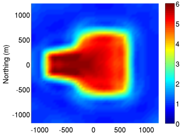
Compressive Conjugate Directions: Linear Theory

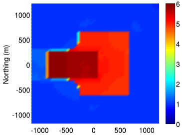
A new tool for sparsity-promoting optimization

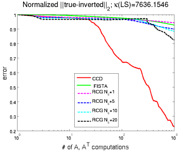
Musa Maharramov and Stewart A. Levin

Department of Geophysics

April 12, 2016







Motivation: regularized large-scale optimization

- Model styling=regularization is critical in applications
- Non-smooth sparsity/blockiness-promoting L_1/TV regularization
- Computationally expensive algorithms exist
- Goal: a computationally feasible and accurate algorithm

Problem: L_1/TV-regularized linear LS fitting

$$\|\mathbf{B}\mathbf{u}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 \rightarrow \min, \tag{1}$$

$$\mathbf{B} = \mathbf{I} \tag{2}$$

or

$$\mathbf{B} = \nabla. \tag{3}$$

A good algorithm: ADMM/"Split Bregman" (Glowinski, Maroco)

ADMM requires solving lots of large LS problems in a loop:

$$\begin{aligned}
\mathbf{u}_{k+1} &= \operatorname{argmin} \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{z}_k - \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2, && \text{very hard!} \\
\mathbf{z}_{k+1} &= \operatorname{argmin} \|\mathbf{z}\|_1 + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\|_2^2, && \text{easy} \\
\mathbf{b}_{k+1} &= \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1},\ k=0,1,2,\ldots && \text{trivial}
\end{aligned} \tag{4}$$

ISTA/FISTA (Beck & Teboulle)—shorten the inner loop

FISTA speeds up gradient descent using Nesterov relaxation:

$$
\begin{aligned}
\mathbf{y}_{k+1} &= \mathbf{u}_k - \gamma\alpha\mathbf{A}^T\left(\mathbf{A}\mathbf{u}_k - \mathbf{d}\right), && \Leftarrow \text{ single iteration!} \\
\mathbf{z}_{k+1} &= \operatorname{shrink}\left\{\mathbf{y}_{k+1}, \gamma\right\}, && \Leftarrow \text{ trivial} \\
\zeta_{k+1} &= \left(1 + \sqrt{1 + 4\zeta_k^2}\right)/2, && \Leftarrow \text{ trivial} \\
\mathbf{u}_{k+1} &= \mathbf{y}_{k+1} + \frac{\zeta_k - 1}{\zeta_{k+1}}\left(\mathbf{y}_{k+1} - \mathbf{y}_k\right), && \Leftarrow \text{ trivial}
\end{aligned}
\tag{5}
$$

Restarted CG does a few CG iterations of each LS problem:

$$
\begin{aligned}
\mathbf{u}_{k+1} &= N \text{ iters of } \operatorname{argmin} \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{z}_k - \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2, \\
\mathbf{z}_{k+1} &= \operatorname{argmin} \|\mathbf{z}\|_1 + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\|_2^2, \\
\mathbf{b}_{k+1} &= \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1},\ k = 0, 1, 2, \ldots
\end{aligned} \tag{6}
$$

- **Problem**: rhs change⇒previous geometry information of standard CG residuals is lost
- **Observation**: the quadratic shape of the objective function (6) **does not change**
- **Idea**: **store the previous conjugate directions GMRES-style**

Compressive Conjugate Directions

$\mathbf{u}_0 \leftarrow \mathbf{0}^N,\ \mathbf{z}_0 \leftarrow \mathbf{0}^K;\ \mathbf{b}_0 \leftarrow \mathbf{0}^K,\ \mathbf{v}_0 \leftarrow \left[\sqrt{\alpha}\mathbf{d}; \sqrt{\lambda}\left(\mathbf{z}_0 + \mathbf{b}_0\right)\right]$

$\mathbf{F} \leftarrow [\mathbf{A}; \mathbf{B}],\ \mathbf{p}_0 \leftarrow \mathbf{F}^T\mathbf{v}_0,\ \mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0,\ \delta_0 \leftarrow \mathbf{q}_0^T\mathbf{q}_0$

for $k = 0, 1, 2, 3, \ldots$ **do**

$\tau_i \leftarrow \mathbf{q}_i^T\mathbf{v}_k/\delta_i,\ i = 0, 1, \ldots, k,\ \mathbf{u}_{k+1} \leftarrow \sum_{i=0}^{k} \tau_i \mathbf{p}_i$

$\mathbf{z}_{k+1} \leftarrow \text{shrink}\left\{\mathbf{B}\mathbf{u}_{k+1} - \mathbf{b}_k, 1/\lambda\right\}$

$\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$

$\mathbf{v}_{k+1} \leftarrow \left[\sqrt{\alpha}\mathbf{d}; \sqrt{\lambda}\left(\mathbf{z}_{k+1} + \mathbf{b}_{k+1}\right)\right]$

$\mathbf{r}_{k+1} \leftarrow \mathbf{v}_{k+1} - \sum_{i=0}^{k} \tau_i \mathbf{q}_i$

$\mathbf{w}_{k+1} \leftarrow \mathbf{F}^T\mathbf{r}_{k+1},\ \mathbf{s}_{k+1} \leftarrow \mathbf{F}\mathbf{w}_{k+1}$

$\beta_i \leftarrow -\mathbf{q}_i^T\mathbf{s}_{k+1}/\delta_i,\ i = 0, 1, \ldots, k$

$\mathbf{p}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i \mathbf{p}_i + \mathbf{w}_{k+1},\ \mathbf{q}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i \mathbf{q}_i + \mathbf{s}_{k+1}$

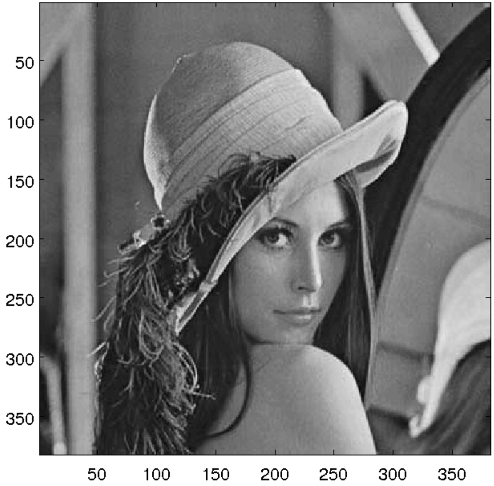
$\delta_{k+1} \leftarrow \mathbf{q}_{k+1}^T\mathbf{q}_{k+1}$

if $\delta_{k+1} = 0,\ \delta_{k+1} \leftarrow 1,\ \mathbf{p}_{k+1} \leftarrow \mathbf{0}^N,\ \mathbf{q}_{k+1} \leftarrow \mathbf{0}^{M+K}$

