Primal Dual Regularized IPM: a Proximal Point perspective

Stefano Cipolla*, J. Gondzio*

Modern Techniques of Very Large Scale Optimization,
Edinburgh, 19th - 20th May 2022

*University of Edinburgh

Overview

The Problem & Investigation Perspectives

- Interior Point Methods
- Convergence & Stability
- Complexity & Rate of Convergence

Our Contribution [CG22]

- Part 1: Inexact Proximal Point Algorithm & Convergence
- Part 2: Replication of Variables & Complexity

The Problem & Investigation Perspectives

The Quadratic Programming (QP) Problem

In this talk we address the solution of the problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d} \ & f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} \\ \text{s.t. } & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{1}$$

using Interior Point Methods (IPM).

$H \in \mathbb{R}^{d\times d}$, $H \succeq 0$, $A \in \mathbb{R}^{m\times d}$, $\mathcal{C} \subset \{1, \dots, d\}$ and $\mathcal{F} := \{1, \dots, d\} \setminus \mathcal{C}$.

A is required to have full rank and we assume that the condition $m \leq d$ holds.

IPM overview [Gon12]

Replace inequality constraints with logarithmic barrier function, i.e.,

$$\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} - \mu \sum_{i\in\mathcal{C}} \ln(x_i)$$

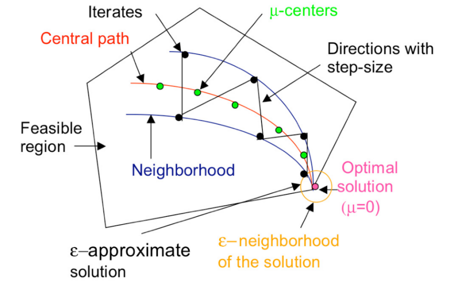
$$\text{s.t. } A\mathbf{x} = \mathbf{b}$$

IPM overview [Gon12]

Replace inequality constraints with logarithmic barrier function, i.e.,

$$\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} - \mu \sum_{i\in\mathcal{C}} \ln(x_i)$$

$$\text{s.t. } A\mathbf{x} = \mathbf{b}$$



- **Feasible region**

$$\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^d \text{ s.t. } A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x}_{\mathcal{C}} \geq 0\};$$

- **Directions** are computed solving a Newton linear system of the form

$$\begin{bmatrix} H + \Theta^{-1} & -A^T \\ A & 0 \end{bmatrix};$$

- $\max \Theta_{ii}^{-1} = O(\frac{1}{\mu})$ and $\min \Theta_{ii}^{-1} = O(\mu)$;
- μ is progressively driven to zero.

Investigation Perspective Part 1: Convergence & Stability

Primal-Dual Regularization of problem (1), i.e.,

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} + \frac{1}{2}\|\mathbf{x}-\mathbf{x}_k\|^2_{P_k} + \frac{1}{2}\|\mathbf{y}\|^2_{D_k} \\
\text{s.t. } & A\mathbf{x} + D_k(\mathbf{y}-\mathbf{y}_k) = \mathbf{b} \\
& \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free },
\end{aligned}
\tag{RP}
$$

has been proposed [AG99] to:

Investigation Perspective Part 1: Convergence & Stability

Primal-Dual Regularization of problem (1), i.e.,

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} + \frac{1}{2}\|\mathbf{x}-\mathbf{x}_k\|^2_{P_k} + \frac{1}{2}\|\mathbf{y}\|^2_{D_k} \\
\text{s.t. } & A\mathbf{x} + D_k(\mathbf{y}-\mathbf{y}_k) = \mathbf{b} \\
& \mathbf{x}_{\mathcal{C}} \geq 0,\ \mathbf{x}_{\mathcal{F}} \text{ free },
\end{aligned} \tag{RP}
$$

has been proposed [AG99] to:

- Alleviate (near) rank deficiency of A;
- Avoid factorization issues of the matrix

$$
\begin{bmatrix} H+\Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix}
$$

(or of its Schur complement).

Investigation Perspective Part 1: Convergence & Stability

Primal-Dual Regularization of problem (1), i.e.,

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^d} \; & \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} + \frac{1}{2}\|\mathbf{x}-\mathbf{x}_k\|^2_{P_k} + \frac{1}{2}\|\mathbf{y}\|^2_{D_k} \\
\text{s.t. } & A\mathbf{x} + D_k(\mathbf{y}-\mathbf{y}_k) = \mathbf{b} \\
& \mathbf{x}_{\mathcal{C}} \geq 0, \; \mathbf{x}_{\mathcal{F}} \text{ free },
\end{aligned} \tag{RP}
$$

has been proposed [AG99] to:

- Alleviate (near) rank deficiency of A;
- Avoid factorization issues of the matrix

$$
\begin{bmatrix} H+\Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix}
$$

(or of its Schur complement).

Indeed, applying the IPM to (RP), the corresponding Newton matrix assumes the form

$$
\begin{bmatrix} H+\Theta_k^{-1}+P_k & -A^T \\ A & D_k \end{bmatrix},
$$

and suitable choices of P_k, D_k may fix the above issues...

Investigation Perspective Part 1: Convergence & Stability

Primal-Dual Regularization of problem (1), i.e.,

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} + \frac{1}{2}\|\mathbf{x}-\mathbf{x}_k\|^2_{P_k} + \frac{1}{2}\|\mathbf{y}\|^2_{D_k} \\ \text{s.t. } & A\mathbf{x} + D_k(\mathbf{y}-\mathbf{y}_k) = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{RP}$$

has been proposed [AG99] to:

- Alleviate (near) rank deficiency of A;
- Avoid factorization issues of the matrix

$$\begin{bmatrix} H+\Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix}$$

(or of its Schur complement).

