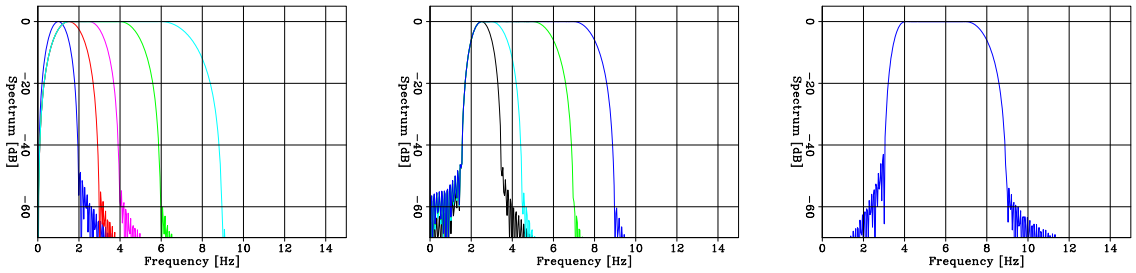


Figure 5.50: Amplitude spectra of the seismic sources employed in this numerical example. (a) Sequence of sources used for the data-space multi-scale FWI workflow using unrealistic low-frequency energy. (b) Sequence of sources used for the data-space multi-scale FWI workflow using energy restricted to the 1.8-9 Hz bandwidth. (c) Source used for the FWIME workflow. [**ER**]

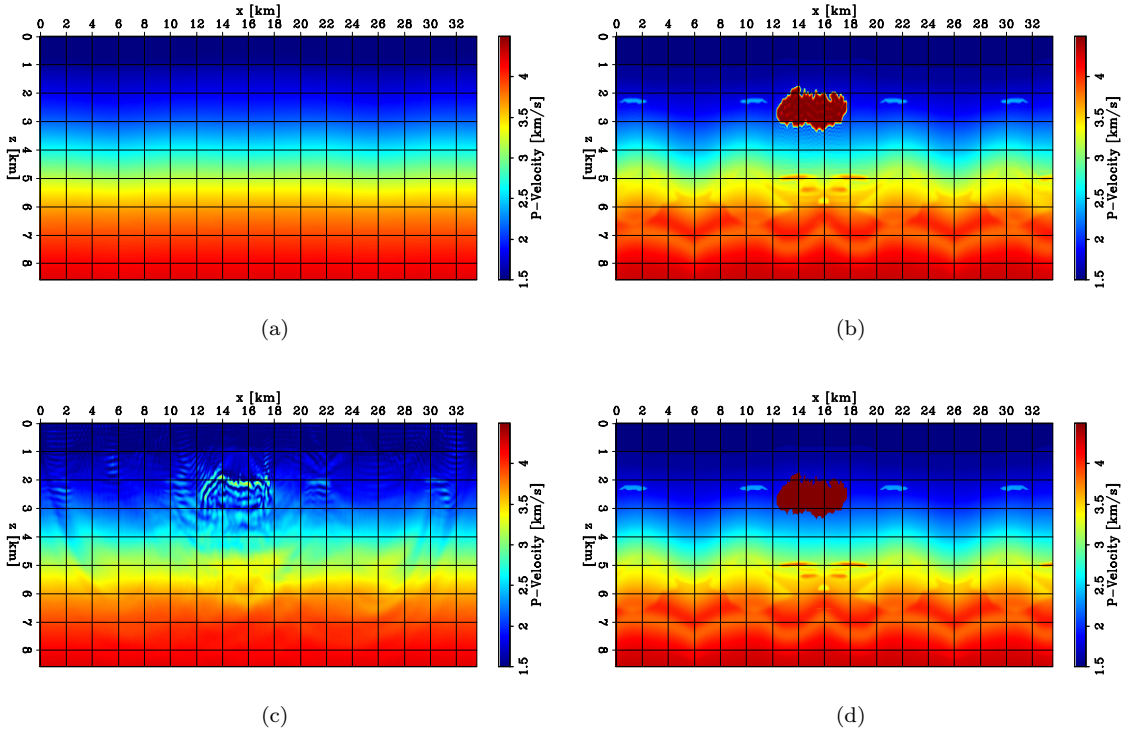


Figure 5.51: 2D panels of velocity models. (a) Initial model used for both the FWI and FWIME schemes. (b) Inverted model after conventional multi-scale FWI using unrealistic low-frequency energy contained within the 0-9 Hz range. (c) Inverted model after conventional multi-scale FWI using energy restricted to the 1.8-9 Hz range. (d) True model. [**CR**]

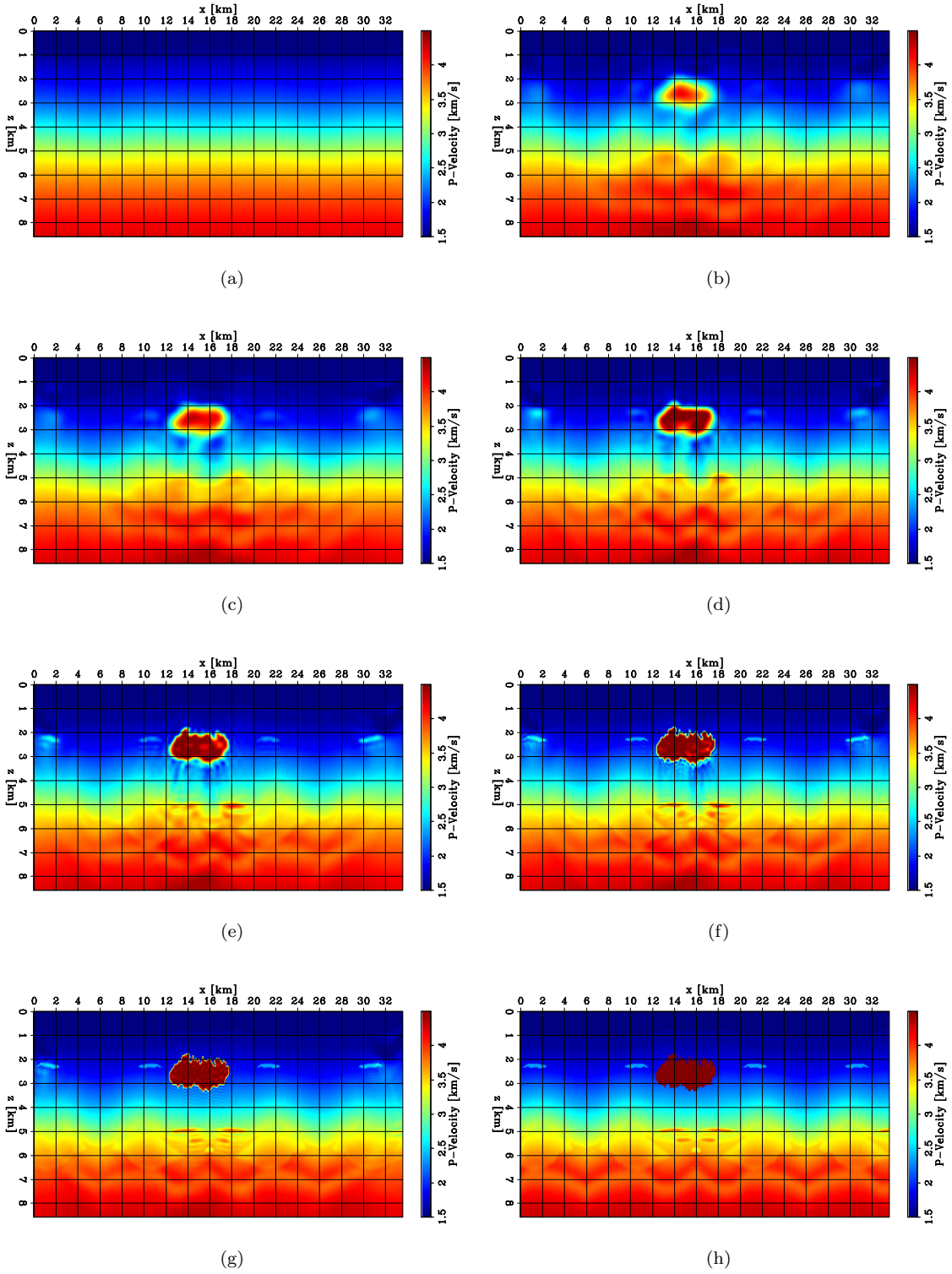
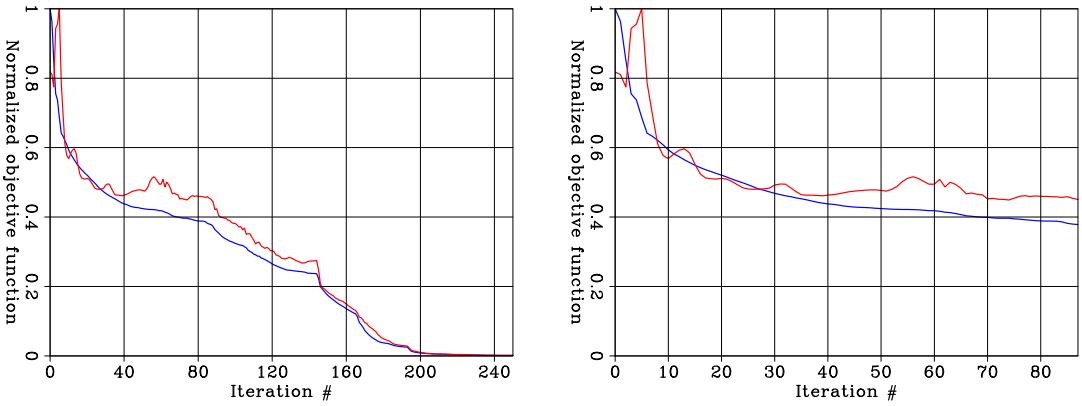


Figure 5.52: 2D panels of velocity models inverted on the various spline grids throughout the FWIME workflow: (a) Initial model (b) first grid, (c) second grid, (d) third grid, (e) fourth grid, (f) fifth grid, and (g) final FWIME inverted model, obtained after a total of 250 iterations of L-BFGS. (h) True model. [**CR**]

Grid number	Δz [km]	Δx [km]
1	0.5	1.2
2	0.3	0.8
3	0.12	0.4
4	0.06	0.12
5	0.04	0.04

Table 5.3: Parameters of the sequence of spline grids used for the model-space multi-scale FWIME scheme. Spline 5 coincides with the finite-difference grid.

focusing on the first 87 iterations, which corresponds to the inversion on the first spline grid. It is interesting to notice that after the sixth iteration, the red curve becomes smoother (i.e., more monotonic) and contains less local minima than its counterpart from section 5.2. This observation potentially indicates that in this example, the inversion on the first spline grid is the most crucial step of the FWIME inversion process. From a physical aspect, the main challenge to overcome is the reconstruction of the low-resolution structure of the salt body from a pure sediment model.

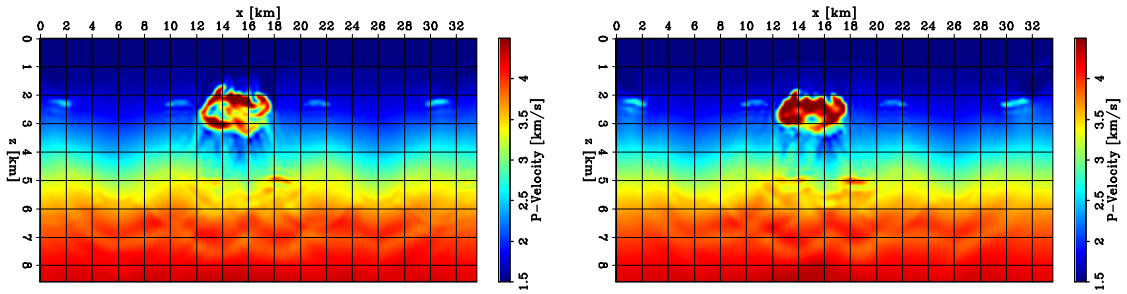


(a) (b)

Figure 5.53: Total normalized FWIME objective function (blue curve), and normalized FWI objective function evaluated at each FWIME inverted model (red curve). (a) Convergence curves plotted for the full inversion workflow (five spline grids for a total of 250 iterations). (b) Convergence curves only plotted for the first 87 iterations (inversion on the first spline grid). [**CR**]

Unlike the numerical examples proposed in sections 5.1, 5.2 and 5.3, I observe here that the quality of the inverted model is mostly affected by the selection of the spline grid sequence, rather than by the trade-off parameter value ϵ. Figure 5.54 shows two FWIME inverted models obtained with a different spline grid sequence (with the same ϵ-value): Figure 5.54a is the result from starting the inversion directly on the second grid, whereas Figure 5.54b is computed by skipping the second grid. In both figures, the contours of salt body have been correctly reconstructed but the inversion

is unable to automatically fill the salt with the correct velocity value. Moreover, by conducting additional tests (not shown here), I notice that the presence of salt (or strong velocity contrasts) may require a slower refinement rate in the spline grid sequence to ensure convergence to an accurate solution.



(a) (b)

Figure 5.54: 2D panels of velocity models inverted with a different spline grid refinement schedule as the one shown in Table 5.3. (a) Inverted model starting from spline 2. (b) Inverted model without the use of spline 2. [**CR**]

Finally, I apply an imaging quality-control step in order to assess the accuracy of the inverted velocity model. Figures 5.55a and b show the migrated images computed with the initial and final velocity models, respectively. As expected, the improvements are extremely significant in the subsalt region, and the thin high-velocity layers are now clearly visible.

5.5 The need for a more flexible model parametrization

I design an experiment based on the BP 2004 benchmark model (Billette and Brandsberg-Dahl, 2005) to assess the limitations of FWIME when imaging in the presence of complex/thick overburdens. I show that FWIME does not manage to recover an accurate solution. However, I anticipate that my method can be conveniently improved by designing and employing a more flexible coarse-model representation, such as radial basis functions (RBF) (Buhmann, 2000; Martin et al., 2015).

The true velocity is 40 km wide and 9 km deep (Figure 5.56d) and focuses on the central part of the original BP 2004 model, which is representative of the geological environments encountered in the Gulf of Mexico and West Africa. Unlike the example shown in section 5.4, the salt body is thicker, deeply rooted, and presents steep flanks, which makes its boundaries challenging to delineate. In addition, a sharp interface between the base salt and a low-velocity zone is present underneath the main salt structure and ranges from an approximate depth of 5 km, down to the bottom of the model. Even with a long-offset acquisition (and diving waves), very little energy can illuminate this region, making its recovery very difficult.

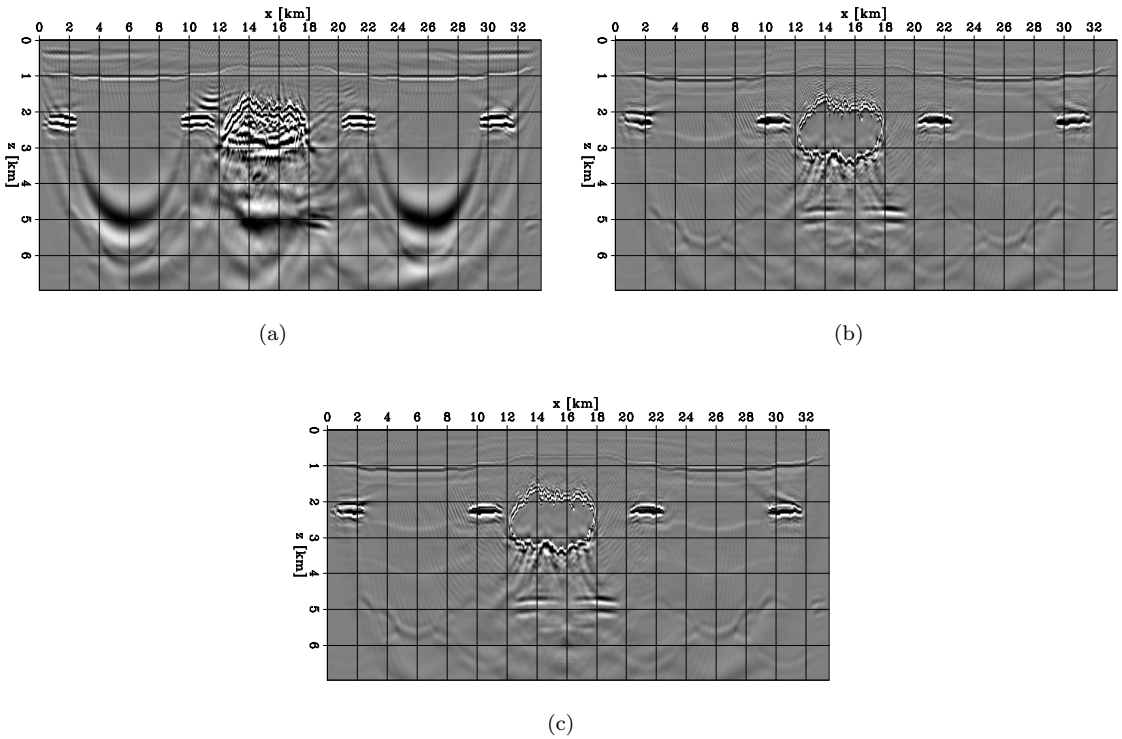


Figure 5.55: RTM images computed with different velocity models. (a) Initial model. (b) FWIME inverted model. (c) True model. [**CR**]

I simulate a long-offset seismic acquisition with 160 sources placed every 250 m (50 m below the surface), and 800 fixed receivers every 50 m (also placed 50 m below the surface). The noise-free data are modeled for 17 s with an acoustic isotropic constant-density propagator and a spatial sampling of 50 m for both vertical and horizontal axes. In order to reduce the difficulty of this test, I use absorbing boundaries in all directions, and no free-surface related multiples are present within the dataset. The initial model is solely composed of sediments (Figure 5.56a). Throughout this example, the maximum frequency used is set to 8 Hz for both FWI and FWIME schemes. Figure 5.58 shows two representative shot gathers for sources placed at $x = 4$ km and $x = 20$ km, respectively. These panels confirm the difficulty to illuminate the low-velocity region underneath the salt. In fact, the only recorded signal coming from this zone is carried by diving waves recorded at very long offsets, as shown by the white arrows in both panels. Moreover, for sources placed above the salt feature, the amplitude of the top-salt reflections dominate the ones generated by the interface between the base of the salt and the low-velocity zone (green arrows in Figure 5.58).

I first conduct FWI using unrealistic low-frequency energy ranging from 0.25 Hz to 8 Hz (Figure 5.57a), which converges to an excellent solution (Figure 5.56b). Then, I limit the available bandwidth to 3-8 Hz (Figure 5.57b) and the final inverted model is unable to recover the salt body (Figure 5.56c), even with a data-space multi-scale strategy.

I apply FWIME with a time-lag extension and I simultaneously invert a dataset generated with a wavelet containing energy within the 3-8 Hz frequency band (Figure 5.57c). Unfortunately, even with a thorough hyper-parameter search (i.e., by testing spline grids with various spatial samplings and a wide range of ϵ-values), FWIME is not able to recover a model as accurate as the one obtained with FWI using unrealistic low frequencies (Figure 5.56b). Figures 5.59 shows the set of the FWIME inverted models obtained by using the most optimal hyper-parameters (a sequence of six spline grids with $\epsilon = 5.0 \times 10^{-4}$). The final inverted model is shown in Figure 5.59g and fails to accurately capture the low-velocity region underneath the salt body, and the right flank is mis-positioned. Moreover, some low-velocity artifacts are present on the left flank of the salt. The blue curve in Figure 5.60 shows the total FWIME objective function corresponding to the optimization sequence, which indicates that the algorithm converges to a local minimum. The red curve in Figure 5.60, which is the FWI objective function evaluated at each FWI inverted model during the sequence, shows that FWIME was able to bypass some - but not all - local minima from the conventional FWI formulation. Nevertheless, FWIME performs better than conventional FWI (for the more realistic scenario where no coherent energy lower than 2 Hz is present) and the salt features in the shallower parts of the model are better imaged (for depths smaller than 4 km).

