Stanford Exploration Project, Report 73, November 18, 1997, pages 59–76

Stanford Exploration Project, Report 73, November 18, 1997, pages 59–76

The seismic data compression limitations of the singular value decomposition

Ray Abma[1]

keywords: *noise, amplitudes, graphics, velocity*

ABSTRACT

Reflection seismic data require large volumes of digital storage at a great cost to the seismic industry. This cost could be reduced if a compression technique suitable to these data was found. Data compression using the singular value decomposition is examined here in an attempt to reduce this cost. When any significant compression factor was used, this compression technique reconstructed the original data poorly because dipping events and high-frequency detail were lost. While other seismic data may be compressed with this method, the compression of reflection seismic data with the singular-value-decomposition technique is not practical.

INTRODUCTION

The cost of storing reflection seismic data is enormous and could be reduced by an effective compression technique. While these data may be compressed with several techniques, the need for very accurate reconstruction of the original data is of special importance with seismic data, since details of amplitude and noise are important to later processing.

A simple image compression technique using the singular value decomposition is examined here with reflection seismic data and is found to reproduce the original data poorly because of lost dips and high-frequency detail. This poor reconstruction is caused by the false assumption that the smallest singular values can be ignored with reflection seismic data.

While simpler seismic datasets such as velocity fields and velocity analyses could be successfully compressed using this technique, compressing reflection seismic data is not practical using singular value decomposition, at least not for cases where very-accurate reconstruction is needed.

[1]**email:** not available

Examples of compression of a photograph, simple synthetics, a shot gather, and a stacked display are shown.

THE NEED FOR COMPRESSION

An effective seismic data compression method would be useful in storing industrial reflection seismic data. These data are stored on hundreds of thousands of digital tapes costing millions of dollars. One particular seismic data acquisition company uses more digital tapes than any concern in the world, except for the U.S. Federal government. Since little original seismic data is ever discarded and since it is best kept in climate-controlled long-term storage, large rooms are cut out of salt domes to store tapes containing old reflection seismic data. Any technique that would reduce seismic data storage space significantly would save the industry millions of dollars.

While smaller data storage requirements would be the main benefit of an effective seismic data compression technique, seismic data could be handled more conveniently in compressed form. Seismic data acquisition now requires shipping thousands of tapes from remote locations; reducing the data volume would lower these shipping costs. Much of the seismic industry's computer resources is used simply to move seismic data from tape to computer memory where they are processed. While compressing and reconstructing data increases computational costs, the cost of data transfer might offset this.

Although new storage technologies such as optical disks and new magnetic storage techniques promise some relief for the information storage problem, the amount of data being acquired is increasing rapidly, especially for 3-D surveys, and new technologies require years before achieving the reliability of standard tape storage and gaining the confidence of the industry. Data compression techniques fit the old storage technology as well as the new technologies and reduces the need for these costly new technologies.

DATA COMPRESSION WITH THE SINGULAR VALUE DECOMPOSITION

A traditional method of compressing images using the singular value decomposition was a reasonable technique to consider for seismic data compression. This method reproduces most photographic images well and allows a significant storage reduction. It is based on simple ideas and readily available software.

The singular value decomposition is based on decomposing a matrix into two matrices, U and V, and a vector, Σ, containing scale factors called singular values. This decomposition of a matrix A is expressed as

$$A = U \Sigma V^t$$

Each singular value in Σ corresponds to an single two-dimensional image built from a single column in U and a single row in V. The reconstructed image is the sum of each partial image scaled by the corresponding singular value in Σ.

The key to compressing an image is recognizing that the smallest singular values and their corresponding images should not significantly contribute to the final image. By ignoring the smallest singular values along with the columns in U and the rows in V that these singular values multiply, the original image should be accurately reconstructed from a dataset much smaller than the original matrix. When more than half these singular values are ignored, the parts of the U and V matrices required to reconstruct the image are smaller than the original image. If only a few singular values are needed, a small fraction of the original U and V matrices is needed to reconstruct the image, and the storage cost is cut significantly.

The reconstruction of compressed images while ignoring small singular values is illustrated by decomposing a photograph with the singular value decomposition and then reconstructing it with the largest singular values. Figure 1 shows the original image, the U and V matrices, and the singular values in Σ, labeled sigma in the plot. Figure 2 shows the reconstruction of the image using first 5, 10, 15, and 30 singular values. Notice that 30 singular values reconstruct an image that is difficult to distinguish from the original. 30 singular values out of 276 produces an excellent compression ratio.

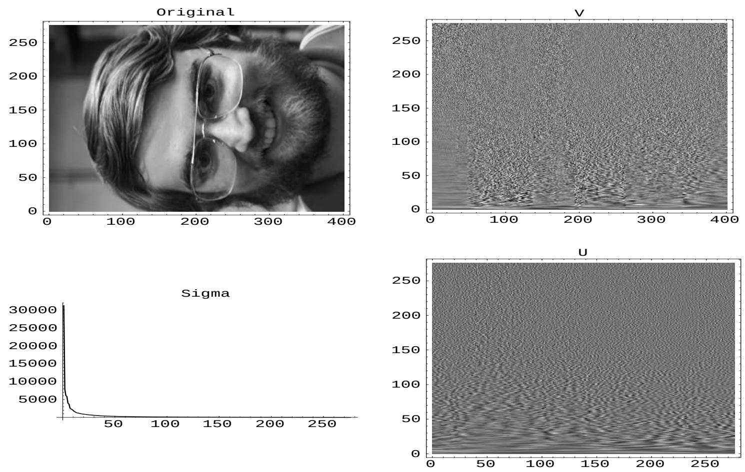

Figure 1: A photograph and its decomposition. Notice the sharp drop in the magnitudes of the singular values shown in Sigma. Also notice the low frequency data in the lower part of the U and V matrices.

