On linear algebra in interior point methods for solving ℓ_1-regularized optimization problems

Marco Viola
Department of Mathematics and Physics
University of Campania "L. Vanvitelli"
marco.viola@unicampania.it

Joint work with
V. De Simone – University of Campania (IT)
D. di Serafino – University of Naples Federico II (IT)
J. Gondzio and S. Pougkakiotis – University of Edinburgh (UK)

Due Giorni di Algebra Lineare e Applicazioni

Centro Congressi Federico II, Napoli (IT)
February 14-15, 2022

Università
degli Studi
della Campania
Luigi Vanvitelli
Dipartimento di Matematica e Fisica

PRIMO

Problem and goal

Efficient solution of a class of optimization problems that are very large and are expected to yield sparse solutions

$$\begin{aligned} \min_x \quad & f(x) + \tau_1 \|x\|_1 + \tau_2 \|Lx\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

$f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable convex function, $L \in \mathbb{R}^{l \times n}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m \leq n$, and $\tau_1, \tau_2 > 0$

$\|x\|_1$ and $\|Lx\|_1$ induce sparsity in x and/or in some dictionary Lx

- Many applications: portfolio optimization, signal/image processing, classification in statistics and machine learning, inverse problems, compressed sensing, ...
- Usually solved by specialized first-order methods, but those methods may be too expensive or struggle with not-so-well conditioned problems

Problem and goal (cont'd)

Non-smooth second-order methods:

- proximal (projected) Newton-type methods
- semi-smooth Newton methods combined with augmented Lagrangian methods

Problem and goal (cont'd)

Non-smooth second-order methods:

- proximal (projected) Newton-type methods
- semi-smooth Newton methods combined with augmented Lagrangian methods

Our goal:

show that Interior Point Methods (IPMs) can be equally or more efficient, robust and reliable than well-assessed first-order methods, by

- exploiting problem features in the linear algebra phase of IPMs
- taking advantage of the expected sparsity of the optimal solution

Outline of this talk

- Interior Point Methods (IPMs) for convex programming
- Interior Point-Proximal Method of Multipliers (IP-PMM)
- Application to TV-based Poisson Image Restoration
- Conclusions

NOTE: more applications in V. De Simone, D. di Serafino, J. Gondzio, S. Pougkakiotis & **MV**, *Sparse Approximations with Interior Point Methods*, to appear on SIAM Review, 2022

Modeling trick

Original formulation

$$\min_x \quad f(x) + \tau_1 \|x\|_1 + \tau_2 \|Lx\|_1 \qquad L \in \mathbb{R}^{l \times n},\ A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m,\ m \le n$$
$$\text{s.t.} \quad Ax = b$$

For any a, let $|a| = a^+ + a^-$, where $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$
Set $d = Lx \in \mathbb{R}^l$

New formulation

$$\begin{array}{cl} \min\limits_{x^+, x^-, d^+, d^-} & f(x^+ - x^-) + \tau_1 (e_n^\top x^+ + e_n^\top x^-) + \tau_2 (e_l^\top d^+ + e_l^\top d^-) \\ \text{s.t.} & A(x^+ - x^-) = b \\ & L(x^+ - x^-) = d^+ - d^- \\ & x^+, x^-, d^+, d^- \ge 0 \end{array}$$

$e_j \in \mathbb{R}^j$ vector of all 1's

Larger smooth problem, but IPMs are able to efficiently handle large sets of linear equality and non-negativity constraints!

(Primal-dual) IPMs for convex programming

Problem in standard form: $\min_x f(x), \quad \text{s.t.} \; Ax = b, \; x \geq 0$

Basic ideas of IPMs

- handle non-negativity constraints with a logarithmic barrier in the objective function
- approximately solve a sequence of barrier problems by using a (possibly inexact) Newton method

(Primal-dual) IPMs for convex programming

Problem in standard form: $\min_x f(x), \quad \text{s.t.} \;\; Ax = b, \;\; x \geq 0$

Basic ideas of IPMs

- handle non-negativity constraints with a logarithmic barrier in the objective function
- approximately solve a sequence of barrier problems by using a (possibly inexact) Newton method