Figure 5.61 shows the observed data, $\mathbf{d}^{obs}$ (first column), predicted data with the final FWIME inverted model $\mathbf{f}(\mathbf{m}_{FWIME})$ (second column), and the final data difference, $\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWIME})$ (third column). The first and second rows correspond to sources located at $x = 4$ km, and $x = 20$ km, respectively. As expected, a large data mismatch can be observed for events occurring at

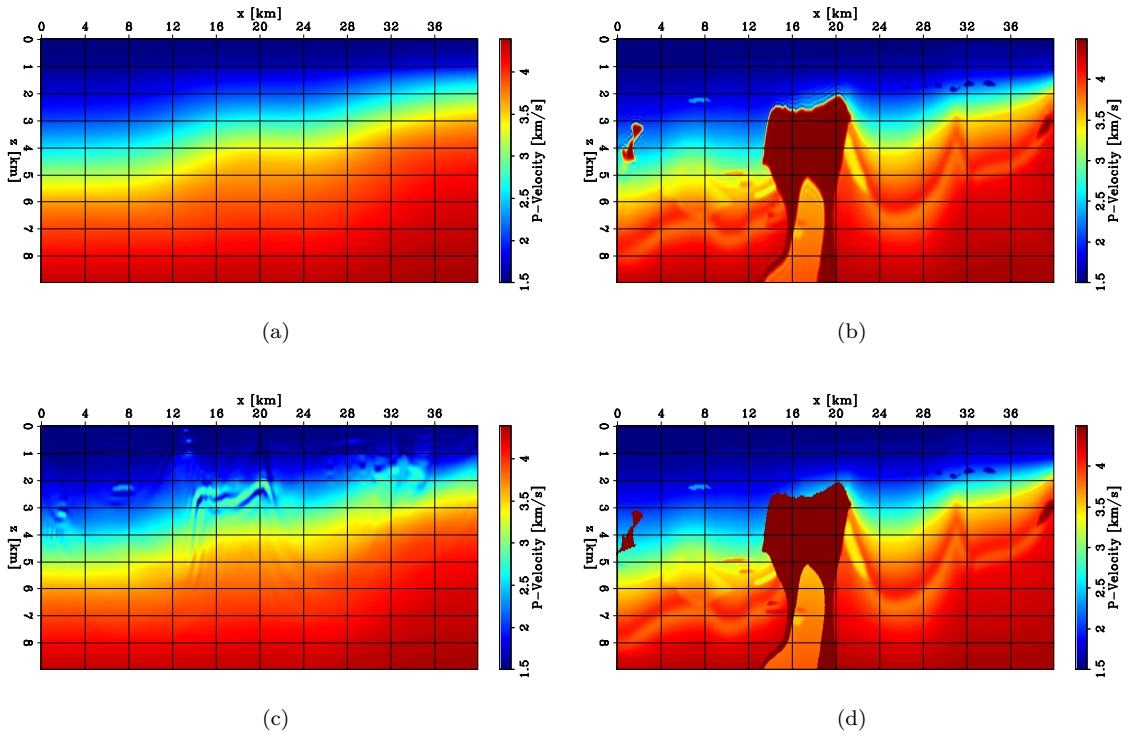


Figure 5.56: 2D panels of velocity models. (a) Initial model. (b) Conventional data-space multi-scale FWI inverted model using a 0-8 Hz dataset. (c) Conventional data-space multi-scale FWI inverted model using a 3-8 Hz dataset. (d) True model. [**CR**]

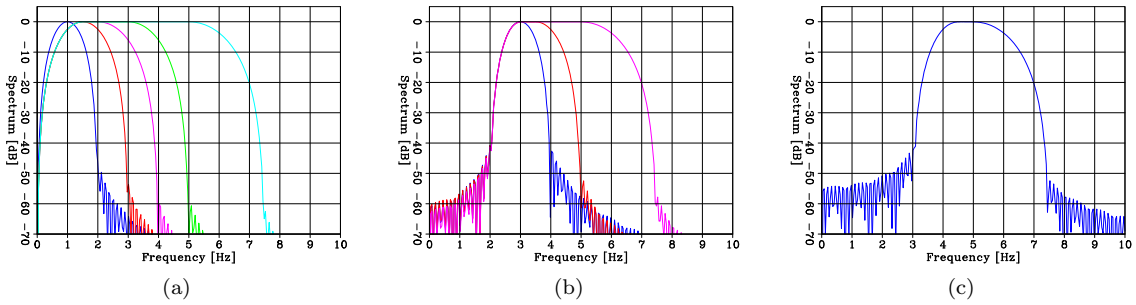
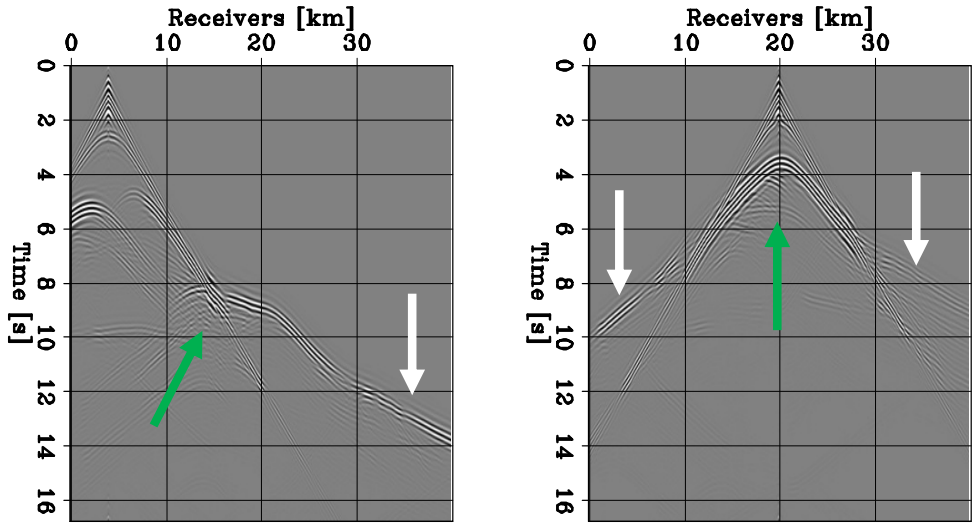


Figure 5.57: Amplitude spectra of the seismic sources employed in this numerical example. (a) Sequence of sources used for the data-space multi-scale FWI workflow using unrealistic low-frequency energy. (b) Sequence of sources used for the data-space multi-scale FWI workflow using energy restricted to the 3-8 Hz bandwidth. (c) Source used for the FWIME workflow. [**ER**]



(a) (b)

Figure 5.58: Representative shot gathers generated with a source wavelet contaning energy restricted to the 3-8 Hz range. (a) Source placed at $x = 4$ km. (b) Source placed at $x = 20$ km. All panels are displayed with the same grayscale. White arrows show the energy recorded from diving waves traveling through the low-velocity zone underneath the salt structure. The green arrows correspond to reflected energy from the base salt. [**CR**]

larger offsets (Figure 5.61c) for receivers' positions greater than 25 km, which correspond to diving waves traveling through the salt body. This prediction error highlights the inability of FWIME to accurately recover for the low-velocity zone underneath the salt. In addition, the last inverted model fails at predicting some reflected and scattered energy coming from the base salt, as shown in Figures 5.61c and 5.61f for near offsets.

In the presence of complex geological structures and overburdens, I observe that the selection of the spline grid spatial sampling and the rate at which I adjust it throughout the optimization process strongly impacts the quality of inverted models. More specifically, the inability to constrain the velocity values to vary less within the salt body (and more rapidly around its edges) seems to hamper the inversion process. Perhaps a more flexible model parametrization may improve the convergence properties of FWIME. When dealing with complex salt bodies, I intend to investigate the use of more adjustable basis functions for coarse model representation, such as radial basis functions (Dahlke et al., 2020). The proposed method would consist in using migrated images to guide the nodes' spacing and their spatial arrangements. The grid density would be increased in regions containing salt edges, and would be reduced within the salt bodies to promote homogeneous velocity values.

5.6 2D computational-cost analysis

The numerical implementation for the 2D examples proposed in this chapter are conducted on a subset of the Stanford Exploration Project (SEP)'s computational resources, which include four compute nodes, each containing four NVIDIA Tesla V100 GPU devices with 16 GB of global memory. Each node is equipped with 24 CPU cores and 512 GB of random access memory (RAM). Each example is solved on one node (four GPU devices) without any domain-decomposition strategy (Micikevicius, 2009): each GPU device simulates one shot throughout the full finite-difference domain of interest.

Table 5.4 provides a breakdown of the FD parameters and computational cost for the five proposed numerical examples. N_z and N_x correspond to the number of spatial samples on the finite-difference grid, which includes additional padding for the absorbing boundaries. N_{ext} is the total number of samples on the extended axis for $\tilde{\mathbf{p}}_{\epsilon}^{opt}$. T_{rec} is the total recording time for each shot, expressed in seconds, and VP_{iter} corresponds to the number of linear conjugate gradient iterations conducted for the variable projection step (minimization of equation 3.6). The quantities t_{FWIME} and t_{FWI} correspond to the approximate computational time taken for one L-BFGS iteration for FWIME and conventional FWI, respectively (expressed in minutes). The last column reports the ratio between the FWIME and the FWI computational times. For these 2D tests, I choose conservative values for N_{ext} and VP_{iter} to ensure accurate convergence for the variable projection step. For the Marmousi2 example, I investigate the effect of using a smaller number of linear iterations for the

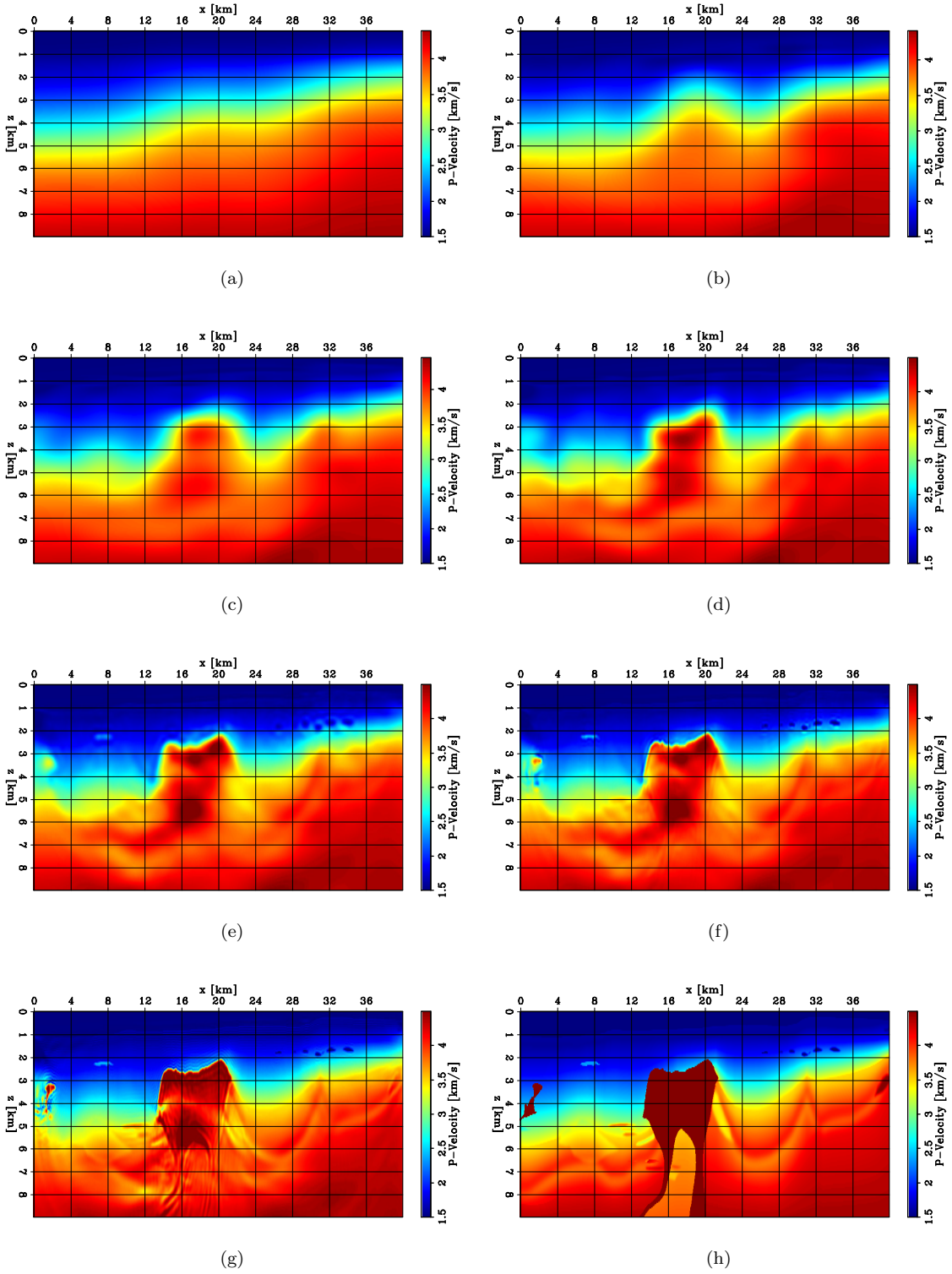


Figure 5.59: 2D panels of FWIME inverted model at various stages of the inversion process. (a) Initial step. (b) Spline 1. (c) Spline 2. (d) Spline 3. (e) Spline 4. (f) Spline 5. (g) FWIME inverted model on the last spline. (h) True model. The last spline grid coincides with the FD grid. Panel (g) is obtained after a total of 258 iteration of L-BFGS. [**CR**]

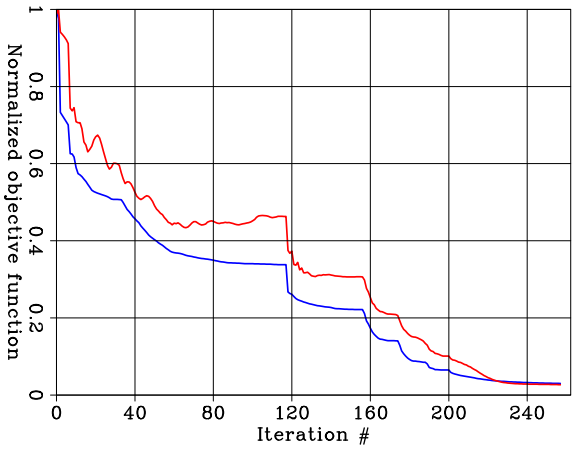


Figure 5.60: Total normalized FWIME objective function (blue curve), and normalized FWI objective function evaluated at each FWIME inverted model (red curve). [**CR**]

variable projection step (section 5.2). I notice that 15 iterations are sufficient to converge to a similar solution, which reduces FWIME's computational cost by a factor of three. For 3D applications, these numbers should indeed be optimized to enhance the overall computational efficiency.

2D Model	N_z	N_x	N_{ext}	T_{rec}	VP_{iter}	t_{FWIME}	t_{FWI}	Ratio
Marmousi2	250	700	101	8 s	60	50 min	0.5 min	100
Marmousi2	250	700	101	8 s	30	26 min	0.5 min	52
Marmousi2	250	700	101	8 s	15	14 min	0.5 min	28
North Sea	250	840	101	13 s	60	80 min	0.75 min	106
Seiscope	220	470	201	6 s	80	15 min	0.2 min	90
Salt	330	1000	101	15 s	100	85 min	0.75 min	110
West Africa	330	1000	101	15 s	100	85 min	0.75 min	110

Table 5.4: Table summarizing the finite-difference modeling parameters and inversion schemes' computational time for the five numerical examples proposed in this paper. Each example is conducted on four NVIDIA Tesla V100 GPU devices. N_z and N_x correspond to the number of samples for the FD grid, which the absorbing boundaries. N_{ext} is the total number of samples on the extended axis. T_{rec} corresponds to the recording time for the simulated data. VP_{iter} is the number of linear conjugate-gradient iterations conducted for the variable projection step in FWIME (equation 3.6). t_{FWIME} and t_{FWI} correspond to the computational time for one L-BFGS iteration of FWIME and FWI, respectively. The last column displays the ratio between t_{FWIME} and t_{FWI}.

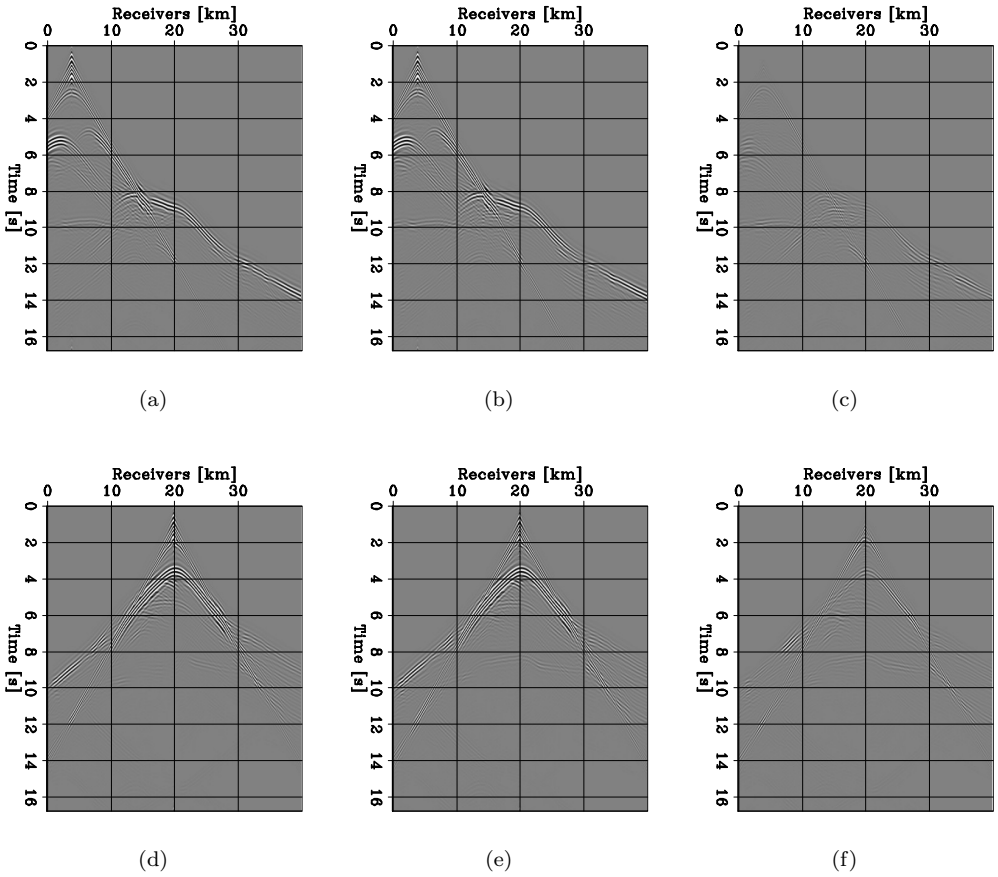


Figure 5.61: Representative shot gathers for sources placed at $x = 4$ km (first row) and $x = 20$ km (second row). Observed data, $\mathbf{d}^{obs}$ (first column), Predicted data with the final FWIME model, $\mathbf{f}(\mathbf{m}_{FWIME})$ (second column), and data-difference, $\Delta\mathbf{d}(\mathbf{m}_{FWIME}) = \mathbf{d}^{obs} - \mathbf{f}(\mathbf{m}_{FWIME})$ (third column). All panels are displayed with the same grayscale. [**CR**]

Chapter 6

3D field-data application

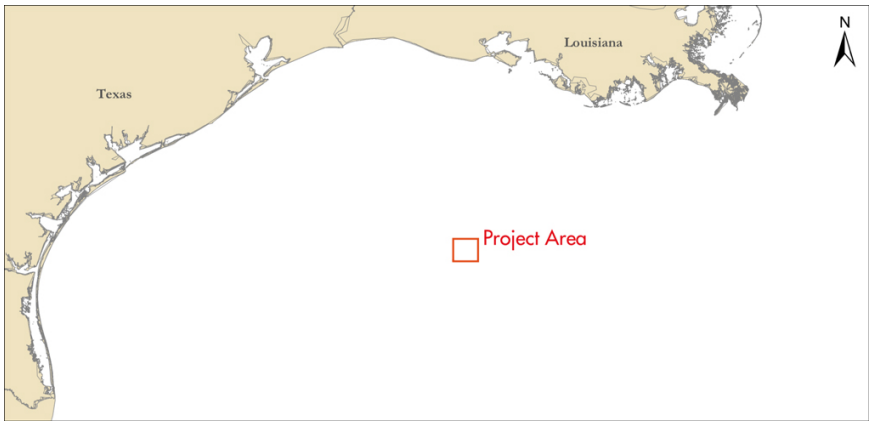
I apply FWIME by inverting the pressure component of a 3D deep-water ocean-bottom node (OBN) survey acquired in the Gulf of Mexico (GoM) by Shell and recently studied by Dahlke (2019); Sang et al. (2019); Vargas (2020); Biondi (2021); Biondi et al. (2021a). I present the geological setting of the survey and the data pre-processing steps. Then, I conduct a conventional multi-scale FWI using an inaccurate initial model, which leads the inversion scheme to converge to a local minimum. Finally, I apply model-space multi-scale FWIME and I improve the accuracy of the inverted model. I assess the quality of the various solutions with high-resolution RTM migrated images.

6.1 Geological setting

This dataset was acquired in 2010 by Shell in the Gulf of Mexico, located in the Garden Banks region, approximately 360 km south-west of New Orleans, Louisiana (Figure 6.1). The goal of this survey was to exploit the wide-azimuth illumination of OBN acquisitions to improve the imaging and the characterization of a 9,500 m-deep oil and gas reservoir, located in the vicinity of a deeply-rooted salt dome. In this thesis, I use this OBN dataset to show the potential of FWIME at recovering satisfactory Earth models, and I focus on improving the image quality of the shallower region of the subsurface, for depths smaller than 4 km.

6.2 Data selection

I extract the hydrophone pressure recordings of 255 node gathers out of the 308 available in the provided multi-component dataset. For each node, the total recording time is approximately 14 s sampled at 2 ms, and no separation between the up- and down-going wavefields is conducted. Figure 6.2a shows the nodes' horizontal positions (yellow dots) overlaid on a depth section extracted at $z = 2.5$ km from the initial velocity map computed with the inputs provided by Shell, which



(a)

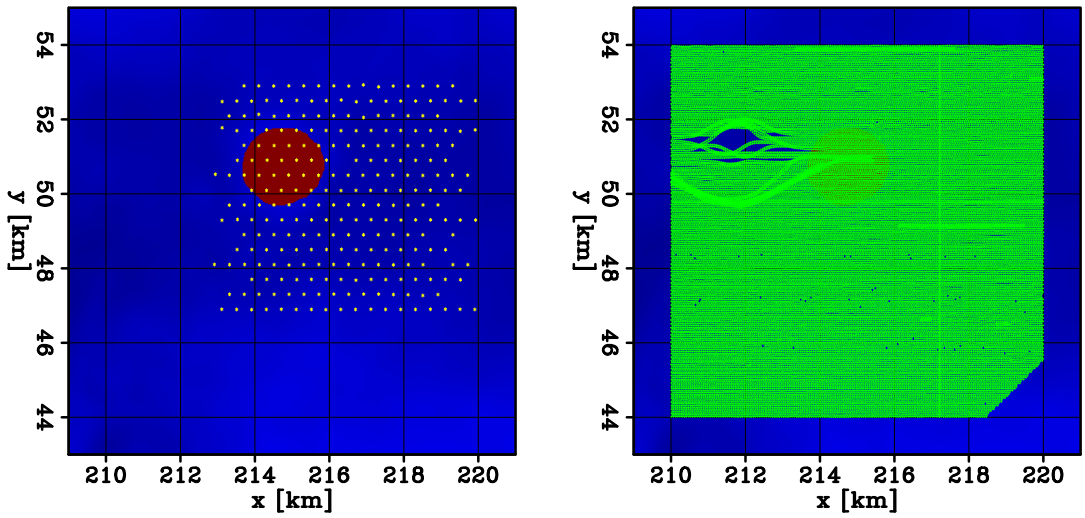
Figure 6.1: Exploration map highlighting the region of the Gulf of Mexico in which the OBN dataset was acquired. [**NR**]

included a model of the vertical salt dome (red circular feature). During acquisition, the nodes are placed at the seabed whose depth ranges from 0.83 km to 1.0 km. The horizontal sampling between two adjacent nodes is approximately 250 m in both x- and y-directions. For each node gather, I use the seismic traces generated by $N_s = 41000$ airgun sources which cover a surface area of approximately 10 km $\times$ 10 km. Even though the sources' positions may differ between each node gather, N_s stays constant across all node gathers. The data were acquired with a flip-flop shooting pattern and a source interval of approximately 25 m, placed at a depth of 9.8 m below the sea level. The sail lines coincide with the x-axis, and the spacing between two consecutive lines in the y-direction is approximately 100 m. Figure 6.2b displays the sources' horizontal positions for a representative node gather. In this study, I reduce the computational cost of my finite-difference (FD) numerical modeling scheme by applying the reciprocity theorem and using the 255 node gathers as shot gathers, and the 41000 shots as receivers (Aki and Richards, 2002).