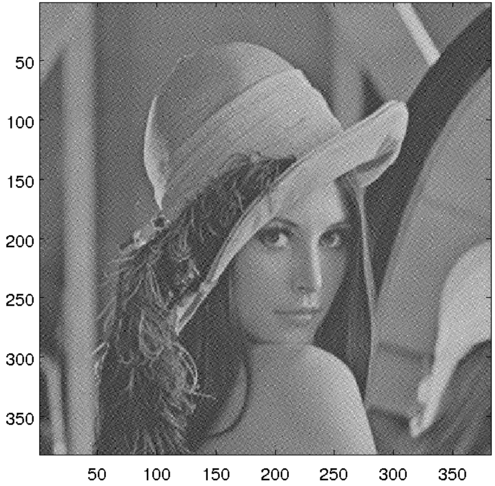
Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2/\|\mathbf{u}_k\|_2 \leq$ target accuracy

end for

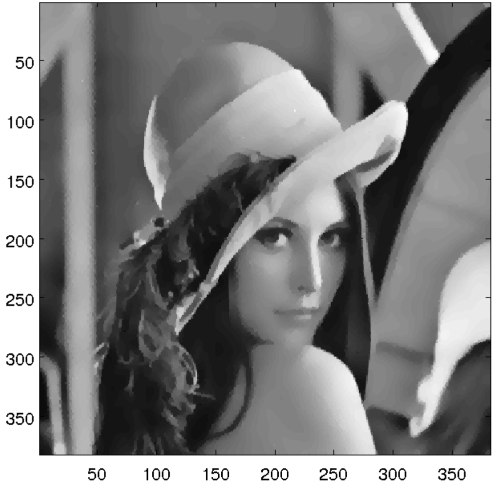
A denoising example ($\mathbf{A} = \mathbf{I}$). True image



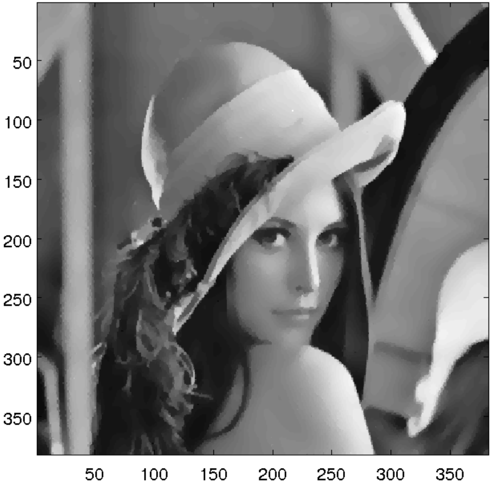
Noisy image



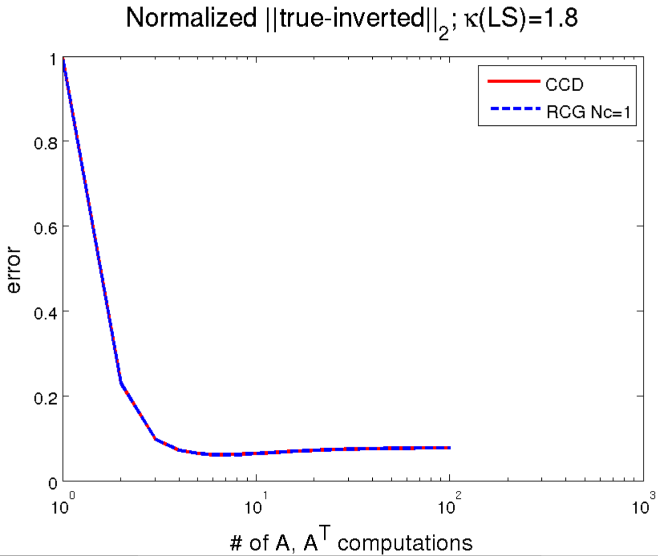
RCG denoising



CCD denoising

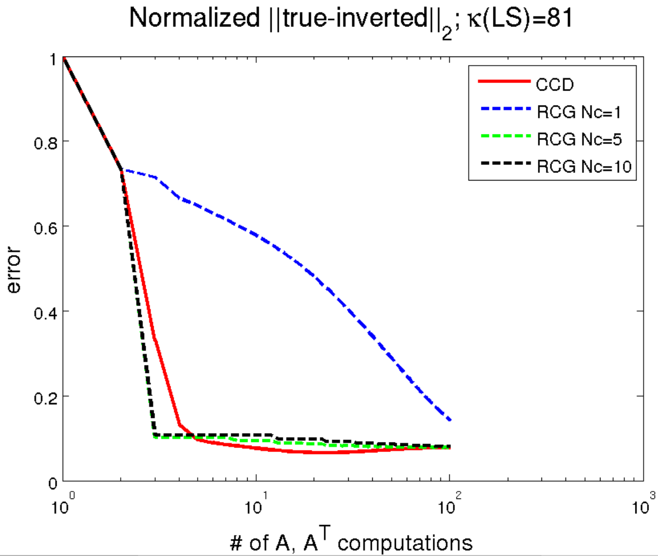


$\lambda = 1$



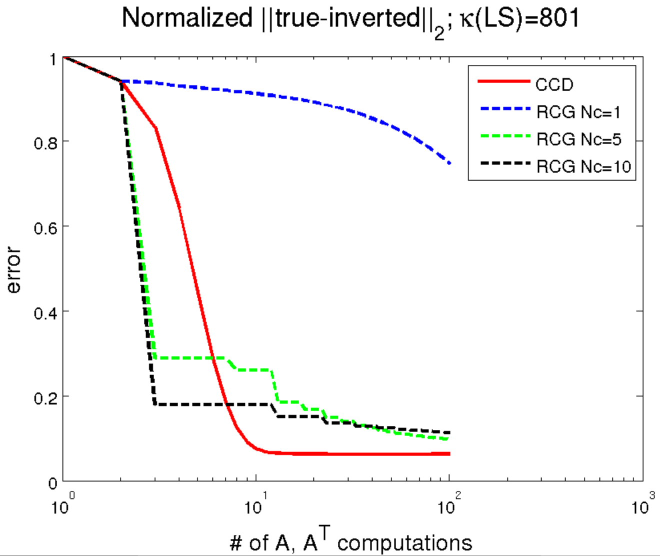
Normalized ||true-inverted||2; κ(LS)=1.8
1
0.8
0.6
0.4
0.2
0
error
10^0
10^1
10^2
10^3
# of A, A^T computations
CCD
RCG Nc=1