At each iteration k

- barrier problem: $\min_x \; f(x) - \mu_k \sum_{j=1}^{n} \ln x^j, \quad \text{s.t.} \;\; Ax = b \quad (\mu_k > 0)$
- apply a Newton step to the first-order optimality conditions, i.e. solve the KKT system (here in augmented form)

$$\begin{bmatrix} -(\nabla^2 f(x_k) + \Theta_k^{-1}) & A^\top \\ A & 0_{m,m} \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \end{bmatrix} = \begin{bmatrix} \nabla f(x_k) - A^\top y_k - \sigma_k \mu_k X_k^{-1} e \\ b - A x_k \end{bmatrix}$$

$\Theta_k = X_k Z_k^{-1}$, $X_k = \text{diag}(x_k)$, $Z_k = \text{diag}(z_k)$, $x_k, z_k > 0$, $\sigma_k > 0$

(Primal-dual) IPMs for convex programming (cont'd)

- The augmented system can be solved either directly (by an appropriate factorization) or iteratively (by an appropriate Krylov subspace method)

  [D'Apuzzo, De Simone & di Serafino, COAP 2010; Gondzio, EJOR 2012; di Serafino & Orban, SISC 2021]

(Primal-dual) IPMs for convex programming (cont'd)

- The augmented system can be solved either directly (by an appropriate factorization) or iteratively (by an appropriate Krylov subspace method) [D'Apuzzo, De Simone & di Serafino, COAP 2010; Gondzio, EJOR 2012; di Serafino & Orban, SISC 2021]
- As $\mu_k \to 0$, an optimal solution of the barrier problem converges to an optimal solution of the original problem [Wright S., book 1997; Forsgren, Gill & Wright M., SIREV 2002]
- Polynomial convergence with respect to the number of variables has been proved for various classes of problems [Nesterov & Nemirovskii, SIAM Studies Appl Math 1994; Zhang, SIOPT 1994]

(Primal-dual) IPMs for convex programming (cont'd)

- The augmented system can be solved either directly (by an appropriate factorization) or iteratively (by an appropriate Krylov subspace method) [D'Apuzzo, De Simone & di Serafino, COAP 2010; Gondzio, EJOR 2012; di Serafino & Orban, SISC 2021]
- As $\mu_k \to 0$, an optimal solution of the barrier problem converges to an optimal solution of the original problem [Wright S., book 1997; Forsgren, Gill & Wright M., SIREV 2002]
- Polynomial convergence with respect to the number of variables has been proved for various classes of problems [Nesterov & Nemirovskii, SIAM Studies Appl Math 1994; Zhang, SIOPT 1994]
- Θ_k contains some very large and some very small elements close to optimality
  $\Longrightarrow$ the KKT matrix becomes increasingly ill-conditioned
  $\Longrightarrow$ regularization is beneficial
  [Friedlander & Tseng, SIOPT 2007; D'Apuzzo, De Simone & di Serafino, COAP 2010; Gondzio, EJOR 2012]

Regularization in IPMs

Use regularization to improve the spectral properties of the KKT matrix

- Dual regularization $\rightarrow$ (2,2) block:

$$0_{m,m} + \delta_k I_m, \quad \delta_k > 0 \quad ([A\ \delta I_m] \text{ full rank})$$

- Primal regularization $\rightarrow$ (1,1) block:

$$\nabla^2 f(x_k) + \Theta_k^{-1} + \rho_k I_n, \quad \rho_k > 0 \quad (\text{eigs bounded away from } 0)$$

A natural way of introducing regularization is through the use of proximal point methods [Altman & Gondzio, OMS 1999; Friedlander & Orban, Math Program Comput 2012; Pougkakiotis & Gondzio, COAP 2021]

Interior Point - Proximal Method of Multipliers (IP-PMM)

Merge IPM with PMM [Pougkakiotis & Gondzio, COAP 2021]

Problem formulation (equivalent to the standard one):

$$\min_x \; f(x), \quad \text{s.t.} \;\; Ax = b, \;\; x^{\mathcal{I}} \geq 0, \;\; x^{\mathcal{F}} \text{ free}$$

$\mathcal{I} \subseteq \{1, \ldots, n\}$, $\mathcal{F} = \{1, \ldots, n\} \setminus \mathcal{I}$

Iteration k: given an estimate ζ_k of a primal solution x^* and an estimate η_k for an optimal Lagrange multiplier vector y^* associated to $Ax = b$