Reconstruction Problems with Seismic Data

Compared to photographs, reflection seismic datasets are poorly compressed with the singular value decomposition. Dipping events extend the range of singular values that

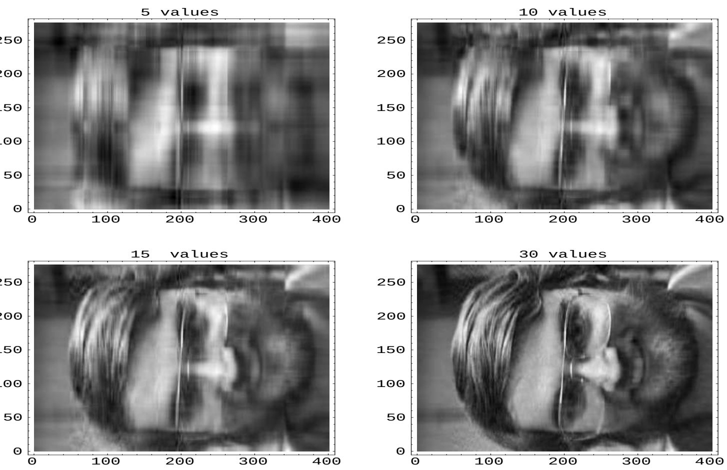

Figure 2: Reconstructed photographs using 5, 10, 15, and 30 singular values. The reconstruction using 30 singular values is practically identical to the original.

contain significant energy, invalidating the assumption that only the small singular values are needed to reconstruct the image. Fine detail, including noise, tends to be associated with the small singular values and is lost using reconstructions with only the large singular values.

Dipping events do not fall along columns or rows of the original image and are spread over a wide range of singular values, violating the assumption that we can represent the input with only the first few singular values. Examples shown in Figures 3 to 5 show the number of significant singular values increases with dip. Figure 3 shows a flat event and its decomposition. This event has only one significant singular value and can be reconstructed with a single row of V and a single column of U. If a dipping event is decomposed as seen in Figure 4, the number of singular values increases and the amount of data needed to reconstruct the image is larger than that needed to reconstruct the image of the flat event. If an event with even more dip is decomposed as seen in Figure 5, all the singular values must be used to reconstruct the image, and the amount of data needed to store the U and V matrices is greater than the size of the original image.

High frequency events in time and space containing the fine detail of an image tend to be distributed in the partial images corresponding to small singular values. The decomposition of the photograph shown in Figure 1 shows high frequency events in the upper rows of the U and V matrices that correspond to the smaller singular values. In photographs like Figure 1, much of the image is made up of large uniform areas; in contrast, seismic images contain mainly fine detail. Figure 6 shows the decomposition of a dipping event in a background of random numbers that simulates a seismic file with one strong event and much small detail. If the U and V matrices

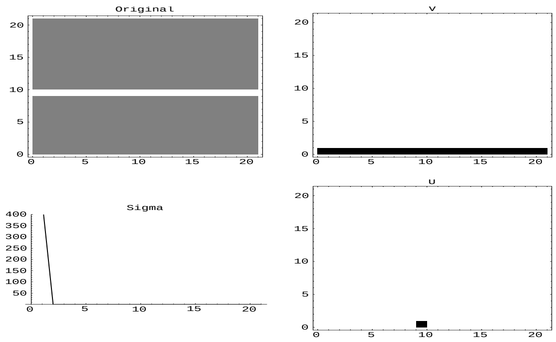

Figure 3: A flat event and its decomposition. Notice that there is only one nonzero singular value.

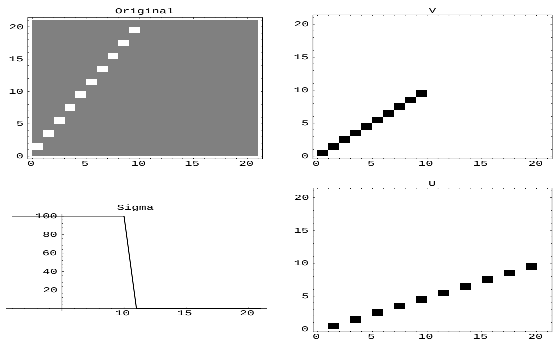

Figure 4: A dipping event and its decomposition. Notice that there are now many more nonzero singular values.

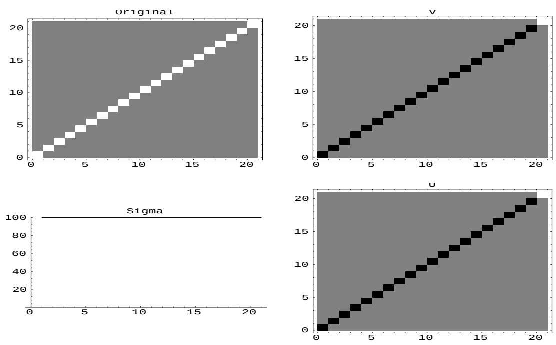

Figure 5: An even steeper dipping event and its decomposition. Every singular value is now equal.

are examined carefully, low frequency signals corresponding to the dipping event are seen in the bottom few rows. When the image is reconstructed in Figure 7 with a few large singular values, the background shows a loss of random noise. Even worse, the strong event is corrupted with noise from outside the event, and the random noise is organized along the columns and rows.

SEISMIC DATA EXAMPLES

The first example of decomposing a seismic record is the shot record shown in Figure 8. The plot of Σ shown in the figure shows the singular values decreasing slowly compared to the those in the photographic case. The U matrix also shows significant detail on the vectors corresponding to the smaller singular values. The reconstructions of the shot gather using a only few of the larger singular values are shown in Figure 9. When only five singular values are used to reconstruct the shot gather, much of the dipping energy is lost, but as more singular values are used, the steeper dips gradually reappear. Even while using 30 singular values, the shot gather is poorly reconstructed. Most of the high amplitude events have reappeared, but the small details are washed out, and noise has corrupted the low amplitude background above the first breaks.

Figure 10 has been included here to give the reader an idea of what the images corresponding to the first few singular values of the shot gather look like. Here vertical and horizontal patterns make it obvious that the images are the outer products of two vectors. The first few images in Figure 10 have the energy concentrated about the central portion of the image. This corresponds roughly to the reconstructions shown in Figure 9, where the flat data in the center is reconstructed first. The steep-dip data in the outer traces associated with the smaller singular values are imaged with the later, smaller singular values.

