The exact worst-case convergence rate of the alternating direction method of multipliers

Moslem Zamani (joint work with Etienne de Klerk and Hadi Abbaszadeh-peivasti)

February 13, 2023

Tilburg University

Alternating direction method of multipliers (ADMM)

ADMM

$$(\mathcal{P}): \qquad \min\ f(x) + g(z)\ \text{s.t.}\ Ax + Bz = b.$$

ADMM

$$(\mathcal{P}) : \qquad \min \ f(x) + g(z) \ \text{s.t.} \ Ax + Bz = b.$$

- Problem $(\mathcal{P})$ appears naturally (or after variable splitting) in many applications.
- The Lasso problem $\min_x \ \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$ may be formulated as

$$\min_{x,z} \ \frac{1}{2}\|Ax - b\|^2 + \lambda\|z\|_1 \ \text{s.t.} \ x - z = 0,$$

Preliminaries

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a closed proper convex function. Suppose that $L \in (0, \infty)$ and $\mu \in [0, \infty)$.

- The function f is called *L-smooth* if for any $x_1, x_2 \in \mathbb{R}^n$,

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\| \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

Preliminaries

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a closed proper convex function. Suppose that $L \in (0, \infty)$ and $\mu \in [0, \infty)$.

- The function f is called *L-smooth* if for any $x_1, x_2 \in \mathbb{R}^n$,

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\| \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

- The function f is called *μ-strongly* convex function if the function $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ is convex.

We denote the set of μ-strongly convex functions by $\mathcal{F}_\mu(\mathbb{R}^n)$.

$$(\mathcal{P}): \quad \min \; f(x) + g(z) \; \text{s.t.} \; Ax + Bz = b.$$

- $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$.
- $A \in \mathbb{R}^{r \times n}, B \in \mathbb{R}^{r \times m}$ and $b \in \mathbb{R}^r$.
- The matrix $\begin{pmatrix} A & B \end{pmatrix}$ has full row rank.
- $(x^\star, z^\star, \lambda^\star)$ is a saddle point.

ADMM

$$(\mathcal{P}): \qquad \min \; f(x) + g(z) \; \text{s.t.} \; Ax + Bz = b.$$

Set: $\lambda^0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m$, number of steps N and step length t. $z^0 = \hat{z}$.

for $k = 1, \ldots, N$

$x^k = \arg\min_x f(x) + \langle \lambda^{k-1}, Ax + Bz^{k-1} - b \rangle + \frac{t}{2} \|Ax + bz^{k-1} - b\|^2$

$z^k = \arg\min_z g(z) + \langle \lambda^{k-1}, Ax^k + Bz - b \rangle + \frac{t}{2} \|Ax^k + bz - b\|^2$

$\lambda^k = \lambda^{k-1} + t \left(Ax^k + Bz^k - b \right)$

Variational inequality format

- Dual objective:

$$D(\lambda) = \min_{x,z} f(x) + g(z) + \langle \lambda, Ax + Bz - b \rangle$$
$$= -\langle \lambda, b \rangle - f^*(-A^T\lambda) - g^*(-B^T\lambda).$$

- $\max_{\lambda \in \mathbb{R}^r} \; D(\lambda)$ (under some mild conditions) is equivalent to

$$0 \in \phi(\lambda) + \psi(\lambda),$$

where $\phi(\lambda) = A\partial f^*(-A^T\lambda) - b$ and $\psi(\lambda) = B\partial g^*(-B^T\lambda)$.

Eckstein, J., & Bertsekas, D. P. (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1), 293-318.

Known results on convergence rate

- Let f and g be strongly convex with moduli $\mu_1 > 0$ and $\mu_2 > 0$, respectively. If $t \le \sqrt[3]{\frac{\mu_1 \mu_2^2}{\lambda_{\max}(A^T A) \lambda_{\max}^2(B^T B)}}$, then

$$D(\lambda^\star) - D(\lambda^N) \le \frac{\|\lambda^1 - \lambda^\star\|^2}{2t(N-1)}.$$

Goldstein, T., O'Donoghue, B., Setzer, S., & Baraniuk, R. (2014). Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 7(3), 1588-1623.