Indeed, applying the IPM to (RP), the corresponding Newton matrix assumes the form

$$\begin{bmatrix} H+\Theta_k^{-1}+P_k & -A^T \\ A & D_k \end{bmatrix},$$

and suitable choices of P_k, D_k may fix the above issues...

...but what about the convergence of the overall method?

Investigation Perspective Part 2: Complexity & Rate of Convergence

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

CONs

- *High computational* cost per iteration: one (or more) linear system(s) involving the iteration dependent matrix

$$\begin{bmatrix} H + \Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix} \qquad (*)$$

has to be so solved.

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

CONs

- *High computational* cost per iteration: one (or more) linear system(s) involving the iteration dependent matrix

$$\begin{bmatrix} H + \Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix} \qquad (*)$$

has to be so solved.

Alternating Direction Method of Multipliers (ADMM) [BPCE11]

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

CONs

- *High computational* cost per iteration: one (or more) linear system(s) involving the iteration dependent matrix

$$\begin{bmatrix} H + \Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix} \qquad (*)$$

has to be so solved.

Alternating Direction Method of Multipliers (ADMM) [BPCE11]

PROs

- Solves Primal-Dual pair;
- *Low computational* cost per iteration: one linear system involving always the same matrix

$$\begin{bmatrix} H + \beta I & -A^T \\ A & 0 \end{bmatrix}$$

has to be solved;

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

CONs

- *High computational* cost per iteration: one (or more) linear system(s) involving the iteration dependent matrix

$$\begin{bmatrix} H + \Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix} \qquad (*)$$

has to be so solved.

Alternating Direction Method of Multipliers (ADMM) [BPCE11]

PROs

- Solves Primal-Dual pair;
- *Low computational* cost per iteration: one linear system involving always the same matrix

$$\begin{bmatrix} H + \beta I & -A^T \\ A & 0 \end{bmatrix}$$

has to be solved;

CONs

- Slow rate of convergence;
- Not Suitable for high accuracy solutions.

Investigation Perspective Part 2: Complexity & Rate of Convergence

Interior Point Methods (IPM)

PROs

- Solves the Primal-Dual pair;
- Overall fast rate of convergence;
- Suitable for high accuracy solutions.

CONs

- *High computational* cost per iteration: one (or more) linear system(s) involving the iteration dependent matrix

$$\begin{bmatrix} H + \Theta_k^{-1} & -A^T \\ A & 0 \end{bmatrix} \qquad (*)$$

has to be so solved.

Alternating Direction Method of Multipliers (ADMM) [BPCE11]

PROs

- Solves Primal-Dual pair;
- *Low computational* cost per iteration: one linear system involving always the same matrix

$$\begin{bmatrix} H + \beta I & -A^T \\ A & 0 \end{bmatrix}$$

has to be solved;

CONs

- Slow rate of convergence;
- Not Suitable for high accuracy solutions.

A method which benefits from the advantages of these two methodologies should be able to **limit the dependence on Θ^{-1} in the Newton IPM matrix (*).**

Our Contribution [CG22]

The Inexact Proximal Point Framework [Luq84]

The Inexact Proximal Point Framework [Luq84]

Introduce

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + I_D(\mathbf{x}, \mathbf{y}),$$

where $I_D(\mathbf{x}, \mathbf{y})$ is the indicator function of the convex closed set

$$D := \mathbb{R}^{|\mathcal{C}|} \times \mathbb{R}_{\geq 0}^{d-|\mathcal{C}|} \times \mathbb{R}^m.$$

The Inexact Proximal Point Framework [Luq84]

Introduce

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + I_D(\mathbf{x}, \mathbf{y}),$$

where $I_D(\mathbf{x}, \mathbf{y})$ is the indicator function of the convex closed set

$$D := \mathbb{R}^{|\mathcal{C}|} \times \mathbb{R}_{\geq 0}^{d-|\mathcal{C}|} \times \mathbb{R}^m.$$

Consider the saddle sub-differential

$$T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) := \begin{bmatrix} \partial_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{y}) \\ \partial_{\mathbf{y}}(-\mathcal{L}(\mathbf{x}, \mathbf{y})) \end{bmatrix} = \begin{bmatrix} H\mathbf{x} + \mathbf{g} - A^T\mathbf{y} + \partial_{\mathbf{x}} I_D(\mathbf{x}, \mathbf{y}) \\ A\mathbf{x} - \mathbf{b} + \partial_{\mathbf{y}} I_D(\mathbf{x}, \mathbf{y}) \end{bmatrix},$$

The Inexact Proximal Point Framework [Luq84]

Introduce

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{g}^T\mathbf{x} - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + I_D(\mathbf{x}, \mathbf{y}),$$

where $I_D(\mathbf{x}, \mathbf{y})$ is the indicator function of the convex closed set

$$D := \mathbb{R}^{|\mathcal{C}|} \times \mathbb{R}_{\geq 0}^{d-|\mathcal{C}|} \times \mathbb{R}^m.$$

Consider the saddle sub-differential

$$T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) := \begin{bmatrix} \partial_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{y}) \\ \partial_{\mathbf{y}}(-\mathcal{L}(\mathbf{x}, \mathbf{y})) \end{bmatrix} = \begin{bmatrix} H\mathbf{x} + \mathbf{g} - A^T\mathbf{y} + \partial_{\mathbf{x}} I_D(\mathbf{x}, \mathbf{y}) \\ A\mathbf{x} - \mathbf{b} + \partial_{\mathbf{y}} I_D(\mathbf{x}, \mathbf{y}) \end{bmatrix},$$

then the Proximal Point Method (PPM) reads as

$$(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) = \mathcal{P}(\mathbf{x}_k, \mathbf{y}_k), \text{ where } \mathcal{P} = (I + \Sigma^{-1} T_{\mathcal{L}})^{-1} \text{ and } \Sigma := blockdiag(\rho I_d, \delta I_m).$$

The Inexact Proximal Point Framework

Evaluating the proximal operator $\mathcal{P}$ is equivalent to finding a solution to the problem

$$0 \in T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) + \Sigma((\mathbf{x}, \mathbf{y}) - (\mathbf{x}_k, \mathbf{y}_k)), \tag{2}$$

which is guaranteed to have a unique solution.