$\lambda = 100$



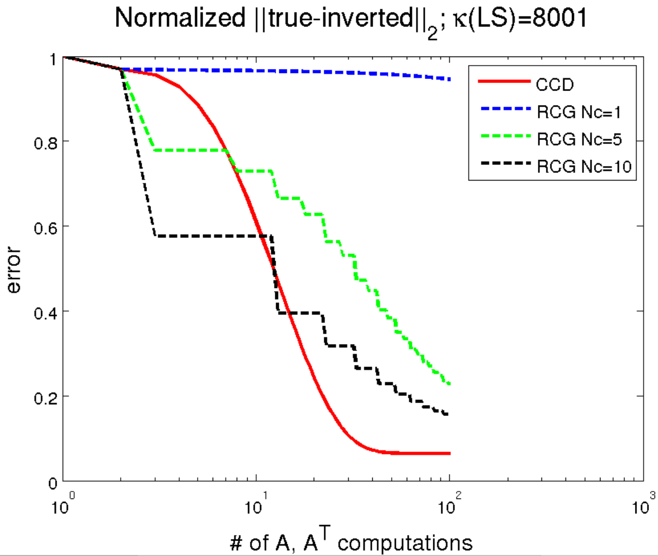
Normalized ||true-inverted||$_2$; κ(LS)=81
error
# of A, A^T computations
CCD
RCG Nc=1
RCG Nc=5
RCG Nc=10
1
0.8
0.6
0.4
0.2
0
10^0
10^1
10^2
10^3

$\lambda = 1000$



Normalized ||true-inverted||$_2$; κ(LS)=801
error
# of A, A^T computations
CCD
RCG Nc=1
RCG Nc=5
RCG Nc=10
1
0.8
0.6
0.4
0.2
0
10^0
10^1
10^2
10^3

$\lambda = 10000$

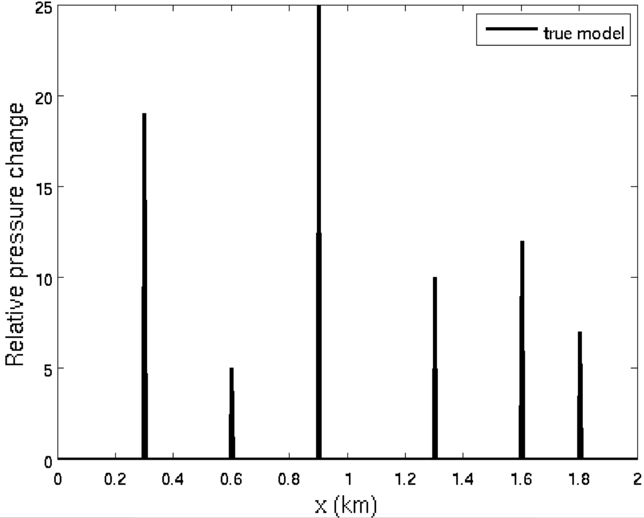


Surface deformation due to distributed dilatational (e.g. pressure) sources:

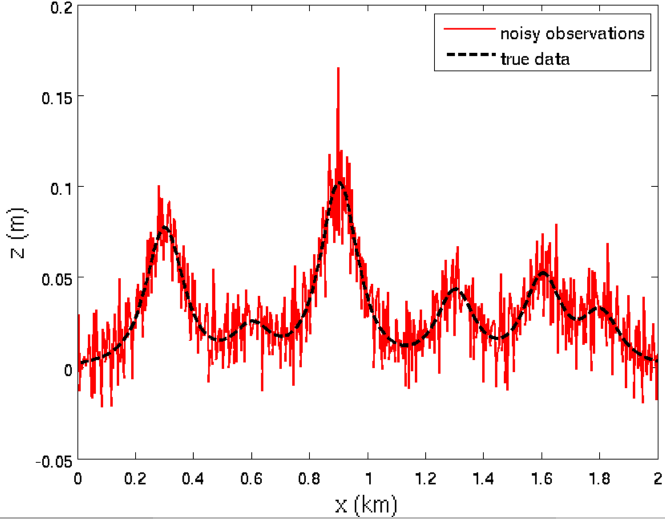
$$\begin{aligned} \mathbf{Au} &= d(x,y), \\ d(x,y) &= c\int_0^A\int_0^A \frac{Du(\xi,\eta)d\xi d\eta}{\left(D^2+(x-\xi)^2+(y-\eta)^2\right)^{3/2}}, \end{aligned} \quad (7)$$

Ill-conditioned inverse problem of resolving pressure from surface deformation...