In an attempt to simplify the input to the singular value decomposition, a two-dimensional Fourier transform was applied to the shot record. The transform, as seen in Figure 11, shows that much of the energy is concentrated in the lower frequency range. While dips are still present, the dips are smaller than those in the original shot gather. It is interesting to see that the plot of the singular values appears to be similar to the plot of the singular values of the original shot gather shown in Figure 8. Although much of the detail in the U and V matrices is lost because the plots show the absolute value of the complex numbers, the U and V matrices appear much simpler than the U and V matrices of the original shot gathers.

More success compressing stacked datasets is expected, since these data contain fewer high-amplitude steep-dip events than typical shot gathers. The stack shown in Figure 12 contains gently sloping events below the water bottom and more complex structures deeper in the section. The plot of the singular values shows a pattern similar to the singular values of the shot gather, which is surprising, since the dips are smaller and the dipping events are of lower amplitude in the stacked case. When

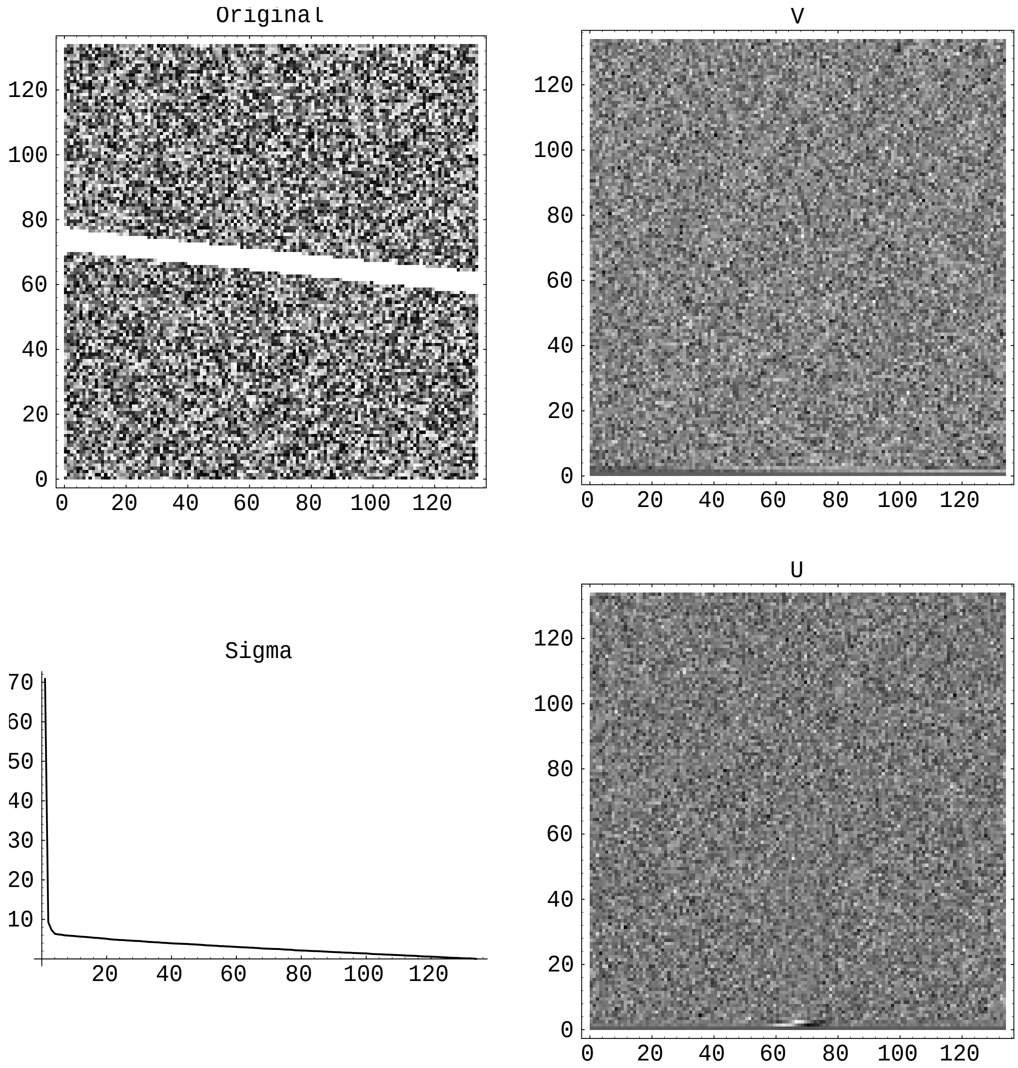

Figure 6: Dip and random noise and its decomposition. The dipping event is contained in the first few singular values, but the noise is distributed throughout the U and V matrices.

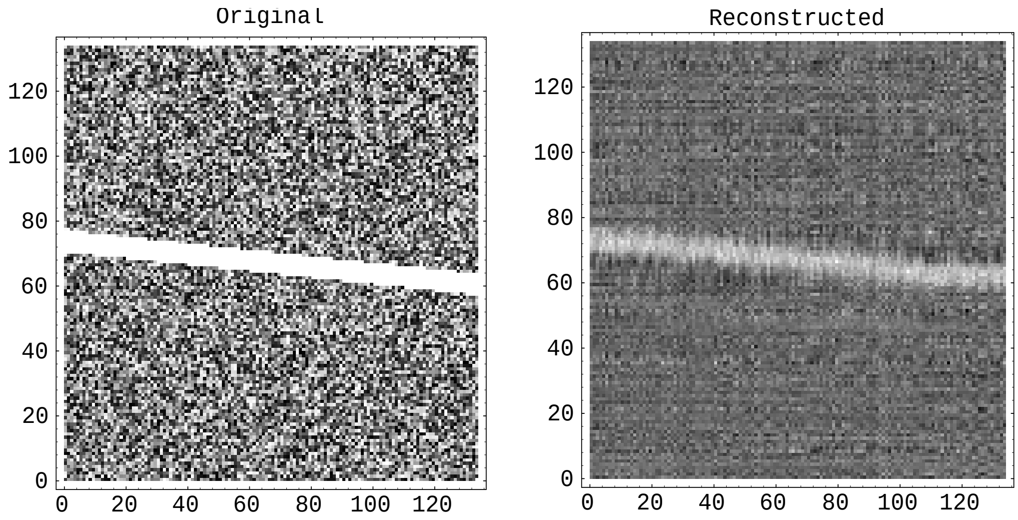

Figure 7: Original and reconstructed images from Figure 6. The noise has been organized along the rows and columns and has contaminated the dipping event.