The Inexact Proximal Point Framework

Evaluating the proximal operator $\mathcal{P}$ is equivalent to finding a solution to the problem

$$0 \in T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) + \Sigma((\mathbf{x}, \mathbf{y}) - (\mathbf{x}_k, \mathbf{y}_k)), \tag{2}$$

which is guaranteed to have a unique solution.

Problem (2) is equivalent to:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H \mathbf{x} + \mathbf{g}^T \mathbf{x} + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 + \frac{\delta}{2}\|\mathbf{y}\|^2 \\ \text{s.t. } & A\mathbf{x} + \delta(\mathbf{y} - \mathbf{y}_k) = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{RP*}$$

The Inexact Proximal Point Framework

Evaluating the proximal operator $\mathcal{P}$ is equivalent to finding a solution to the problem

$$0 \in T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) + \Sigma((\mathbf{x}, \mathbf{y}) - (\mathbf{x}_k, \mathbf{y}_k)), \tag{2}$$

which is guaranteed to have a unique solution.

Problem (2) is equivalent to:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H \mathbf{x} + \mathbf{g}^T \mathbf{x} + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 + \frac{\delta}{2}\|\mathbf{y}\|^2 \\ \text{s.t. } & A\mathbf{x} + \delta(\mathbf{y} - \mathbf{y}_k) = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{RP*}$$

i.e., we need to solve problem (RP) where $P_k \equiv \rho I$ and $D_k \equiv \delta I$.

The Inexact Proximal Point Framework

Evaluating the proximal operator $\mathcal{P}$ is equivalent to finding a solution to the problem

$$0 \in T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) + \Sigma((\mathbf{x}, \mathbf{y}) - (\mathbf{x}_k, \mathbf{y}_k)), \tag{2}$$

which is guaranteed to have a unique solution.

Problem (2) is equivalent to:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H \mathbf{x} + \mathbf{g}^T \mathbf{x} + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 + \frac{\delta}{2}\|\mathbf{y}\|^2 \\ \text{s.t. } & A\mathbf{x} + \delta(\mathbf{y} - \mathbf{y}_k) = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{RP*}$$

i.e., we need to solve problem (RP) where $P_k \equiv \rho I$ and $D_k \equiv \delta I$.

Sufficient condition for convergence of PPM is that (RP*) **is solved exactly, but it is not necessary!**

The Inexact Proximal Point Framework

Evaluating the proximal operator $\mathcal{P}$ is equivalent to finding a solution to the problem

$$0 \in T_{\mathcal{L}}(\mathbf{x}, \mathbf{y}) + \Sigma((\mathbf{x}, \mathbf{y}) - (\mathbf{x}_k, \mathbf{y}_k)), \tag{2}$$

which is guaranteed to have a unique solution.

Problem (2) is equivalent to:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \ & \frac{1}{2}\mathbf{x}^T H \mathbf{x} + \mathbf{g}^T \mathbf{x} + \frac{\rho}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 + \frac{\delta}{2}\|\mathbf{y}\|^2 \\ \text{s.t. } & A\mathbf{x} + \delta(\mathbf{y} - \mathbf{y}_k) = \mathbf{b} \\ & \mathbf{x}_{\mathcal{C}} \geq 0, \ \mathbf{x}_{\mathcal{F}} \text{ free }, \end{aligned} \tag{RP*}$$

i.e., we need to solve problem (RP) where $P_k \equiv \rho I$ and $D_k \equiv \delta I$.

Sufficient condition for convergence of PPM is that (RP*) is solved exactly, but it is not necessary!

Inexact versions of the PPM are well understood [Luq84].

Proximal Stabilized Interior Point Method (PS-IPM) [CG22]

Input: $tol > 0$, $\sigma_r \in (0, 1)$.

Initialization: Iteration counter $k = 0$; initial point $(\mathbf{x}_0, \mathbf{y}_0)$

while *Stopping Condition False* **do**

  Use IPM with starting point $(\mathbf{x}_k^0, \mathbf{y}_k^0) = (\mathbf{x}_k, \mathbf{y}_k)$ to find $(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})$ s.t.

$$\|r_k(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\| < C\sigma_r^k \min\{1, \|(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - (\mathbf{x}_k, \mathbf{y}_k)\|$$

  Update the iteration counter: $k := k + 1$.

end

Algorithm 1: PS-IPM for QP

where $r_k(\mathbf{x}, \mathbf{y})$ is a computable residual associated with the *variational formulation of the problem*.

Proximal Stabilized Interior Point Method (PS-IPM) [CG22]

Input: $tol > 0$, $\sigma_r \in (0, 1)$.