- PMM: minimize the proximal penalty function ($\rho_k, \delta_k > 0$)

  $$\mathcal{L}^{PMM}_{\rho_k,\delta_k}(x; \zeta_k, \eta_k) = f(x) - \eta_k^\top (Ax - b) + \frac{1}{2\delta_k}\|Ax - b\|_2^2 + \frac{\rho_k}{2}\|x - \zeta_k\|_2^2$$

- IP-PMM: solve the PMM subproblem by applying one or more iters of IPM, i.e. alter the proximal penalty function with a barrier

  $$\mathcal{L}^{IP-PMM}_{\rho_k,\delta_k}(x; \zeta_k, \eta_k) = \mathcal{L}^{PMM}_{\rho_k,\delta_k}(x; \zeta_k, \eta_k) - \mu_k \sum_{j \in \mathcal{I}} \ln x^j$$

IP-PMM: Newton system

By writing the optimality conditions, applying a Newton step and performing straightforward computations we get the (symmetric indefinite) regularized augmented system

$$\begin{bmatrix} -(\nabla^2 f(x_k) + \Xi_k + \rho_k I_n) & A^\top \\ A & \delta_k I_m \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_{1,k} \\ r_{2,k} \end{bmatrix}$$

$$\Xi_k = \begin{bmatrix} 0_{|\mathcal{F}|,|\mathcal{F}|} & 0_{|\mathcal{I}|,|\mathcal{F}|} \\ 0_{|\mathcal{F}|,|\mathcal{I}|} & (X_k^{\mathcal{I}})^{-1}(Z_k^{\mathcal{I}}) \end{bmatrix}$$

NOTE: The (algorithmic) regularization in IP-PMM allows one to retrieve the solution of the original problem

TV-based Poisson image restoration

$$\min_{w} \; D_{KL}(w) + \lambda \|Lw\|_1$$
$$\text{s.t.} \;\; e_n^\top w = r, \quad w \geq 0$$

$$D_{KL}(w) = \sum_{j=1}^{m} \left(g^j \ln \frac{g^j}{(Dw+a)^j} + (Dw + a)^j - g^j \right)$$
$$L \in \mathbb{R}^{l \times n} \text{ discrete TV operator}, \;\; r = \sum_{j=1}^{m} (g^j - a^j)$$

- Object to be restored: $w \in \mathbb{R}^n$, measured data: $g \in \mathbb{N}_0^m$, with entries g^j that are realizations of m independent random variables $G^j \sim \mathit{Poisson}((Dw + a)^j)$
- $D \in \mathbb{R}^{m \times n}$ modeling the imaging system, $d^{ij} \geq 0$ for all i, j, $\sum_{i=1}^{m} d^{ij} = 1$ for all j; we assume periodic boundary conditions $\Rightarrow$ BCCB structure
- $a \in \mathbb{R}_+^m$ modeling the background radiation detected by the sensors
- Maximum-likelihood approach $\Longrightarrow$ minimization of Kullback-Leibler (KL) divergence (highly ill-conditioned problem) $\Longrightarrow$ TV regularization
- Non-negative image intensity, total image intensity preserved $\Longrightarrow$ non-negativity + single linear constraint

TV-based Poisson image restoration (cont'd)

Smooth problem reformulation

$$\min_{x} \quad f(x) \equiv D_{KL}(w) + c^\top u,$$
$$\text{s.t.} \quad Ax = b, \quad x \geq 0$$

$$d = Lw, \quad u = [(d^+)^\top,\ (d^-)^\top]^\top, \quad x = [w^\top,\ u^\top]^\top$$

$$A = \begin{bmatrix} e_n^\top & 0_l^\top & 0_l^\top \\ L & -I_l & I_l \end{bmatrix}$$

TV-based Poisson image restoration: Newton system

- $$\underbrace{\begin{bmatrix} -H_k & A^\top \\ A & \delta_k I \end{bmatrix}}_{M_k} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_{1,k} \\ r_{2,k} \end{bmatrix}, \qquad H_k = (\nabla^2 f(x_k) + \Theta_k^{-1} + \rho_k I)$$