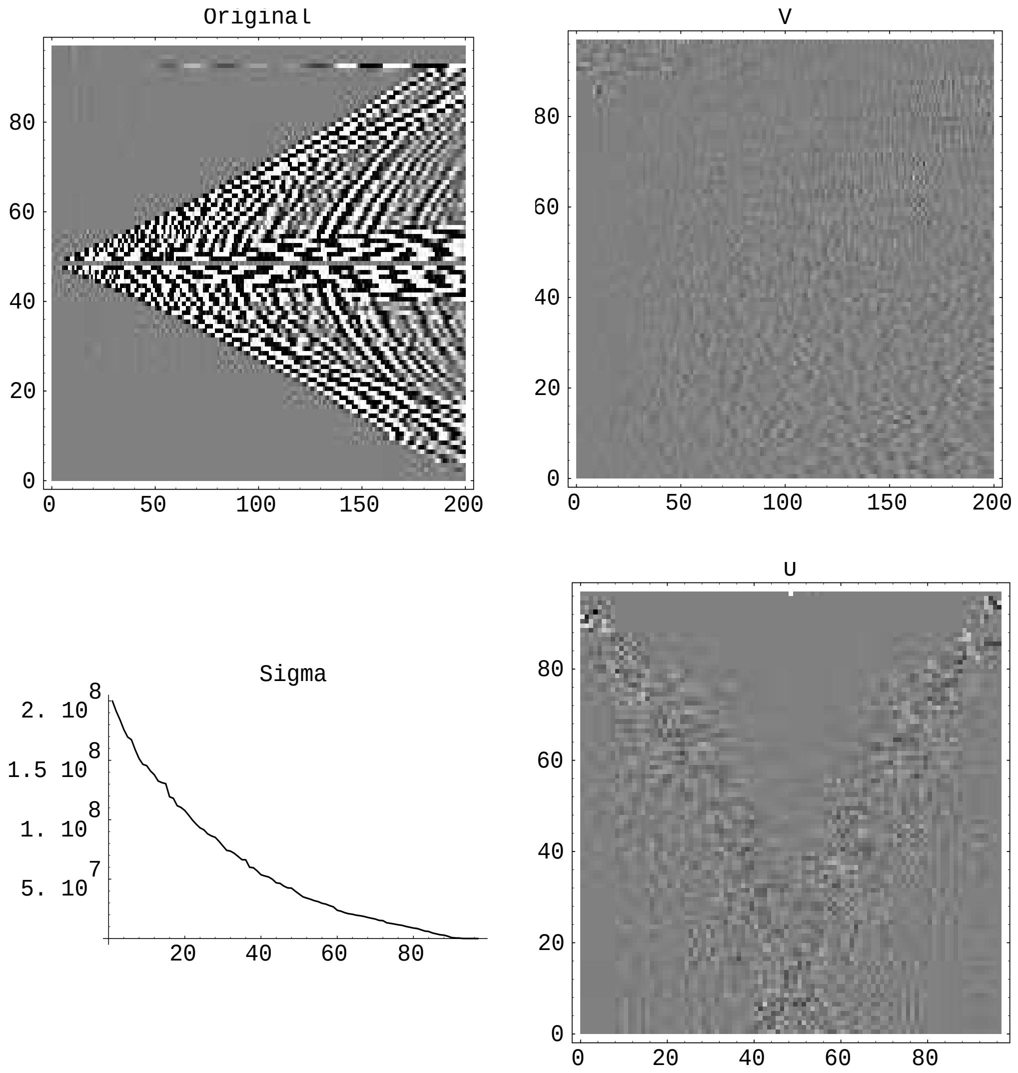

Figure 8: A shot gather and its decomposition. Notice the slow decrease in the size of the singular values and the structure evident in the U matrix.

