# Matrix-free Interior Point Method for Compressed Sensing Problems

Kimon Fountoulakis Jacek Gondzio Pavel Zhlobich

School of Mathematics
University of Edinburgh

3rd Conference On Optimization Methods and Software

May 13, 2012

## Existing solvers

1. FPC_AS by Wen, Yin, Goldfarb, Y. Zhang, H. Zhang.
2. SPGL1 by van den Berg, Friedlander.
3. GPSR by Figueiredo, Nowak, Wright
4. $\ell_1\_\ell_s$ by Kim, Koh, Lustig, Boyd, Gorinevsky.
5. PDCO by Saunders.
6. many others ...

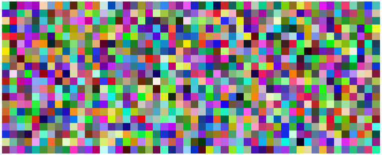
# Outline

# Objects

| Object | CS terminology | Optimization terminology |
|---|---|---|
| $x \in \mathbb{R}^n$ | sparse signal | decision variables |
| $b \in \mathbb{R}^m$ | noiseless measured signal | rhs, equality constraints |
| $e \in \mathbb{R}^m$ | vector of noise | slacks, inequality constraints |
| $b + e = \tilde{b} \in \mathbb{R}^m$ | noisy measured signal | rhs, inequality constraints |
| $A \in \mathbb{R}^{m \times n}$ | measurement matrix | constraint matrix, Eq./Ineq. |

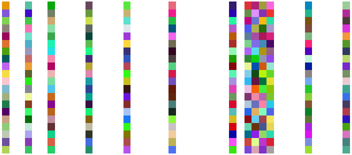
# CS objects and their properties

- $x$: $\|x\|_0 = k \ll n$ nonzero components.
- $A$: Satisfies the Restricted Isometry Property (RIP).

$A =$ 

# CS objects and their properties

- $x$: $\|x\|_0 = k \ll n$ nonzero components.
- $A$: Satisfies the Restricted Isometry Property (RIP).

$A =$ 

# CS objects and their properties

- $x$: $\|x\|_0 = k \ll n$ nonzero components.
- $A$: Satisfies the Restricted Isometry Property (RIP).

$\bar{A} =$ 

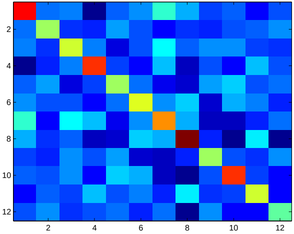
RIP

$$\left\| \frac{n}{m} \bar{A}^T \bar{A} - I_m \right\|_2 \leq \delta_k$$

# CS objects and their properties

- $x$: $\|x\|_0 = k \ll n$ nonzero components.
- $A$: Satisfies the Restricted Isometry Property (RIP).

$\bar{A} =$ 

$\bar{A}^T \bar{A} =$ 

# CS Objective

Find the k-sparse $x \in \mathbb{R}^n$, having $b \in \mathbb{R}^m$ with $m \ll n$.

**Noiseless measurements** $b$

**BP:**
$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \qquad \textit{subject to:} \quad Ax = b$$

**Noisy measurements** $\tilde{b}$

**BPDN:**
$$\min_{x \in \mathbb{R}^n} \quad \tau \|x\|_1 + \frac{1}{2}\|Ax - \tilde{b}\|_2^2$$

# BPDN for IPMs

$$\textbf{BPDN:} \quad \min_{x\in\mathbb{R}^n} \quad \tau\|x\|_1 + \frac{1}{2}\|Ax - \tilde{b}\|_2^2$$

| replace | with | using |
|---|---|---|
| $\|x\|_1$ | $\|x\|_1 = 1_{2n}^T z$ | $\lvert x_i\rvert = z_i + z_{i+n}$ |
| | $z \in \mathbb{R}^{2n} \geq 0$ | $z_i \geq 0, \ \forall i$ |
| $\|Ax - \tilde{b}\|_2^2$ | $F\tilde{b} + z^T F F^T z$ | $x_i = z_i - z_{i+n}$ |

$$F = \begin{bmatrix} A^T \\ -A^T \end{bmatrix} \in \mathbb{R}^{2n\times m}$$

$$FF^T = B = \begin{bmatrix} A^T \\ -A^T \end{bmatrix} \begin{bmatrix} A & -A \end{bmatrix} = \begin{bmatrix} A^TA & -A^TA \\ -A^TA & A^TA \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$$