  $\Longrightarrow$ use preconditioned MINimum RESidual (MINRES) method

- Block-diagonal preconditioner:

$$\widetilde{M}_k = \begin{bmatrix} \widetilde{H}_k & 0 \\ 0 & A\,\widetilde{H}_k^{-1} A^\top + \delta_k I \end{bmatrix}, \qquad \widetilde{H}_k \text{ diagonal approx of } H_k$$

Theorem

The eigenvalues of $\widetilde{M}_k^{-1} M_k$ *lie in the union of the intervals*

$$I_- = \Big[-\beta_H - 1, -\alpha_H \Big], \qquad I_+ = \Big[\frac{1}{1+\beta_H}, 1 \Big],$$

where $\alpha_H = \lambda_{\min}(\widehat{H}_k)$, $\beta_H = \lambda_{\max}(\widehat{H}_k)$ *and* $\widehat{H}_k = \widetilde{H}_k^{-\frac{1}{2}} H_k \widetilde{H}_k^{\frac{1}{2}}$.

[Bergamaschi, Gondzio, Martínez, Pearson & Pougkakiotis, NLAA 2021]

If $\widetilde{H}_k = \mathrm{diag}(H_k)$, then $\alpha_H \le 1 \le \beta_H$

TV-based Poisson image restoration: Newton sys (cont'd)

- $\begin{bmatrix} -H_k & A^\top \\ A & \delta_k I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_{1,k} \\ r_{2,k} \end{bmatrix}, \quad H_k = (\nabla^2 f(x_k) + \Theta_k^{-1} + \rho_k I)$

- $\nabla^2 f(x) = \begin{bmatrix} \nabla^2 D_{KL}(w) & 0 \\ 0 & 0 \end{bmatrix}, \quad \nabla^2 D_{KL}(w) = D^\top U(w)^2 D$

  where $U(w) = \text{diag}\left(\frac{\sqrt{g}}{Dw + a}\right)$

  D may be dense, but its action on a vector can be computed via FFT

- $\widetilde{H}_k = U(w_k)^2 + \Theta_k^{-1} + \rho_k I$, in practice performs better than $\widetilde{H}_k = \text{diag}(H_k)$

TV-based Poisson image restoration: test setting

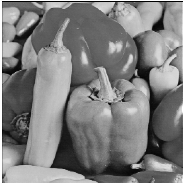
Test images

- 256×256, grayscale

cameraman

house

peppers



- Poisson noise and Gaussian blur (GB), motion blur (MB), out-of-focus blur (OF)

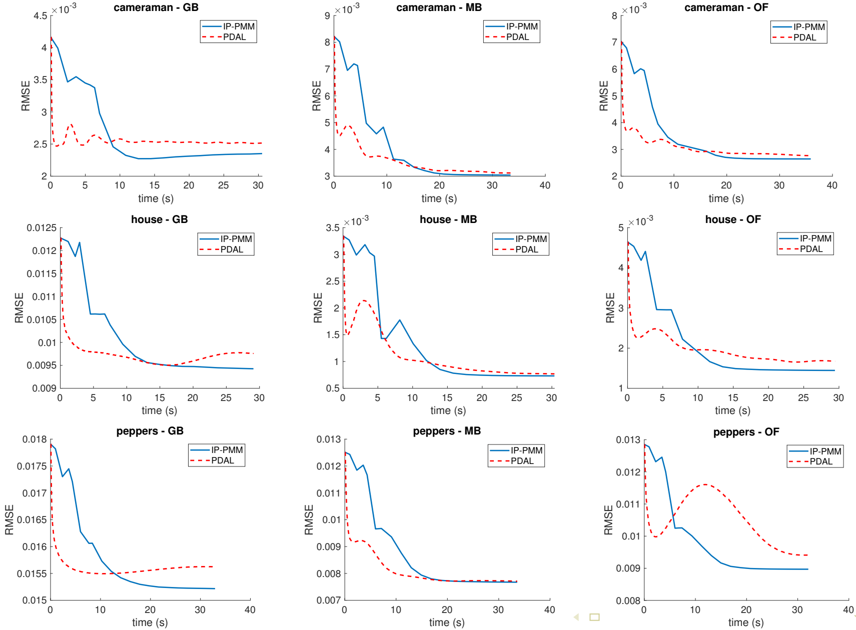
Comparison of IP-PMM with Primal-Dual Algorithm with Linesearch (PDAL)