Initialization: Iteration counter $k = 0$; initial point $(\mathbf{x}_0, \mathbf{y}_0)$

while *Stopping Condition False* **do**

Use IPM with starting point $(\mathbf{x}_k^0, \mathbf{y}_k^0) = (\mathbf{x}_k, \mathbf{y}_k)$ to find $(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})$ s.t.

$$\|r_k(\mathbf{x}_{k+1}, \mathbf{y}_{k+1})\| < C\sigma_r^k \min\{1, \|(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}) - (\mathbf{x}_k, \mathbf{y}_k)\|$$

Update the iteration counter: $k := k + 1$.

end

Algorithm 2: PS-IPM for QP

where $r_k(\mathbf{x}, \mathbf{y})$ is a computable residual associated with the *variational formulation of the problem*.

Primal-Dual CONVERGENCE IS GUARANTEED (if the problem is feasible)!

Important Observations on Algorithm 1

Important Observations on Algorithm 1

1. Asymptotic rate of convergence: (a is the "*Lipschitz constant* of $T_{\mathcal{L}}^{-1}(0)$")

$$\limsup_{k\to\infty} \frac{dist((\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), T_{\mathcal{L}}^{-1}(0))}{dist((\mathbf{x}_k, \mathbf{y}_k), T_{\mathcal{L}}^{-1}(0))} \leq \frac{a}{(a^2 + (1/\max\{\rho, \delta\})^2)^{1/2}} < 1,$$

i.e., ρ and δ should be chosen small for problems with *large* a.

Important Observations on Algorithm 1

1. Asymptotic rate of convergence: (a is the "*Lipschitz constant* of $T_{\mathcal{L}}^{-1}(0)$")

$$\limsup_{k\to\infty} \frac{dist((\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), T_{\mathcal{L}}^{-1}(0))}{dist((\mathbf{x}_k, \mathbf{y}_k), T_{\mathcal{L}}^{-1}(0))} \le \frac{a}{(a^2 + (1/\max\{\rho, \delta\})^2)^{1/2}} < 1,$$

   i.e., ρ and δ should be chosen small for problems with *large* a.

2. Warm Starting (the proximal operator is Lipschitz) and:

$$\begin{aligned} \|\mathcal{P}(\mathbf{t}_k) - \mathbf{t}_k\| &\le \|\mathcal{P}(\mathbf{t}_k) - \mathcal{P}(\mathbf{t}_{k-1})\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \\ &\le \eta\|\mathbf{t}_k - \mathbf{t}_{k-1}\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\|, \end{aligned}$$

   and using the convergence, we have

$$\|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \to 0 \text{ and } \|\mathbf{t}_k - \mathbf{t}_{k-1}\| \to 0.$$

Important Observations on Algorithm 1

1. Asymptotic rate of convergence: (a is the "*Lipschitz constant* of $T_{\mathcal{L}}^{-1}(0)$")

$$\limsup_{k\to\infty} \frac{dist((\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), T_{\mathcal{L}}^{-1}(0))}{dist((\mathbf{x}_{k}, \mathbf{y}_{k}), T_{\mathcal{L}}^{-1}(0))} \leq \frac{a}{(a^2 + (1/\max\{\rho, \delta\})^2)^{1/2}} < 1,$$

   i.e., ρ and δ should be chosen small for problems with *large* a.

2. Warm Starting (the proximal operator is Lipschitz) and:

$$\begin{aligned} \|\mathcal{P}(\mathbf{t}_k) - \mathbf{t}_k\| &\leq \|\mathcal{P}(\mathbf{t}_k) - \mathcal{P}(\mathbf{t}_{k-1})\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \\ &\leq \eta\|\mathbf{t}_k - \mathbf{t}_{k-1}\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\|, \end{aligned}$$

   and using the convergence, we have

$$\|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \to 0 \text{ and } \|\mathbf{t}_k - \mathbf{t}_{k-1}\| \to 0.$$

   $\Rightarrow$ The proximal sub-problems will need a non-increasing number of IPM iterations to be solved.

Important Observations on Algorithm 1

1. Asymptotic rate of convergence: (a is the "*Lipschitz constant* of $T_{\mathcal{L}}^{-1}(0)$")

$$\lim\sup_{k\to\infty} \frac{dist((\mathbf{x}_{k+1}, \mathbf{y}_{k+1}), T_{\mathcal{L}}^{-1}(0))}{dist((\mathbf{x}_k, \mathbf{y}_k), T_{\mathcal{L}}^{-1}(0))} \leq \frac{a}{(a^2 + (1/\max\{\rho, \delta\})^2)^{1/2}} < 1,$$

   i.e., ρ and δ should be chosen small for problems with *large* a.

2. Warm Starting (the proximal operator is Lipschitz) and:

$$\begin{aligned}\|\mathcal{P}(\mathbf{t}_k) - \mathbf{t}_k\| &\leq \|\mathcal{P}(\mathbf{t}_k) - \mathcal{P}(\mathbf{t}_{k-1})\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \\ &\leq \eta\|\mathbf{t}_k - \mathbf{t}_{k-1}\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\|,\end{aligned}$$

   and using the convergence, we have

$$\|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \to 0 \text{ and } \|\mathbf{t}_k - \mathbf{t}_{k-1}\| \to 0.$$

   $\Rightarrow$ The proximal sub-problems will need a non-increasing number of IPM iterations to be solved.

Item 2. justifies the computational evidence that *changing the reference point* at every IPM iteration works very well in practice [AG99; FO12], but convergence was not clear and proved only under the strong assumption of the uniform boundedness of the computed Newton directions [FO12]...