SDP performance analysis

Worst-case analysis

$$
\begin{aligned}
\max\ & D(\lambda^\star) - D(\lambda^N) \\
\text{s.t. } & \{x^k, z^k, \lambda^k\}_1^N \text{ is generated by ADMM w.r.t. } f, g, A, B, b, \lambda^0, \hat{z}, t \\
& (x^\star, z^\star) \text{ is an optimal solution with Lagrangian multipliers } \lambda^\star \\
& \|\lambda^0 - \lambda^\star\|^2 + t^2 \left\|B\left(\hat{z} - z^\star\right)\right\|^2 = \Delta \\
& f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n), g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m) \\
& \lambda_{\max}(A^T A) = \nu_1, \lambda_{\max}(B^T B) = \nu_2 \\
& \lambda_0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r,
\end{aligned}
$$

- Variables: $f, g, A, B, b, \lambda^0, \hat{z}, x^\star, z^\star, \lambda^\star$;
- Parameters: $\mu_1, \mu_2, \nu_1, \nu_2, t, \Delta$.

Worst-case analysis

$$
\begin{aligned}
\max\ & D(\lambda^\star) - D(\lambda^N) \\
\text{s.t. } & \{x^k, z^k, \lambda^k\}_1^N \text{ is generated by ADMM w.r.t. } f, g, A, B, b, \lambda^0, \hat{z}, t \\
& (x^\star, z^\star) \text{ is an optimal solution with Lagrangian multipliers } \lambda^\star \\
& \|\lambda^0 - \lambda^\star\|^2 + t^2 \left\|B\left(\hat{z} - z^\star\right)\right\|^2 = \Delta \\
& f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n), g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m) \\
& \lambda_{\max}(A^T A) = \nu_1, \lambda_{\max}(B^T B) = \nu_2 \\
& \lambda_0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r,
\end{aligned}
$$

- Variables: $f, g, A, B, b, \lambda^0, \hat{z}, x^\star, z^\star, \lambda^\star$;
- Parameters: $\mu_1, \mu_2, \nu_1, \nu_2, t, \Delta$.

Key idea: This can be solved using semidefinite programming (SDP) by representing $\mathcal{F}_\mu(\mathbb{R}^n)$ via interpolation.

Semidefinite programming performance estimation

Main tool: Semidefinite programming (SDP) performance estimation, introduced in:

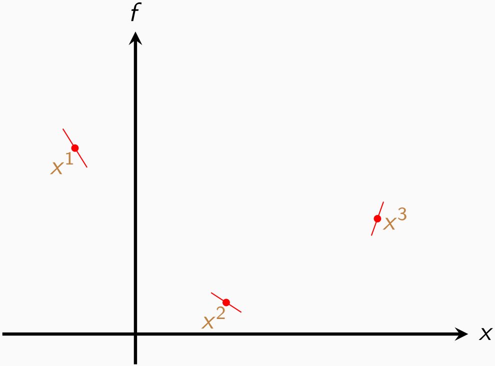
Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1-2):451–482, 2014.

Interpolation Problem

Consider a finite index set I, and given triple $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.