True spiky pressure model

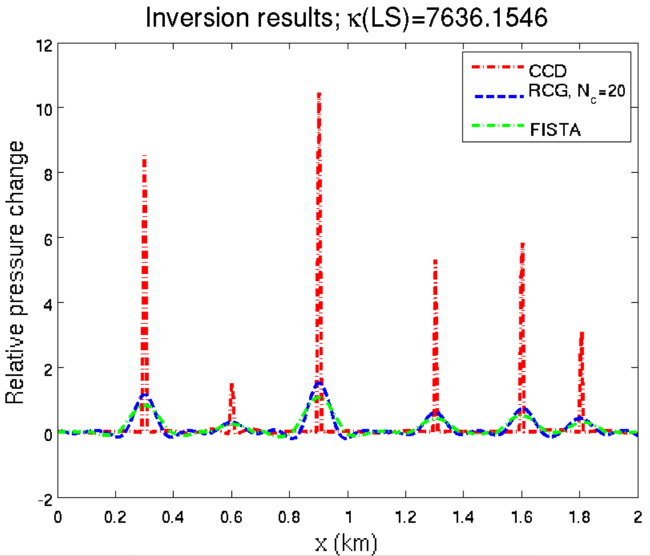


Noisy data

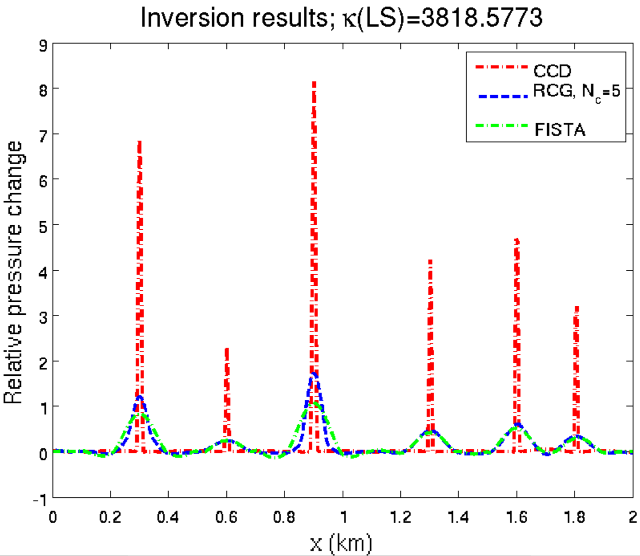


noisy observations
true data
0.2
0.15
0.1
0.05
0
-0.05
z (m)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
x (km)