6.3 Data pre-processing

I begin by determining the useful frequency range from the raw hydrophone pressure data. The lowest frequency employed for my analysis is set to 3 Hz because no clear coherent signal is found below that level (Figure 6.3). On the higher end of the spectrum, I observe a good signal-to-noise ratio (SNR) for the signal up to 40 Hz (Biondi, 2021). However, in order to reduce the modeling computational cost for this study, I restrict the maximum frequency content of the data to 12 Hz for the velocity model building phase, and 30 Hz for the RTM imaging step.

In order to mitigate potential amplitude and waveform inconsistencies between recorded and predicted signals (caused by the acoustic isotropic constant-density approximation made in the



(a) (b)

Figure 6.2: Map view of the acquisition geometry corresponding to the subset of the data selected for my analysis. The background displays the extent of the sediments (blue) and the Auger salt dome (red circular anomaly) extracted at a depth of 2.5 km from the velocity model provided by Shell. (a) Nodes' horizontal positions (yellow dots). (b) Sources' horizontal positions (green dots). [**NR**]

modeling and the unknown sources' signature), I compute and I apply a trace-by-trace filter $\mathbf{f}_i \in \mathbb{R}^{N_t}$ to every raw observed node gathers, $\mathbf{d}_i^{raw} \in \mathbb{R}^{N_t \times N_s}$, where i corresponds to the node gather index ranging from 1 to 255, and N_t is the number of time samples per trace. Rather than conducting an iterative wavelet-estimation scheme throughout the optimization process (Rickett, 2013; Sun et al., 2014), this approach only requires to compute and apply $\mathbf{f}_i$ to the raw data at the initial step. The observed data corresponding to node gather i, $\mathbf{d}_i^{obs} \in \mathbb{R}^{N_t \times N_s}$ is given by the following expression,

$$\mathbf{d}_i^{obs} \quad = \quad \mathbf{f}_i * \mathbf{d}_i^{raw}, \tag{6.1}$$

where * denotes the time-convolution operator, and $\mathbf{f}_i$ is applied trace by trace. To compute $\mathbf{f}_i$, I begin by modeling the predicted data using the initial P-wave velocity model provided by Shell. For both observed and predicted signals, I obtain a proxy for the direct arrival by applying a hyperbolic moveout (HMO) correction and stacking all the traces for that node into a single trace, $\mathbf{s}_i^{raw} \in \mathbb{R}^{N_t}$ and $\mathbf{s}_i^{pred} \in \mathbb{R}^{N_t}$. Then, I use these two traces to estimate $\mathbf{f}_i$ by solving the following equation in the frequency domain,

$$\mathbf{s}_i^{pred} \quad = \quad \mathbf{f}_i * \mathbf{s}_i^{raw}. \tag{6.2}$$

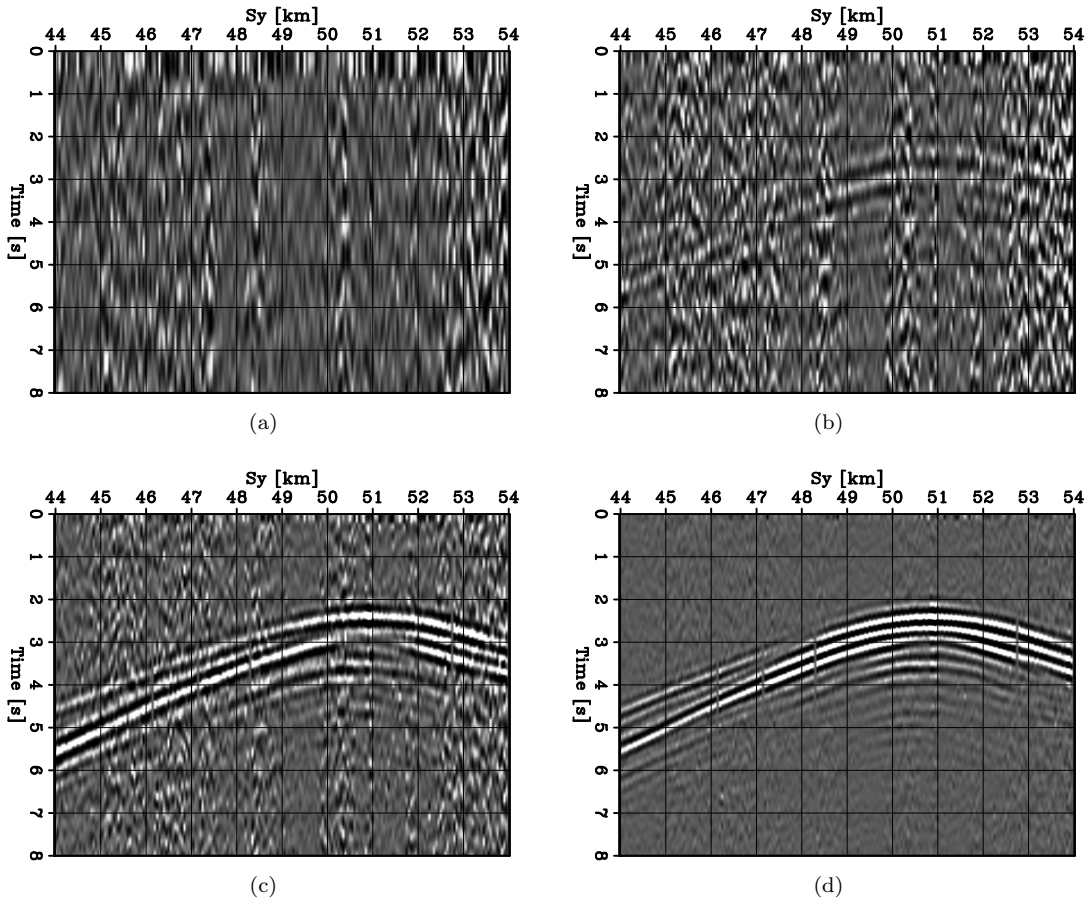


Figure 6.3: Representative shot-binned common-receiver gather for multiple sources with a fixed x-position $S_x = 214.8$ km (aligned with the cross-line direction) on which a high-cut filter was applied with various cut-off frequencies. (a) 1 Hz, (b) 2 Hz, (c) 3 Hz, and (d) 4 Hz. [**ER**]

Figure 6.4 shows a representative node gather of the raw observed data, $\mathbf{d}_i^{raw}$, the initial prediction, $\mathbf{d}_i^{pred}$, and the observed data $\mathbf{d}_i^{obs}$ used all waveform inversion schemes described in this chapter. As expected, the application of $\mathbf{f}_i$ removes the bubble response from the observed data commonly generated by airgun sources (Watson et al., 2019), and largely improves the consistency with the initial prediction.

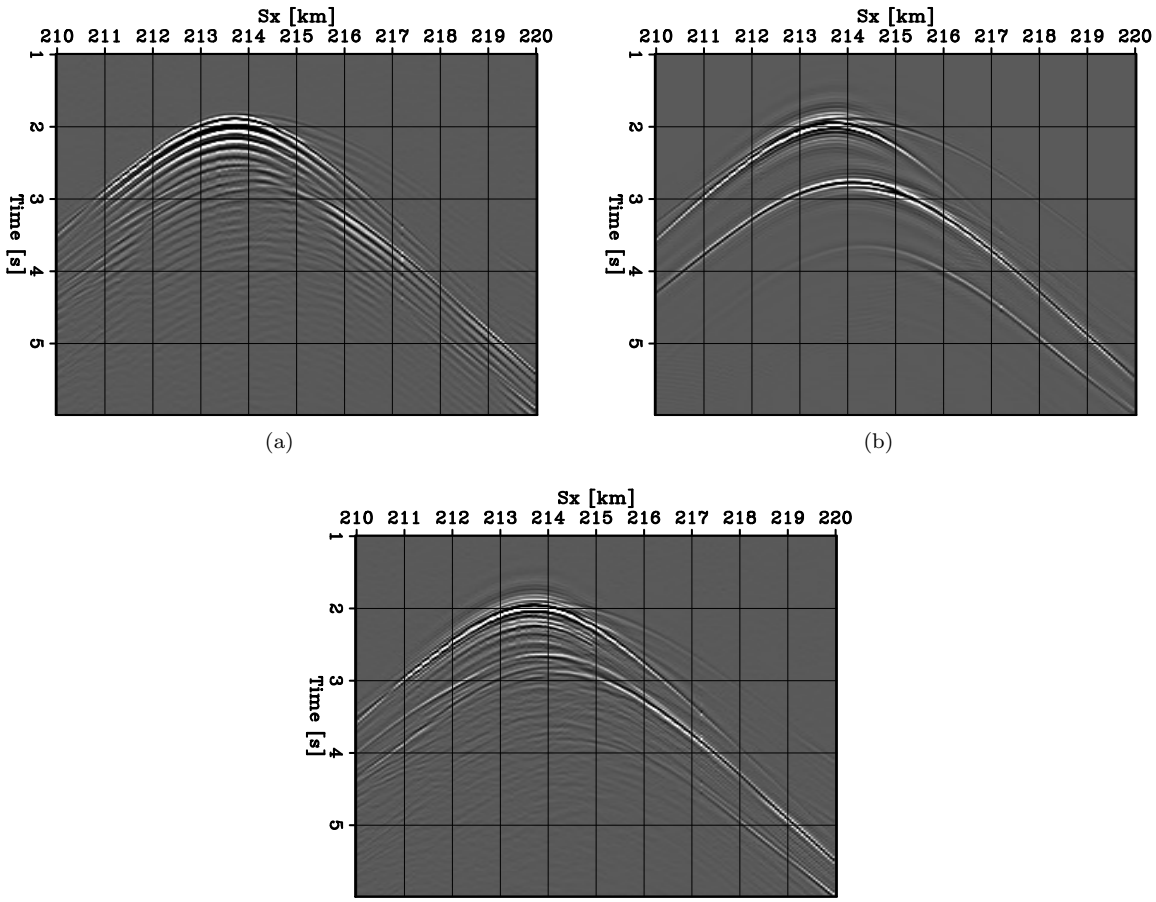


Figure 6.4: Representative common-receiver gathers for node #100 generated by sources placed along the in-line direction at $y = 49$ km. (a) Raw data $\mathbf{d}^{raw}$, (b) initial prediction $\mathbf{d}^{pred}$, and (c) observed data after processing, $\mathbf{d}^{obs}$. [**CR**]

6.4 Finite-difference modeling

I select an area of study which is approximately 10 km x 10 km wide and 4 km deep with a water-layer thickness of 1 km. The inline and crossline positions range from $x = 209.75$ km to $x = 220.22$ km and from $y = 43.7$ km to $y = 54.2$ km. I design a GPU-based acoustic isotropic constant-density 3D

finite-difference modeling scheme with a free-surface boundary condition at the water surface, and absorbing boundaries in all other directions. The free-surface condition allows me to use both up- and down-going signals recorded by the nodes at the water bottom (Robertsson, 1996; Biondi, 2021). For both FWI and FWIME, the frequency content of the inverted data is restricted to the 3-12 Hz range, and the wavefield propagations are modeled for 8 s. To reduce the computational cost and memory usage, all scattering and imaging operations are performed with a time sampling of 24 ms (which is different from the time sampling used for the FD propagation, set to 2 ms). For the imaging step (RTM) which is usually conducted with higher-frequency data, I use a maximum frequency of 30 Hz. I modify the FD spatial/time samplings accordingly to ensure numerical stability and to avoid numerical dispersion (Dablain, 1986). Table 6.1 summarizes the FD parameters selected for the different imaging algorithms described in this chapter.

Parameter	**FWI/FWIME (12 Hz)**	**RTM 30 Hz**
N_z	170	296
N_x	400	760
N_y	400	746
$\Delta x/\Delta y/\Delta z$	35 m	15 m
N_t	330	500
Δt	24 ms	15 ms

Table 6.1: Summary of the FD parameters used for the velocity estimation and imaging steps shown in this chapter.

6.5 Inversion strategy

I conduct two separate FWI schemes using two initial velocity models with various accuracy. Model 1 is derived from the inputs provided by Shell, which include the elastic stiffness tensor coefficients C_{11}, C_{13}, C_{33}, C_{44}, and C_{66}, a density model of the sediments, and a mask for the salt regions (Mah and Schmitt, 2003; Biondi, 2021). The velocity value in the salt is initially assumed to be homogeneous and is set to 4.5 km/s (Zong et al., 2015). The left column of Figures 6.5 shows representative cross-sections in the crossline direction extracted from model 1 for various x-positions. The background sediment velocity model is smooth and does not contain any distinct geological characteristics. As shown by the FWI study conducted in Biondi (2021), model 1 seems to be accurate enough for conventional data-space multi-scale FWI to converge to a useful solution down to an approximate depth of 3.5 km. However, its design requires considerable amount of human input and prior geological knowledge.

I then create a second model (model 2) purposely chosen to be inaccurate, which largely underestimates the velocity values from model 1 for depths greater than 1.5 km (left column of Figure 6.6).

The same mask to delineate the salt diapir (also provided by Shell) is employed for both models 1 and 2. I first show that conventional FWI converges to a unsatisfactory solution when model 2 is used as initial guess (compared to model 1). Then, I demonstrate FWIME's ability to recover a solution much more accurate than the one obtained by applying conventional FWI with model 1, thereby bypassing the tedious and user-intensive steps required to build a good-quality initial model. For both FWI and FWIME inversions, I use an acoustic isotropic constant-density operator to model wavefields. To reduce for amplitude mismatch between predicted and observed data, I add the trace-by-trace normalization operator $\mathbf{n}$ defined in section 3.2.5.

6.6 FWI starting with model 1

I apply conventional data-space multi-scale FWI with model 1 as initial guess. I use three frequency bands whose amplitude spectra are shown in Figure 6.7. The final inverted model after a total of 300 iterations of L-BFGS (100 iterations for each band) is shown in the right column of Figure 6.5, and seems to recover an accurate and geologically consistent solution down to approximately 3.5 km. Beyond this depth, low-velocity zones are introduced throughout the inversion, which may correspond to non-physical values caused by the presence of a local minimum within the FWI objective function.

One potential explanation for this issue is the presence of a thick horizontal layer with strong anisotropy, ranging from an approximate depth of 3 km down to 6 km. Figures 6.8a and 6.8b show two cross-line sections from the Thomsen anisotropy parameter ϵ (not to be confused with the FWIME trade-off parameter) computed with the initial stiffness coefficients provided by Shell, and assuming vertical transverse isotropy (VTI) (Thomsen, 1986; Tsvankin, 2012). The Shell-provided inputs indicate that the entire subsurface is an elliptically anisotropic medium (both Thomsen anisotropy parameters ϵ and δ are equal), and thus the δ-model is not shown here. Figure 6.9 displays a vertical profile extracted from the ϵ model at $x = 214.8$ km and $y = 48$ km, which shows that ϵ reaches values as high as 0.11. Figures 6.10a and 6.10b show a closeup of the ϵ model focusing on my area of study. It is interesting to notice that the depth at which the FWI scheme seems to converge to a local minimum coincides with the top of the strongly-anisotropic layer. Perhaps incorporating anisotropic modeling into the FWI scheme could improve the accuracy of the inverted result (Le, 2019; Kamath et al., 2021).

The quality of the FWI model can be analyzed by examining the RTM image computed with a maximum frequency content of 30 Hz and a spatial sampling of 15 m in all directions. In addition, a source illumination compensation is applied to mitigate acquisition artifacts present in the vicinity of the OBN positions (Kaelin and Guitton, 2006). Figure 6.11 shows four cross-sections extracted at $x = 214.8$ km, $x = 215.5$ km, $x = 216$ km, and $x = 217$ km from the RTM images computed with model 1 (left column) and with the final FWI inverted model (right column). The deeper

reflectors (below 2 km) in the right column show clear improvement in terms of lateral coherency, focusing, and continuity, especially near the flanks of the salt dome (Figures 6.11b and 6.11d). In the sections shown in Figures 6.5g and 6.18d, which solely contains sedimentary rocks, the FWI scheme dramatically improves the image quality and manages to recover well-focused reflectors at depths greater than 3 km that were not visible on the initial image (Figure 6.11g).

In the following, I use this FWI result as a target model: my goal is to apply FWIME using model 2 as initial guess and recover a solution whose quality is comparable to the FWI output starting with model 1.

6.7 FWI starting with model 2

I conduct an analogous FWI workflow starting with model 2. The right column of Figure 6.6 displays cross-sections of the recovered model obtained after 60 iterations of L-BFGS and by inverting the 3-6 Hz frequency band from the data. In the shallow regions (for $z < 1.75$ km), the inversion scheme converges to a similar solution as for the inversion with model 1. However, a non-physical low-velocity layer is introduced for depths ranging from $z = 2$ km to $z = 2.5$ km, which likely indicates that the inversion has converged to a local minimum. Consequently, I stop the FWI workflow after the first frequency band. As expected, the accuracy of the recovered model is much lower than the one starting with model 1, which is shown by the RTM images displayed in Figure 6.12. The low-velocity layer generates a spurious sharp reflection in the image, and multiple reflectors below that event lack coherency. In some regions, the image computed with the final FWI model is less focused and the overall image quality decreases compared to the initial RTM image, as shown by the reflector near the salt flank located at $z = 2.8$ km and ranging between $y = 49$ km and $y = 50.25$ km in Figure 6.12f.

6.8 FWIME

I use model 2 as initial guess and I simultaneously invert all available events within the 3-12 Hz frequency band from the observed data. The amplitude spectrum of the source wavelet used throughout the inversion is shown in Figure 6.7 (pink curve). I begin by describing the hyper-parameter selection process, and I analyze the initial FWIME search direction. Then, I present and assess the quality of the results from the FWIME scheme. Finally, I discuss the computational cost for this study.

6.8.1 Hyper-parameter selection

For the optimal extended perturbation $\tilde{\mathbf{p}}_{\epsilon}^{opt}$, I use a time-lag extension with a total of 41 points sampled at 24 ms, allowing τ to range from -0.48 s to 0.48 s (a very conservative range). The

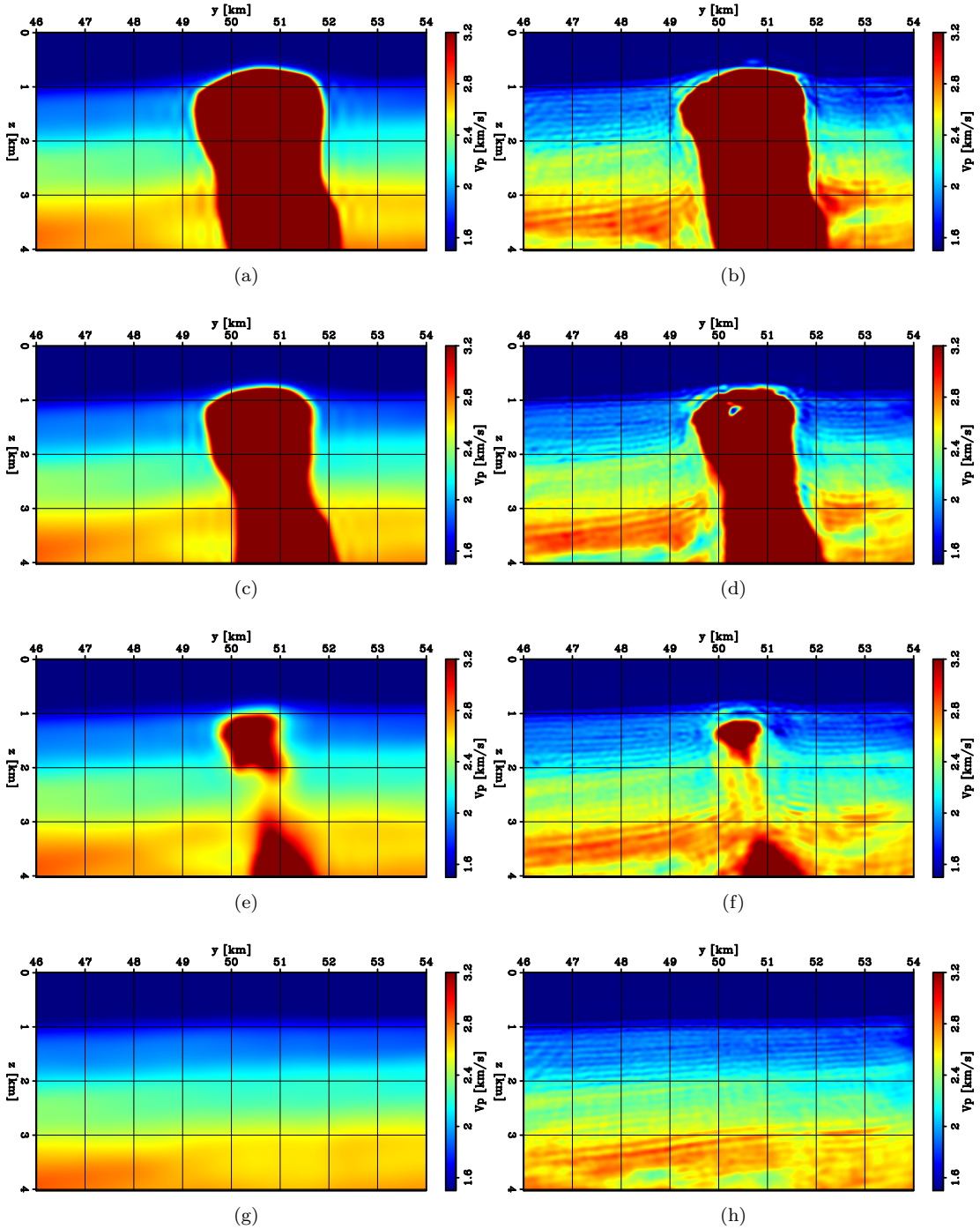


Figure 6.5: Cross-line sections extracted from velocity models at four different x-positions. The left panels displays the initial velocity model provided by Shell (model 1). The right panel corresponds to the FWI inverted model after 300 iterations of L-BFGS using three frequency bands spanning the 3-12 Hz range. The first, second, third and fourth rows correspond to $x = 214.8$ km, $x = 215.5$ km, $x = 216$ km, and $x = 217$ km, respectively. [**CR**]