Interpolation Problem

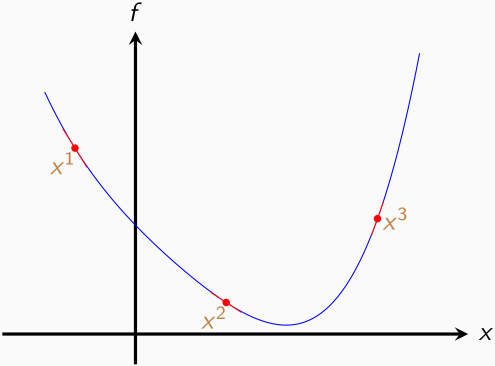
Consider a finite index set I, and given triple $\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\}_{k \in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.



$$?\exists \;\; f \in \mathcal{F}_\mu(\mathbb{R}^n):\; f(\mathbf{x}^k) = f^k, \quad \text{and} \quad \mathbf{g}^k \in \partial f(\mathbf{x}^k), \qquad \forall k \in I.$$

Interpolation Problem

Consider a finite index set I, and given triple $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ where $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{g}^k \in \mathbb{R}^n$ and $f^k \in \mathbb{R}$.



$?\exists\ \ f \in \mathcal{F}_\mu(\mathbb{R}^n)\colon\ f(\mathbf{x}^k) = f^k, \quad \text{and} \quad \mathbf{g}^k \in \partial f(\mathbf{x}^k), \qquad \forall k \in I.$

If yes, we say $\left\{(\mathbf{x}^k, \mathbf{g}^k, f^k)\right\}_{k\in I}$ is $\mathcal{F}_\mu(\mathbb{R}^n)$-interpolable.

L-smooth and μ-strongly Interpolation

Theorem (Taylor, Hendrickx, and Glineur (2017))

The following statements are equivalent:

1. $\left\{(\mathbf{x}^i, \mathbf{g}^i, f^i)\right\}_{i\in I}$ *is* $\mathcal{F}_\mu(\mathbb{R}^n)$*-interpolable;*
2. $\forall i, j \in I$:

$$\frac{\mu}{2}\left\|g^i - g^j\right\|^2 \leq f^i - f^j - \left\langle g^j, x^i - x^j\right\rangle.$$

A.B. Taylor, J.M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming* 161.1-2, 307–345 (2017)

Finite dimensional formulation

$$
\begin{aligned}
\max\ & D(\lambda^\star) - D(\lambda^N) \\
\text{s. t. } & \{(x^k; \xi^k; f^k)\}_1^N \cup \{(x^\star; \xi^\star; f^*)\} \text{ satisfy interpolation constraints} \\
& \{(z^k; \eta^k; g^k)\}_0^N \cup \{(z^\star; \eta^\star; g^*)\} \text{ satisfy interpolation constraints} \\
& (x^\star, z^\star) \text{ is an optimal solution with Lagrangian multipliers } \lambda^\star \\
& \|\lambda^0 - \lambda^\star\|^2 + t^2 \|B(\hat{z} - z^\star)\|^2 = \Delta \\
& \xi^1 = -A^T\lambda^0 - tA^TAx^1 - tA^TB\hat{z}, \\
& \xi^k = -A^T\lambda^{k-1} - tA^TAx^k - tA^TBz^{k-1}, \quad k \in \{2, ..., N\} \\
& \eta^k = -B^T\lambda^{k-1} - tB^TAx^k - tB^TBz^k, \quad k \in \{1, ..., N\} \\
& \lambda^k = \lambda^{k-1} + t(Ax^k + Bz^k - b), \quad k \in \{1, ..., N\} \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& \lambda_0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r.
\end{aligned}
$$

- ADMM is invariant under translation. We may assume w.l.o.g.

$$x^\star = 0, \ z^\star = 0, \ b = 0.$$

- ADMM is invariant under translation. We may assume w.l.o.g.

$$x^\star = 0,\ z^\star = 0,\ b = 0.$$

- Since $\begin{pmatrix} A & B \end{pmatrix}$ has full row rank

$$\lambda^\star = A\bar{x} + B\bar{z},\ \lambda^0 = Ax^0 + Bz^0,$$

for some $\bar{x}, \bar{z}, x^0, z^0$.

- ADMM is invariant under translation. We may assume w.l.o.g.

$$x^\star = 0,\ z^\star = 0,\ b = 0.$$

- Since $\begin{pmatrix} A & B \end{pmatrix}$ has full row rank

$$\lambda^\star = A\bar{x} + B\bar{z},\ \lambda^0 = Ax^0 + Bz^0,$$

for some $\bar{x}, \bar{z}, x^0, z^0$.

