ICLR 2025

ADMM for Structured Fractional Minimization

Ganzhao Yuan
Peng Cheng Laboratory, China

1. Nonconvex Nonsmooth Fractional Minimization Problem

$$\min_{\mathbf{x}} F(\mathbf{x}) \triangleq \frac{u(\mathbf{x})}{d(\mathbf{x})}, \text{ where } u(\mathbf{x}) \triangleq f(\mathbf{x}) + \delta(\mathbf{x}) - g(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

2. Assumptions

Assumption 3.1. *There exists a universal positive constant* $\overline{\mathrm{x}}$ *such that* $\|\mathbf{x}\| \leq \overline{\mathrm{x}}$ *for all* $\mathbf{x} \in \mathrm{dom}(F)$.

Assumption 3.2. *The function* $f(\cdot)$ *is* L_f*-smooth such that* $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}')\| \leq L_f\|\mathbf{x} - \mathbf{x}'\|$ *holds for all* $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$. *This implies that:* $|f(\mathbf{x}) - f(\mathbf{x}') - \langle \nabla f(\mathbf{x}'), \mathbf{x} - \mathbf{x}'\rangle| \leq \frac{L_f}{2}\|\mathbf{x} - \mathbf{x}'\|_2^2$ *(cf. Lemma 1.2.3 in (Nesterov, 2003)).*

Assumption 3.3. *Let* $\mu > 0$, $\mathbf{x}' \in \mathbb{R}^n$, *and* $\mathbf{y}' \in \mathbb{R}^m$. *Both proximal operators,* $\mathrm{Prox}(\mathbf{y}'; h, \mu) \triangleq \arg\min_{\mathbf{y}} \frac{1}{2\mu}\|\mathbf{y} - \mathbf{y}'\|_2^2 + h(\mathbf{y})$ *and* $\mathrm{Prox}(\mathbf{x}'; \delta, \mu) \triangleq \arg\min_{\mathbf{x}} \frac{1}{2\mu}\|\mathbf{x} - \mathbf{x}'\|_2^2 + \delta(\mathbf{x})$, *can be computed efficiently and exactly.*

Assumption 3.4. *The function* $d(\mathbf{x})$ *is* C_d*-Lipschitz continuous with* $C_d \geq 0$, *and meets one of the following conditions for some* $W_d \geq 0$: ***(a)*** $d(\mathbf{x})$ *is* W_d*-weakly convex.* ***(b)*** $\sqrt{d(\mathbf{x})}$ *is* W_d*-weakly convex.*

Assumption 5.1. *Let* $\{\mathbf{x}^t\}_{t=0}^{\infty}$ *be generated by Algorithm 1. For all t, there exist constants* $\{\underline{\mathrm{d}}, \overline{\mathrm{d}}\}$ *such that* $0 < \underline{\mathrm{d}} \leq d(\mathbf{x}^t) \leq \overline{\mathrm{d}}$, *and constants* $\{\underline{\mathrm{F}}, \overline{\mathrm{F}}\}$ *such that* $0 < \underline{\mathrm{F}} \leq F(\mathbf{x}^t) \leq \overline{\mathrm{F}}$.

Assumption 5.7. *Assume* $\Delta \triangleq \beta^0 - \underline{\mathrm{v}}/(\underline{\mathrm{F}} \cdot \underline{\mathrm{d}}) > 0$.

3. Comparisons With Existing Works

- **Algorithms in Limited Scenarios**
  - **Dinkelbach's Parametric Algorithm (DPA)**

$$\mathbf{x}^{t+1} \in \arg\min_{\mathbf{x}} \; u(\mathbf{x}) - \lambda^t d(\mathbf{x}) \qquad \lambda^t = F(\mathbf{x}^t)$$

  - **Quadratic Transform Algorithm (QTA)**

$$\mathbf{x}^{t+1} \in \arg\min_{\mathbf{x}} (\alpha^t)^2 u(\mathbf{x}) - 2\alpha^t d(\mathbf{x})^{1/2}$$
$$\alpha^t = d(\mathbf{x}^t)^{1/2} \cdot u(\mathbf{x}^t)^{-1}$$

  - **Linearized DPA and Linearized DPA**

- **General Algorithms for Solving the Main Problem**
  - **Subgradient Projection Methods**
  - **Smoothing Proximal Gradient Methods**
  - **Full Splitting Algorithm (FSA)**

4. Applications

- **Sparse Fisher Discriminant Analysis**
- **Sharpe ratio maximization**
- **Signal-to-noise ratio maximization**