MATLAB, implementation details in [De Simone, di Serafino, Gondzio, Pougkakiotis & **MV**, to appear on SIREV, 2022]

Performance metrics

- $\text{RMSE}(w) = \frac{1}{\sqrt{n}}\|w - \overline{w}\|_2$, $\overline{w}$ original image
- $\text{PSNR}(w) = 20\log_{10}(\max_i \overline{w}^i / \text{RMSE}(w))$
- MSSIM = structural similarity measure, the higher the better

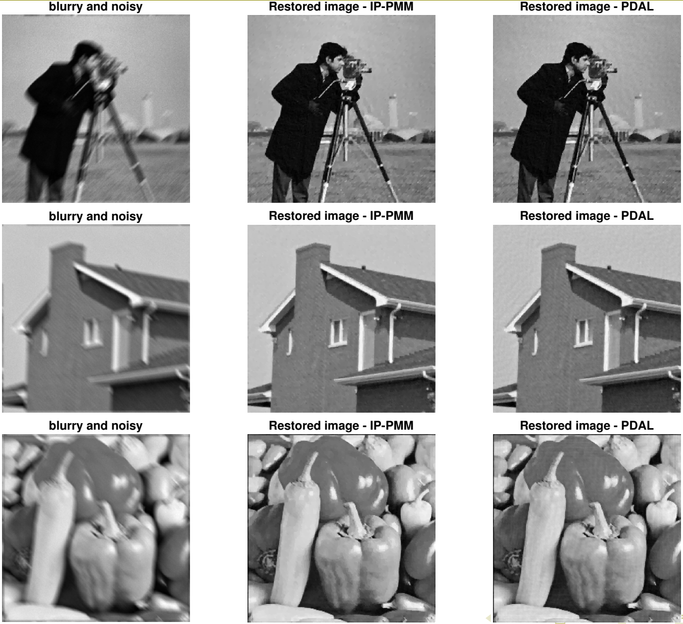
TV-based Poisson image restoration: results



TV-based Poisson image restoration: results (cont'd)

	IP-PMM			PDAL		
Problem	**RMSE**	**PSNR**	**MSSIM**	**RMSE**	**PSNR**	**MSSIM**
cameraman - GB	*4.85e−2*	*2.63e+1*	*8.33e−1*	5.02e−2	2.60e+1	8.22e−1
cameraman - MB	*5.52e−2*	*2.52e+1*	*8.11e−1*	5.59e−2	2.51e+1	7.77e−1
cameraman - OF	*5.14e−2*	*2.58e+1*	*7.98e−1*	5.26e−2	2.56e+1	7.62e−1
house - GB	*9.71e−2*	*2.03e+1*	*7.51e−1*	9.88e−2	2.01e+1	6.92e−1
house - MB	*2.70e−2*	*3.14e+1*	*8.67e−1*	2.77e−2	3.11e+1	8.43e−1
house - OF	*3.80e−2*	*2.84e+1*	*8.33e−1*	4.09e−2	2.78e+1	7.70e−1
peppers - GB	*1.23e−1*	*1.82e+1*	*7.46e−1*	1.25e−1	1.81e+1	6.57e−1
peppers - MB	*8.76e−2*	*2.12e+1*	*8.90e−1*	8.78e−2	2.11e+1	8.72e−1
peppers - OF	*9.47e−2*	*2.05e+1*	*8.01e−1*	9.70e−2	2.03e+1	6.60e−1

TV-based Poisson image restoration: results (cont'd) - MB



Conclusions

- Specialized IPMs for quadratic and general convex nonlinear optimization problems with sparse solutions have been developed
- By a proper choice of linear algebra solvers, IPMs can efficiently solve larger but smooth optimization problems coming from a standard reformulation of the original ones
- Computational experiments on diverse applications provide evidence that IPMs can offer a noticeable advantage over state-of-the-art first-order methods, especially when dealing with not-so-well conditioned problems
- This work may provide a basis for an in-depth analysis of the application of IPMs to many sparse approximation problems

Main reference: V. De Simone, D. di Serafino, J. Gondzio, S. Pougkakiotis, **MV**, *Sparse Approximations with Interior Point Methods*, to appear on SIAM Review, 2022

Thanks for your attention!

Enjoy your stay in Naples!