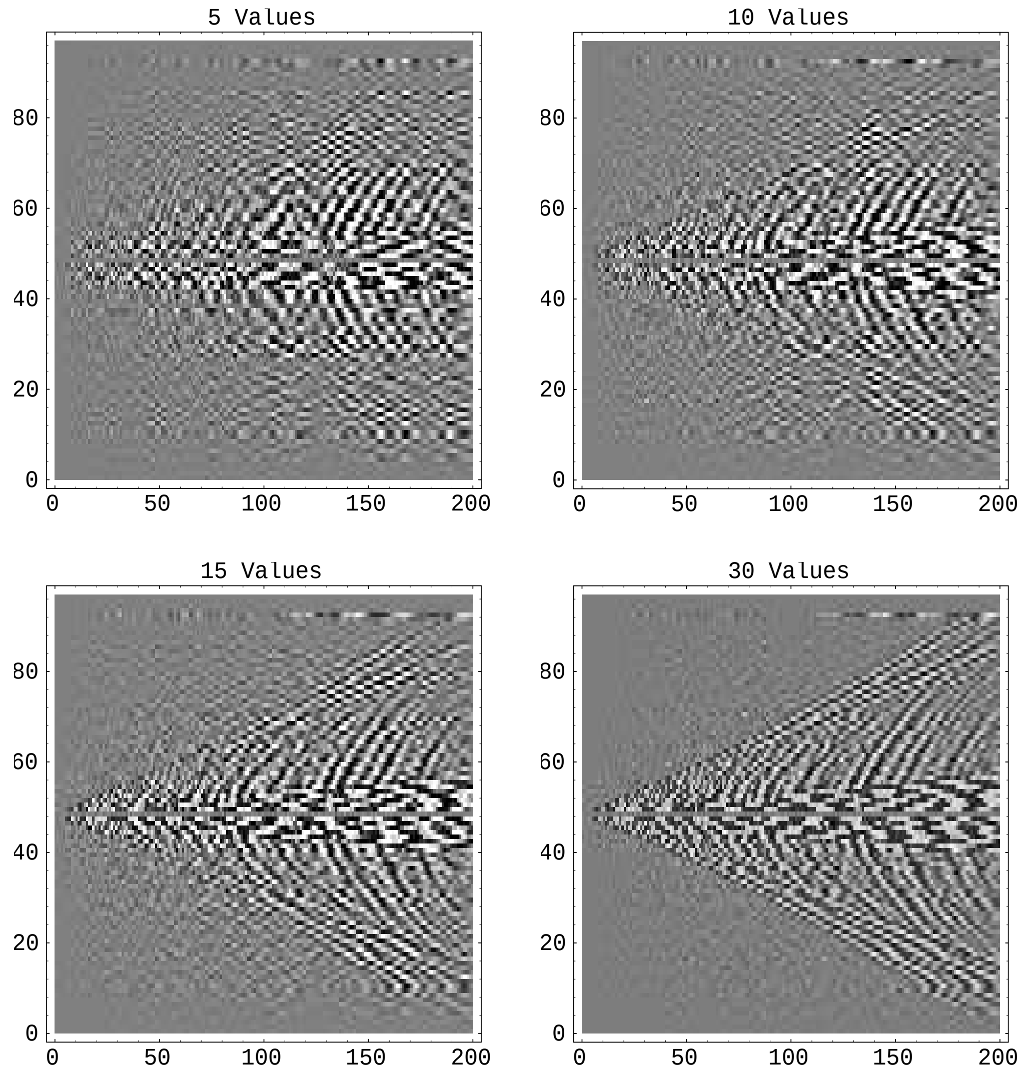

Figure 9: The shot gather reconstructed using 5, 10, 15, and 30 singular values. Events with low dips are reconstructed with the first few singular values, and the higher dips appear with the smaller singular values added later.

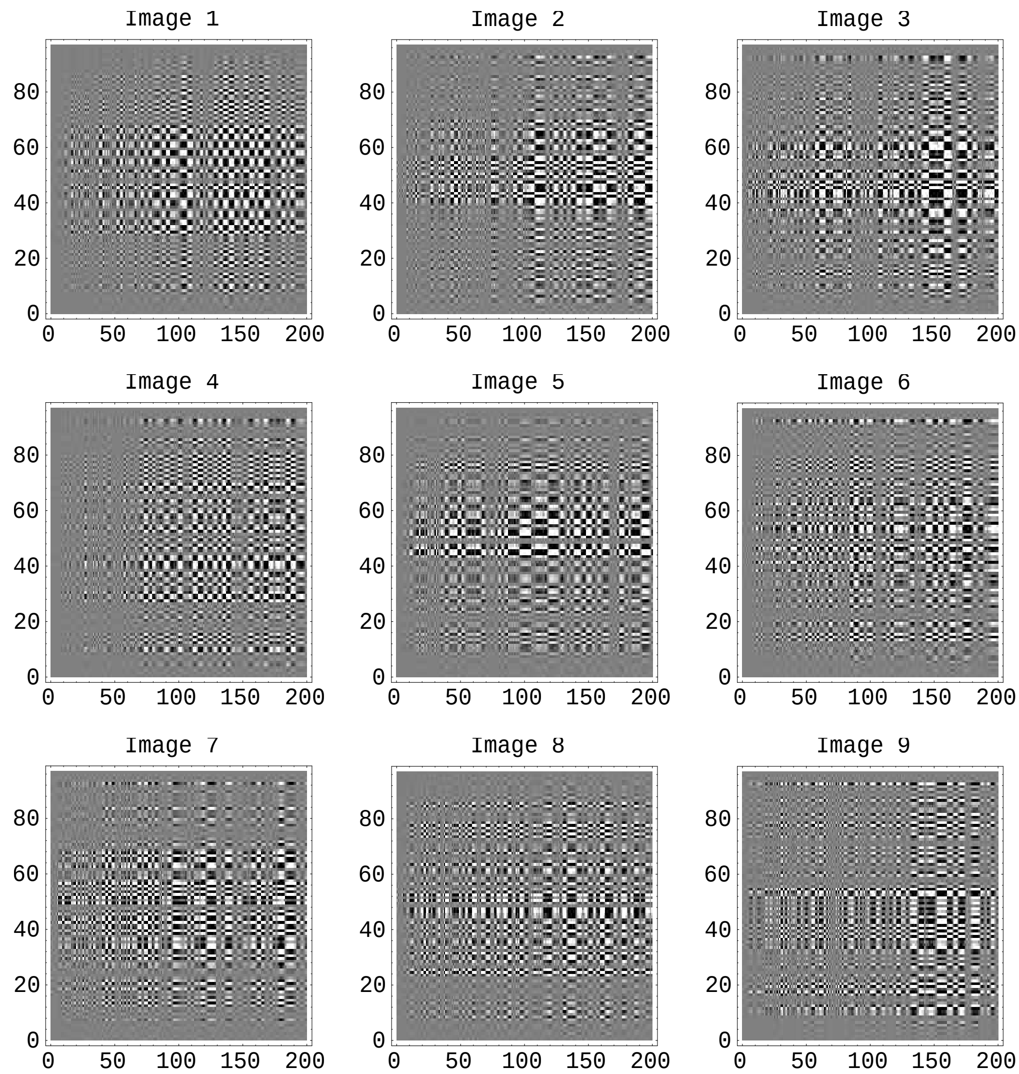

Figure 10: The images of a shot gather associated with the first nine singular values. Here it becomes obvious that the images are constructed from the outer product of two vectors.

