Matrix-free Interior Point Method for Compressed Sensing Problems

Kimon Fountoulakis Jacek Gondzio Pavel Zhlobich

School of Mathematics
University of Edinburgh

21st International Symposium On Mathematical Programming

October 24, 2013

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 & terminology

$x \in \mathbb{R}^n$: sparse signal

$b \in \mathbb{R}^m$: noiseless measured signal

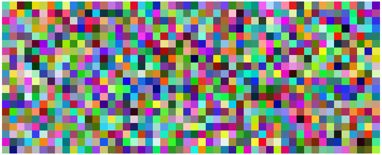
$e \in \mathbb{R}^m$: vector of noise

$\tilde{b} \in \mathbb{R}^m$: noisy measured signal

$A \in \mathbb{R}^{m \times n}$: measurement matrix

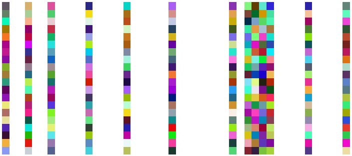
Objects and their properties

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

$A =$ 

Objects and their properties

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

$A =$ 

Objects and their properties

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

$\bar{A} =$ 

RIP

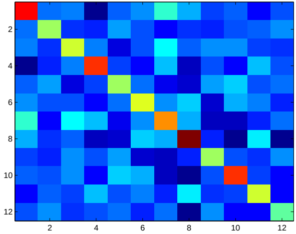
$$\left\| \frac{n}{m} \bar{A}^T \bar{A} - I_m \right\|_2 \leq \delta_k$$

Objects and their properties

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

$\bar{A} =$

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



Objective of CS

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

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

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

Primal-Dual BPDN Programs

Primal BPDN

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

Dual BPDN

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

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

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:

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

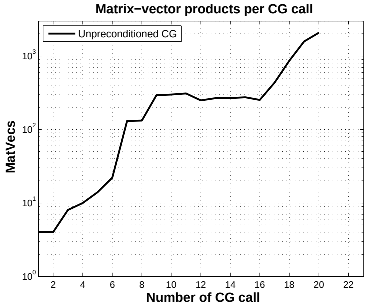
$\Theta_i \in \mathbb{R}^{n \times n}$, $i = 1, 2$, are the (i,i) blocks of $\Theta = S^{-1} Z$.

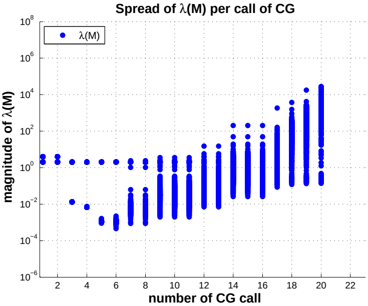
The fast convergence of the CG method depends on:

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

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

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 $\lambda(P^{-1}M)$

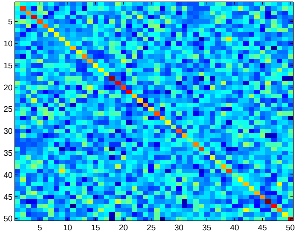
Proposed preconditioner

Approximate M:

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

with P:

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

$A^TA =$ 

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

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

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

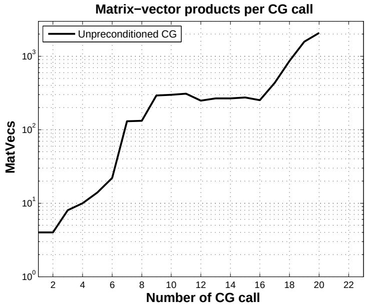
Theorem

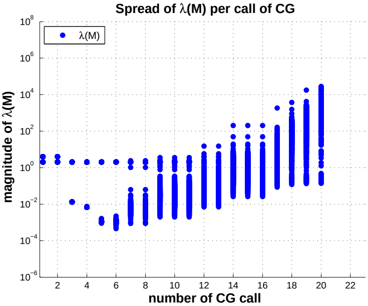
- Exactly n eigenvalues of $P^{-1}M$ are 1.
- The remaining n satisfy $|\lambda(P^{-1}M) - 1| \leq \delta_k + \frac{n}{m\delta_k L}$

where $L = \mathcal{O}(\max_i(\Theta_1 + \Theta_2)^{-1}) \to \infty$ and δ_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.