5. The Proposed Algorithm

- **Smoothing the nonsmooth function *h*(*y*):**

$$\min_{\mathbf{x},\mathbf{y}}\{f(\mathbf{x}) + \delta(\mathbf{x}) - g(\mathbf{x}) + h_\mu(\mathbf{y})\}/d(\mathbf{x}), \;\; \mathbf{A}\mathbf{x} = \mathbf{y}$$

- **The resulting augmented Lagrangian function:**

FADMM-D variant: $\mathcal{L}(\mathbf{x}, \mathbf{y}; \mathbf{z}; \beta, \mu) \triangleq \frac{\mathcal{U}(\mathbf{x},\mathbf{y};\mathbf{z};\beta,\mu)}{d(\mathbf{x})}$

FADMM-Q variant: $\mathcal{K}(\alpha, \mathbf{x}, \mathbf{y}; \mathbf{z}; \beta, \mu) \triangleq -2\alpha\sqrt{d(\mathbf{x})} + \alpha^2\mathcal{U}(\mathbf{x}, \mathbf{y}; \mathbf{z}; \beta, \mu)$

$$\mathcal{U}(\mathbf{x}, \mathbf{y}; \mathbf{z}; \beta, \mu) \triangleq \underbrace{f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x} - \mathbf{y}, \mathbf{z}\rangle + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{\triangleq \mathcal{S}(\mathbf{x},\mathbf{y};\mathbf{z};\beta)} + \delta(\mathbf{x}) - g(\mathbf{x}) + h_\mu(\mathbf{y}).$$

- **Our Key Strategy: Increasing Penalization and Decreasing Smoothing**

$$\beta^t = \mathcal{O}(\sqrt[3]{t}), \quad \mu^t = \mathcal{O}(1/\sqrt[3]{t})$$

- **The Proposed FADMM Algorithm**
  - **Majorization Technique**

(a) $s(\mathbf{x}) \leq \mathcal{U}^t(\mathbf{x}; \mathbf{x}^t) \triangleq s(\mathbf{x}^t) + \langle \mathbf{x} - \mathbf{x}^t, \nabla s(\mathbf{x}^t)\rangle + \frac{1}{2}(L_f + \beta^t\|\mathbf{A}\|_2^2)\|\mathbf{x} - \mathbf{x}^t\|_2^2$.

(b) $-g(\mathbf{x}) \leq \mathcal{R}(\mathbf{x}; \mathbf{x}^t) \triangleq -g(\mathbf{x}^t) - \langle \mathbf{x} - \mathbf{x}^t, \boldsymbol{\xi}\rangle, \forall \boldsymbol{\xi} \in \partial g(\mathbf{x}^t)$.

(c) $-d(\mathbf{x}) \leq \dot{\mathcal{V}}(\mathbf{x}; \mathbf{x}^t) \triangleq -d(\mathbf{x}^t) - \langle \mathbf{x} - \mathbf{x}^t, \boldsymbol{\xi}\rangle + \frac{W_d}{2}\|\mathbf{x} - \mathbf{x}^t\|_2^2, \forall \boldsymbol{\xi} \in \partial d(\mathbf{x}^t)$.

(d) $-\sqrt{d(\mathbf{x})} \leq \ddot{\mathcal{V}}(\mathbf{x}; \mathbf{x}^t) \triangleq -\sqrt{d(\mathbf{x}^t)} - \langle \mathbf{x} - \mathbf{x}^t, \boldsymbol{\xi}\rangle + \frac{W_d}{2}\|\mathbf{x} - \mathbf{x}^t\|_2^2, \forall \boldsymbol{\xi} \in \partial\sqrt{d(\mathbf{x}^t)}$.

  - **Majorization Technique**

$$\min_{\mathbf{x}} \dot{\mathcal{M}}^t(\mathbf{x}; \mathbf{x}^t, \lambda^t) \triangleq \delta(\mathbf{x}) + \mathcal{U}^t(\mathbf{x}; \mathbf{x}^t) + \mathcal{R}(\mathbf{x}; \mathbf{x}^t) + \lambda^t\dot{\mathcal{V}}(\mathbf{x}; \mathbf{x}^t) + \frac{\theta-1}{2}\ell(\beta^t)\|\mathbf{x} - \mathbf{x}^t\|_2^2$$

$$\min_{\mathbf{x}} \ddot{\mathcal{M}}^t(\mathbf{x}; \mathbf{x}^t, \alpha^{t+1}) \triangleq \delta(\mathbf{x}) + \mathcal{U}^t(\mathbf{x}; \mathbf{x}^t) + \mathcal{R}(\mathbf{x}; \mathbf{x}^t) + \frac{2}{\alpha^{t+1}}\ddot{\mathcal{V}}(\mathbf{x}; \mathbf{x}^t) + \frac{\theta-1}{2}\ell(\beta^t)\|\mathbf{x} - \mathbf{x}^t\|_2^2$$

  - **The Main Algorithm**

Algorithm 1: FADMM: The Proposed ADMM using Dinkelbach's Parametric Method or the Quadratic Transform Method for Solving Problem (1).