- Let $U = \begin{pmatrix} x^0 & x^1 & \dots & x^{N+1} & \bar{x} \end{pmatrix}$,
$V = \begin{pmatrix} z^0 & z^1 & \dots & z^N & \bar{z} & \hat{z} \end{pmatrix}$. Consider matrix variable

$$X = U^T U, \qquad Z = V^T V,$$

$$Y = \begin{pmatrix} AU & BV \end{pmatrix}^T \begin{pmatrix} AU & BV \end{pmatrix}.$$

SDP formulation

$$
\begin{aligned}
\max\ & f^\star + g^\star - f^{N+1} - g^N - \operatorname{tr}(L_o Y) \\
\text{s.t.}\ & \operatorname{tr}(L^f_{i,j} Y) + \operatorname{tr}(O^f_{i,j} X) \le f^i - f^j, \quad i,j \in \{1, \dots, N+1, \star\} \\
& \operatorname{tr}(L^g_{i,j} Y) + \operatorname{tr}(O^g_{i,j} Z) \le g^i - g^j, \quad i,j \in \{1, \dots, N, \star\} \\
& \operatorname{tr}(L_0 Y) + Z_{N+3,N+3} = \Delta \\
& X \succeq 0, Y \succeq 0, Z \succeq 0, \\
& \nu_1 X \succeq Y_{11}, \\
& \nu_2 Z \succeq Y_{22}.
\end{aligned}
$$

Performance estimation technique

- We employ *weak duality* to bound the optimal value of the last problem by constructing a dual feasible solution of SDP.

Performance estimation technique

- We employ *weak duality* to bound the optimal value of the last problem by constructing a dual feasible solution of SDP.
- The dual feasible solution is constructed empirically by doing *numerical experiments* with fixed values of the parameters $\Delta, N, \mu_1, L_1, \mu_2, L_2$.

Performance estimation technique

- We employ *weak duality* to bound the optimal value of the last problem by constructing a dual feasible solution of SDP.
- The dual feasible solution is constructed empirically by doing *numerical experiments* with fixed values of the parameters $\Delta, N, \mu_1, L_1, \mu_2, L_2$.
- The *analytical expressions* of the dual multipliers and optimal value *are guessed* and the guess is verified analytically.

New results

Convergence rate in terms of dual objective value

Theorem 1.

Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_0(\mathbb{R}^m)$ with $\mu_1 > 0$. If $t \leq \frac{\mu_1}{\lambda_{\max}(A^T A)}$ and $N \geq 2$, then

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{\|\lambda^0 - \lambda^\star\|^2 + t^2 \|B(\hat{z} - z^\star)\|^2}{4Nt}.$$

Exactness of the bound

- Let $\mu_1 > 0$, $N \geq 2$ and $t \in (0, \mu_1]$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be given as,

  $$f(x) = \frac{1}{2}|x| + \frac{\mu_1}{2}x^2, \quad g(z) = \frac{1}{2}\max\{\frac{N-1}{N}(z - \frac{1}{2Nt}) - \frac{1}{2Nt}, -z\}.$$

Exactness of the bound

- Let $\mu_1 > 0$, $N \geq 2$ and $t \in (0, \mu_1]$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be given as,

  $f(x) = \frac{1}{2}|x| + \frac{\mu_1}{2}x^2, \;\; g(z) = \frac{1}{2}\max\{\frac{N-1}{N}(z - \frac{1}{2Nt}) - \frac{1}{2Nt}, -z\}.$

- Consider the optimization problem

$$\min f(x) + g(z), \text{ s.t. } x + z = 0.$$

  $(x^\star, z^\star) = (0, 0)$ and $\lambda^\star = \frac{1}{2}$.