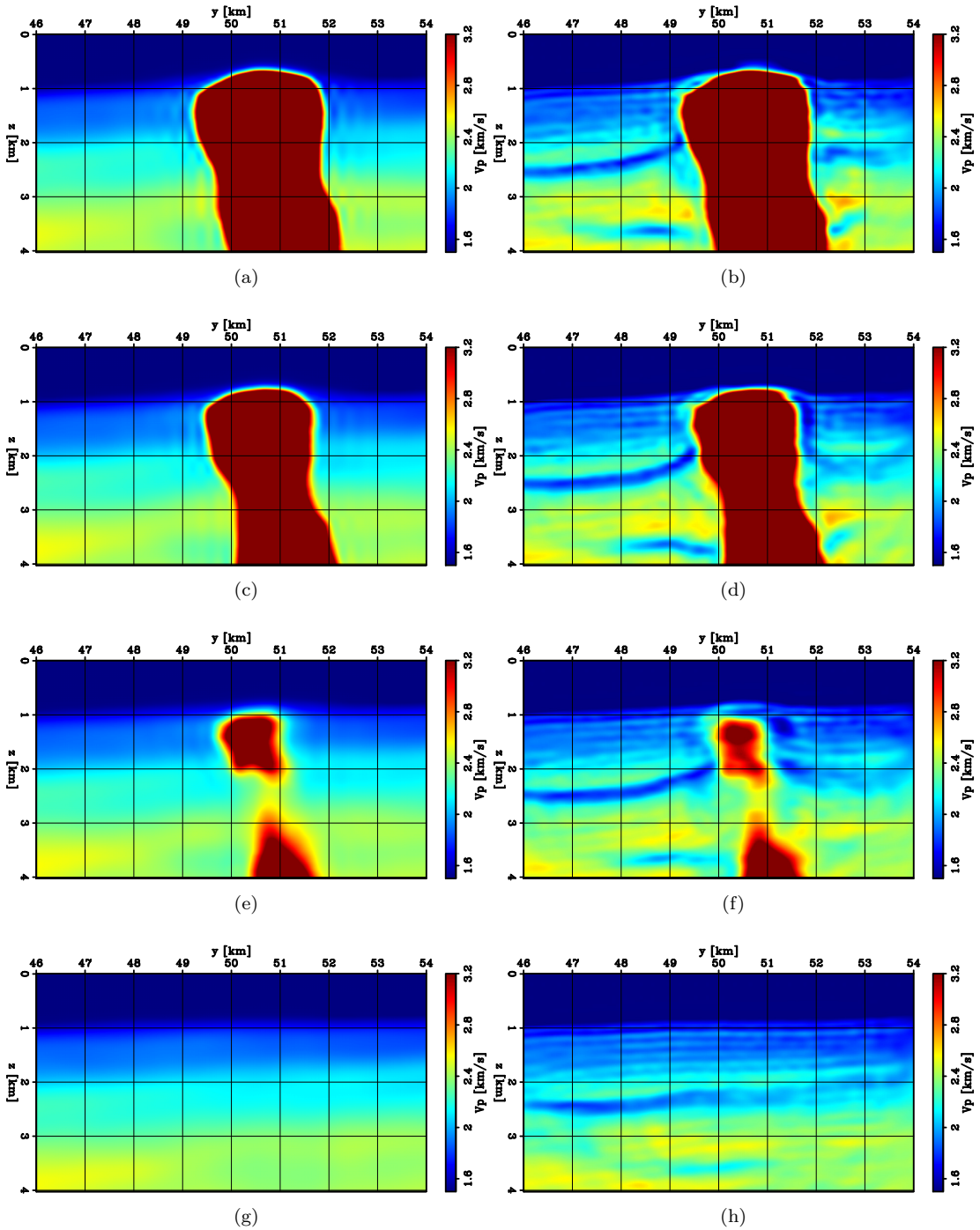


Figure 6.6: Cross-line sections extracted from velocity models at four different x-positions. The left panels displays initial velocity model 2. The right panel corresponds to the FWI inverted model after 60 iterations of L-BFGS using data whose frequency content is restricted to the 3-6 Hz range (and model 2 as initial guess). The first, second, third and fourth rows correspond to $x = 214.8$ km, $x = 215.5$ km, $x = 216$ km, and $x = 217$ km, respectively. [**CR**]

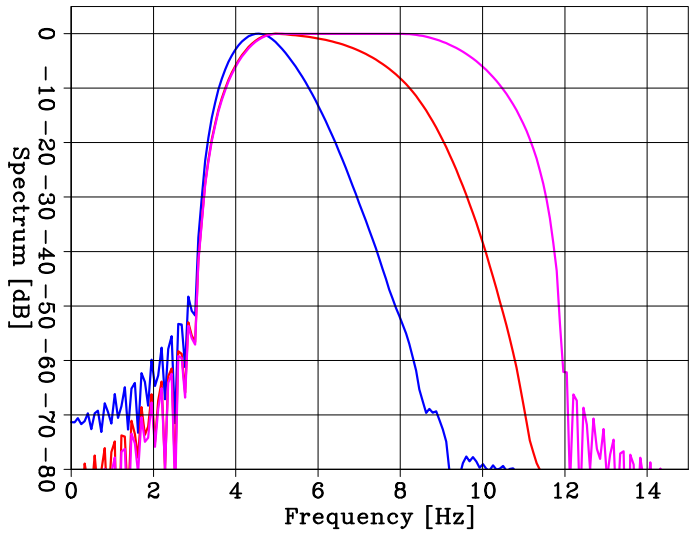
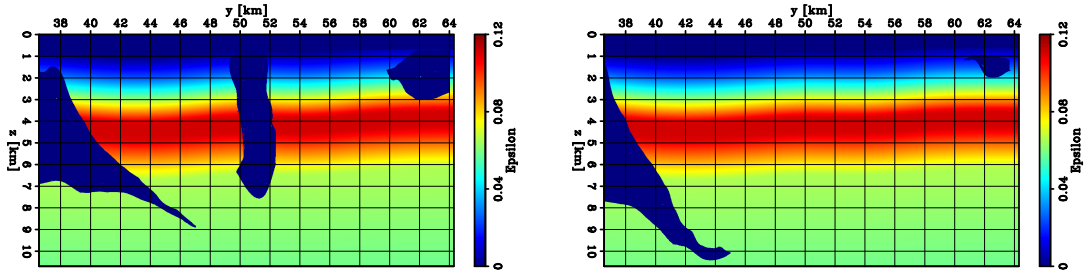


Figure 6.7: Amplitude spectra of the three seismic sources used for the data-space multi-scale FWI. The third band (pink curve) is also employed for the FWIME workflow. [**ER**]



(a) (b)

Figure 6.8: Cross-line sections extracted at different x-positions from the ϵ model computed with the stiffness tensor coefficients provided by Shell, and assuming an elliptical vertical transversely isotropic (VTI) medium. (a) $x = 214.8$ km, (b) $x = 217$ km. [**CR**]

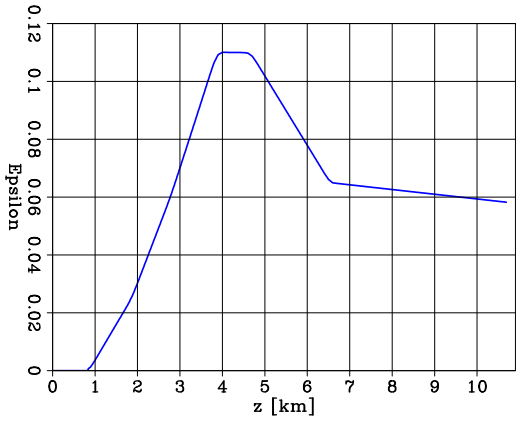
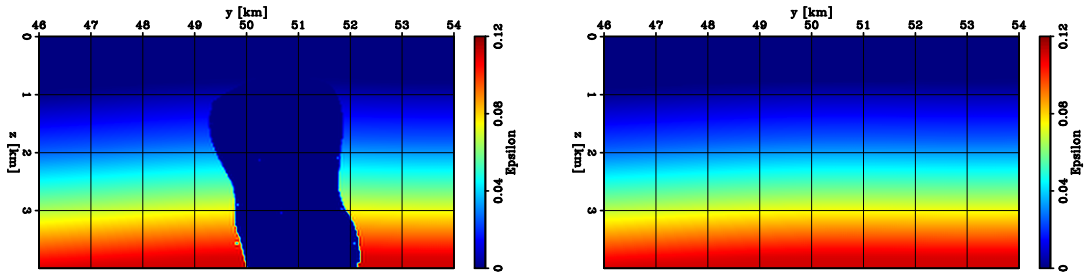


Figure 6.9: Depth profile of the ϵ-model extracted at $x = 214.8$ km and $y = 48$ km. [**CR**]



(a) (b)

Figure 6.10: Close-up of the analogous panels in Figure 6.8 focusing on the area of study. (a) $x = 214.8$ km, (b) $x = 217$ km. [**CR**]

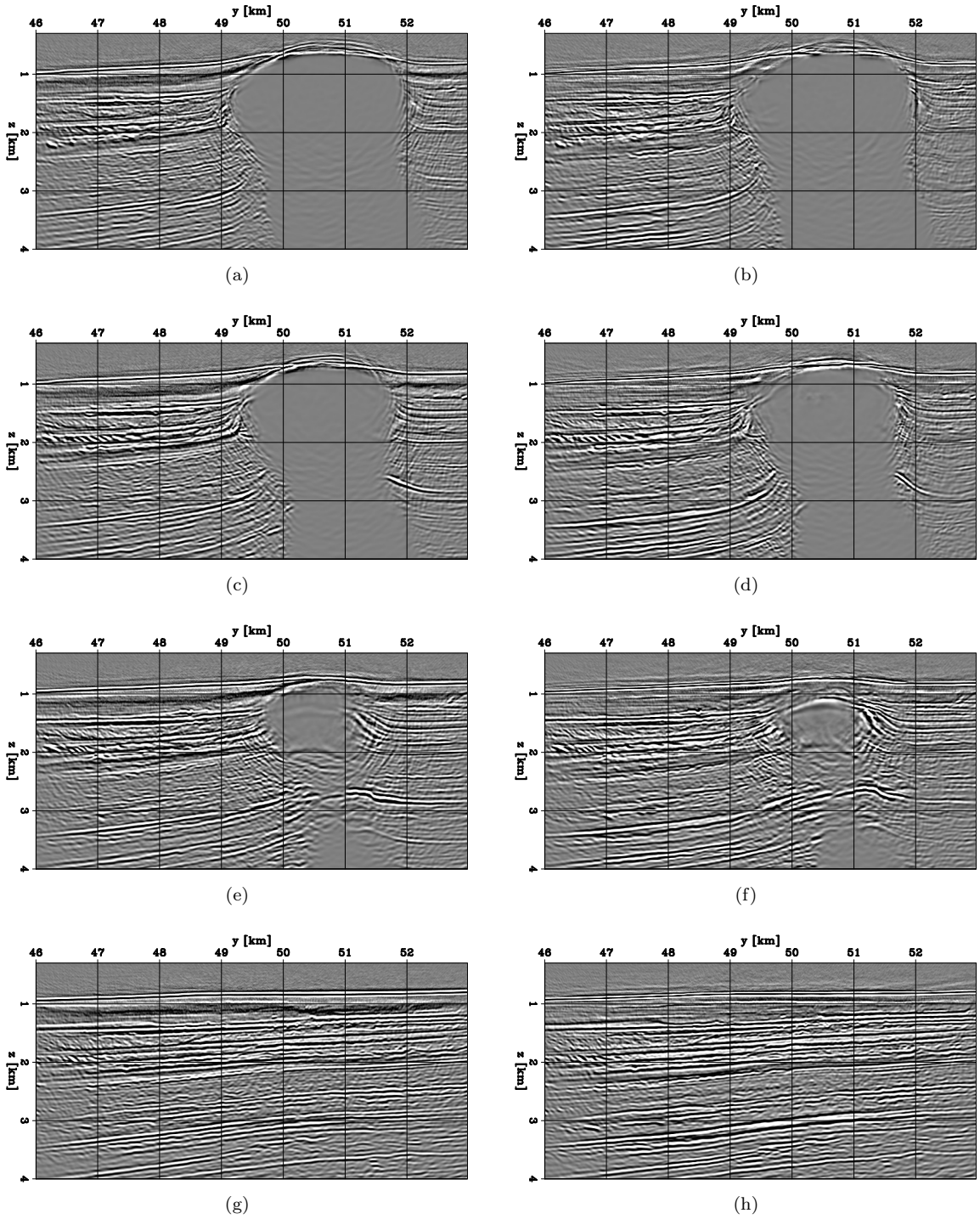


Figure 6.11: Cross-line sections extracted from 30 Hz RTM images at four different x-positions. The left panels correspond to the image migrated with the initial velocity model provided by Shell (model 1). The right panels correspond to the image migrated with the FWI inverted model obtained with model 1 as initial guess. The first, second, third and fourth rows correspond to $x = 214.8$ km, $x = 215.5$ km, $x = 216$ km, and $x = 217$ km, respectively. All panels are displayed with the same grayscale. [**CR**]

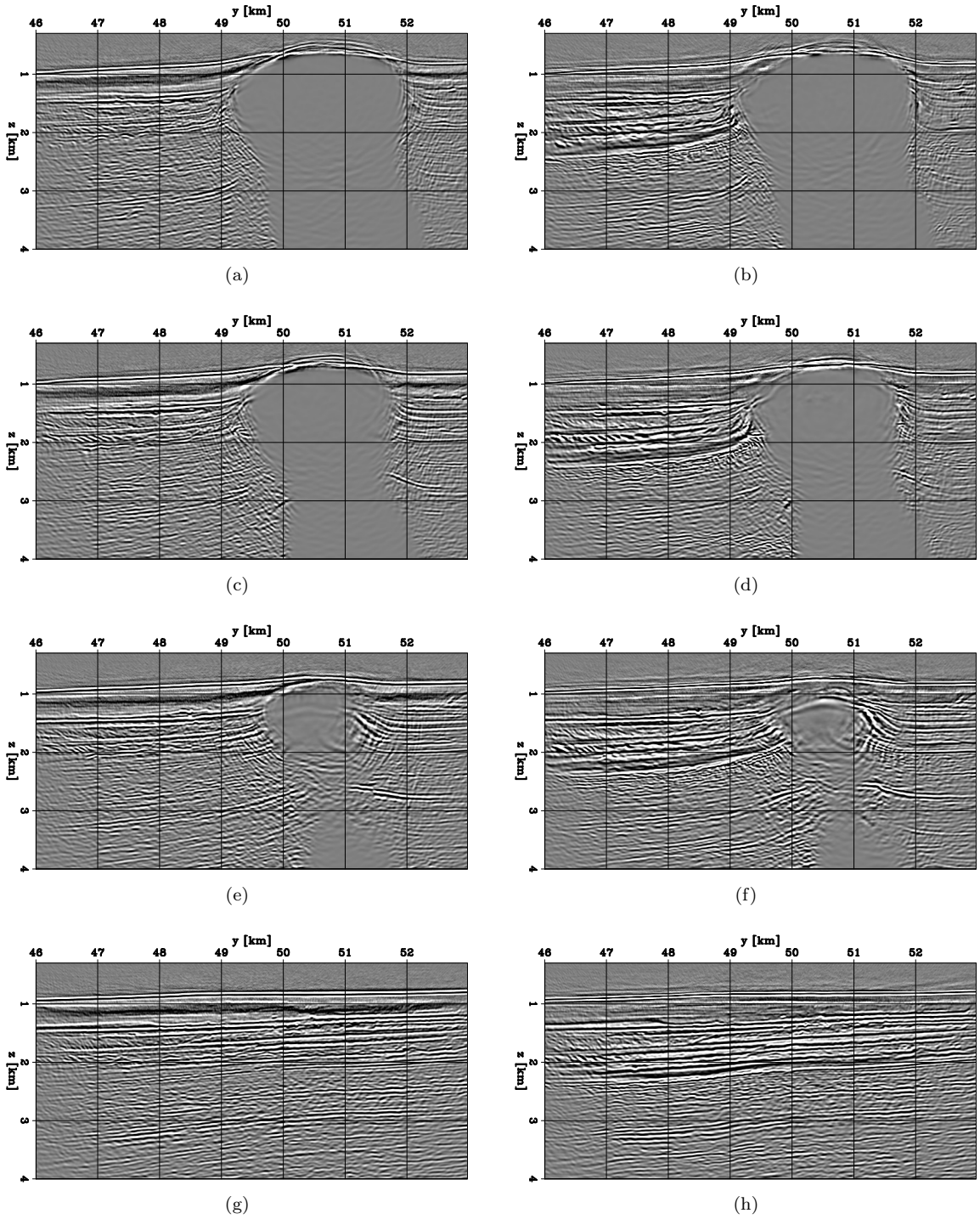


Figure 6.12: Cross-line sections extracted from 30 Hz RTM images at four different x-positions. The left panels correspond to the image migrated with initial velocity model 2. The right panels correspond to the image migrated with the FWI inverted model (using model 2 as initial guess). The first, second, third and fourth rows correspond to $x = 214.8$ km, $x = 215.5$ km, $x = 216$ km, and $x = 217$ km, respectively. All panels are displayed with the same grayscale. [**CR**]

optimal order of magnitude for the trade-off parameter ϵ is computed by selecting a value that approximately balances the amplitude of the two components from the FWIME objective function defined in equation 3.15 (Claerbout, 2014). To further refine the selection of ϵ, I compute $\tilde{\mathbf{p}}_\epsilon^{opt}$ by minimizing equation 3.18 for five ϵ-values ranging from $\epsilon_1 = 1 \times 10^{-6}$ to $\epsilon_5 = 1 \times 10^{-5}$ and with 80 iterations of linear conjugate gradient (using model 2 for the fixed background velocity model). Figure 6.13 shows the corresponding objective functions, which indicate that approximately 50 iterations of linear conjugate gradient are likely sufficient to minimize equation 3.18. Moreover, I observe (in a test not shown here) that the extended axis can be reduced to 31 points without affecting the convergence properties for $\tilde{\mathbf{p}}_\epsilon^{opt}$, which allows me to further decrease the computational cost of this study.

I examine the effect of the ϵ-value on $\tilde{\mathbf{p}}_\epsilon^{opt}$ by showing representative TLCIGs extracted at two locations from each of the five computed $\tilde{\mathbf{p}}_\epsilon^{opt}$ (Figure 6.14). The top and bottom panels correspond to TLCIGs at $M_1(x = 214.8$ km, y $= 49$ km) and $M_2(x = 216.5$ km, y $= 50$ km), respectively. In a similar behavior as for the synthetic examples previously shown in section 5.2, low ϵ-values allow for a considerable amount of energy to be mapped into the extended space and do not impose enough constraints on $\tilde{\mathbf{p}}_\epsilon^{opt}$ (first column in Figure 6.14). For high ϵ-values, most of the events mapped disappear and the information they carry is lost (last column in Figure 6.14), which may lead to cycle-skipping. By examining a larger subset of TLCIGs extracted over the area of study, I arbitrarily select $\epsilon = \epsilon_3 = 5 \times 10^{-6}$ for the FWIME inversion scheme.

The panels in the left column of Figures 6.15 and 6.16 show the initial search directions for conventional FWI (first row), for FWIME computed with $\epsilon = \epsilon_3$ (second row), and the target search direction (last row), all computed using the 3-12 Hz frequency band from the data. The target search direction is defined as the difference between the FWI inverted model using model 1 as initial guess (assumed to be a satisfactory solution), and model 2.

By examining Figures 6.15a and 6.16a, it is clear that conventional FWI promotes a velocity decrease in the region located between $z = 1.5$ km and $z = 2.5$ km, which may lead to the spurious low-velocity layer introduced throughout the FWI workflow and discussed in section 6.7. However, FWIME appears to be guiding the inversion towards a velocity increase in that same region, which seems more promising (Figures 6.15c and 6.16c). Finally, for the model-space multi-scale approach, I design an initial spline grid with a sampling of 70 m in the vertical direction, 200 m in both horizontal directions, and its corresponding spline mapping operator $\mathbf{S}_0$ (equation 4.12). Panels on the right column of Figures 6.15 and 6.16 are obtained by applying $\mathbf{S}_0\mathbf{S}_0^*$ to the panels on the left column. As expected, the application of the spline mapping removes some of the high-wavenumber artifacts and produces a much smoother and coherent update direction (Figures 6.15d and 6.16d).

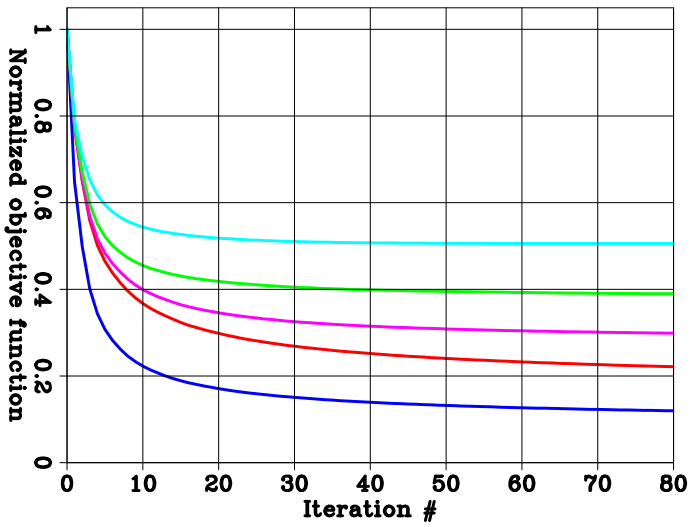


Figure 6.13: Normalized convergence curves corresponding to the minimization of equation 3.18 (initial variable projection step) with various ϵ-values: $\epsilon_1 = 1 \times 10^{-6}$ (dark-blue curve), $\epsilon_2 = 2.5 \times 10^{-6}$ (red curve), $\epsilon_3 = 5 \times 10^{-6}$ (pink curve), $\epsilon_4 = 7.5 \times 10^{-6}$ (green curve), and $\epsilon = 1 \times 10^{-5}$ (light-blue curve). All curves are normalized by the same value. [**CR**]

6.8.2 FWIME inversion results and next steps

I conduct 60 L-BFGS iterations of FWIME on the initial grid $\mathbf{S}_0$ and I observe a decrease of approximately 65% of the initial objective function value. Figure 6.17 shows the total (blue curve), data-fitting component (red curve), and annihilating component (pink curve) of the FWIME objective function.

Figures 6.18-6.21 show four cross-sections extracted from model 2 (panel a), the FWI model starting with model 2 (panel b), the FWIME model starting with model 2 (panel c), and the FWI model starting with model 1 (panel d). By comparing panels b with panels c, it is clear that the FWIME scheme does not introduce the spurious low-velocity layer present within the FWI model. For depths greater than 2.5 km, the FWIME inverted model shows a clear improvement compared to the FWI output starting with model 2, and seems to potentially converge to a similar solution as the accurate FWI model (Figure 6.18d). However, at this point, the inversion has not fully converged and more iterations should be conducted on the first spline grid to further improve my results and compare them with the accurate FWI.

Figures 6.22-6.24 show TLCIGs extracted from $\tilde{\mathbf{p}}_\epsilon^{opt}$ computed at the initial step (first row), and after 60 iterations of L-BFGS. Each column corresponds to a different y-position. As expected, a considerable amount of energy is initially present in the extended space (top row). At the final step, the velocity model is more accurate, so the energy focuses toward the physical plane and begins to vanish (bottom row).