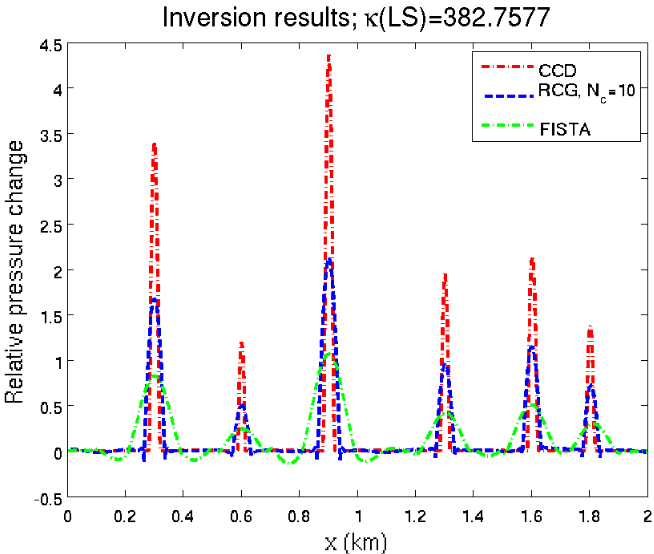
$\lambda = .05$, 100 iterations



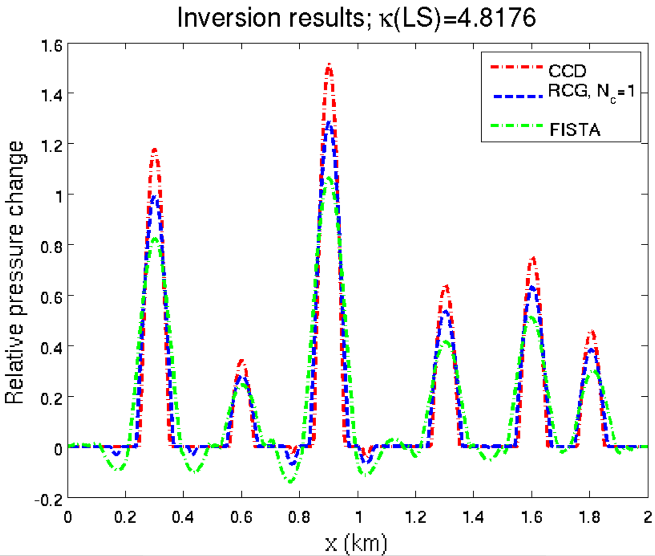
$\lambda = .1$, 100 iterations



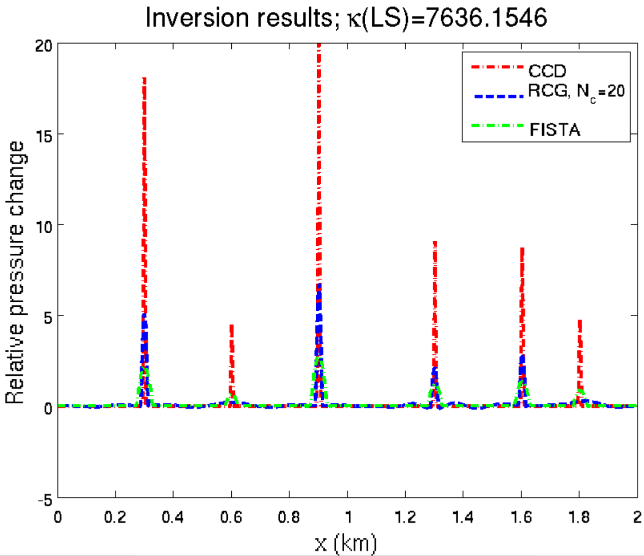
$\lambda = 1$, 100 iterations



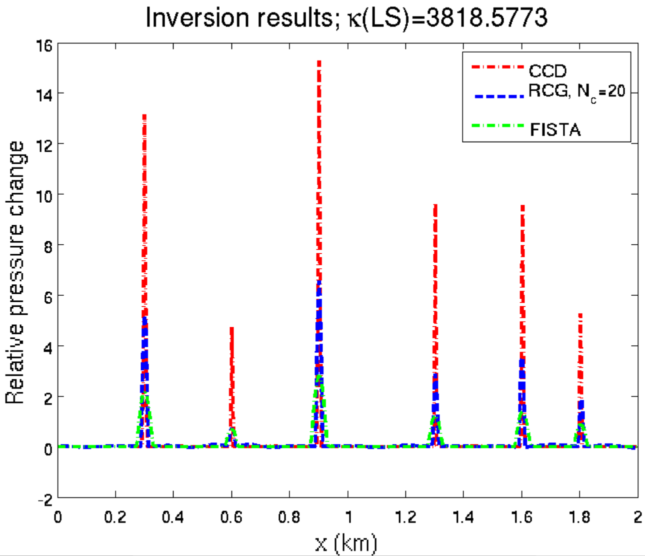
$\lambda = 100$, 1000 iterations



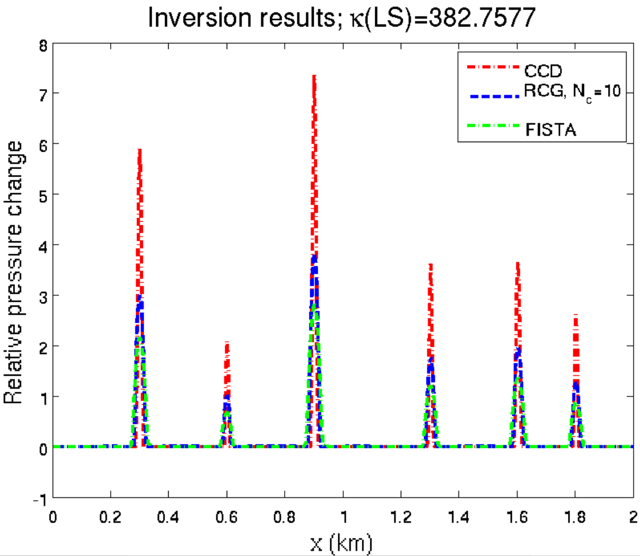
$\lambda = .05$, 1000 iterations



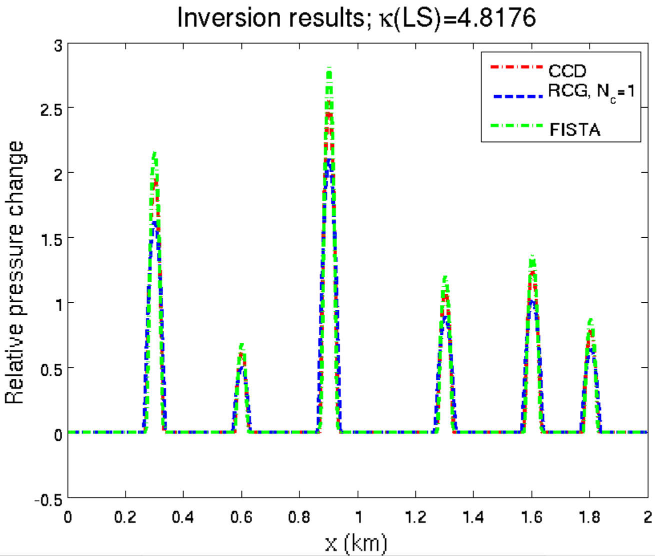
$\lambda = .1$, 1000 iterations



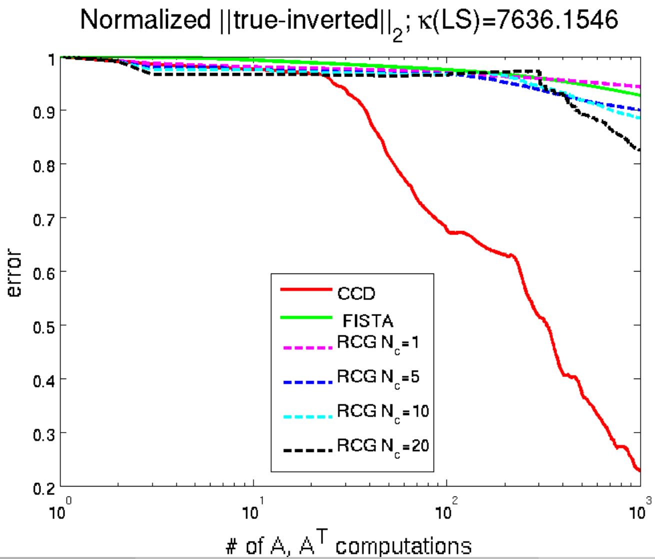
$\lambda = 1$, 1000 iterations



$\lambda = 100$, 1000 iterations

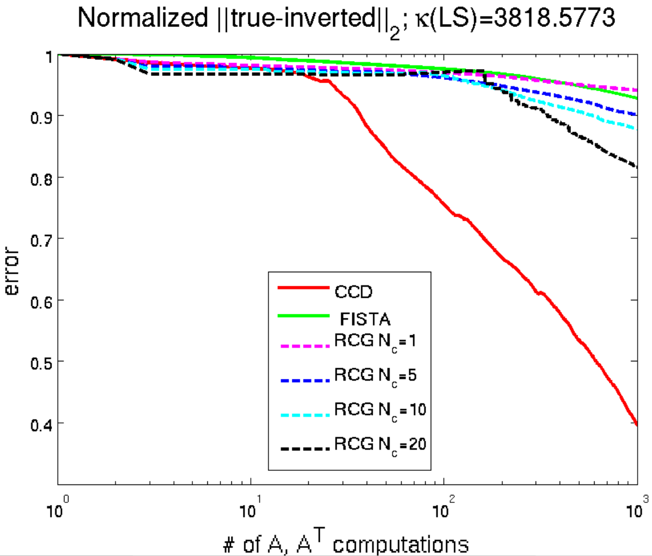


$\lambda = .05$

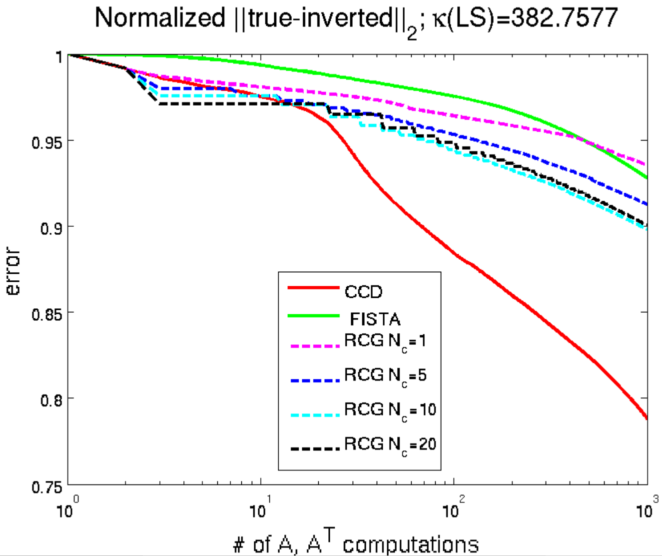


Normalized ||true-inverted||₂; κ(LS)=7636.1546
error
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
10^0
10^1
10^2
10^3
# of A, A^T computations
CCD
FISTA
RCG N_c=1
RCG N_c=5
RCG N_c=10
RCG N_c=20