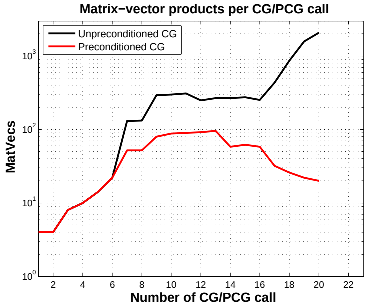
Practical performance

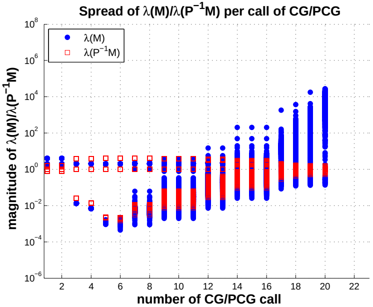




Required accuracy: 1.0e-6

Practical performance





Required accuracy: 1.0e-6

Results I: Sparco test suite

Table : Number of matrix-vector products

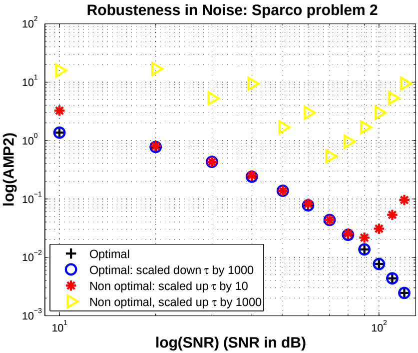
ID	rhs	Accuracy	mfipm	ℓ_1_ℓ_s	pdco	fpc_as cg	spgl1
2	$\tilde{b}$	3.0e-04	61	48	687	9	40000
	b	1.0e-11	65	98	40007	40002	22
3	$\tilde{b}$	7.0e-04	241	462	4941	106	40000
	b	1.0e-08	415	1612	40157	212	148
5	$\tilde{b}$	2.0e-03	5991	9842	28203	521	40000
	b	2.0e-05	7953	19684	41283	874	2567
7	$\tilde{b}$	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.

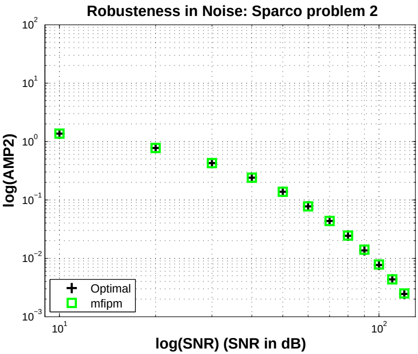


Noise robustness

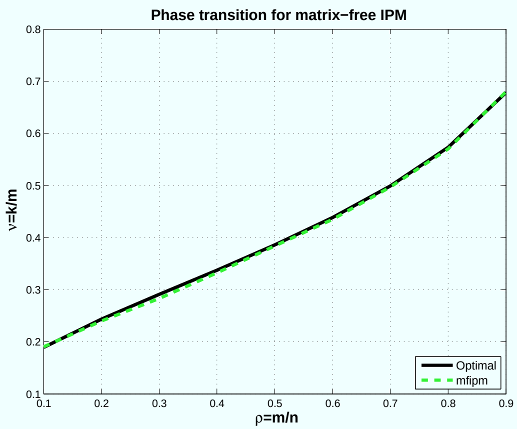
- SNR$= 10, 20, \ldots, 120$ dB.
- The quality of reconstruction is measured by the amplification factor:

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

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



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.
- Revision of the MATLAB software
  - Implementation of sparse directions.
  - Exploit k-sparse operators, i.e sparse FFT `http://groups.csail.mit.edu/netmit/sFFT/`.
  - Implementation of regularization techniques.
- Extend to $\ell_1 - \textit{analysis}$ ($\|Wx\|_1$) 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.
Submitted to Mathematical Programming Computation.