# Primal-Dual BPDN Programs

**Primal BPDN**

$$\min_{z \in \mathbb{R}^{2n}} \quad c^T z + \tfrac{1}{2} z^T B^T z$$
$$\textit{subject to:} \quad z \geq 0$$

**Dual BPDN**

$$\max_{z,s \in \mathbb{R}^{2n}} \quad -\tfrac{1}{2} z^T B^T z$$
$$\textit{subject to:} \quad s - B^T z = -c$$
$$z, s \geq 0$$

$$B = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix} \in \mathbb{R}^{2n \times 2n}, \quad c = \begin{bmatrix} \tau 1_n - A^T \tilde{b} \\ \tau 1_n + A^T \tilde{b} \end{bmatrix} \in \mathbb{R}^{2n}$$

# Matrix-free IPM

Calculate the direction by solving:

**Reduced Newton system for Primal-Dual BPDN**

$$(B + \Theta^{-1}) \times \Delta z = *$$

$\Theta \in \mathbb{R}^{2n \times 2n} = S^{-1}Z$ and $Z, S \in \mathbb{R}^{2n \times 2n}$ are diagonal matrices

Matrix-free Regime

- Inexact solution using an iterative process, i.e CG.
- Only matrix-vector products are allowed.
- If the matrix $B + \Theta^{-1}$ is an operator then the process is memoryless.

# Solution of linear Systems for IPMs

Use the CG method to solve linear systems with the matrix:

$$M = B + \Theta^{-1} = \begin{bmatrix} A^TA & -A^TA \\ -A^TA & A^TA \end{bmatrix} + \begin{bmatrix} \Theta_1^{-1} & \\ & \Theta_2^{-1} \end{bmatrix}$$

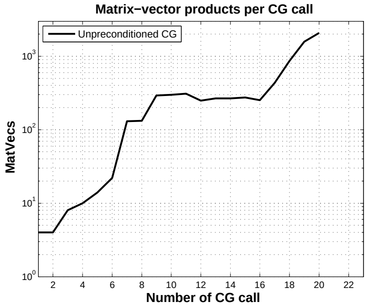
$\Theta_i \in \mathbb{R}^{n \times n}$, $i = 1, 2$, are the (i,i) blocks of $\Theta = S^{-1}Z$.

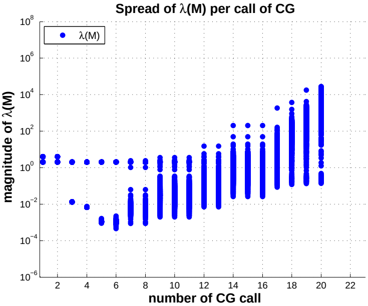
The fast convergence of the CG method depends on:

- The clustering of $\lambda(M)$. Better clustering $\Rightarrow$ faster CG.

How the $\lambda(M)$ behave as the IPM progresses?

# Slow convergence of the CG for IPMs




Required accuracy: 1.0e-6

# Objectives of Preconditioning

Reduce the computational efforts of the CG method with preconditioning.