$\lambda = .1$

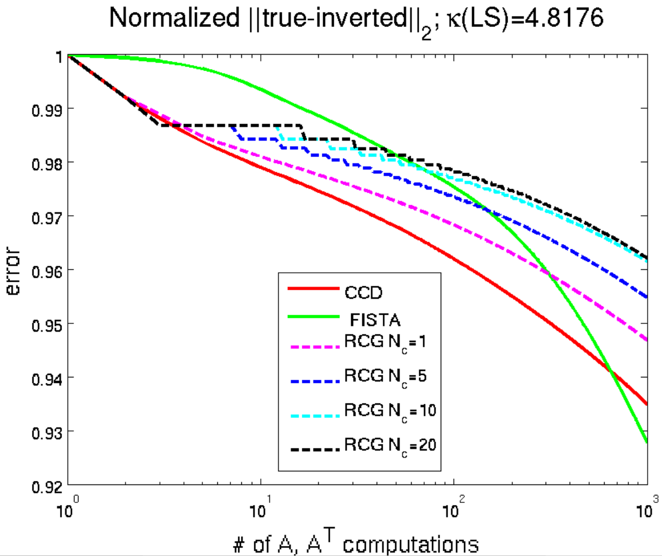

Normalized ||true-inverted||$_2$; κ(LS)=3818.5773
error
1
0.9
0.8
0.7
0.6
0.5
0.4
10^0
10^1
10^2
10^3
# of A, A^T computations
CCD
FISTA
RCG N_c=1
RCG N_c=5
RCG N_c=10
RCG N_c=20

$\lambda = 1$



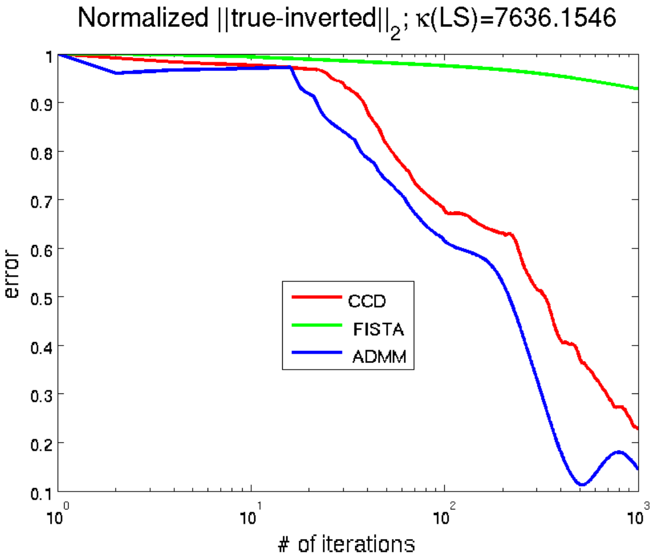
Normalized ||true-inverted||$_2$; κ(LS)=382.7577
error
# of A, A^T computations
1
0.95
0.9
0.85
0.8
0.75
10^0
10^1
10^2
10^3
CCD
FISTA
RCG N$_c$=1
RCG N$_c$=5
RCG N$_c$=10
RCG N$_c$=20

$\lambda = 100$

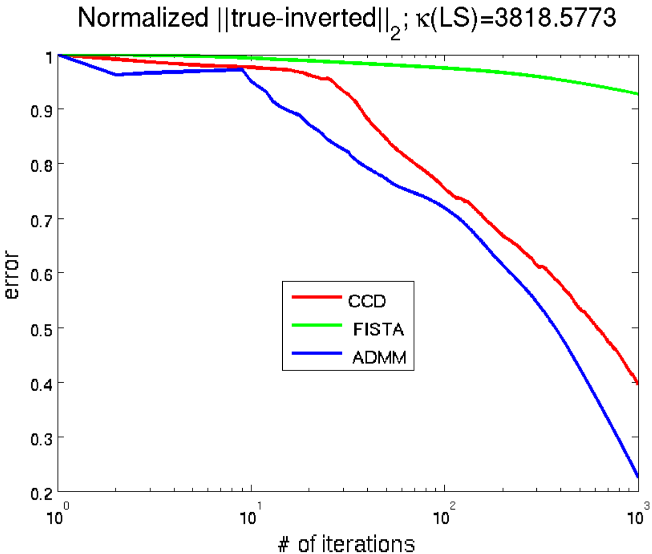


Normalized ||true-inverted||2; κ(LS)=4.8176
error
# of A, A^T computations
1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
10^0
10^1
10^2
10^3
CCD
FISTA
RCG Nc=1
RCG Nc=5
RCG Nc=10
RCG Nc=20

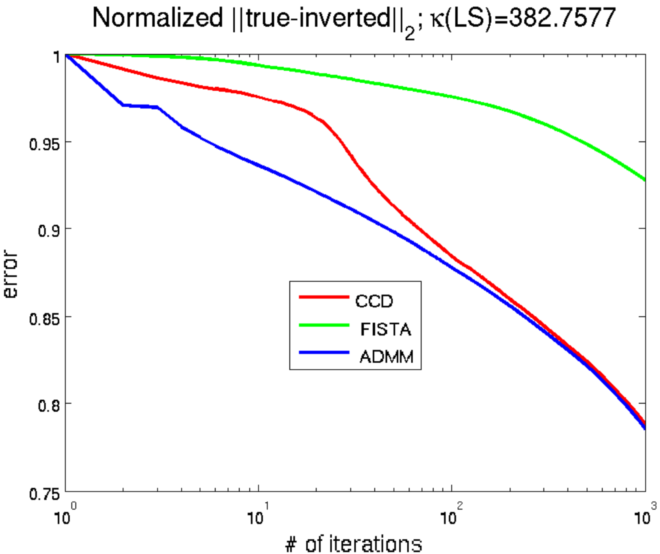
$\lambda = .05$, comparison with ADMM and exact solver



