
Preconditioning Linear Systems in Interior Point Methods for Sparse Signal Reconstruction

Kimon Fountoulakis Jacek Gondzio Pavel Zhlobich

School of Mathematics
University of Edinburgh

3rd IMA Conference on Numerical Linear Algebra and Optimization

September 11, 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 l_1 - l_s by Kim, Koh, Lustig, Boyd, Gorinevsky.
- 5 PDCO by Saunders.
- 6 many others ...

Outline

Sparse signal reconstruction basics

- Objects & terminology

- Properties

- Objective

IPM basics for sparse signal reconstruction

- Formulations

- Matrix-free IPM

Sparse signal reconstruction and linear systems for IPMs

- Solution of linear systems for IPMs

- Convergence of the CG method for IPMs

- Preconditioning

Results

- Sparco

- Noise robustness

- Phase transition

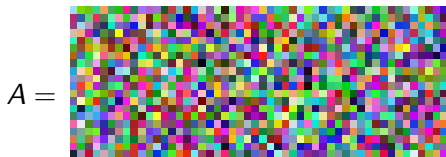
Conclusions & further research

Objects & terminology

Object	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.

Objects and their properties

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



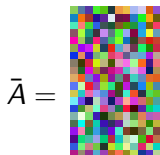
Objects and their properties

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



Objects and their properties

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

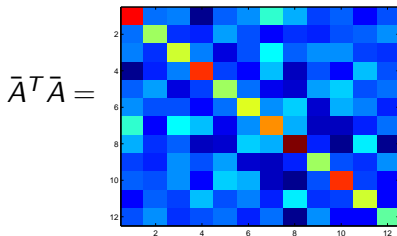


RIP

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

Objects and their properties

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



Objective of Sparse Signal Reconstruction

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

Noiseless
measurements b

BP:

$$\min_{x \in \mathbb{R}^n} \|x\|_1$$

subject to: $Ax = b$

Noisy
measurements \tilde{b}

BPDN:

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

BPDN for IPMs

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

replace

$$\|x\|_1$$

with

$$\begin{aligned} \|x\|_1 &= \mathbf{1}_{2n}^T z \\ z &\in \mathbb{R}^{2n} \geq 0 \end{aligned}$$

using

$$\begin{aligned} |x_i| &= z_i + z_{i+n} \\ z_i &\geq 0, \forall i \end{aligned}$$

$$\|Ax - \tilde{b}\|_2^2$$

$$F\tilde{b} + z^T F F^T z$$

$$x_i = z_i - z_{i+n}$$

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

$$F F^T = B = \begin{bmatrix} A^T \\ -A^T \end{bmatrix} [A \quad -A] = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$$

Primal-Dual BPDN Programs

Primal BPDN

$$\min_{z \in \mathbb{R}^{2n}} \quad c^T z + \frac{1}{2} z^T B^T z$$

subject to: $z \geq 0$

Dual BPDN

$$\max_{z, s \in \mathbb{R}^{2n}} \quad -\frac{1}{2} z^T B^T z$$

subject to: $B^T z - s = -c$
 $z, s \geq 0$

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

Matrix-free IPM

Calculate the direction by solving

Newton system for primal-dual BPDN

$$\begin{bmatrix} B & -I_{2n} \\ S & Z \end{bmatrix} \times \begin{bmatrix} \Delta z \\ \Delta s \end{bmatrix} = *$$

$\xrightarrow{\text{eliminate}}$

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:

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

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

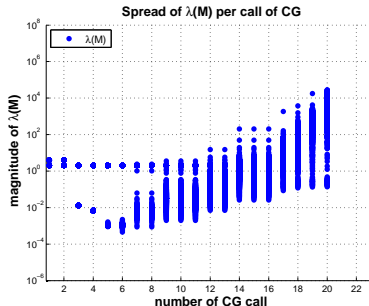
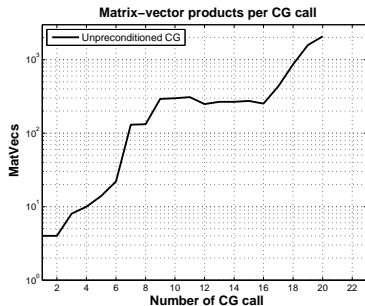
The fast convergence of the CG method depends on:

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

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

- └ Sparse signal reconstruction and linear systems for IPMs
- └ Convergence of the CG method for IPMs

Slow convergence of the CG for IPMs



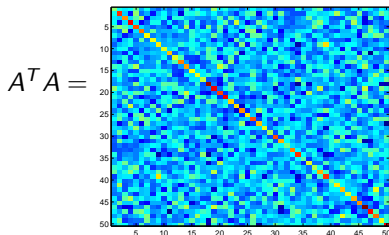
Required accuracy: $1.0e-6$

Preconditioning is necessary!