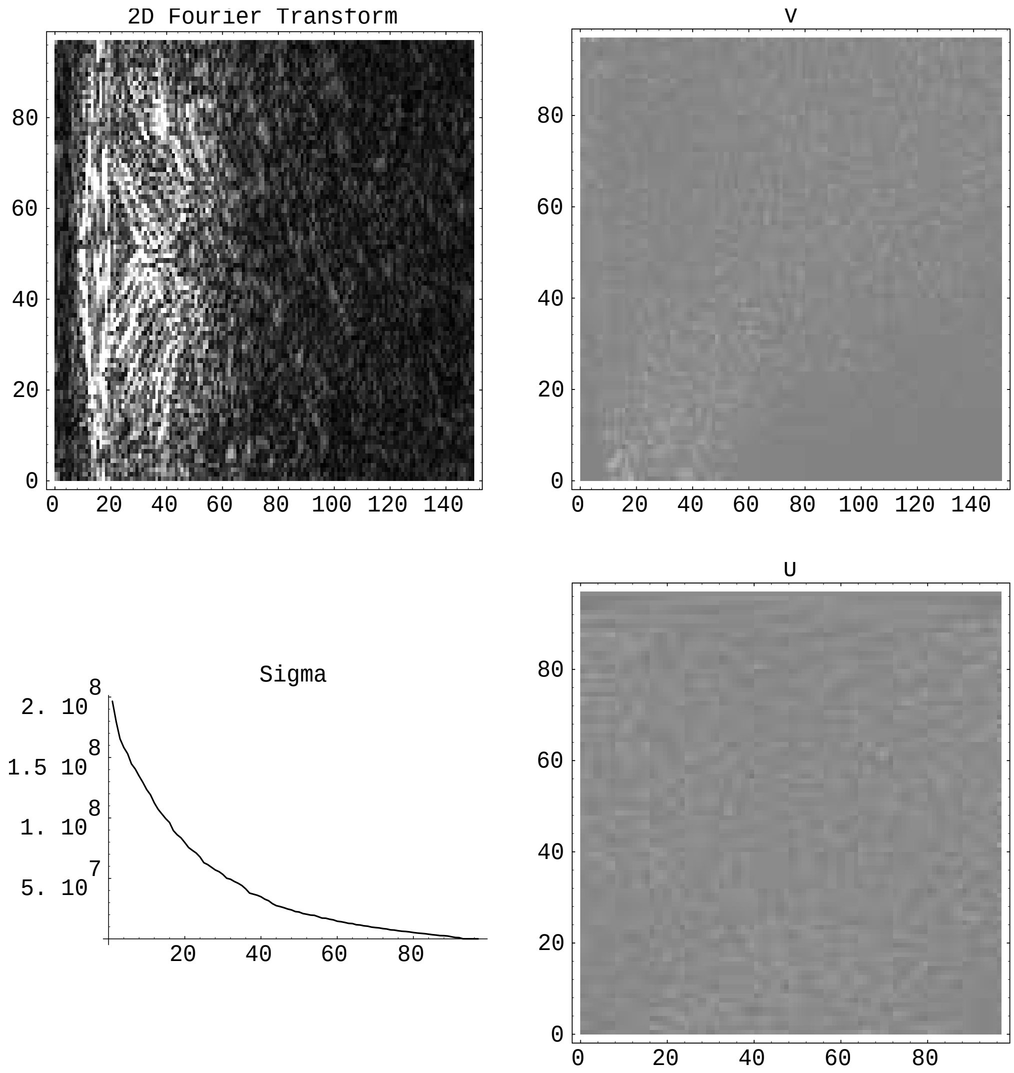

Figure 11: The two-D Fourier transform of the shot gather and its decomposition. Notice the slow decrease in the size of the singular values and compare these singular values to the singular values shown in Figure 8.

the original data is reconstructed from the stack's decomposition using the first few singular values, discontinuities can be seen in the first few reconstructions shown in Figure 13. The steeper dips missing on the first few reconstructions do not appear until the last reconstruction using 30 singular values.

Figure 14 shows the images associated with the first few singular values and gives some insight into the discontinuities seen in the reconstructions shown in Figure 13. The reflections are broken up into events consisting of alternating polarities with spatial periods that generally decrease as the size of the singular values decrease. As more singular values are used, more detail is seen in the reconstruction.

While it might be hoped that increasing or decreasing the number of traces put into the singular value decomposition might change the trend of the singular values, this effect is not seen with the stack image. Figure 15 shows the singular values for various numbers of traces from the stack in Figure 12. The slow decrease in amplitude of the singular values is similar for all these cases.

CONCLUSIONS

While the problems seen above prevent the use of the singular value decomposition technique as a general-purpose compression method, there is still a place for it in the seismic industry. Some derived seismic images, such as velocity fields and velocity analyses, do not contain the high frequency detail seen in the reflection data and therefore will compress well with this technique. Also, for low resolution seismic displays, small details are unimportant and may be ignored. Although these displays are typically already reduced from the original sizes, they might be further compressed using the singular value decomposition.

Reflection seismic data cannot be reconstructed accurately from compressed forms using the singular value decomposition because dipping events and high frequency detail map into the small singular values, violating the assumption that these small singular values can be ignored in reconstructing the compressed image. While other applications of singular value decomposition compression exist in the seismic industry, this technique fails to compress the largest amount of data stored, the seismic reflection data.

ACKNOWLEDGEMENTS

I would like to thank Western Geophysical for permission to continue working on this topic and Steve Cole of the Stanford Exploration Project for permission and help in using his picture for the photographic example.