Exactness of the bound

- Let $\mu_1 > 0$, $N \geq 2$ and $t \in (0, \mu_1]$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be given as,

  $f(x) = \frac{1}{2}|x| + \frac{\mu_1}{2}x^2, \;\; g(z) = \frac{1}{2}\max\{\frac{N-1}{N}(z - \frac{1}{2Nt}) - \frac{1}{2Nt}, -z\}.$

- Consider the optimization problem

$$\min f(x) + g(z), \text{ s.t. } x + z = 0.$$

  $(x^\star, z^\star) = (0, 0)$ and $\lambda^\star = \frac{1}{2}$.

- ADMM with initial point $\lambda^0 = \frac{-1}{2}$ and $\hat{z} = 0$ generates

$$x^k = 0, \; z^k = \frac{1}{2Nt}, \; \lambda^k = \frac{-1}{2} + \frac{k}{2N} \quad k \in \{1, \ldots, N\}.$$

Exactness of the bound

- Let $\mu_1 > 0$, $N \geq 2$ and $t \in (0, \mu_1]$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be given as,

$$f(x) = \frac{1}{2}|x| + \frac{\mu_1}{2}x^2, \quad g(z) = \frac{1}{2}\max\{\frac{N-1}{N}(z - \frac{1}{2Nt}) - \frac{1}{2Nt}, -z\}.$$

- Consider the optimization problem

$$\min f(x) + g(z), \text{ s.t. } x + z = 0.$$

$(x^\star, z^\star) = (0, 0)$ and $\lambda^\star = \frac{1}{2}$.

- ADMM with initial point $\lambda^0 = \frac{-1}{2}$ and $\hat{z} = 0$ generates

$$x^k = 0, \ z^k = \frac{1}{2Nt}, \ \lambda^k = \frac{-1}{2} + \frac{k}{2N} \quad k \in \{1, \ldots, N\}.$$

- At λ^N, we have $D(\lambda^N) = \frac{-1}{4Nt} = \frac{\|\lambda^0 - \lambda^\star\|^2 + t^2\|B(\hat{z} - z^\star)\|^2}{4Nt}$.

Linear convergence rate

PŁ inequality

- The function D is said to satisfy the PŁ inequality if there exists an $L_p > 0$ such that for any $\lambda \in \mathbb{R}^r$ we have

$$D(\lambda^\star) - D(\lambda) \leq \frac{1}{2L_p}\|\xi\|^2, \quad \xi \in b - A\partial f^*(-A^T\lambda) - B\partial g^*(-B^T\lambda).$$

- For $D(\lambda) = -\langle \lambda, b \rangle - f^*(-A^T\lambda) - g^*(-B^T\lambda)$:

$$b - A\partial f^*(-A^T\lambda) - B\partial g^*(-B^T\lambda) \subseteq \partial\left(-D(\lambda)\right).$$

Linear convergence rate

Theorem 2.

Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$ with $\mu_1, \mu_2 > 0$, and let D satisfies the PŁ inequality with L_p. Suppose that $t \leq \sqrt{c_1 c_2}$, where $c_1 = \frac{\mu_1}{\lambda_{\max}(A^T A)}$ and $c_2 = \frac{\mu_2}{\lambda_{\max}(B^T B)}$.

(i) If $c_1 \geq c_2$, then

$$\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \leq \frac{2c_1c_2 - t^2}{2c_1c_2 - t^2 + L_p t\left(4c_1c_2 - c_2 t - 2t^2\right)},$$

in particular, if $t = \sqrt{c_1 c_2}$,

$$\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \leq \frac{1}{1 + L_p\left(2\sqrt{c_1c_2} - c_2\right)}.$$

(ii) If $c_1 < c_2$, then

$$\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \leq$$

$$\frac{4c_2^2 - 2c_2\sqrt{c_1c_2} - t^2}{4c_2^2 - 2c_2\sqrt{c_1c_2} - t^2 + L_p t\left(8c_2^2 + 5c_2 t - 2\sqrt{c_1c_2}\left(1 + \frac{t}{c_1}\right)(2c_2 + t)\right)}.$$

Sufficient conditions for the PŁ inequality

Scenario	Strong convexity	Lipschitz continuity	Full row rank
1	f, g	∇f	A
2	f, g	$\nabla f, \nabla g$	-

Deng, W., & Yin, W. (2016). On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66(3), 889-916.