Important Observations on Algorithm 1

1. Asymptotic rate of convergence: (a is the "*Lipschitz constant* of $T_{\mathcal{L}}^{-1}(0)$")

$$\limsup_{k\to\infty} \frac{dist((\mathbf{x}_{k+1},\mathbf{y}_{k+1}), T_{\mathcal{L}}^{-1}(0))}{dist((\mathbf{x}_{k},\mathbf{y}_{k}), T_{\mathcal{L}}^{-1}(0))} \leq \frac{a}{(a^2 + (1/\max\{\rho,\delta\})^2)^{1/2}} < 1,$$

i.e., ρ and δ should be chosen small for problems with *large* a.

2. Warm Starting (the proximal operator is Lipschitz) and:

$$\begin{aligned}\|\mathcal{P}(\mathbf{t}_k) - \mathbf{t}_k\| &\leq \|\mathcal{P}(\mathbf{t}_k) - \mathcal{P}(\mathbf{t}_{k-1})\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \\ &\leq \eta\|\mathbf{t}_k - \mathbf{t}_{k-1}\| + \|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\|,\end{aligned}$$

and using the convergence, we have

$$\|\mathcal{P}(\mathbf{t}_{k-1}) - \mathbf{t}_k\| \to 0 \text{ and } \|\mathbf{t}_k - \mathbf{t}_{k-1}\| \to 0.$$

$\Rightarrow$ The proximal sub-problems will need a non-increasing number of IPM iterations to be solved.

Item 2. justifies the computational evidence that *changing the reference point* at every IPM iteration works very well in practice [AG99; FO12], but convergence was not clear and proved only under the strong assumption of the uniform boundedness of the computed Newton directions [FO12]...

...ASSUMPTION NOT NEEDED IN OUR APPROACH!

Item 2.:

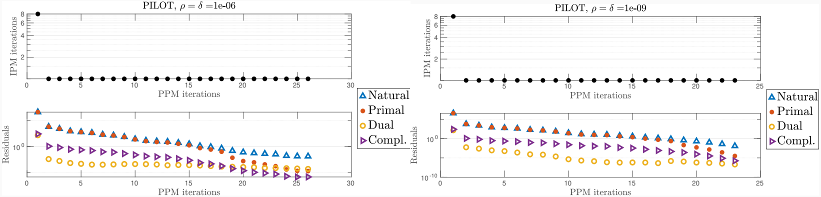


Figure 1: Problem PILOT. Upper panels: PPM Iterations & IPM Iterations. Lower panels: Behaviour of Residuals.

Item 2.:

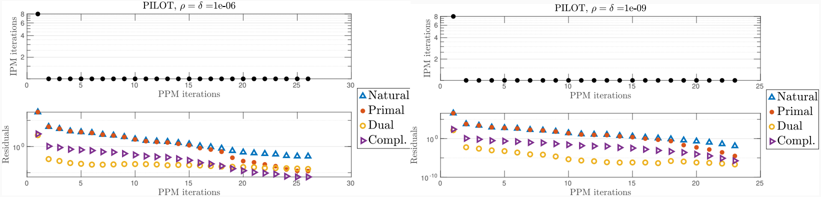


Figure 1: Problem PILOT. Upper panels: PPM Iterations & IPM Iterations. Lower panels: Behaviour of Residuals.

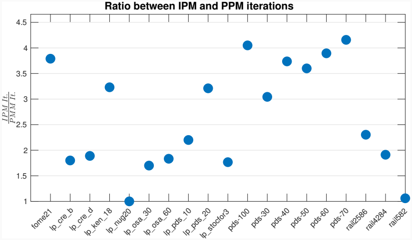


Figure 2: Average IPM sweeps per PPM iteration, Large Scale LP problems

Item 1.:

It is better to choose small regularization parameters rather than driving regularization to zero!

Item 1.:

It is better to choose small regularization parameters rather than driving regularization to zero!

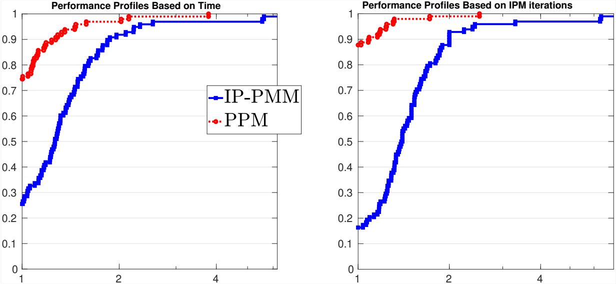


Figure 3: Performance Profiles for Netlib's LP problems.

Comparison with [PG21], where regularization is driven to zero.

Item 1.:

It is better to choose small regularization parameters rather than driving regularization to zero!

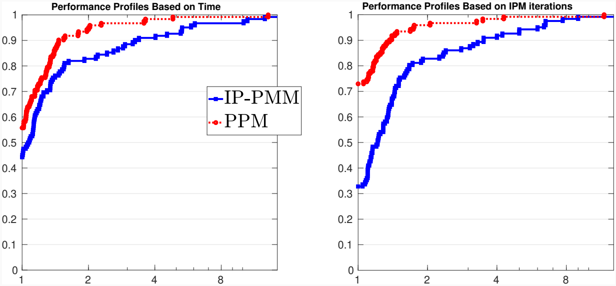


Figure 4: Performance Profiles for Maros–Mészáros test set.

Comparison with [PG21], where regularization is driven to zero.

It is better to choose small regularization parameters rather than driving regularization to zero!