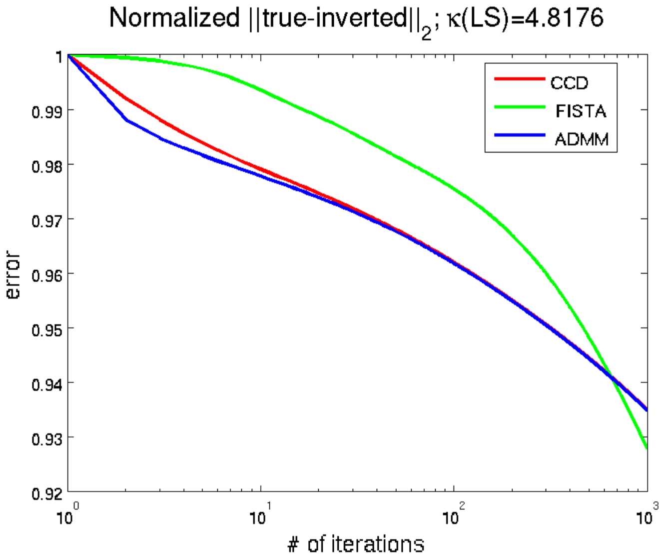
$\lambda = .1$, comparison with ADMM and exact solver



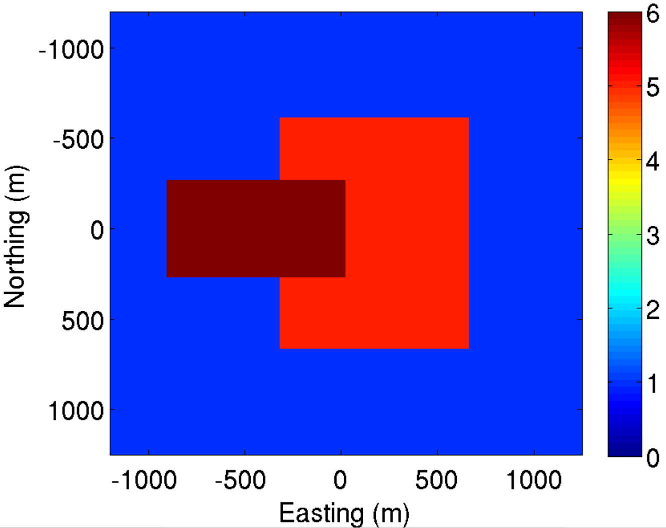
$\lambda = 1$, comparison with ADMM and exact solver



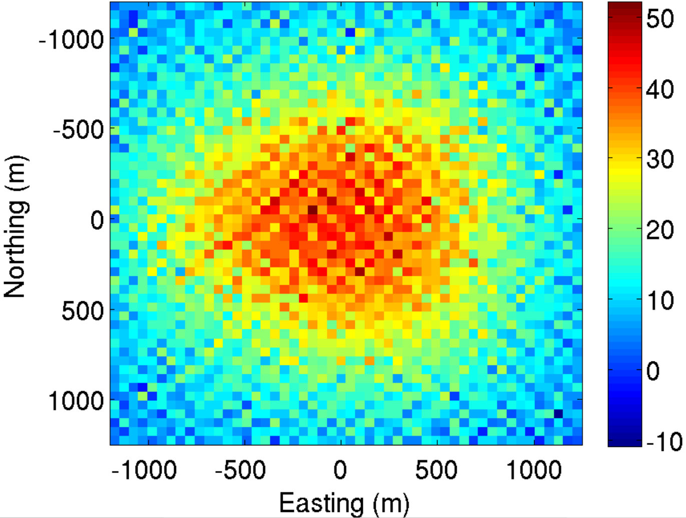
$\lambda = 100$, comparison with ADMM and exact solver