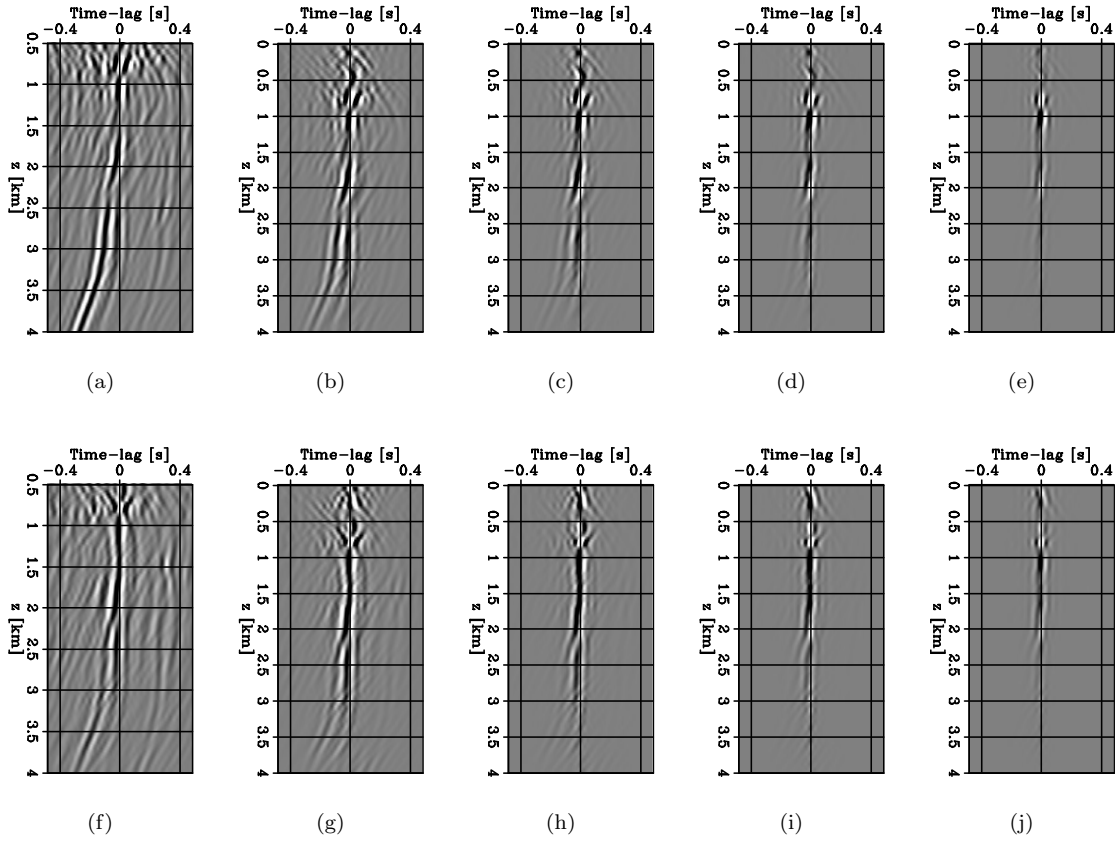


Figure 6.14: TLCIGs extracted at two locations from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed with model 2 and with five ϵ-values used for the inversions shown in Figure 6.13: $\epsilon_1 = 1 \times 10^{-6}$ (first column), $\epsilon_2 = 2.5 \times 10^{-6}$ (second column), $\epsilon_3 = 5 \times 10^{-6}$ (third column), $\epsilon_4 = 7.5 \times 10^{-6}$ (fourth column), and $\epsilon_5 = 1 \times 10^{-5}$ (last column). The top and bottom rows correspond to $M_1(x = 214.8$ km, y $= 49$ km) and $M_2(x = 216.5$ km, y $= 50$ km). All panels are displayed with the same grayscale. [**CR**]

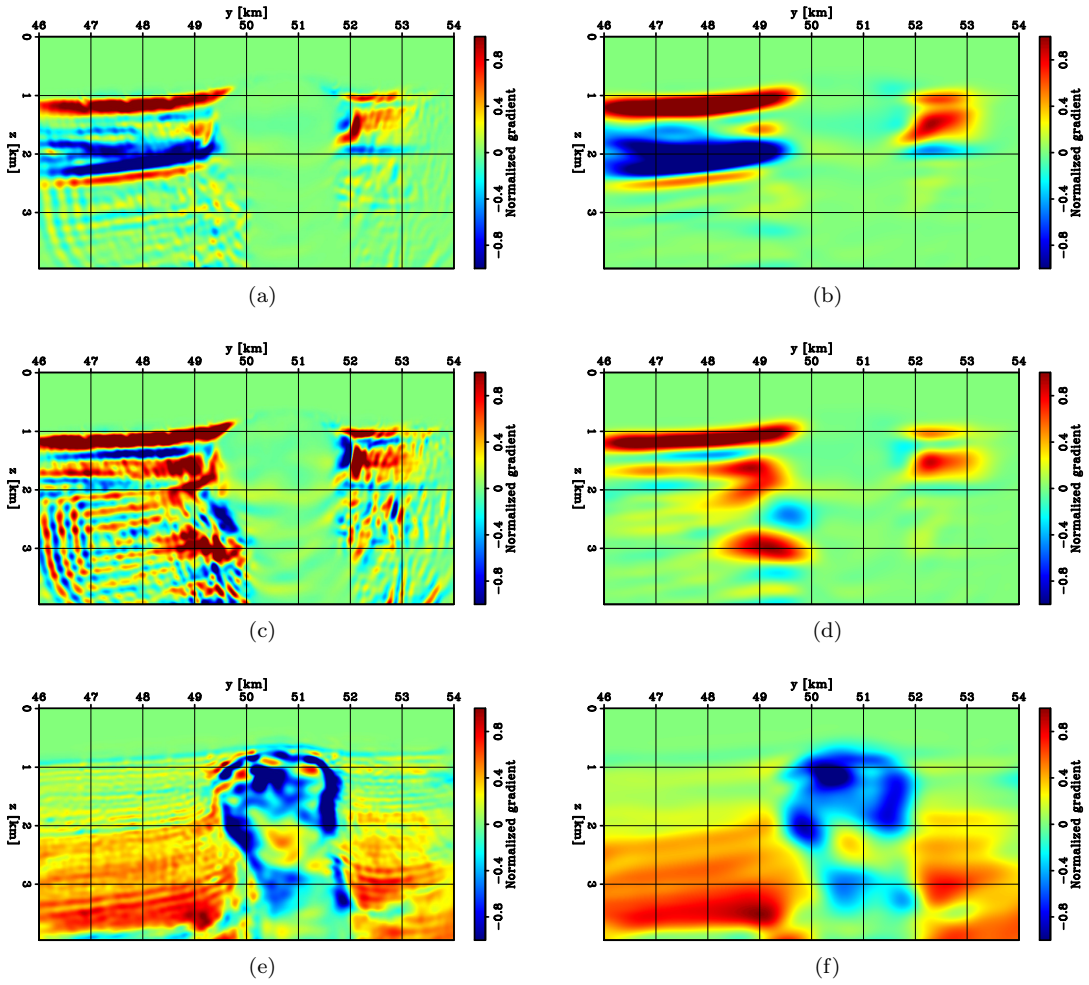


Figure 6.15: Cross-line sections showing normalized initial search directions, extracted at $x = 215.5$ km. Panels in the left column show the search direction before its mapping onto the initial spline grid. Panels in the right column are obtained by applying $\mathbf{S}_0\mathbf{S}_0^*$ to the panels in the left column ($\mathbf{S}_0$ is the initial spline mapping operator). The top row corresponds to the conventional FWI search direction. The third row is the FWIME search direction computed for $\epsilon_3 = 5 \times 10^{-6}$. The last row displays the target search direction, based on the FWI inverted model using model 1 as initial guess (section 6.6). [**CR**]

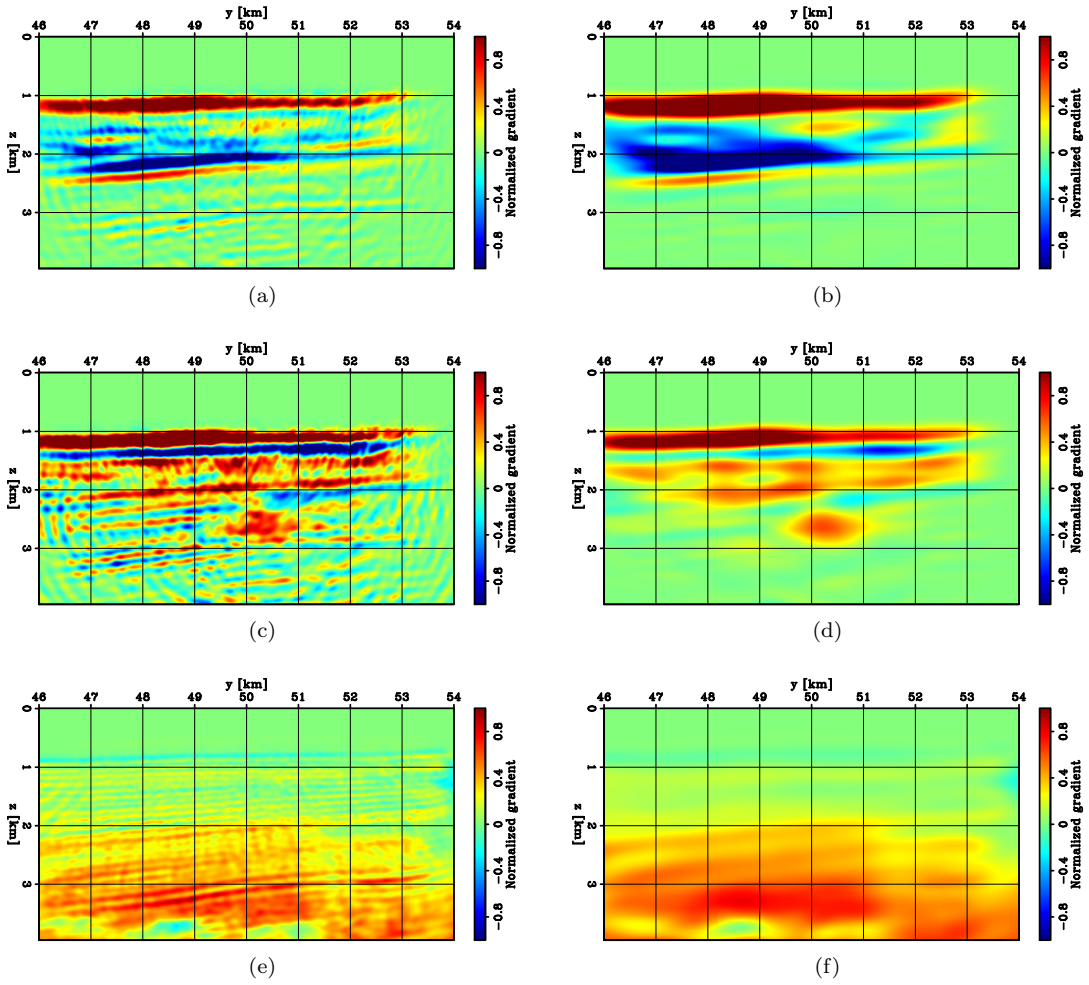


Figure 6.16: Cross-line sections showing normalized initial search directions, extracted at $x = 217$ km. Panels in the left column show the search direction before its mapping onto the initial spline grid. Panels in the right column are obtained by applying $\mathbf{S}_0\mathbf{S}_0^*$ to the panels in the left column ($\mathbf{S}_0$ is the initial spline mapping operator). The top row corresponds to the conventional FWI search direction. The third row is the FWIME search direction computed for $\epsilon_3 = 5 \times 10^{-6}$. The last row displays the target search direction, based on the FWI inverted model using model 1 as initial guess (section 6.6). [**CR**]

After full convergence of FWIME, I intend to use the inverted model as initial guess for a conventional data-space multi-scale workflow in order to efficiently retrieve a solution as accurate as the FWI starting with model 1. Then, I will QC the model quality with a 30 Hz RTM image.

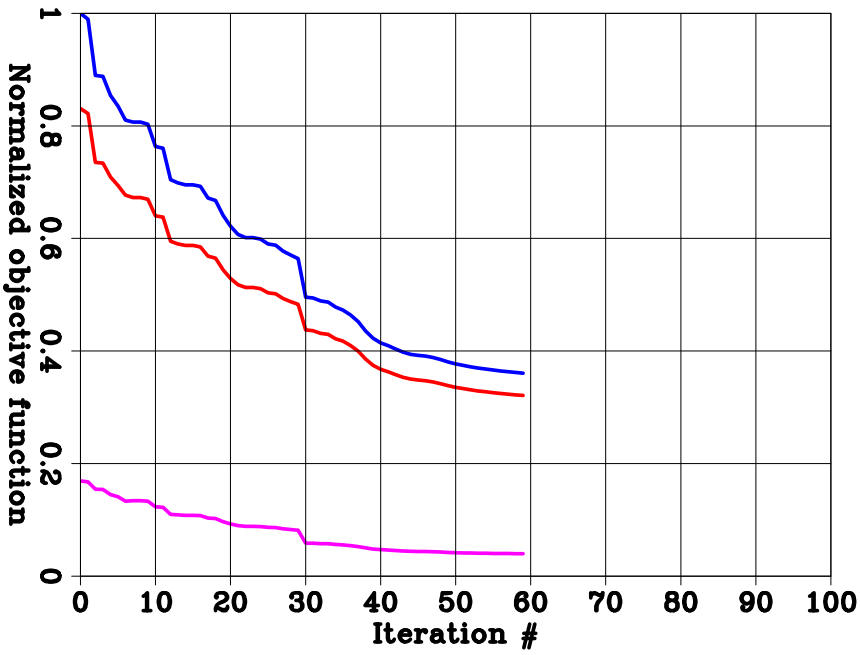


Figure 6.17: Components of the normalized FWIME objective function: total (blue curve), data-fitting component (red curve), and annihilating component (pink curve). [**CR**]

6.8.3 Numerical implementation and computational cost

Figure 6.25 summarizes the computational cost for 3D FWIME using 12×A100 Tesla NVIDIA GPU devices. The third column corresponds to the number of points on the extended axis, and the fifth column is the number of linear conjugate gradient iterations used for the variable projection step. The variable projection step remains FWIME's main bottleneck, and accounts for approximately 95% of the total computational time. For this specific application, I notice that setting the number of linear iterations to 30 produced similar results (in an analogous manner as the analysis from section 5.2), which reduces the computational time of one FWIME iteration from 15 hours to approximately 9 hours. In future work, the development of more quantitative methods to improve the convergence of the variable project step, such as the algorithm proposed by Hou and Symes (2017), must be investigated. Finally, Figure 6.26 summarizes FWIME's memory footprint on both the CPU and GPU devices.

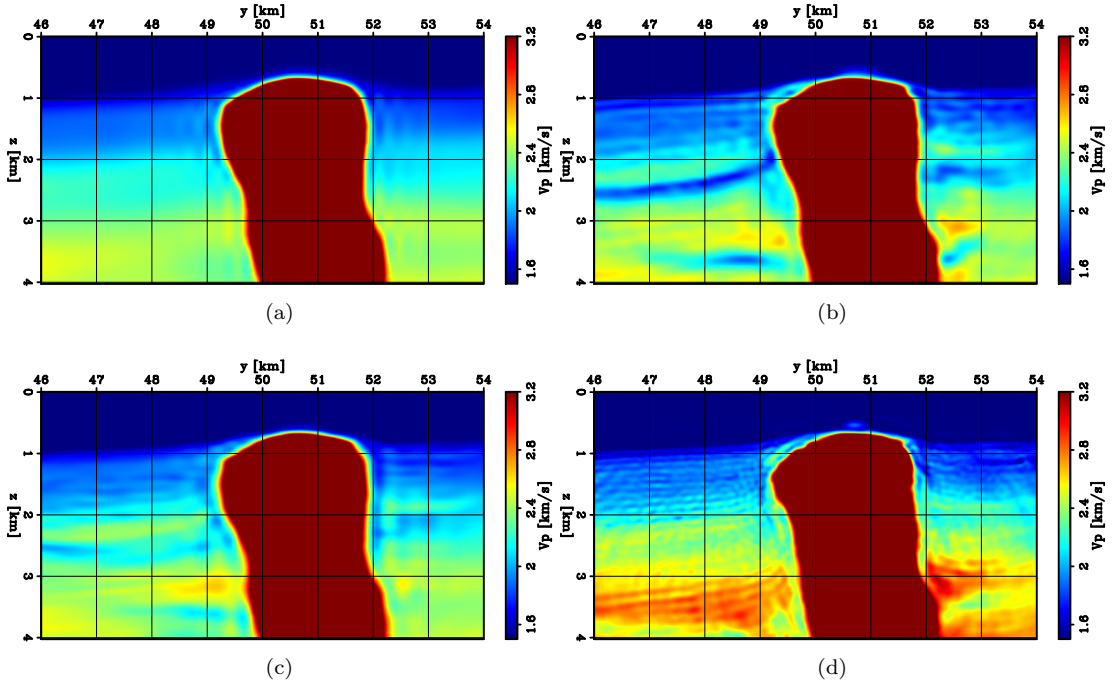


Figure 6.18: Cross-line sections extracted from various velocity models at $x = 214.8$ km. (a) Initial velocity model (model 2). (2) FWI inverted model using model 2 as initial guess and the 3-6 Hz frequency band from the data. (3) FWIME inverted model using model 2 as initial guess using the full 3-12 Hz frequency band from the data. (4) Conventional data-space multi-scale FWI inverted model using model 1 as initial guess using three bands spanning the 3-12 Hz frequency range. [**CR**]

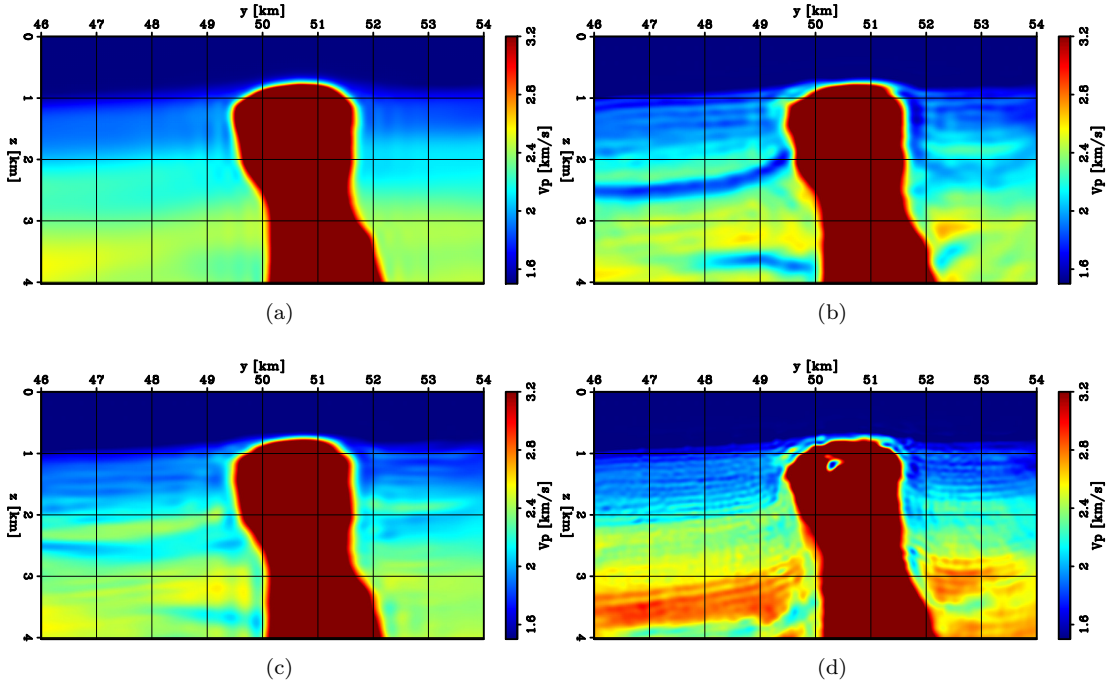


Figure 6.19: Cross-line sections extracted from various velocity models at $x = 215.5$ km. (a) Initial velocity model (model 2). (2) FWI inverted model using model 2 as initial guess and the 3-6 Hz frequency band from the data. (3) FWIME inverted model using model 2 as initial guess using the full 3-12 Hz frequency band from the data. (4) Conventional data-space multi-scale FWI inverted model using model 1 as initial guess using three bands spanning the 3-12 Hz frequency range. [**CR**]

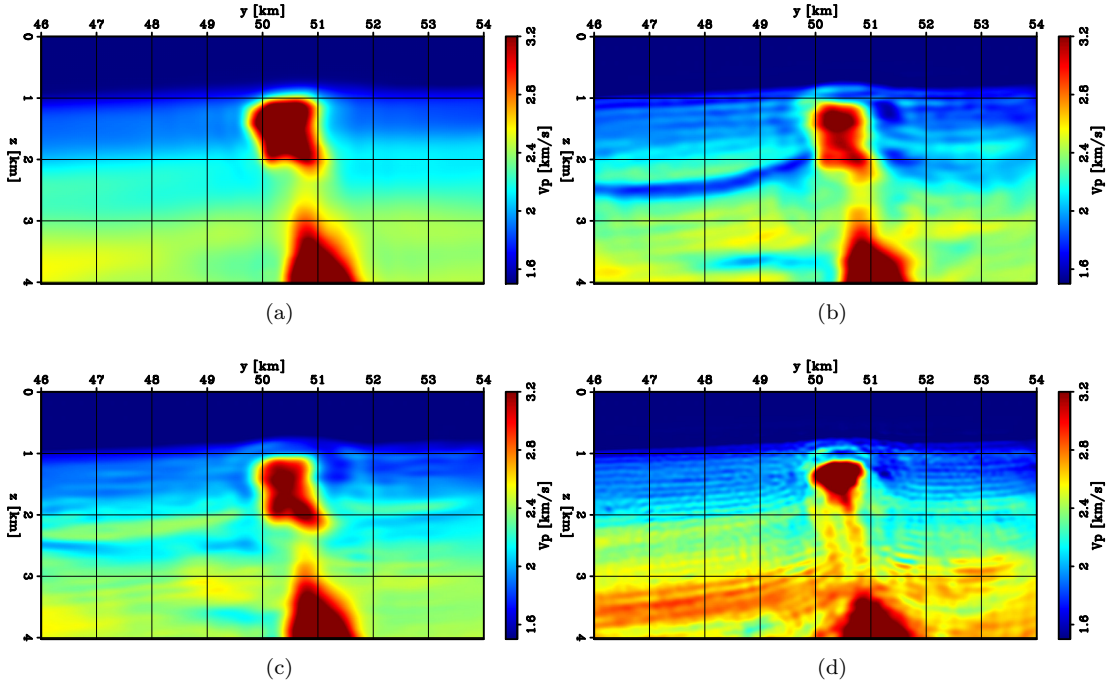


Figure 6.20: Cross-line sections extracted from various velocity models at $x = 216$ km. (a) Initial velocity model (model 2). (2) FWI inverted model using model 2 as initial guess and the 3-6 Hz frequency band from the data. (3) FWIME inverted model using model 2 as initial guess using the full 3-12 Hz frequency band from the data. (4) Conventional data-space multi-scale FWI inverted model using model 1 as initial guess using three bands spanning the 3-12 Hz frequency range. [**CR**]