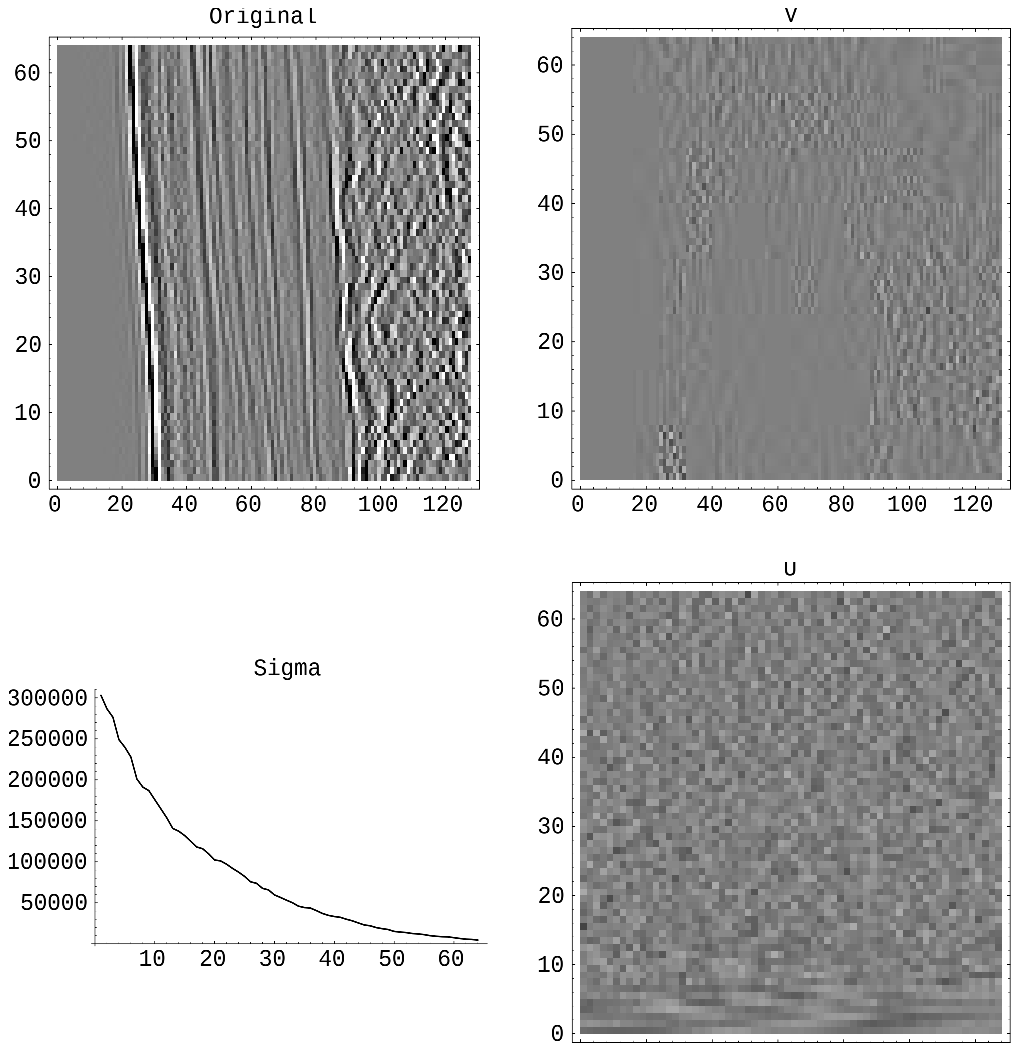

Figure 12: A stack and it's decomposition. While the stack appears simpler than the shot gather shown in Figure 8, the slow decrease of the singular values is similar.

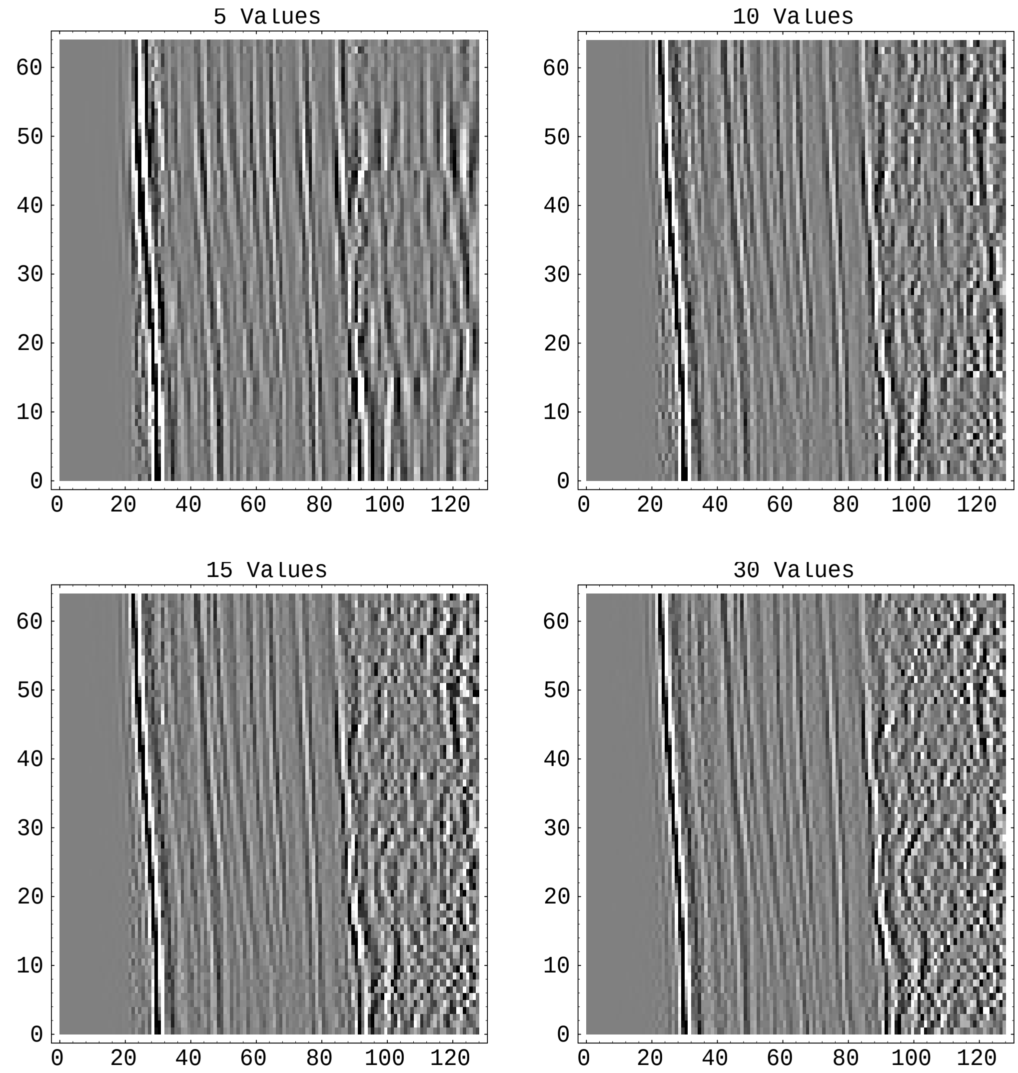

Figure 13: The stack reconstructed using 5, 10, 15, and 30 singular values. Notice the discontinuities in the shallow zone and the lost diffraction tails in the deeper zone for the reconstructions using 5 and 10 values.