- Introduce a matrix $P \in \mathbb{R}^{2n \times 2n}$ and solve instead:

$$P^{-1}M = P^{-1}*$$

An efficient preconditioner should:

1. Be easily invertible.
2. Cluster the eigenvalues $\lambda(P^{-1}M)$

# Proposed Preconditioner

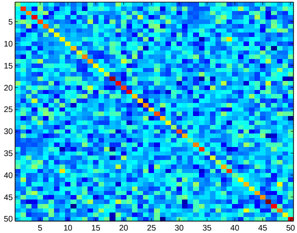
Approximate M:

$$M = \begin{bmatrix} A^TA & -A^TA \\ -A^TA & A^TA \end{bmatrix} + \begin{bmatrix} \Theta_1^{-1} & \\ & \Theta_2^{-1} \end{bmatrix}$$

with P:

$$P = \rho \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} + \begin{bmatrix} \Theta_1^{-1} & \\ & \Theta_2^{-1} \end{bmatrix}$$

---

$A^TA =$ 


$k$ entries of $\Theta^{-1} \to 0$

$2n - k$ entries of $\Theta^{-1} \to \infty$

# Spectral Properties of $P^{-1}M$

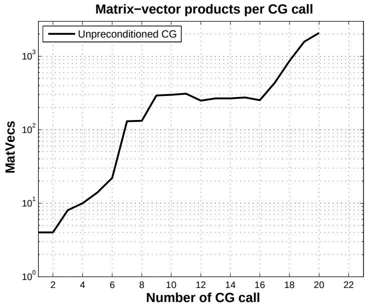
### Theorem

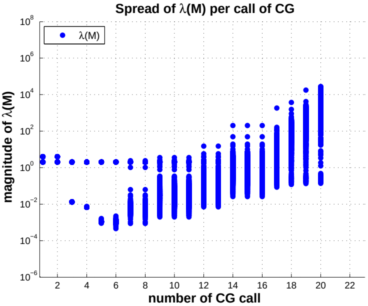
- Exactly $n$ eigenvalues of $P^{-1}M$ are 1.
- The remaining $n$ satisfy $|\lambda(P^{-1}M) - 1| \leq \delta_k + \frac{n}{m\delta_k L}$

where $L = \mathcal{O}(\max_i(\Theta_1 + \Theta_2)^{-1}) \to \infty$ and $\delta_k$ is the RIP-constant.

1. Universality of the proof! It holds $\forall A \in \mathbb{R}^{m\times n} \vdash$ RIP.
2. $P^{-1}$ works if $\|x\|_0 \leq$ maximum $k$ such that $A \vdash$ RIP.
3. $P^{-1}$ is not efficient if the problem is unsolvable.

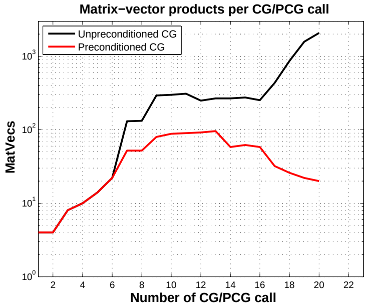
# Practical Performance

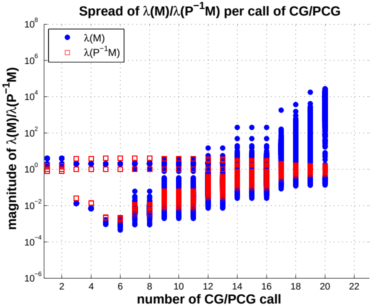



Required accuracy: 1.0e-6

# Practical Performance




Required accuracy: 1.0e-6

# Results I: Sparco Test Suite

Subset of noiseless and noisy Sparco problems.