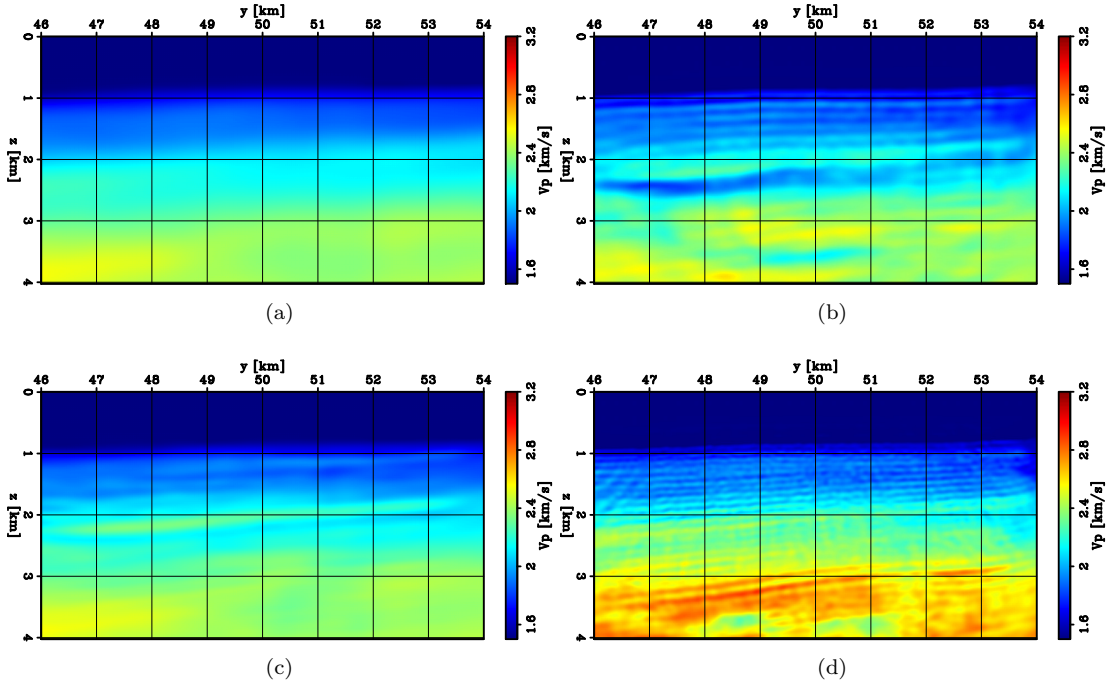


Figure 6.21: Cross-line sections extracted from various velocity models at $x = 217$ km. (a) Initial velocity model (model 2). (2) FWI inverted model using model 2 as initial guess and the 3-6 Hz frequency band from the data. (3) FWIME inverted model using model 2 as initial guess using the full 3-12 Hz frequency band from the data. (4) Conventional data-space multi-scale FWI inverted model using model 1 as initial guess using three bands spanning the 3-12 Hz frequency range. [**CR**]

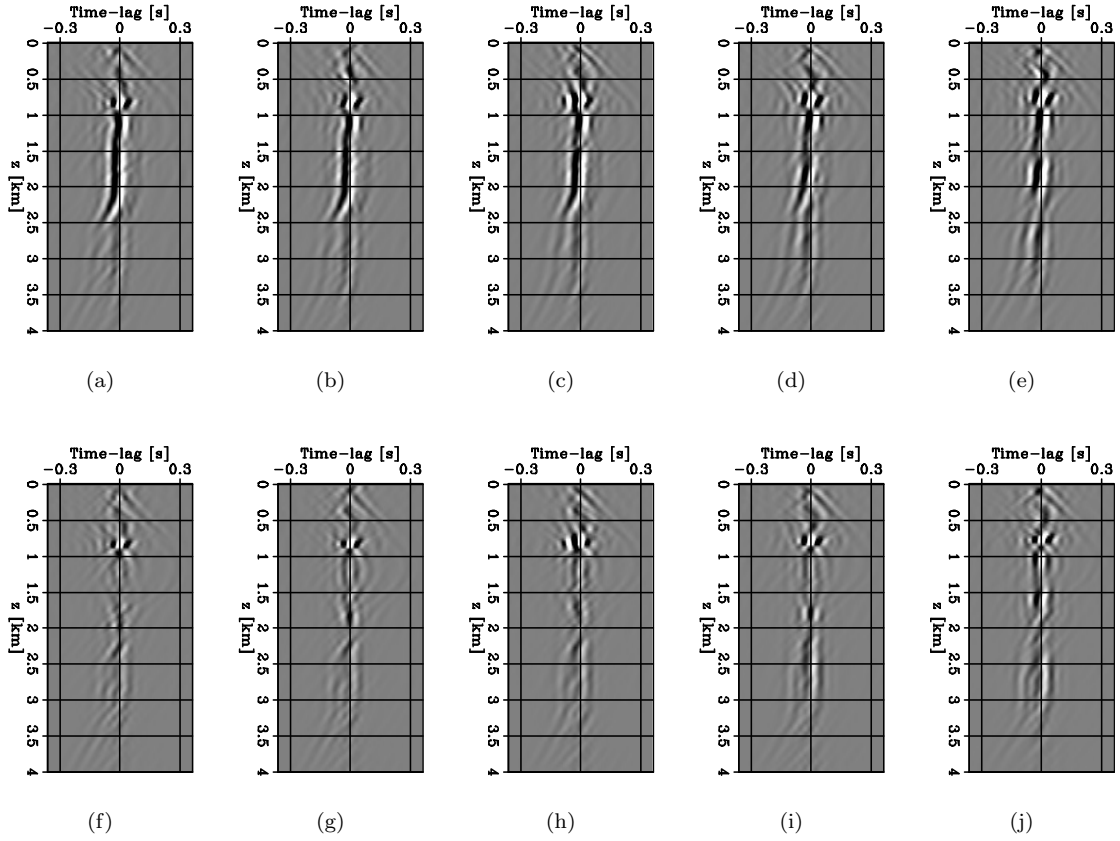


Figure 6.22: TLCIGs extracted at $x = 214.8$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed at the initial step (first row), and after 30 iterations of L-BFGS. Each column corresponds to a different y-position: $y = 48$ km (first column), $y = 48.25$ km (second column), $y = 48.5$ km (third column), $y = 48.75$ km (fourth column), and $y = 49$ km (fifth column). All panels are displayed with the same grayscale. [**CR**]

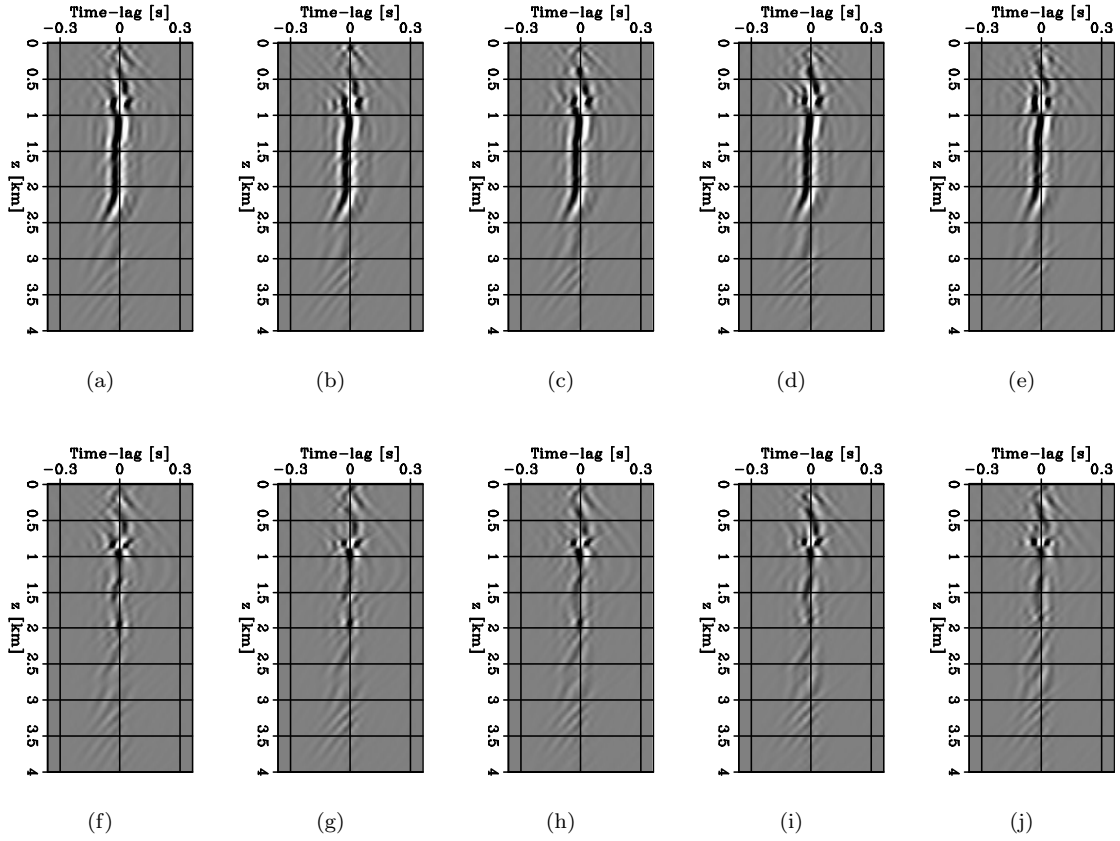


Figure 6.23: TLCIGs extracted at $x = 216.5$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed at the initial step (first row), and after 30 iterations of L-BFGS. Each column corresponds to a different y-position: $y = 48$ km (first column), $y = 48.25$ km (second column), $y = 48.5$ km (third column), $y = 48.75$ km (fourth column), and $y = 49$ km (fifth column). All panels are displayed with the same grayscale. [**CR**]

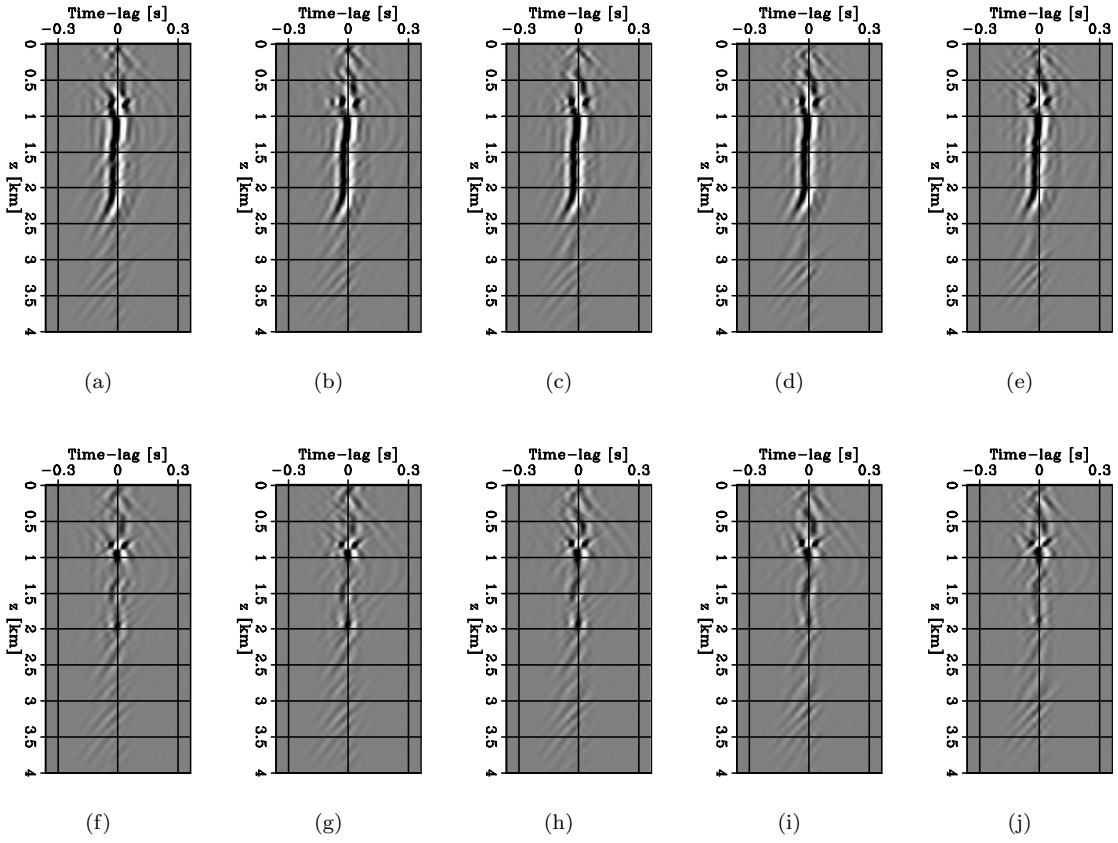


Figure 6.24: TLCIGs extracted at $x = 217$ km from $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ computed at the initial step (first row), and after 30 iterations of L-BFGS. Each column corresponds to a different y-position: $y = 48$ km (first column), $y = 48.25$ km (second column), $y = 48.5$ km (third column), $y = 48.75$ km (fourth column), and $y = 49$ km (fifth column). All panels are displayed with the same grayscale. [**CR**]

Technique	Model dim.	Ext.	# Shots	# VP iter.	Time/VP iter.	Time/iter.
FWI	400 x 400 x 170	N/A	255	N/A	N/A	11 min
FWIME	400 x 400 x 170	31	255	50	17 min	15 hours
FWIME	400 x 400 x 170	31	255	30	17 min	9 hours

Figure 6.25: 3D FWIME computational cost analysis conducted with 12×A100 Tesla NVIDIA GPU devices. [**NR**]

Parameter	Value
Model dimensions ($N_x \times N_y \times N_z$)	400 x 400 x 170
Extension type	Time lags
Extension dimension (N_{ext})	31
GPU memory usage	**10 GB**
# GPU per CPU (host)	4
CPU RAM pinned memory usage	**260 GB**

Figure 6.26: 3D FWIME CPU and GPU memory footprint. [**NR**]

Chapter 7

Conclusions

In this thesis, I present the theory of my new velocity-model building technique, namely FWIME, that successfully combines WEMVA with FWI, thereby leveraging the robust convergence properties of the former with the accuracy and high-resolution nature of the latter. The first novelty of my approach resides in the way I merge the two conventional algorithms, which makes FWIME more powerful and efficient than applying WEMVA and FWI sequentially. I devise a new cost function formulation where I modify the original FWI problem by adding a data-correcting term based on extended modeling to control the level of data fitting. I add an annihilating component to gradually penalize this data-correcting term and guide the inversion towards the optimal solution. The use of the variable projection method automatically handles the coupling between the two terms of my objective function, and reduces the number of adjustable hyper-parameters.

I create a new model-space multi-scale inversion workflow based on re-parametrizing the unknown velocity model on spline grids using B-spline basis functions. I simultaneously invert the full bandwidth of the observed data, and I control the resolution content of the model updates by adjusting the spline grid spatial sampling and its refinement rate. This new approach is crucial for FWIME.

I successfully apply FWIME on many realistic synthetic numerical tests that replicate some of the most challenging environments encountered when exploring for hydrocarbons: inaccurate initial model, lack of low-frequency data, lack of illumination, and presence of complex overburdens. I show that FWIME outperforms conventional approaches and possesses much more robust convergence properties than the FWI formulation.

I develop an industry-quality high-performance computing (HPC) numerical implementation using general-purpose graphics processing units (GPU). As a result, I make FWIME computationally tractable for 3D field applications. Then, I successfully apply my method to a 3D OBN dataset from the Gulf of Mexico, and I show that FWIME converges to more accurate solution than FWI.

Even though FWIME seems to be a powerful and promising technique, there are still many opportunities to improve its efficiency and accuracy. They include reducing the computational cost of the variable projection step, embedding more realistic physics into the seismic wave modeling operators (e.g., elasticity, anisotropy, and attenuation), and designing a more flexible model parametrization (e.g., radial basis functions) to better receover for complex overburdens, such as salt bodies. From a more theoretical standpoint, a formal mathematical proof showing that FWIME produces convex descent paths towards the optimal solution must be investigated.

Finally, it seems that computational cost will continue to dramatically decrease over time. Therefore, there is great value in developing new algorithms that optimize the need for human input, which is exactly what FWIME is trying to achieve. As the world is moving towards more sustainable forms of energy, FWIME will soon play a significant role not only for hydrocarbon exploration, but also for various applications such as geothermal energy, and CO_2 sequestration and monitoring.

Appendix A

FWIME and FWI share the same global minimum

We show that FWIME and FWI share the same global minimum under the following assumptions,

1. There exists a unique global minimum to the FWI objective function, $\mathbf{m}_t$
2. The observed data are acoustic and noise free, and $\mathbf{f}(\mathbf{m}_t) = \mathbf{d}^{obs}$
3. $\epsilon > 0$ and $\alpha > 0$, where α is defined in Figure 3.17.

First, the existence of a global minimizer is straightforward to verify because $\Phi_\epsilon(\mathbf{m}_t) = 0$. To prove its uniqueness, we assume there exists another $\mathbf{m}^*$ such that $\Phi_\epsilon(\mathbf{m}^*) = 0$ and $\mathbf{m}^* \neq \mathbf{m}_t$. Then, both components of the FWIME objective function (equation 3.5) must vanish:

$$\begin{cases} \mathbf{f}(\mathbf{m}^*) + \tilde{\mathbf{B}}(\mathbf{m}^*)\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}^*) - \mathbf{d}^{obs} = \mathbf{0}, \\ \mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}^*) = \mathbf{0}. \end{cases} \tag{A.1}$$

Moreover, since $\mathbf{D}$ is invertible, $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}^*) = \mathbf{0}$. Therefore, $\mathbf{f}(\mathbf{m}^*) - \mathbf{d}^{obs} = \mathbf{0}$, which contradicts the assumption of a unique global minimizer for the FWI objective function.

Appendix B

Convergence of FWIME towards conventional FWI

I show that the FWIME objective function converges pointwise in $\mathbf{m} \in \mathbb{R}^M$ towards the FWI objective function (equation 3.1) when $\epsilon \to +\infty$. To emphasize the fact that ϵ is now a variable, I re-write $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})$ and $\Phi_\epsilon(\mathbf{m})$ as

- $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) = \tilde{\mathbf{p}}^{opt}(\mathbf{m}, \epsilon)$, and
- $\Phi_\epsilon(\mathbf{m}) = \Phi(\mathbf{m}, \epsilon)$.

I conduct my proof in two steps:

1. I first prove that $\forall\, \mathbf{m},\ \lim_{\epsilon \to \infty} \frac{\epsilon^2}{2} \|\mathbf{D}\tilde{\mathbf{p}}^{opt}(\mathbf{m}, \epsilon)\|_2^2 = 0$, and
2. I conclude that $\forall\, \mathbf{m},\ \lim_{\epsilon \to \infty} \Phi(\mathbf{m}, \epsilon) = \Phi_{\text{FWI}}(\mathbf{m})$,

where Φ is the FWIME objective function. The optimal extended perturbation $\tilde{\mathbf{p}}^{opt}$ is given by

$$\begin{aligned} \tilde{\mathbf{p}}^{opt}(\mathbf{m}, \epsilon) &= \left[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2 \mathbf{D}^*\mathbf{D}\right]^{-1} \tilde{\mathbf{B}}^*(\mathbf{m}) \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right) \\ &= \frac{1}{\epsilon^2}\left[\frac{1}{\epsilon^2}\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \mathbf{D}^*\mathbf{D}\right]^{-1} \tilde{\mathbf{B}}^*(\mathbf{m}) \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right). \end{aligned} \tag{B.1}$$

When $\epsilon \to +\infty$, I can make the following approximation

$$\frac{1}{\epsilon^2}\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \mathbf{D}^*\mathbf{D} \approx \mathbf{D}^*\mathbf{D}. \tag{B.2}$$

Therefore,

$$\begin{aligned}\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon) &\approx \frac{1}{\epsilon^2}\left[\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs}-\mathbf{f}(\mathbf{m})\right) \\ &\approx \frac{1}{\epsilon^2}\tilde{\mathbf{q}}(\mathbf{m}),\end{aligned} \tag{B.3}$$

where $\tilde{\mathbf{q}}(\mathbf{m}) = \left[\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs}-\mathbf{f}(\mathbf{m})\right)$. Assuming that $\tilde{\mathbf{q}}(\mathbf{m})$ is bounded, I can deduce that for any $\mathbf{m}$,

- $\|\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$, and that
- $\frac{\epsilon^2}{2}\|\mathbf{D}\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$.

From equations 3.1 and 3.5, I can write

$$\begin{aligned}&\Phi(\mathbf{m},\epsilon)-\Phi_{\mathrm{FWI}}(\mathbf{m}) = \\ &\quad\frac{1}{2}\|\mathbf{f}(\mathbf{m})+\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)-\mathbf{d}^{obs}\|_2^2+\frac{\epsilon^2}{2}\|\mathbf{D}\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2-\frac{1}{2}\|\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\|_2^2.\end{aligned} \tag{B.4}$$

Moreover,

$$\begin{aligned}&\|\mathbf{f}(\mathbf{m})+\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{opt}(\mathbf{m})-\mathbf{d}^{obs}\|_2^2 = \\ &\quad\|\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\|_2^2+\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{opt}(\mathbf{m})\|_2^2+2\,\left(\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\right)^*\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon).\end{aligned} \tag{B.5}$$

$\|\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$ implies that $\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$, and by the Cauchy-Schwarz inequality,

$$\left|\left(\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\right)^*\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\right| \le \|\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\|_2\,\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2. \tag{B.6}$$

Therefore,

- $\left(\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\right)^*\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon) \underset{\epsilon\to\infty}{\to} 0$, and
- $\|\mathbf{f}(\mathbf{m})+\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)-\mathbf{d}^{obs}\|_2^2-\|\mathbf{f}(\mathbf{m})-\mathbf{d}^{obs}\|_2^2 \underset{\epsilon\to\infty}{\to} 0$.