Proposed preconditioner

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

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



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

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

Spectral properties of $P^{-1}M$

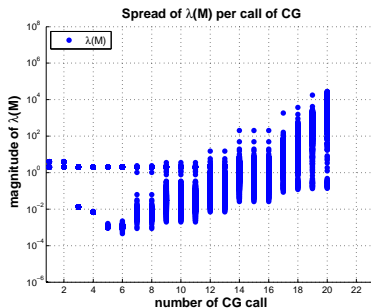
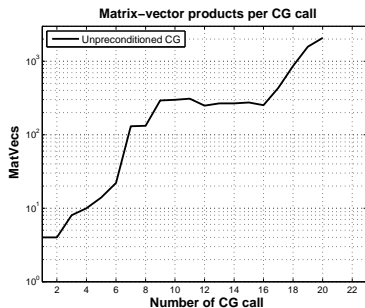
Theorem

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

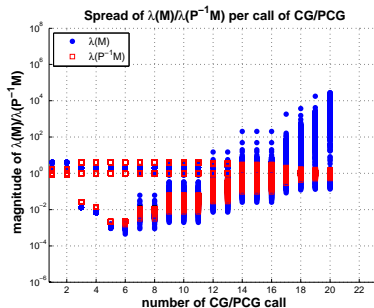
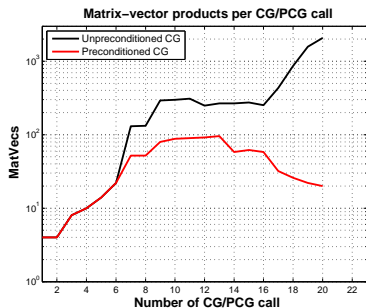
- 1 **Universality of the proof!** It holds $\forall A \in \mathbb{R}^{m \times n} \vdash \text{RIP}$.
- 2 P^{-1} works if $\|x\|_0 \leq \text{maximum } k \text{ such that } A \vdash \text{RIP}$.
- 3 P^{-1} is not efficient if the problem is unsolvable.

Practical performance



Required accuracy: $1.0e-6$

Practical performance



Required accuracy: $1.0e-6$

Results I: Sparco test suite

Table: Number of matrix-vector products

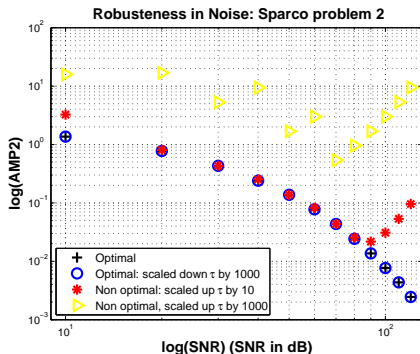
ID	rhs	Accuracy	mfipm	ℓ_1 - ℓ_s	pdco	fpc.as cg	spg11
2	\tilde{b}	3.0e-04	61	48	687	9	40000
	b	1.0e-11	65	98	40007	40002	22
3	\tilde{b}	7.0e-04	241	462	4941	106	40000
	b	1.0e-08	415	1612	40157	212	148
5	\tilde{b}	2.0e-03	5991	9842	28203	521	40000
	b	2.0e-05	7953	19684	41283	874	2567
7	\tilde{b}	4.0e-03	179	272	425	62	39
	b	1.0e-06	255	850	601	76	81
9	\tilde{b}	1.0e-03	689	1546	7065	1680	40000
	b	5.0e-12	649	1886	6845	40016	40000
10	\tilde{b}	1.0e-03	4775	8529	6203	40002	40000
	b	9.0e-10	4567	8192	41227	40161	40000
701	\tilde{b}	2.0e-02	947	1794	5967	1049	40000
	b	7.0e-09	1341	2656	42041	40017	15239
702	\tilde{b}	4.0e-03	809	1574	3341	40001	40000
	b	1.0e-07	1123	3030	49563	40157	11089
902	\tilde{b}	3.0e-04	181	588	261	40	40000
	b	2.0e-06	225	675	275	42	59
903	\tilde{b}	7.0e-03	2201	4944	5939	11395	8941
	b	3.0e-06	4173	25114	29901	40161	40000

Noise robustness

- ▶ SNR= 10, 20, ..., 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.

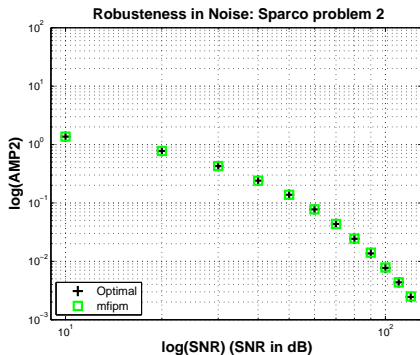


Noise robustness

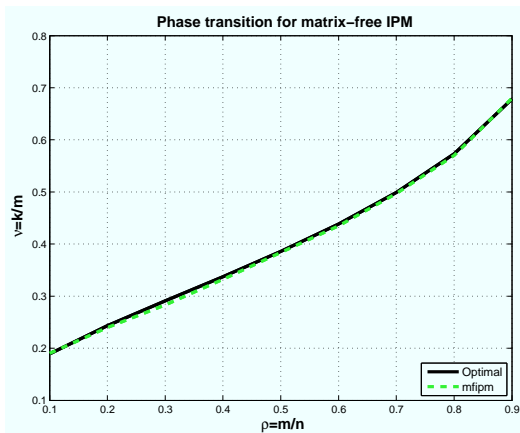
- ▶ SNR= 10, 20, ..., 120 dB.
- ▶ The quality of reconstruction is measured by the amplification factor:

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

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



Optimal phase transition



$A \in \mathbb{R}^{m \times n}$: partial Discrete Cosine Transform.

$x \in \mathbb{R}^n$: randomly placed ± 1 's at $k \ll n$ positions.

Conclusions

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

Further research

- ▶ Substantial decrement of the computational complexity via reduction of Newton's systems size (i.e fewer rows/columns).
 - ▶ Pros: The computational complexity will be greatly reduced.
 - ▶ Cons: The complexity of the matrix-free IPM will increase.

- ▶ Extend to ℓ_1 - *analysis* ($\|Wx\|_1$) and total variation ($\|x\|_{TV}$).

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.