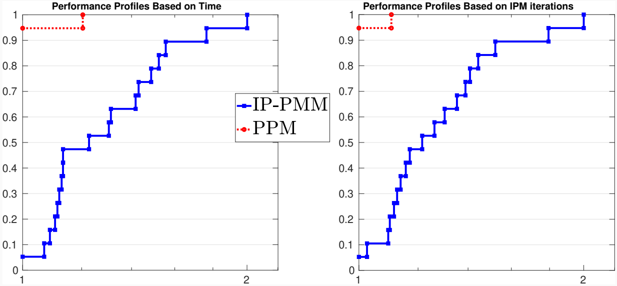


Figure 5: Performance Profiles for Large Scale LP Problems.

Comparison with [PG21], where regularization is driven to zero.

Replication of variables: the ADMM trick

Suppose we want to solve the problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^{d_1}} \ & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{s.t. } & B\mathbf{x} = \mathbf{f}, \\ & \mathbf{x} \geq 0, \end{aligned}$$

Replication of variables: the ADMM trick

Suppose we want to solve the problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^{d_1}} \ & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{s.t. } & B\mathbf{x} = \mathbf{f}, \\ & \mathbf{x} \geq 0, \end{aligned}$$

which can be reformulated as

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^{d_1}} \ & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{s.t. } & B\mathbf{x} = \mathbf{f}, \ \mathbf{x} - \mathbf{z} = 0 \\ & \mathbf{z} \geq 0. \end{aligned} \tag{VR}$$

This reformulation fits the original framework setting

$$d = 2d_1, \ m = d_1 + m_1, \ H = blockdiag(Q, 0), \ \mathbf{g} = \begin{bmatrix} \mathbf{c} \\ 0 \end{bmatrix}, \ A := \begin{bmatrix} B & 0 \\ I & -I \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}.$$

Applying the PS-IPM method to the reformulation (VR) we need to solve a Newton system of the form

$$\begin{bmatrix} Q+\rho I & 0 & -B^T & -I \\ 0 & \Theta^{-1}+\rho I & 0 & I \\ B & 0 & \delta I & 0 \\ I & -I & 0 & \delta I \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{z} \\ \Delta \mathbf{y}_1 \\ \Delta \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \xi_d^1 \\ \xi_d^2 + Z^{-1}\xi_{\mu,\sigma} \\ \xi_p^1 \\ \xi_p^2 \end{bmatrix},$$

LP/QP part vs IPM contribution & Variables Replication.

Applying the PS-IPM method to the reformulation (VR) we need to solve a Newton system of the form

$$\begin{bmatrix} Q+\rho I & 0 & -B^T & -I \\ 0 & \Theta^{-1}+\rho I & 0 & I \\ B & 0 & \delta I & 0 \\ I & -I & 0 & \delta I \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{z} \\ \Delta \mathbf{y}_1 \\ \Delta \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \xi_d^1 \\ \xi_d^2 + Z^{-1}\xi_{\mu,\sigma} \\ \xi_p^1 \\ \xi_p^2 \end{bmatrix},$$

LP/QP part vs IPM contribution & Variables Replication.

This system can be transformed as:

$$\underbrace{\begin{bmatrix} \Theta^{-1}+\rho I & -I & 0 & 0 \\ -I & -\delta I & I & 0 \\ 0 & I & Q+\rho I & B^T \\ 0 & 0 & B & -\delta I \end{bmatrix}}_{=:\mathcal{N}(\Theta)} \begin{bmatrix} \Delta \mathbf{z} \\ -\Delta \mathbf{y}_2 \\ \Delta \mathbf{x} \\ -\Delta \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} \xi_d^2 + Z^{-1}\xi_{\mu,\sigma} \\ \xi_p^2 \\ \xi_d^1 \\ \xi_p^1 \end{bmatrix}.$$

Applying the PS-IPM method to the reformulation (VR) we need to solve a Newton system of the form

$$\begin{bmatrix} Q+\rho I & 0 & -B^T & -I \\ 0 & \Theta^{-1}+\rho I & 0 & I \\ B & 0 & \delta I & 0 \\ I & -I & 0 & \delta I \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{z} \\ \Delta \mathbf{y}_1 \\ \Delta \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \xi_d^1 \\ \xi_d^2 + Z^{-1}\xi_{\mu,\sigma} \\ \xi_p^1 \\ \xi_p^2 \end{bmatrix},$$

LP/QP part vs IPM contribution & Variables Replication.

This system can be transformed as:

$$\underbrace{\begin{bmatrix} \Theta^{-1}+\rho I & -I & 0 & 0 \\ -I & -\delta I & I & 0 \\ 0 & I & Q+\rho I & B^T \\ 0 & 0 & B & -\delta I \end{bmatrix}}_{=:\mathcal{N}(\Theta)} \begin{bmatrix} \Delta \mathbf{z} \\ -\Delta \mathbf{y}_2 \\ \Delta \mathbf{x} \\ -\Delta \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} \xi_d^2 + Z^{-1}\xi_{\mu,\sigma} \\ \xi_p^2 \\ \xi_d^1 \\ \xi_p^1 \end{bmatrix}.$$

Since linear systems involving the red part are easy to solve, its solution can be obtained by reducing it further to the Schur complement.