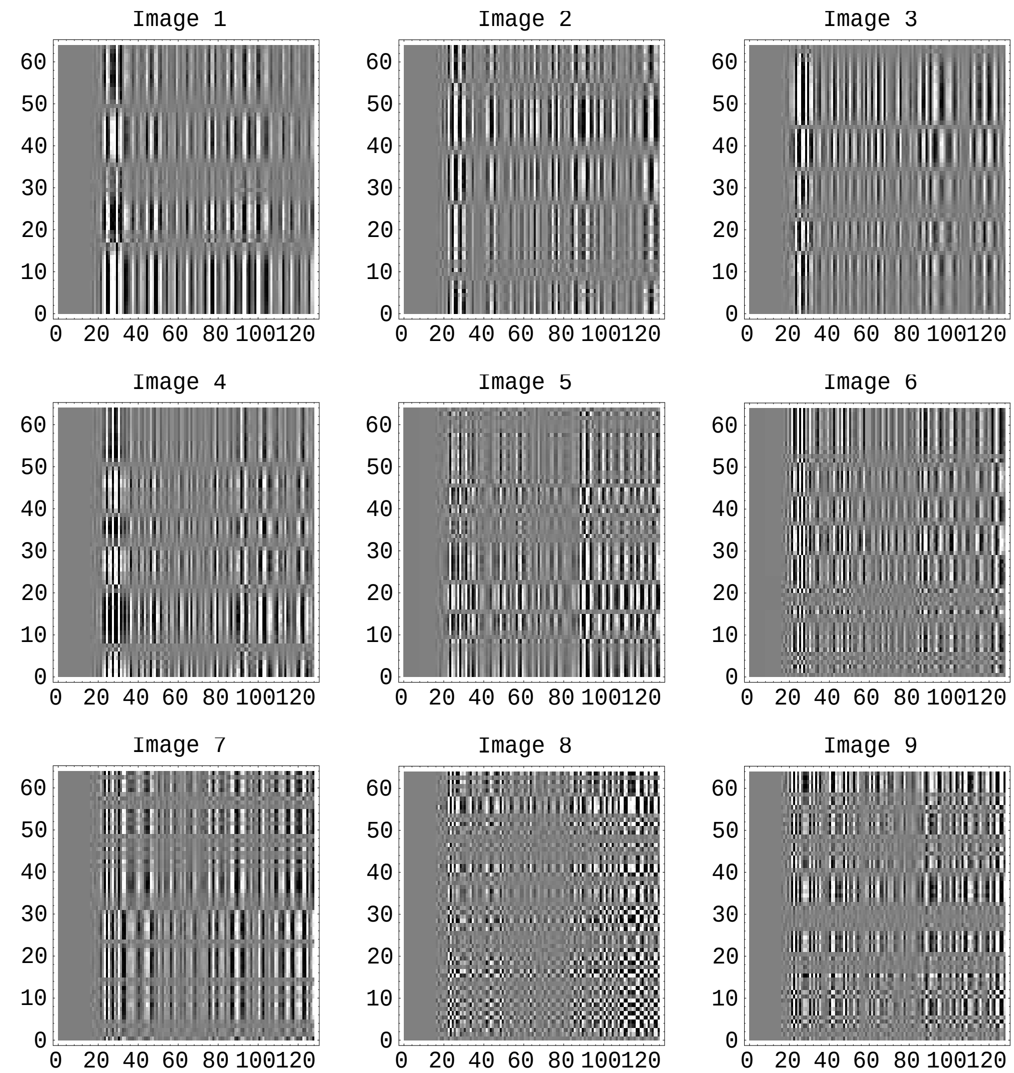

Figure 14: The stack images associated with the first nine singular values.

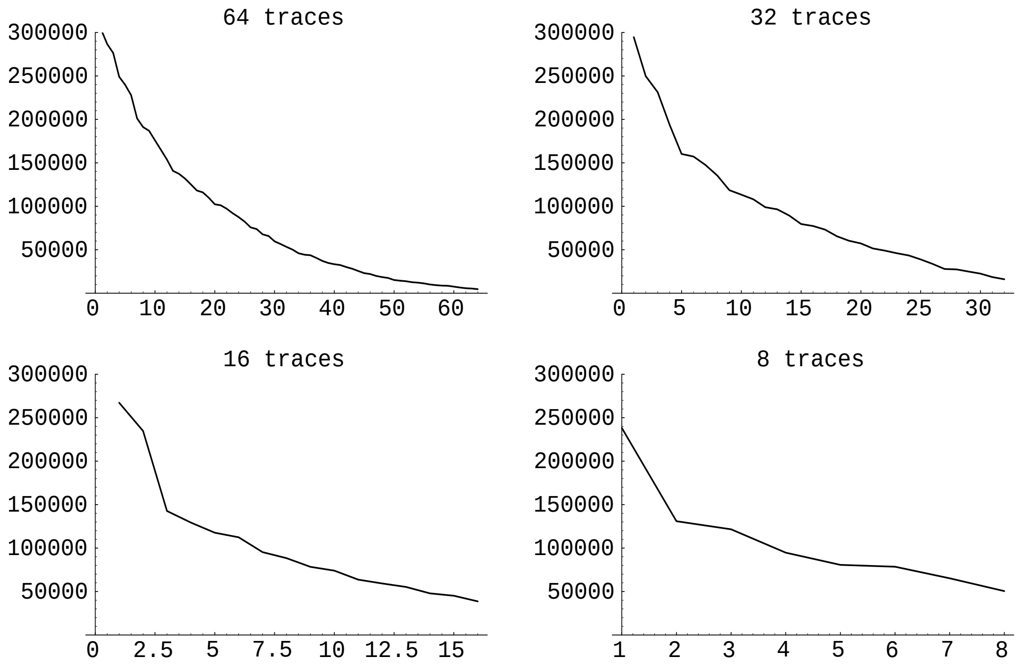

Figure 15: Sizes of the singular values using 64, 32, 16, and 8 traces of a stack image. The slow decrease in amplitude is similar for these cases.