Finally, we I deduce that

$$\forall\,\mathbf{m},\ \lim_{\epsilon\to\infty}\Phi(\mathbf{m},\epsilon)=\Phi_{FWI}(\mathbf{m}). \tag{B.7}$$

Appendix C

Derivation of the FWIME gradient

C.1 Conventional formulation

I derive the gradient of the objective function expressed in equation 4.1. First, I define

$$\mathbf{r}_d(\mathbf{m}) = \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs} \tag{C.1}$$

$$\mathbf{r}_{\tilde{p}}(\mathbf{m}) = \mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}), \tag{C.2}$$

where $\mathbf{r}_d \in \mathbb{R}^{N_d}$ is the FWIME data residual, and $\mathbf{r}_{\tilde{p}} \in \mathbb{R}^{N_{\tilde{p}}}$ is the argument of the annihilating component. Therefore, the FWIME objective function can be written as

$$\Phi_\epsilon(\mathbf{m}) = \frac{1}{2}\|\mathbf{r}_d(\mathbf{m})\|_2^2 + \frac{\epsilon^2}{2}\|\mathbf{r}_{\tilde{p}}(\mathbf{m})\|_2^2. \tag{C.3}$$

The gradient of Φ_ϵ is given by

$$\nabla\Phi_\epsilon(\mathbf{m}) = \left(\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{r}_d(\mathbf{m}) + \epsilon^2 \left(\frac{\partial \mathbf{r}_{\tilde{p}}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{r}_{\tilde{p}}(\mathbf{m}). \tag{C.4}$$

where $\nabla\Phi_\epsilon(\mathbf{m}) \in \mathbb{R}^{N_m}$. The Jacobian of $\mathbf{r}_d(\mathbf{m})$ is an operator mapping velocity perturbations into data perturbations,

$$\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}} : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_d}, \tag{C.5}$$

and its expression is given by

$$\begin{aligned}\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}} &= \frac{\partial \mathbf{f}(\mathbf{m})}{\partial \mathbf{m}} + \frac{\partial \left(\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\right)}{\partial \mathbf{m}} \\ &= \mathbf{B}(\mathbf{m}) + \frac{\partial \left(\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\right)}{\partial \mathbf{m}}\Big|_{\tilde{\mathbf{p}}_\epsilon^{opt}} + \tilde{\mathbf{B}}(\mathbf{m})\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}. \end{aligned} \tag{C.6}$$

The first term of the right side of equation C.6, is the conventional non-extended Born modeling operator, $\mathbf{B}(\mathbf{m}) : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_d}$. The second term of the right side of equation C.6 characterizes linear variations of the Born-modeled data (using an extended Born modeling operator) with variations in the velocity model $\mathbf{m}$, given a fixed extended perturbation $\mathbf{p}_\epsilon^{opt}$. Therefore,

$$\mathbf{T}(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}) = \frac{\partial \left(\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\right)}{\partial \mathbf{m}}\Big|_{\tilde{\mathbf{p}}_\epsilon^{opt}} \tag{C.7}$$

where $\mathbf{T}(\mathbf{m}, \tilde{\mathbf{p}}_\epsilon^{opt}) : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_d}$ is the data-space tomographic operator (equation 4.9). In the following, to simplify notations, I do not explicitly write the dependency of $\mathbf{T}$ with respect to $\tilde{\mathbf{p}}_\epsilon^{opt}$. The last term of right side of equation C.6 is the composition of the following two operators:

$$\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}} : \mathbb{R}^{N_m} \mapsto \mathbb{R}^{N_{\tilde{p}}} \tag{C.8}$$

$$\tilde{\mathbf{B}}(\mathbf{m}) : \mathbb{R}^{N_{\tilde{p}}} \mapsto \mathbb{R}^{N_d}. \tag{C.9}$$

Therefore, the Jacobian of the FWIME data residual is given by

$$\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{B}(\mathbf{m}) + \mathbf{T}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}, \tag{C.10}$$

and its adjoint is expressed by

$$\left(\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}}\right)^* = \mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})^*, \tag{C.11}$$

where $\left(\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}}\right)^* : \mathbb{R}^{N_d} \mapsto \mathbb{R}^{N_m}$. For the second term of right side of equation C.4, I have

$$\left(\frac{\partial \mathbf{r}_{\tilde{p}}(\mathbf{m})}{\partial \mathbf{m}}\right)^* = \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{D}^*, \tag{C.12}$$

where $\left(\frac{\partial \mathbf{r}_{\tilde{p}}(\mathbf{m})}{\partial \mathbf{m}}\right)^{*} : \mathbb{R}^{N_d} \mapsto \mathbb{R}^{N_m}$. Therefore, equation C.4 becomes

$$\begin{aligned}\nabla\Phi_{\epsilon}(\mathbf{m}) &= \left[\mathbf{B}^{*}(\mathbf{m}) + \mathbf{T}^{*}(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^{*} \tilde{\mathbf{B}}(\mathbf{m})^{*}\right]\mathbf{r}_d(\mathbf{m}) + \epsilon^2 \left(\frac{\partial \tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^{*} \mathbf{D}^{*}\mathbf{r}_{\tilde{p}}(\mathbf{m}) \\ &= \left[\mathbf{B}^{*}(\mathbf{m}) + \mathbf{T}^{*}(\mathbf{m})\right]\mathbf{r}_d(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^{*} \left[\tilde{\mathbf{B}}^{*}(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{r}_{\tilde{p}}(\mathbf{m})\right]. \quad \text{(C.13)}\end{aligned}$$

Now, I am going to show that

$$\tilde{\mathbf{B}}^{*}(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{r}_{\tilde{p}}(\mathbf{m}) = \mathbf{0} \quad \text{(C.14)}$$

Since $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ satisfies equation 3.8 (variable projection step), I can write

$$\left[\tilde{\mathbf{B}}^{*}(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{D}\right]\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) = \tilde{\mathbf{B}}^{*}(\mathbf{m})\left(\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m})\right). \quad \text{(C.15)}$$

Therefore,

$$\begin{aligned}\tilde{\mathbf{B}}^{*}(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{r}_{\tilde{p}}(\mathbf{m}) &= \tilde{\mathbf{B}}^{*}(\mathbf{m})\left(\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}}\right) + \epsilon^2 \mathbf{D}^{*}\mathbf{D}\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) \\ &= \left[\tilde{\mathbf{B}}^{*}(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{D}\right]\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) - \tilde{\mathbf{B}}^{*}(\mathbf{m})(\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m})) \\ &= \mathbf{0}. \quad \text{(C.16)}\end{aligned}$$

The result from Equation C.14 is very useful because it allows me to avoid computing the following term:

$$\left(\frac{\partial \tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^{*} \left[\tilde{\mathbf{B}}^{*}(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2 \mathbf{D}^{*}\mathbf{r}_{\tilde{p}}(\mathbf{m})\right]. \quad \text{(C.17)}$$

Notice that Equation C.14 is true provided that equation 3.8 is satisfied, which highlights the importance to conduct enough linear conjugate gradient iterations during the FWIME variable projection step. Finally, equation C.13 reduces to

$$\begin{aligned}\nabla\Phi(\mathbf{m}) &= \left[\mathbf{B}^{*}(\mathbf{m}) + \mathbf{T}^{*}(\mathbf{m})\right]\left(\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}}\right) \\ &= \left[\mathbf{B}^{*}(\mathbf{m}) + \mathbf{T}^{*}(\mathbf{m})\right]\mathbf{r}_d(\mathbf{m}). \quad \text{(C.18)}\end{aligned}$$

C.2 Phase-only formulation

For the phase-only formulation, the FWIME data residuals are modified modified as follows,

$$\mathbf{r}_d^n(\mathbf{m}) \quad = \quad \mathbf{n}(\mathbf{f}(\mathbf{m})) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{n}(\mathbf{d}^{obs}). \tag{C.19}$$

Therefore, the Jacobian of $\mathbf{r}_d^n$ is given by

$$\frac{\partial \mathbf{r}_d^n(\mathbf{m})}{\partial \mathbf{m}} \quad = \quad \mathbf{N}(\mathbf{f}(\mathbf{m}))\mathbf{B}(\mathbf{m}) + \mathbf{T}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}, \tag{C.20}$$

where $\mathbf{N} : \mathbb{R}^{N_d} \mapsto \mathbb{R}^{N_d}$ is the Jacobian of the trace-normalization operator $\mathbf{n}$ (derived in Appendix C.3). Hence,

$$\left(\frac{\partial \mathbf{r}_d^n(\mathbf{m})}{\partial \mathbf{m}}\right)^* \quad = \quad \mathbf{B}^*(\mathbf{m})\mathbf{N}^*(\mathbf{f}(\mathbf{m})) + \mathbf{T}^*(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})^*. \tag{C.21}$$

In an analogous derivation as the one conducted in equation C.16, I can express the gradient of the phase-only FWIME objective function,

$$\nabla \Phi_\epsilon^n(\mathbf{m}) \quad = \quad \Big[\mathbf{B}^*(\mathbf{m})\mathbf{N}^*(\mathbf{f}(\mathbf{m})) + \mathbf{T}^*(\mathbf{m})\Big]\Big(\mathbf{n}(\mathbf{f}(\mathbf{m})) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{n}(\mathbf{d}^{\mathbf{obs}})\Big) \tag{C.22}$$

$$= \quad \Big[\mathbf{B}^*(\mathbf{m})\mathbf{N}^*(\mathbf{f}(\mathbf{m})) + \mathbf{T}^*(\mathbf{m})\Big]\mathbf{r}_d^n(\mathbf{m}). \tag{C.23}$$

C.3 Linearization of the trace-normalization operator

I derive the application of the Jacobian of the trace-normalization operator $\mathbf{n}$. The analysis is conducted for a single trace $\mathbf{x} \in \mathbb{R}^{N_d}$ (i.e., the data corresponding to as single source/receiver pair). Recall the definition of $\mathbf{n}$ from equation 3.16,

$$\begin{aligned} \mathbf{n} : \mathbb{R}^{N_d} &\mapsto \mathbb{R}^{N_d} \\ \mathbf{x} &\mapsto \frac{\mathbf{x}}{\|\mathbf{x}\|_2 + c}. \end{aligned} \tag{C.24}$$

The application of the Jacobian matrix of $\mathbf{n}$ evaluated about vector $\mathbf{x}_0$ on an input $\mathbf{x}$ is given by

$$\begin{aligned} \mathbf{N}(\mathbf{x}_0) : \mathbb{R}^{N_d} &\mapsto \mathbb{R}^{N_d} \\ \mathbf{x} &\mapsto \left.\frac{\partial \mathbf{n}(\mathbf{u})}{\partial \mathbf{u}}\right|_{\mathbf{u}=\mathbf{x}_0} \mathbf{x}. \end{aligned} \tag{C.25}$$

After differentiation of equation 3.16, the (i, j) entry of $\mathbf{N}(\mathbf{x}_0)$ is given by

$$\big(\mathbf{N}(\mathbf{x}_0)\big)_{i,j} = \frac{\delta_{ij}}{\|\mathbf{x}_0\|_2 + c} - \frac{(\mathbf{x}_0)_i \, (\mathbf{x}_0)_j}{(\|\mathbf{x}_0\|_2 + c)^2 \|\mathbf{x}_0\|_2}, \tag{C.26}$$

where $(\mathbf{x}_0)_i$ is the i-th component of $\mathbf{x}_0$. $\mathbf{N}(\mathbf{x_0})$ is a real symmetric matrix and its corresponding linear mapping is self-adjoint. The application of $\mathbf{N}(\mathbf{x_0})$ on a vector $\mathbf{x} \in \mathbb{R}^{N_d}$ is given by

$$\mathbf{y} = \mathbf{N}(\mathbf{x}_0)\mathbf{x}, \tag{C.27}$$

where

$$(\mathbf{y})_i = \frac{(\mathbf{x})_i}{\|\mathbf{x}_0\|_2 + c} - \frac{(\mathbf{x}_0)_i}{(\|\mathbf{x}_0\|_2 + c)^2 \|\mathbf{x}_0\|_2} \, \mathbf{x}_0^T \mathbf{x}. \tag{C.28}$$

Appendix D

High-performance computing implementation for 3D applications

I describe my FWIME high-performance computing (HPC) implementation using general-purpose graphics processing units (GPU) for 3D field applications. Many authors have previously proposed useful discussions and tutorials on how to efficiently program seismic imaging algorithms on GPUs by taking advantage of their unique hardware and software features (Micikevicius, 2009; Leader, 2015; Le, 2019). Therefore, in this Appendix, my goal is to provide an exhaustive list of parameters and explain my software-design choices for the 3D implementation (for reproducibility purposes), which is specifically tailored to the 3D OBN dataset application (Chapter 6).

D.1 Software solution

Due to its algorithmic complexity and high computational cost for 3D field-data applications, FWIME must be efficiently executed on a computer (and computer clusters). Like many seismic imaging algorithms based on finite-difference modeling, FWIME is a highly scalable and parallelizable algorithm which makes it eligible to benefit from an optimized GPU implementation (Leader, 2015). In this thesis, I leverage the aforementioned property and I design a fully-fledged software solution with the use of GPUs.

Figure D.1 shows an abstract schematic representation of my proposed software architecture, which is composed of three layers. (1) The first layer handles FWIME's most computationally intensive and parallelizable tasks (e.g., wavefield propagations, extended imaging/scattering operations, etc.), which are performed directly on the GPU card (referred to as the device) and are programmed in the parallel computing platform referred to as compute unified device architecture (CUDA). (2) The collection of CUDA functions (referred to as kernels) are then wrapped into a

C++ framework which allows me to dispatch the various modeling tasks on multiple GPU cards available within a compute node (referred to as the host). (3) A Python-based abstraction layer is then added to easily build, define, and solve large-scale complex inverse problems on multi-node GPU clusters with the use of the objected-oriented optimization framework proposed by Biondi et al. (2021b). Finally, the conversion between C++ and Python user-defined classes is achieved with the pybind11 library (Jakob et al., 2017).

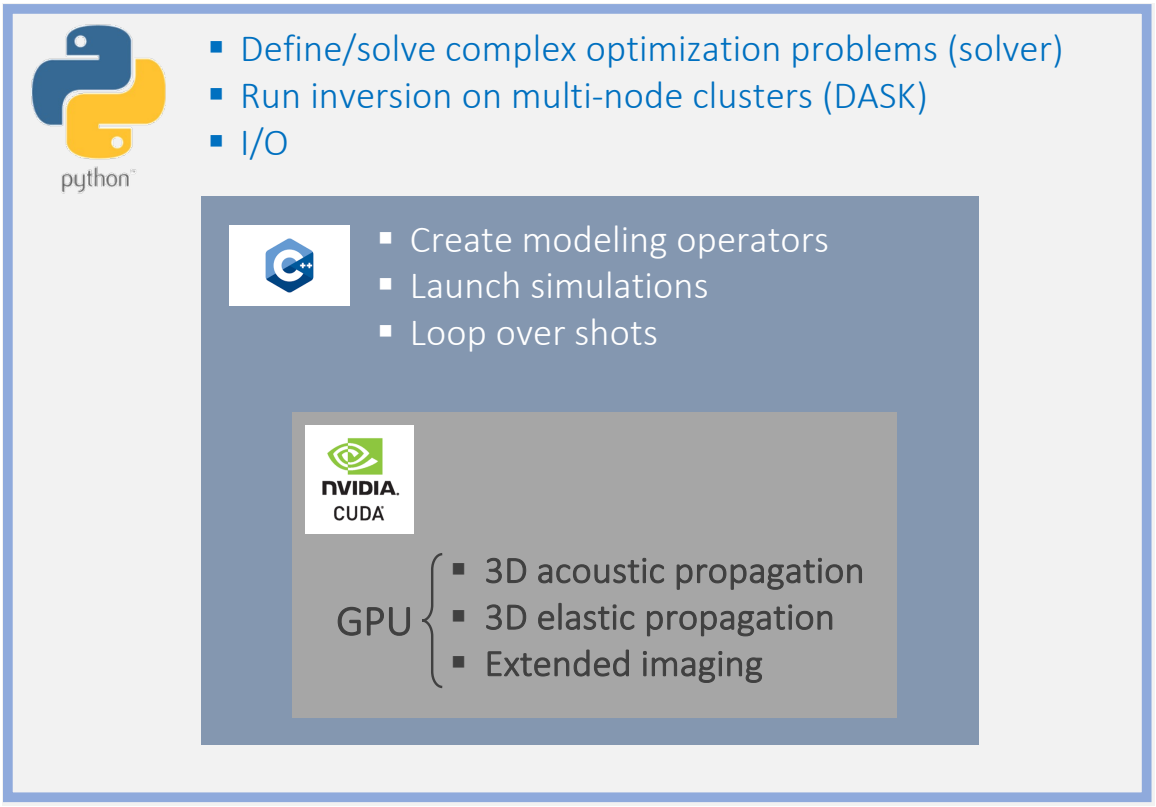


Figure D.1: Schematic representation of the three abstract layers composing the software solution developed for 3D FWIME applications. (1) The first layer consists of a set of CUDA kernels performing wave propagations, imaging and scattering operations on the GPU card. (2) A C++ wrapper is then added to dispatch the simulations on multiple cards within one compute node. (3) In the third layer, a Python-based interface is implemented to build, define, and solve large-scale inverse problems on multi-node GPU clusters (Biondi et al., 2021b). [**NR**]

D.2 3D finite-difference GPU kernel

I solve the scalar acoustic isotropic constant-density wave equation (equation 2.1) with a numerical time-domain finite-difference scheme (second-order in time, 8^{th}-order in space), and I implement the

time-stepping kernel in a similar fashion as the one proposed in Micikevicius (2009): the (z, x)-plane (where z and x correspond to the fastest and second fastest axes for memory access, respectively) is divided into a 2D grid of 2D thread blocks. Each block advances through the y-direction (the slowest axis) to process spatial derivatives in the (z, x)-plane while the spatial derivative along the y-direction is handled by an array of registers local to each thread (Le, 2019).

For the application to the Cardamom 3D OBN dataset, I obtain the best performance by using a 2D thread block size of 32 in both z- and x-directions, which requires the execution of 1024 threads per block. Moreover, my implementation is based on the assumption that the computational volume (i.e., the velocity model) is able to fit in the global memory of a single GPU card, thus no domain decomposition is needed (Leader, 2015). I use a free-surface boundary condition to simulate the water-air interface (Robertsson, 1996), and an absorbing boundary condition in the other directions.

I benchmark the performance of my finite-difference propagation code with a single NVIDIA Tesla V100 GPU card. Figure D.2 shows the processing rate r_{proc} of my finite-difference kernel as a function of model size, which is obtained by propagating one single wavefield in the computational volume and by varying the volume size. More specifically, r_{proc} is defined as follows,

$$r_{proc} = \frac{N_m \times N_t}{\Delta t_{prop}}, \tag{D.1}$$

where $N_m = N_x \times N_y \times N_z$ is the number of voxels in the computational volume, $N_t = 2400$ is the number of time steps for the propagation (which is kept fixed), and Δt_{prop} corresponds to the propagation time taken by the GPU to complete the simulation. Hence, the processing rate r_{proc} represents the number of cells per unit time processed by the finite-difference propagator. The blue and orange curves in Figure D.2 correspond to the double- and single-precision floating-point format implementations, respectively. In both cases, the performance stabilizes for computational volumes greater than 150 Bn voxels.

Even though all field-data applications are performed in single precision, I develop a double-precision version of my GPU code solely for quality-control purposes. I test the numerical implementation accuracy of the various Jacobian operators (and their adjoints) employed in FWIME and defined in Chapter 2 with the so-called dot-product test (Claerbout, 1985). In this thesis, all operators satisfy the dot-product test up to a relative error of $\epsilon^d_{rel} = 10^{-15}$ and $\epsilon^s_{rel} = 10^{-6}$ for the double- and single-precision implementations, respectively.

D.3 Compute resources

The results shown in Chapter 6 for the application of FWIME to the Cardamom 3D OBN dataset are conducted on two different compute systems,

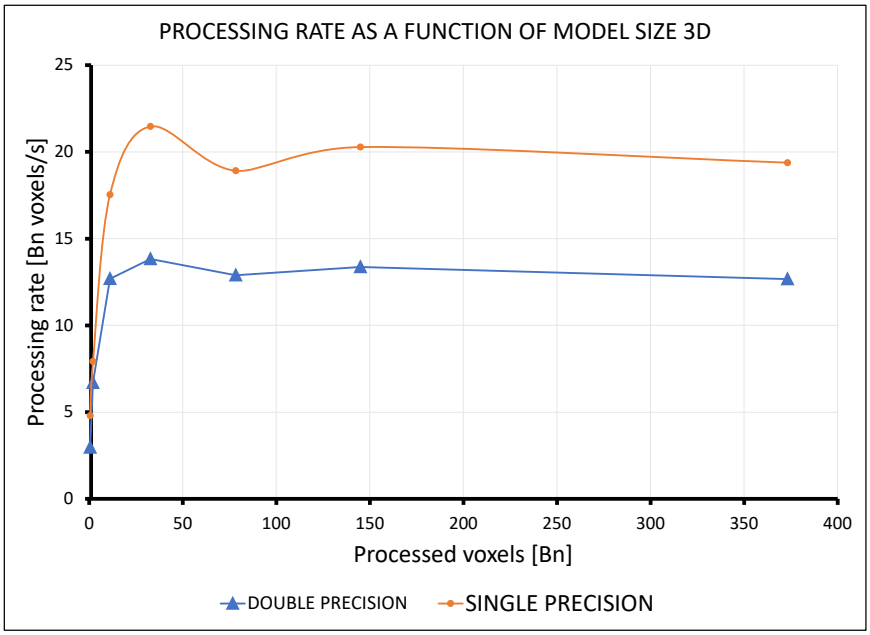


Figure D.2: Processing rate r_{proc} as a function of model size for the 3D finite-difference scheme. [**CR**]

- The Stanford Exploration Project (SEP) GPU cluster, which includes three compute nodes, each containing 4 NVIDIA V100 GPU cards (one with 32 GB of global memory, two with 16 GB), and each with 24 CPU cores with 500 GB of CPU RAM.
- The Stanford University Sherlock GPU cluster, on which I can access seven compute nodes. Each node contains two NVIDIA A100 GPU cards, each with 40 GB of global memory and 24 cores with 500 GB of CPU RAM.

The properties of each card are summarized in Table D.1.

D.4 3D waveform inversion modeling parameters

I describe the modeling parameters employed for both 3D waveform inversions presented in this thesis, namely FWI and FWIME. For both schemes, the maximum frequency present within the data is set to 12 Hz.

D.4.1 Nonlinear modeling operator

The computational volume (i.e., velocity model) for the Cardamom dataset is composed of $N_m = 400 \times 400 \times 170$ voxels in the x-, y-, and z-directions, respectively, giving a total size of 0.1 GB.