The Schur complement of $\mathcal{N}(\Theta)$:

$$S(\Theta) := \begin{bmatrix} Q + \rho I + (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1} & B^T \\ B & -\delta I \end{bmatrix},$$

The Schur complement of $\mathcal{N}(\Theta)$:

$$S(\Theta) := \begin{bmatrix} Q + \rho I + (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1} & B^T \\ B & -\delta I \end{bmatrix},$$

which has exactly the same sparsity pattern as the Newton matrix of PS-IPM obtained without replication of variables, i.e.,

$$\mathcal{N}_C(\Theta) := \begin{bmatrix} Q + \rho I + \Theta^{-1} & A^T \\ A & -\delta I \end{bmatrix}.$$

The Schur complement of $\mathcal{N}(\Theta)$:

$$S(\Theta) := \begin{bmatrix} Q + \rho I + (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1} & B^T \\ B & -\delta I \end{bmatrix},$$

which has exactly the same sparsity pattern as the Newton matrix of PS-IPM obtained without replication of variables, i.e.,

$$\mathcal{N}_C(\Theta) := \begin{bmatrix} Q + \rho I + \Theta^{-1} & A^T \\ A & -\delta I \end{bmatrix}.$$

...

What advantages do we get from the new reformulation using VR?

$S(\Theta)$ "***varies less***" than $\mathcal{N}_C(\Theta)$

Consider $\widehat{\Theta}^{-1}$ and Θ^{-1} two IPM matrices obtained, respectively, in two different IPM iterations.

$S(\Theta)$ "***varies less***" than $\mathcal{N}_C(\Theta)$

Consider $\widehat{\Theta}^{-1}$ and Θ^{-1} two IPM matrices obtained, respectively, in two different IPM iterations.

Lemma ([CG22])

Define

$$D_A := (\delta I + (\widehat{\Theta}^{-1} + \rho I)^{-1})^{-1} - (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1}$$

and

$$D_C := \widehat{\Theta}^{-1} - \Theta^{-1}.$$

Then,

$$\|S(\widehat{\Theta}) - S(\Theta)\|_2 = \|D_A\|_2 < \|D_C\|_2 = \|\mathcal{N}_C(\widehat{\Theta}) - \mathcal{N}_C(\Theta)\|_2. \quad (3)$$

$S(\Theta)$ "*varies less*" than $\mathcal{N}_C(\Theta)$

Consider $\widehat{\Theta}^{-1}$ and Θ^{-1} two IPM matrices obtained, respectively, in two different IPM iterations.

Lemma ([CG22])

Define

$$D_A := (\delta I + (\widehat{\Theta}^{-1} + \rho I)^{-1})^{-1} - (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1}$$

and

$$D_C := \widehat{\Theta}^{-1} - \Theta^{-1}.$$

Then,

$$\|S(\widehat{\Theta}) - S(\Theta)\|_2 = \|D_A\|_2 < \|D_C\|_2 = \|\mathcal{N}_C(\widehat{\Theta}) - \mathcal{N}_C(\Theta)\|_2. \tag{3}$$

Suppose we computed a preconditioner for $S(\Theta)$, e.g., an incomplete factorization. Equation (3) shows that any accurate preconditioner for $S(\Theta)$ approximates $S(\widehat{\Theta})$ better than an accurate preconditioner for $\mathcal{N}_C(\Theta)$ would approximate $\mathcal{N}_C(\widehat{\Theta})$.

$S(\Theta)$ "***varies less***" than $\mathcal{N}_C(\Theta)$

Consider $\widehat{\Theta}^{-1}$ and Θ^{-1} two IPM matrices obtained, respectively, in two different IPM iterations.

Lemma ([CG22])

Define

$$D_A := (\delta I + (\widehat{\Theta}^{-1} + \rho I)^{-1})^{-1} - (\delta I + (\Theta^{-1} + \rho I)^{-1})^{-1}$$

and

$$D_C := \widehat{\Theta}^{-1} - \Theta^{-1}.$$

Then,

$$\|S(\widehat{\Theta}) - S(\Theta)\|_2 = \|D_A\|_2 < \|D_C\|_2 = \|\mathcal{N}_C(\widehat{\Theta}) - \mathcal{N}_C(\Theta)\|_2. \tag{3}$$

Suppose we computed a preconditioner for $S(\Theta)$, e.g., an incomplete factorization. Equation (3) shows that any accurate preconditioner for $S(\Theta)$ approximates $S(\widehat{\Theta})$ better than an accurate preconditioner for $\mathcal{N}_C(\Theta)$ would approximate $\mathcal{N}_C(\widehat{\Theta})$.

COMPUTED PRECONDITIONERS ARE *REUSABLE*!!

More in details:

$$(D_A)_{ii} = \frac{(D_C)_{ii}}{1 + \delta^2(\Theta_{ii}^{-1} + \rho)(\widehat{\Theta}_{ii}^{-1} + \rho) + \delta(\Theta_{ii}^{-1} + \rho) + \delta(\widehat{\Theta}_{ii}^{-1} + \rho)}.$$

More in details:

$$(D_A)_{ii} = \frac{(D_C)_{ii}}{1 + \delta^2(\Theta_{ii}^{-1} + \rho)(\widehat{\Theta}_{ii}^{-1} + \rho) + \delta(\Theta_{ii}^{-1} + \rho) + \delta(\widehat{\Theta}_{ii}^{-1} + \rho)}.$$

For large regularization parameters ρ, δ the IPM diagonal in the VR reformulation is expected to have limited changes.

Trade-off between PS-IPM rate of convergence and computational footprint related to the re-computation of preconditioners.