Necessary conditions for the linear convergence

Theorem 3.

Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$. If ADMM is linearly convergent with respect to the dual objective value, then D satisfies the PŁ inequality.

R-linear convergence

Known results on convergence rate

Nishihara et al. showed the R-linear convergence of ADMM in terms of $\{x^k, z^k, \lambda^k\}$ under the following conditions:

i) The function $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ is L-smooth with $\mu_1 > 0$;

ii) The matrix A is invertible and that B has full column rank.

Nishihara, R., Lessard, L., Recht, B., Packard, A., & Jordan, M. (2015). A general analysis of then convergence of ADMM. *International Conference on Machine Learning (PMLR)*, 343-352.

Necessary conditions for the linear convergence

Theorem 4.

Let $N \geq 2$ and let A has full row rank. Suppose that $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ is L-smooth with $\mu_1 > 0$ and $g \in \mathcal{F}_0(\mathbb{R}^m)$. If

$$t < \min\{\tfrac{\mu_1}{\lambda_{\max}(A^TA)}, \sqrt{\tfrac{\mu_1 L}{\lambda_{\min}(AA^T)\lambda_{\max}(A^TA)}}\},$$

then

$$D(\lambda^\star) - D(\lambda^N) \leq \tfrac{V^0}{8t}\left(1 - \tfrac{2\lambda_{\min}(AA^T)\mu_1 t}{L\mu_1 + 2\lambda_{\min}(AA^T)\mu_1 t + \lambda_{\min}(AA^T)\lambda_{\max}(A^TA)t^2}\right)^{N-2},$$

where

$$V^0 = \|\lambda^0 - \lambda^\star\|^2 + t^2\left\|B\left(z^0 - z^\star\right)\right\|^2.$$

Necessary conditions for the linear convergence

Theorem 5.

Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ with $\mu_1 > 0$ and let $g \in \mathcal{F}_0(\mathbb{R}^m)$ be L-smooth. Assume that $N \geq 3$ and B has full row rank. If $t < \min\{\frac{\mu_1}{2\lambda_{\max}(A^T A)}, \frac{L}{2\lambda_{\min}(BB^T)}\}$, then

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{V^0}{8t} \left(\frac{L}{L + t\lambda_{\min}(BB^T)} \right)^{2N-6}, \tag{5.1}$$

where

$$V^0 = \|\lambda^0 - \lambda^\star\|^2 + t^2 \left\| B\left(z^0 - z^\star\right) \right\|^2 .$$

Conclusion

Discussion

- The computation of the tight bound for linear convergence rate.

Discussion

- The computation of the tight bound for linear convergence rate.
- The performance analysis of fast ADMM.

Discussion

- The computation of the tight bound for linear convergence rate.
- The performance analysis of fast ADMM.
- The strong convexity of f (or g) can be replaced with the convexity of $f - c_1\|\cdot\|_A^2$ (or $g - c_2\|\cdot\|_B^2$) for some $c_1 \geq 0$ (or $c_2 \geq 0$).

Preprint at `arxiv.org/abs/2206.09865`

The End

Omar Khayyam (1048-1131)

ابر آمد و زار بر سر سبزه گریست   بی بادهٔ گلرنگ نمی‌شاید زیست

این سبزه که امروز تماشاگه ماست   تا سبزهٔ خاک ما تماشاگه کیست

And we, that now make merry in the Room
They left, and Summer dresses in new Bloom.
Ourselves must we beneath the Couch of Earth
Descend, ourselves to make a Couch - for whom?