Table: Number of matrix-vector products

| ID | rhs | Accuracy | mfipm | $\ell_1$_$\ell_s$ | pdco | fpc_as cg | spgl1_bp |
|---|---|---|---|---|---|---|---|
| 3 | $\tilde{b}$ | $1.0e-4$ | 761 | 518 | 8185 | 106 | 30000 |
| | $b$ | $1.0e-8$ | 1643 | 1246 | 31941 | 151 | 114 |
| 9 | $\tilde{b}$ | $1.0e-2$ | 1339 | 180 | 257 | 541 | 3054 |
| | $b$ | $1.0e-7$ | 2007 | 540 | 1341 | 293 | 29482 |
| 10 | $\tilde{b}$ | $1.0e-3$ | 3879 | 7829 | 5843 | 30002 | 30000 |
| | $b$ | $1.0e-8$ | 3783 | 27622 | 30785 | 9 | 22 |
| 701 | $\tilde{b}$ | $1.0e-2$ | 347 | 30202 | 10287 | 30002 | 30000 |
| | $b$ | $1.0e-5$ | 873 | 996 | 11541 | 30001 | 9523 |
| 702 | $\tilde{b}$ | $1.0e-3$ | 709 | 650 | 1897 | 30001 | 30000 |
| | $b$ | $1.0e-7$ | 1105 | 1576 | 3955 | 30002 | 7024 |
| 902 | $\tilde{b}$ | $1.0e-3$ | 333 | 348 | 185 | 40 | 29 |
| | $b$ | $1.0e-7$ | 913 | 871 | 351 | 42 | 65 |
| 903 | $\tilde{b}$ | $1.0e-2$ | 3643 | 17412 | 30189 | 30019 | 8529 |
| | $b$ | $1.0e-8$ | 16087 | 27912 | 21005 | 1764 | 30000 |

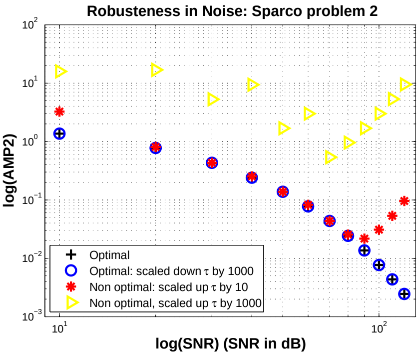
- The problems for which mfipm was first are denoted with red color.
- The problems for which mfipm was second are denoted with blue color.

# Noise robustness

- SNR$= 10, 20, \ldots, 120$ dB.
- The quality of reconstruction is measured by the amplification factor:

$$\text{AMP2} = \frac{\sqrt{\frac{1}{n}\|x - \hat{x}\|_2}}{\sqrt{\frac{1}{m}\|e\|_2}}$$

- The FPC_AS CG solver is used as a reference algorithm which produces the minimum AMP2.

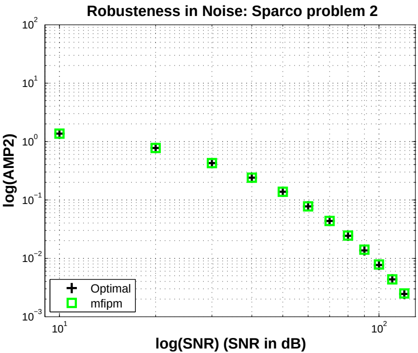


# Noise robustness

- SNR= $10, 20, \ldots, 120$ dB.
- The quality of reconstruction is measured by the amplification factor:

$$\text{AMP2} = \frac{\sqrt{\frac{1}{n}\|x - \hat{x}\|_2}}{\sqrt{\frac{1}{m}\|e\|_2}}$$

- The FPC_AS CG solver is used as a reference algorithm which produces the minimum AMP2.



# Conclusions & Further Research

Conclusions

- Matrix-free IPM
- Implements universal and robust preconditioner.
- Robust reconstruction for noisy measurements.
- Faster on some signal processing applications

Further Research

- Preconditioner for the non-asymptotic phase of an IPM.
- Exploit sparsity using heuristics.
- Exploit k-sparse operators, i.e sparse FFT
  `http://groups.csail.mit.edu/netmit/sFFT/`

# Thank you!

Kimon Fountoulakis, Jacek Gondzio, and Pavel Zhlobich.
Matrix-free interior point method for compressed sensing problems.
Technical report, School of Mathematics, The University of Edinburgh, 2012.