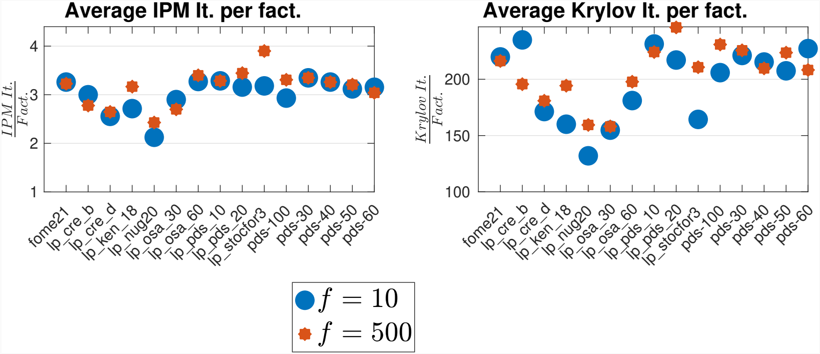
Numerical Results

Details:

- **Regularization** [PG21]: $reg = \max\{\frac{1}{\max\{\|A\|_\infty, \|Q\|_\infty\}}, 10^{-10}\}$ and we consider $\delta = \rho = f * reg$;
- linear systems involving $S(\widehat{\Theta})$ are solved using `GMRES(100,1)`;
- as preconditioner is used Matlab's `ldl` factorization of a previous $S(\Theta)$;
- the factorization is recomputed if GMRES needs more $Maxit/2$ iter.

Table 4: Large Scale Problems $f = 10$

Problem	PPM It.	IPM It.	Kryl. It.	Fact.	Time(s)	Obj Val	Reg. Par.	Status
Mittelmann/fome21	20	75	5057	23	700.63	47346318912.00	5.43e-09	opt
LPnetlib/lp_cre_b	23	48	3760	16	81.57	23129639.89	5.00e-09	opt
LPnetlib/lp_cre_d	22	46	3084	18	59.21	24454969.78	5.00e-09	opt
LPnetlib/lp_ken_18	14	38	2241	14	215.48	-52217025287.38	5.00e-09	opt
Qaplib/lp_nug20	17	17	1056	8	310.74	2181.64	1.25e-07	opt
LPnetlib/lp_osa_30	19	29	1548	10	42.96	2142139.87	5.00e-09	opt
LPnetlib/lp_osa_60	17	36	1992	11	121.06	4044072.51	5.00e-09	opt
LPnetlib/lp_pds_10	19	46	3239	14	80.81	26727094976.01	5.43e-09	opt
LPnetlib/lp_pds_20	19	60	4125	19	339.66	23821658640.00	5.43e-09	opt
LPnetlib/lp_stocfor3	32	35	1808	11	19.82	-39976.78	5.00e-09	opt
Mittelmann/pds-100	20	85	5971	29	5638.99	10928229968.00	5.00e-09	opt
Mittelmann/pds-30	22	77	5087	23	709.16	21385445736.00	5.43e-09	opt
Mittelmann/pds-40	20	75	4953	23	1265.16	18855198824.08	5.43e-09	opt
Mittelmann/pds-50	19	78	5188	25	1666.61	16603525724.02	5.43e-09	opt
Mittelmann/pds-60	19	82	5909	26	2655.46	14265904407.03	5.43e-09	opt
Mittelmann/pds-70	20	80	5763	26	3511.44	12241162812.00	5.43e-09	opt
Mittelmann/rail2586	34	84	5734	33	2412.17	936.55	5.00e-09	opt
Mittelmann/rail4284	35	76	5353	27	2892.35	1054.89	5.00e-09	opt
Mittelmann/rail582	35	35	2461	11	56.05	209.75	5.00e-09	opt



RED $\geq$ BLUE $\Rightarrow$ the computed factorization has been "*exploited more*".

Conclusions

- Primal-Dual Regularized IPM can be naturally framed in the context of the Inexact Proximal Point Method;

Conclusions

- Primal-Dual Regularized IPM can be naturally framed in the context of the Inexact Proximal Point Method;
- in this framework convergence and rate of convergence of the Primal Dual Regularized IPMs are clear;

Conclusions

- Primal-Dual Regularized IPM can be naturally framed in the context of the Inexact Proximal Point Method;
- in this framework convergence and rate of convergence of the Primal Dual Regularized IPMs are clear;
- using the trick of *replication of variables*, the computational footprint related to the re-computation of preconditioners can be greatly reduced;

Conclusions

- Primal-Dual Regularized IPM can be naturally framed in the context of the Inexact Proximal Point Method;
- in this framework convergence and rate of convergence of the Primal Dual Regularized IPMs are clear;
- using the trick of *replication of variables*, the computational footprint related to the re-computation of preconditioners can be greatly reduced;
- if *large* regularization parameters are allowed, virtually, one factorization would be enough!

THANK YOU FOR YOUR ATTENTION!

Minimal Bibliography from the talk:

[AG99] A. Altman and J. Gondzio. "Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization". In: vol. 11/12. 1-4. Interior point methods. 1999, pp. 275–302

[BPCE11] S. Boyd, N. Parikh, E. Chu, and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011

[CG22] S. Cipolla and J. Gondzio. "Proximal stabilized Interior Point Methods for quadratic programming and *low-frequency-updates* preconditioning techniques". In: *Preprint: https://arxiv.org/abs/2205.01775* (2022)

[FO12] M. P. Friedlander and D. Orban. "A primal-dual regularized interior-point method for convex quadratic programs". In: *Math. Program. Comput.* 4.1 (2012), pp. 71–107

[Gon12] J. Gondzio. "Interior point methods 25 years later". In: *European J. Oper. Res.* 218.3 (2012), pp. 587–601

[Luq84] F. J. Luque. "Asymptotic convergence analysis of the proximal point algorithm". In: *SIAM J. Control Optim.* 22.2 (1984), pp. 277–293

[PG21] S. Pougkakiotis and J. Gondzio. "An interior point-proximal method of multipliers for convex quadratic programming". In: *Comput. Optim. Appl.* 78.2 (2021), pp. 307–351

QUESTIONS?