GPU specifications	**V100 (SEP)**	**A100 (Sherlock)**
Architecture	Volta	Ampere
Global memory size	16/32 GB	40 GB
GPU memory bandwidth	900 GB/s	1,555 GB/s
L2 cache size / card	6 MB	40 MB
Number of streaming multiprocessors (SM)	80	108
L1 + shared memory / SM	128 KB	192 KB
Number of 32-bit registers / SM	65,536	65,536
Constant memory size / card	64 KB	64 KB
Maximum number of registers / thread	255	255
Maximum number of threads / block	1024	1024
Maximum number of concurrent warps / SM	64	64
Maximum number of concurrent blocks / SM	32	32
Warp size	32 threads	32 threads
GPU/CPU bandwidth (using streams)	12.5 GB/s	25 GB/s

Table D.1: Summary of the specifications for the two types of NVIDIA's GPU cards employed for the 3D FWIME application to the Cardamom OBN dataset.

The spatial sampling in uniform and set to 35 m (Tables D.2 and D.3). Each shot is simulated on a single GPU card, and is propagated with the finite-difference scheme for $N_{t_{CFL}} = 2600$ time steps, sampled at $\Delta t_{CFL} = 3$ ms in order to satisfy the Courant Friedrichs Lewy (CFL) stability condition. For one shot, the associated wavefield stored and used for the (extended) Born and tomographic operators (as described in sections 2.2, 2.3 and 2.4, respectively) is a 4D hypercube with total of $N_w = N_m \times N_{t_s}$ elements and where $N_{t_s} = 330$, giving a total size of approximately 32 GB (first row of Table D.4). To optimize memory usage, the sampling rate at which the wavefield and the seismic data are stored is set to $\Delta t_s = 24$ ms.

D.4.2 Born operator

The conventional Born forward and adjoint modeling operations are both conducted in two steps, as explained in section 2.4.4. First, a source wavefield is computed and then stored on the host's RAM. By using the CUDA stream feature which allows asynchronous data transfers from the device to the host (and vice-versa), the computation (on the device) of the source wavefield time slice at time t_n is overlapped with the transfer (to the host) of the time slice at time t_{n-1}, thereby hiding the latency associated with this communication process. In the second step, the source wavefield must be gradually transferred back to the device to generate a scattered wavefield (Born forward) or an image (Born adjoint). Both scattering and imaging conditions are performed directly on the

Parameter	Value
N_x	400 samples
N_y	400 samples
N_z	170 samples
$\Delta x/\Delta y/\Delta z$	35 m
$N_{t_{CFL}}$	2600 samples
Δt_{CFL}	3 ms
N_{t_s}	330 samples
Δt_s	24 ms

Table D.2: Summary of the spatial and time modeling parameters employed for the 3D FWIME application to the Cardamom OBN dataset.

Model extension	GPU memory usage
No extension	0.1 GB
Time ($N_\tau = 31$)	10 GB
Offset ($N_{h_x} = N_{h_x} = 20$)	43 GB

Table D.3: Summary of FWIME's GPU memory footprint for the different extension types.

Method	CPU (pinned) RAM usage
Single wavefield	32.5 GB
FWIME (1 GPU/node)	65 GB
FWIME (4 GPUs/node)	260 GB

Table D.4: Summary of FWIME's CPU memory footprint.

GPU. In this step, an analogous approach can also be used with the stream feature to overlap data transfer and computations on the device.

For NVIDIA's GPU hardware (programmed with CUDA), the ability to overlap host-device data transfer and computations requires the use of the page-locked (or "pinned") memory on the host's RAM. Pinned memory is physical RAM that is set aside and not allowed to be paged out by the operating system (OS). At a particular time, the maximum pinnable memory is determined by the amount of system memory that other processes are competing for (Jasjuang, 2021). After conducting numerical tests, I observe that keeping approximately 50 GB of free RAM on the host (while allocating the remaining amount as pinned) is sufficient to prevent any loss of performance from the OS. Using a node containing 4 GPU cards, the total memory footprint for the source wavefields' CPU storage in FWIME is thus $4 \times 32.5 = 130$ GB (second row of Table D.4), which allows me to safely store all the computed auxiliary wavefields on the host's pinned memory.

D.4.3 Extended Born and tomographic operators

In this thesis, I implement the extended Born forward and adjoint operators for both time-lag and horizontal-subsurface offset extensions (but both extensions can not be simultaneously employed). In an analogous fashion as for the non-extended case, efficient overlap of source wavefield computation and transfer is implemented. Both extended scattering and imaging conditions are also performed directly on the device and extended models are fully stored on the device, which can potentially considerably increase the GPU memory footprint of the program.

For the Cardamom application, I use a time-lag extension with a maximum value of $N_\tau = 31$ points, giving an extended-model size of 3 GB. However, the multi-threadable implementation of the time-lag extended scattering and imaging conditions (equations 2.97 and 2.98) require the allocation and pre-loading of an additional $2 \times N_\tau$ time slices on the device, thereby bringing the total device memory footprint to approximately 10 GB (second row of Table D.3). For horizontal subsurface offsets, no additional time-slices are required for the extended scattering/imaging conditions but the double spatial extension rapidly increases the GPU memory usage. As a reference, an extension of 21 points on each horizontal direction leads to a device memory footprint of 44 GB (third row of Table D.3), which exceeds the global memory capacity on both available types of GPU cards (Table D.1).

Finally, the numerical implementation of the extended tomographic forward/adjoint operators follows the method described in section 2.4.4, which is similar to the one developed for extended Born. While the device memory usage does not change from the Born implementation, an additional difficulty arises from the need of storing two intermediate auxiliary wavefields on the host (compared to a single one for Born) for both forward and adjoint modeling operations. For the Cardamom dataset application (and assuming 4 GPU cards per node), this requirement leads to a pinned memory allocation of $4 \times 2 \times 32.5 = 260$ GB (third row of Table D.4) on the host's pinned memory.

Bibliography

Aghamiry, H. S., Gholami, A., and Operto, S., 2019, Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers: Geophysics, **84**, no. 1, R125–R148.

Aki, K., and Richards, P. G., 2002, Quantitative seismology:.

Almomin, A. A. M., 2016, Tomographic full waveform inversion: Ph.D. thesis, Stanford University.

Aster, R. C., Borchers, B., and Thurber, C. H., 2018, Parameter estimation and inverse problems: Elsevier.

Baeten, G., de Maag, J. W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., ten Kroode, F., and Rujie, Z., 2013, The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study: Geophysical Prospecting, **61**, no. 4, 701–711.

Barnier, G., and Biondi, E., 2020, Full waveform inversion by model extension using a model-space multi-scale approach: SEG Technical Program Expanded Abstracts 2020, 646–650.

Barnier, G., Biondi, E., and Biondi, B., 2018, Full waveform inversion by model extension: SEG Technical Program Expanded Abstracts 2018, 1183–1187.

Barnier, G., Biondi, E., and Clapp, R., 2019, Waveform inversion by model reduction using spline interpolation: Waveform inversion by model reduction using spline interpolation:, SEG International Exposition and Annual Meeting.

Bate, D., and Mike Perz, T., 2021, Ultra-long offsets signal a bright future for obn:.

Baumstein, A., and Hadidi, M. T., 2006, 3d surface-related multiple elimination: Data reconstruction and application to field data: Geophysics, **71**, no. 3, E25–E33.

Baysal, E., Kosloff, D. D., and Sherwood, J. W., 1983, Reverse time migration: Geophysics, **48**, no. 11, 1514–1524.

Billette, F., and Brandsberg-Dahl, S., 2005, The 2004 bp velocity benchmark: The 2004 bp velocity benchmark:, 67th EAGE Conference & Exhibition, cp–1.

Biondi, B., and Almomin, A., 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129–WA140.

Biondi, E., and Barnier, G., 2020, Elastic-parameter estimation by combining full-waveform inversion by model extension and target-oriented elastic inversion: SEG Technical Program Expanded Abstracts 2020, 735–739.

Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts 1999, 1723–1726.

Biondi, B., and Symes, W. W., 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, no. 5, 1283–1298.

Biondi, E., Biondi, B., and Barnier, G., 2018, Target-oriented elastic full-waveform inversion through extended-migration redatuming: SEG Technical Program Expanded Abstracts 2018, 1228–1232.

Biondi, E., Meadows, M., and Biondi, B., 2019, Amplitude preserving migration through extended acoustic least-squares rtm: SEG Technical Program Expanded Abstracts 2019, 4196–4200.

Biondi, E., Barnier, G., Biondi, B., and Clapp, R., 2021a, Target-oriented elastic full-waveform inversion: Deep-water ocean-bottom-node field-data application: Target-oriented elastic full-waveform inversion: Deep-water ocean-bottom-node field-data application:, First International Meeting for Applied Geoscience & Energy, 627–631.

——— 2021b, An object-oriented optimization framework for large-scale inverse problems: Computers & Geosciences, **154**, 104790.

Biondi, B. L., 2006, 3d seismic imaging: Society of Exploration Geophysicists.

Biondi, E., 2021, Target-oriented elastic full-waveform inversion: Ph.D. thesis, Stanford University.

Brenders, A., and Pratt, R. G., 2007, Waveform tomography of marine seismic data: What can limited offset offer?: Waveform tomography of marine seismic data: What can limited offset offer?:, 2007 SEG Annual Meeting.

Brenders, A., Dellinger, J., Kanu, C., Li, Q., and Michell, S., 2018, The wolfspar® field trial: Results from a low-frequency seismic survey designed for fwi: SEG Technical Program Expanded Abstracts 2018, 1083–1087.

Brittan, J., and Jones, I., 2019, Fwi evolution—from a monolith to a toolkit: The Leading Edge, **38**, no. 3, 179–184.

Brossier, R., Operto, S., and Virieux, J., 2015, Velocity model building from seismic reflection data by full-waveform inversion: Geophysical Prospecting, **63**, no. 2, 354–367.

Buhmann, M. D., 2000, Radial basis functions: Acta numerica, **9**, 1–38.

Buland, A., and Omre, H., 2003, Bayesian linearized avo inversion: Geophysics, **68**, no. 1, 185–198.

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, no. 5, 1457–1473.

Chen, Y., Yuan, J., Zu, S., Qu, S., and Gan, S., 2015, Seismic imaging of simultaneous-source data using constrained least-squares reverse time migration: Journal of Applied Geophysics, **114**, 32–35.

Claerbout, J. F., 1985, Imaging the earth's interior: Blackwell scientific publications Oxford.

Claerbout, J., 2014, Geophysical image estimation by example: Lulu. com.

Dablain, M., 1986, The application of high-order differencing to the scalar wave equation: Geophysics, **51**, no. 1, 54–66.

Dahlke, T., Biondi, B., and Clapp, R., 2020, Applied 3d salt body reconstruction using shape optimization with level sets: Geophysics, **85**, no. 5, R437–R446.

Dahlke, T., 2019, Velocity model building using shape optimization applied to level sets: Stanford University.

De Boor, C., B (asic)-spline basics.:, Technical report, WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER, 1986.

Dellinger, J., Ross, A., Meaux, D., Brenders, A., Gesoff, G., Etgen, J., Naranjo, J., Openshaw, G., and Harper, M., 2016, Wolfspar®, an "fwi-friendly" ultralow-frequency marine seismic source: SEG Technical Program Expanded Abstracts 2016, 4891–4895.

Dragoset, B., Verschuur, E., Moore, I., and Bisley, R., 2010, A perspective on 3d surface-related multiple elimination: Geophysics, **75**, no. 5, 75A245–75A261.

Díaz Pantin, E. F., 2016, Extended imaging and tomography under two-way operators: Ph.D. thesis, Colorado School of Mines. Arthur Lakes Library.

Fei, T. W., Luo, Y., Aramco, S., and Schuster, G. T., 2010, De-blending reverse-time migration: SEG Technical Program Expanded Abstracts 2010, 3130–3134.

Fichtner, A., 2010, Full seismic waveform modelling and inversion: Springer Science & Business Media.

Fletcher, R. F., Fowler, P., Kitchenside, P., and Albertin, U., 2005, Suppressing artifacts in prestack reverse time migration: SEG Technical Program Expanded Abstracts 2005, 2049–2051.

Fleury, C., and Perrone, F., 2012, Bi-objective optimization for the inversion of seismic reflection data: Combined fwi and mva: SEG Technical Program Expanded Abstracts 2012, 1–6.

Fu, L., and Symes, W. W., 2017, A discrepancy-based penalty method for extended waveform inversion: Geophysics, **82**, no. 5, R287–R298.

Golub, G. H., and Pereyra, V., 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on numerical analysis, **10**, no. 2, 413–432.

Grana, D., and Della Rossa, E., 2010, Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion: Geophysics, **75**, no. 3, O21–O37.

Guasch, L., Warner, M., and Ravaut, C., 2019, Adaptive waveform inversion: Practice: Geophysics, **84**, no. 3, R447–R461.

Guitton, A., Kaelin, B., and Biondi, B., 2007, Least-squares attenuation of reverse-time-migration artifacts: Geophysics, **72**, no. 1, S19–S23.

Guo, P., Guan, H., and McMechan, G. A., 2020, Data-and model-domain up/down wave separation for reverse-time migration with free-surface multiples: Geophysical Journal International, **223**, no. 1, 77–93.

Hou, J., and Symes, W. W., 2017, An alternative formula for approximate extended born inversion: Geophysics, **82**, no. 1, S1–S8.

Huang, Y., and Symes, W. W., 2015, Born waveform inversion via variable projection and shot record model extension: SEG Technical Program Expanded Abstracts 2015, 1326–1331.

Huang, G., Nammour, R., and Symes, W., 2017, Full-waveform inversion via source-receiver extension: Geophysics, **82**, no. 3, R153–R171.

Jakob, W., Rhinelander, J., and Moldovan, D. pybind11 – seamless operability between c++11 and python:, 2017. https://github.com/pybind/pybind11.

Jasjuang. What is pinned memory in cuda:, 2021.

Kaelin, B., and Guitton, A., 2006, Imaging condition for reverse time migration: SEG Technical Program Expanded Abstracts 2006, 2594–2598.

Kamath, N., Brossier, R., Métivier, L., Pladys, A., and Yang, P., 2021, Multiparameter full-waveform inversion of 3d ocean-bottom cable data from the valhall field: Geophysics, **86**, no. 1, B15–B35.

Lailly, P., and Bednar, J., 1983, The seismic inverse problem as a sequence of before stack migrations:.

Lazaratos, S., Chikichev, I., and Wang, K., 2011, Improving the convergence rate of full wavefield inversion using spectral shaping: SEG Technical Program Expanded Abstracts 2011, 2428–2432.

Le, H. Q., 2019, Anisotropic full-waveform inversion with pore pressure constraints: Stanford University.

Leader, C., 2015, The separation and imaging of continuously recorded seismic data: Ph.D. thesis, Stanford University.

Liu, F., Zhang, G., Morton, S. A., and Leveille, J. P., 2011, An effective imaging condition for reverse-time migration using wavefield decomposition: Geophysics, **76**, no. 1, S29–S39.

Mah, M., and Schmitt, D. R., 2003, Determination of the complete elastic stiffnesses from ultrasonic phase velocity measurements: Journal of Geophysical Research: Solid Earth, **108**, no. B1, ECV–6.

Martin, G. S., Marfurt, K. J., and Larsen, S., 2002, Marmousi-2: An updated model for the investigation of avo in structurally complex areas: SEG Technical Program Expanded Abstracts 2002, 1979–1982.

Martin, G. S., Wiley, R., and Marfurt, K. J., 2006, Marmousi2: An elastic upgrade for marmousi: The leading edge, **25**, no. 2, 156–166.

Martin, B., Fornberg, B., and St-Cyr, A., 2015, Seismic modeling with radial-basis-function-generated finite differences: Geophysics, **80**, no. 4, T137–T146.

Métivier, L., and Brossier, R., 2021, Receiver-extension strategy for time-domain full-waveform inversion using a relocalization approach: Geophysics, **87**, no. 1, R13–R33.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J., 2016, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, **205**, no. 1, 345–377.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J., 2018, Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach: Geophysics, **83**, no. 5, R515–R540.

Micikevicius, P., 2009, 3d finite difference computation on gpus using cuda: 3d finite difference computation on gpus using cuda:, Proceedings of 2nd workshop on general purpose processing on graphics processing units, 79–84.

Mora, P., 1989, Inversion= migration+ tomography: Geophysics, **54**, no. 12, 1575–1586.

Moré, J. J., and Thuente, D. J., 1994, Line search algorithms with guaranteed sufficient decrease: ACM Transactions on Mathematical Software (TOMS), **20**, no. 3, 286–307.

Nocedal, J., and Wright, S., 2006, Numerical optimization: Springer Science & Business Media.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The leading edge, **32**, no. 9, 1040–1054.

Perrone, F., Sava, P., Andreoletti, C., and Bienati, N., 2014, Linearized wave-equation migration velocity analysis by image warping: Geophysics, **79**, no. 2, S35–S46.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., , et al. Numerical recipes:, 1989.

Prucha, M. L., 2004, Imaging under salt: Illumination compensation by regularized inversion.:.

Rickett, J. E., and Sava, P. C., 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, no. 3, 883–889.

Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, no. 4, 874–881.

Robertsson, J. O., 1996, A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography: Geophysics, **61**, no. 6, 1921–1934.

Routh, P., Krebs, J., Lazaratos, S., Baumstein, A., Chikichev, I., Lee, S., Downey, N., Hinkley, D., and Anderson, J., 2011, Full-wavefield inversion of marine streamer data with the encoded simultaneous source method: Full-wavefield inversion of marine streamer data with the encoded simultaneous source method:, 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011, cp–238.

Sang, L., Klein-Helmkamp, U., Cook, A., and Jimenez, J. R., 2019, A practical workflow using quantitative interpretation of seismic amplitudes to derisk infill wells in deepwater gulf of mexico: The auger example: The Leading Edge, **38**, no. 10, 754–761.

Sava, P., and Biondi, B., 2004, Wave-equation migration velocity analysis. i. theory: Geophysical Prospecting, **52**, no. 6, 593–606.

Sava, P., and Fomel, S., 2006, Time-shift imaging condition in seismic migration: Geophysics, **71**, no. 6, S209–S217.

Shen, P., and Symes, W. W., 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, no. 5, VE49–VE59.

Shen, X., Ahmed, I., Brenders, A., Dellinger, J., Etgen, J., and Michell, S., 2018, Full-waveform inversion: The next leap forward in subsalt imaging: The Leading Edge, **37**, no. 1, 67b1–67b6.

Shen, P., 2005, Wave equation migration velocity analysis by differential semblance optimization: Ph.D. thesis.

Shen, X., 2014, Early-arrival waveform inversion for near-surface velocity estimation: Stanford University.

Shene, C.-K. Introduction to computing with geometry notes:, May 2011.

Siahkoohi, A., Verschuur, D. J., and Herrmann, F. J., 2019, Surface-related multiple elimination with deep learning: SEG Technical Program Expanded Abstracts 2019, 4629–4634.

Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, no. 1, 231–248.

Sirgue, L., Barkved, O., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J., 2010, Thematic set: Full waveform inversion: The next leap forward in imaging at valhall: First Break, **28**, no. 4.

Squires, L. J., Stoffa, P. L., and Cambois, G., 1994, Borehole transmission tomography for velocity plus statics: Geophysics, **59**, no. 7, 1028–1036.

Stolk, C. C., and Symes, W. W., 2002, Smooth objective functionals for seismic velocity inversion: Inverse problems, **19**, no. 1, 73.

Sun, D., Jiao, K., Vigh, D., and Coates, R., 2014, Source wavelet estimation in full waveform inversion: Source wavelet estimation in full waveform inversion:, 76th EAGE Conference and Exhibition 2014, 1–5.

Symes, W., and Carazzone, J. J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, no. 5, 654–663.

Symes, W. W., and Kern, M., 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples 1: Geophysical Prospecting, **42**, no. 6, 565–614.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical prospecting, **56**, no. 6, 765–790.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259–1266.

Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, no. 10, 1954–1966.

Tsvankin, I., 2012, Seismic signatures and analysis of reflection data in anisotropic media: Society of Exploration Geophysicists.

Van Leeuwen, T., and Herrmann, F. J., 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, no. 1, 661–667.

van Leeuwen, T., Herrmann, F. J., and Peters, B., 2014, A new take on fwi-wavefield reconstruction inversion: A new take on fwi-wavefield reconstruction inversion:, 76th EAGE Conference and Exhibition 2014, 1–5.

Vargas, A. C., 2020, Joint inversion of reflectivity and background subsurface components: Stanford University.

Verschuur, D. J., Berkhout, A., and Wapenaar, C., 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, no. 9, 1166–1177.

Versteeg, R., and Grau, G., 1990, Practical aspects of seismic data inversion, the marmousi experience: Proceedings of 1990 eaeg workshop: Practical aspects of seismic data inversion, the marmousi experience: Proceedings of 1990 eaeg workshop:, 52nd Annual Meeting, Soc. Expl. Geophys.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

Warner, M., and Guasch, L., 2014, Adaptive waveform inversion-fwi without cycle skipping-theory: Adaptive waveform inversion-fwi without cycle skipping-theory:, 76th EAGE Conference and Exhibition 2014, 1–5.

Warner, M., and Guasch, L., 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, no. 6, R429–R445.

Watson, L. M., Werpers, J., and Dunham, E. M., 2019, What controls the initial peak of an air-gun source signature?: Geophysics, **84**, no. 2, P27–P45.

Xu, S., Wang, D., Chen, F., Zhang, Y., and Lambare, G., 2012, Full waveform inversion for reflected seismic data: Full waveform inversion for reflected seismic data:, 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, cp–293.

Yang, T., 2013, Wavefield tomography using extended images: Ph.D. thesis, Colorado School of Mines. Arthur Lakes Library.

Yao, G., da Silva, N. V., Warner, M., Wu, D., and Yang, C., 2019, Tackling cycle skipping in full-waveform inversion with intermediate data: Geophysics, **84**, no. 3, R411–R427.

Youn, O. K., and Zhou, H.-w., 2001, Depth imaging with multiples: Geophysics, **66**, no. 1, 246–255.

Zhang, Y., 2015, Velocity model building using residual-moveout-based wave-equation migration velocity analysis: Stanford University.

Zhou, W., Brossier, R., Operto, S., and Virieux, J., 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, no. 3, 1535–1554.

Zong, J., Stewart, R. R., Dyaur, N., and Myers, M. T., 2015, Elastic properties of rock salt: Lab measurements and well log analysis in the gulf of mexico: Elastic properties of rock salt: Lab measurements and well log analysis in the gulf of mexico:, 2015 SEG Annual Meeting.