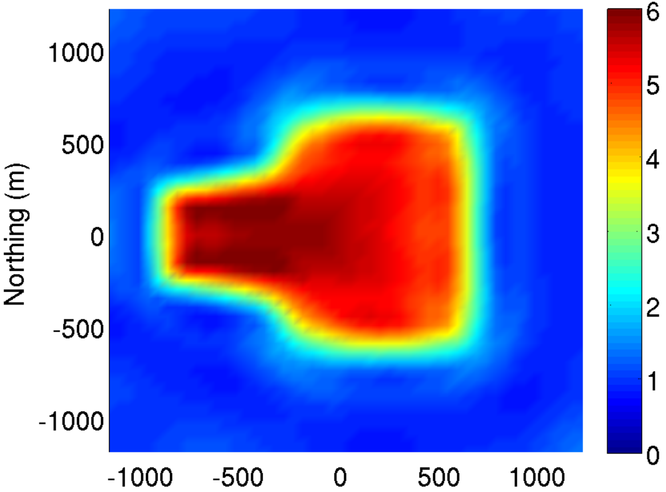
True blocky pressure model



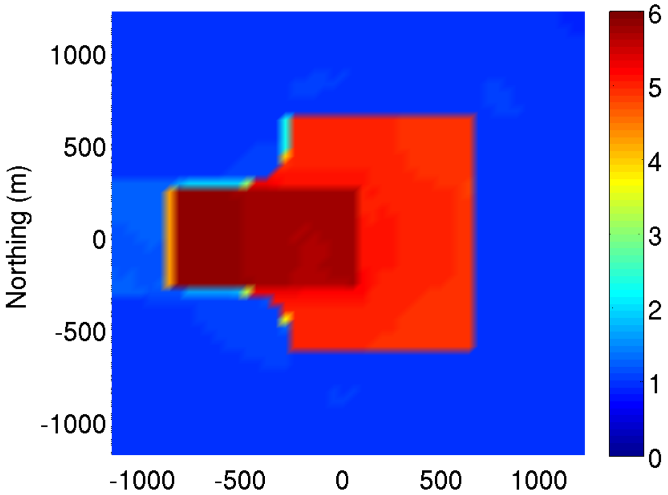
Noisy surface uplift data



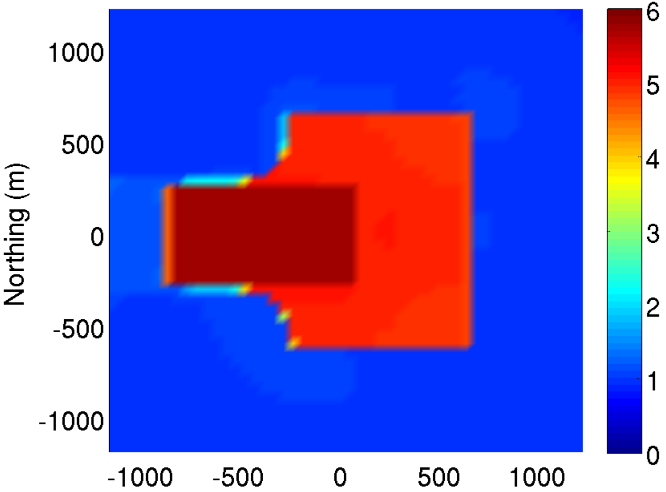
$\lambda = 10$, 100 iterations of RCG



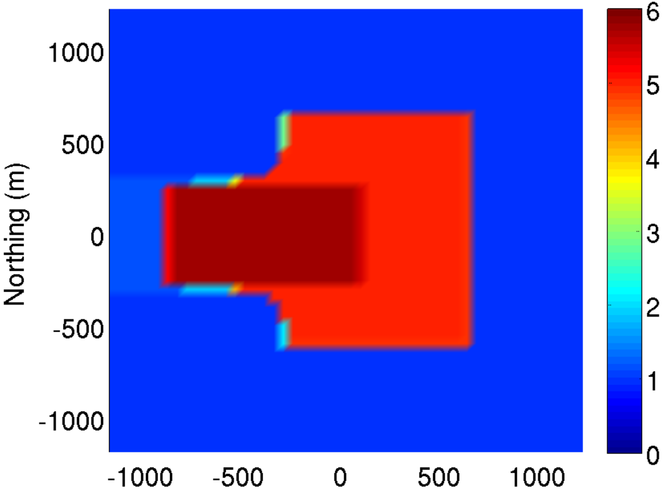
$\lambda = 10$, 100 iterations of CCD



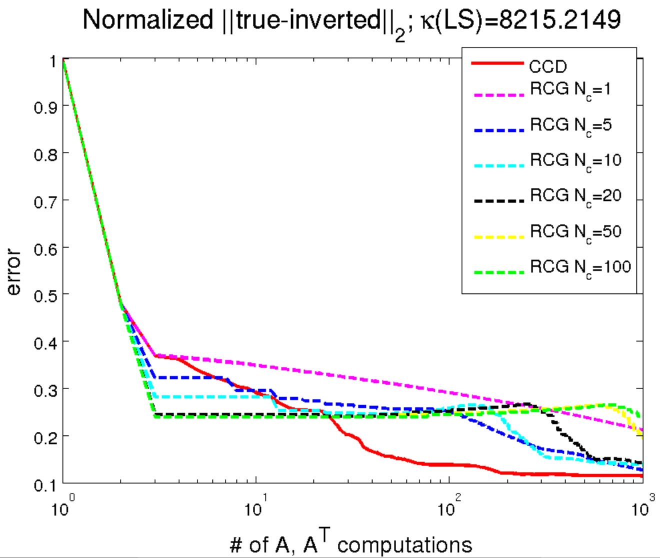
$\lambda = 10$, 1000 iterations of RCG



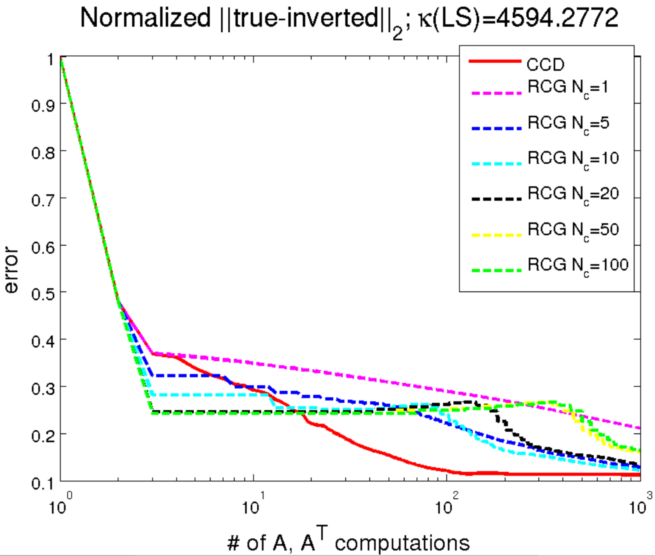
$\lambda = 10$, 1000 iterations of CCD



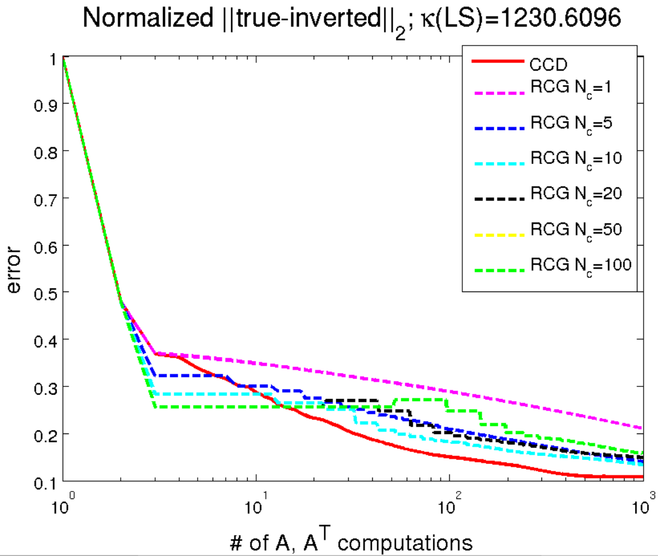
$\lambda = 5$, convergence comparison



$\lambda = 10$, convergence comparison



$\lambda = 50$, convergence comparison



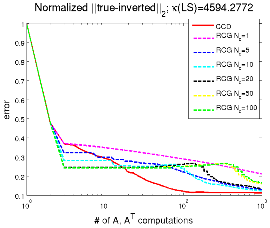
Conclusions and perspectives

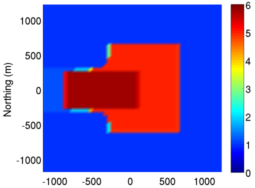
1. Adresses a **nasty** trade-off of classic methods: in practical problems **good** condition numbers of intermediate LS$\Leftrightarrow$**slow** ADMM convergence (Glowinski, 1982)
2. CCD beats alternatives when the intermediate LS is **ill-conditioned**!
3. Trades applications of the modeling operator for increased memory usage
4. A fast nonlinear extension has been developed and can be used with FWI
5. What's the expected ADMM convergence vs worst case?

Acknowledgements

The authors thank **Biondo Biondi** for a number of useful discussions, and the **affiliate members** of Stanford Exploration Project for their support.

Q&A


Normalized ||true-inverted||$_2$; κ(LS)=4594.2772
error
# of A, A^T computations
CCD
RCG N_c=1
RCG N_c=5
RCG N_c=10
RCG N_c=20
RCG N_c=50
RCG N_c=100


Northing (m)

Q&A