(S0) Initialize $\{\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0\}$.
(S1) Choose $\xi \in (0, \infty)$, $\theta \in (1, \infty)$, $p \in (0, 1)$, and $\chi \in (2\sqrt{1+\xi}, \infty)$.
(S2) Choose β^0 large enough such that $\beta^0 > \underline{\mathrm{v}}/(\underline{\mathrm{F}}\underline{\mathrm{d}})$, satisfying Assumption 5.7.
for t from 0 to T **do**
  (S3) $\beta^t = \beta^0(1 + \xi t^p)$, $\mu^t = \chi/\beta^t$.
  (S4) Solve the $\mathbf{x}$-subproblem using FADMM-D or FADMM-Q:
  if FADMM-D **then**
    Set $\lambda^t = \mathcal{U}(\mathbf{x}^t, \mathbf{y}^t; \mathbf{z}^t; \beta^t, \mu^t)/d(\mathbf{x}^t)$, and $\mathbf{x}^{t+1} \in \arg\min_{\mathbf{x}} \dot{\mathcal{M}}^t(\mathbf{x}; \mathbf{x}^t, \lambda^t)$.
  end
  if FADMM-Q **then**
    Set $\alpha^{t+1} = \sqrt{d(\mathbf{x}^t)}/\mathcal{U}(\mathbf{x}^t, \mathbf{y}^t; \mathbf{z}^t; \beta^t, \mu^t)$, and $\mathbf{x}^{t+1} \in \arg\min_{\mathbf{x}} \ddot{\mathcal{M}}^t(\mathbf{x}; \mathbf{x}^t, \alpha^{t+1})$.
  end
  (S5) $\mathbf{y}^{t+1} = \arg\min_{\mathbf{y}} h_{\mu^t}(\mathbf{y}) + \frac{\beta^t}{2}\|\mathbf{y} - \mathbf{b}^t\|_2^2$, where $\mathbf{b}^t \triangleq \mathbf{y}^t - \nabla_{\mathbf{y}}\mathcal{S}(\mathbf{x}^{t+1}, \mathbf{y}^t; \mathbf{z}^t; \beta^t)/\beta^t$. It can be solved as $\mathbf{y}^{t+1} = \frac{\check{\mathbf{y}}^{t+1} + \beta^t\mu^t\mathbf{b}^t}{1+\beta^t\mu^t}$, where $\check{\mathbf{y}}^{t+1} \triangleq \mathrm{Prox}(\mathbf{b}^t; h, \mu^t + 1/\beta^t)$.
  (S6) $\mathbf{z}^{t+1} = \mathbf{z}^t + \beta^t(\mathbf{A}\mathbf{x}^{t+1} - \mathbf{y}^{t+1})$.
end

6. Iteration Complexity

- **First-Order Optimality Conditions**

$$\mathbf{z}^{t+1} = \nabla h_{\mu^t}(\mathbf{y}^{t+1}) \in \partial h(\check{\mathbf{y}}^{t+1})$$

- **Controlling Dual using Primal**

$$\|\mathbf{z}^{t+1} - \mathbf{z}^t\|_2^2 \leq 2\frac{(\beta^t)^2}{\chi^2}\|\mathbf{y}^{t+1} - \mathbf{y}^t\|_2^2 + 2C_h^2\left(\frac{6}{t} - \frac{6}{t+1}\right)$$

- **The Associated Lyapunov Function**
  - **FADMM-D:**

$$\mathbb{P}^t \triangleq \underbrace{\mathcal{L}(\mathbf{x}^t, \mathbf{y}^t; \mathbf{z}^t; \beta^t, \mu^t)}_{\triangleq \mathbb{L}^t} + \underbrace{12(1+\xi)C_h^2/(\beta^0\underline{\mathrm{d}}t)}_{\triangleq \mathbb{T}^t} + \underbrace{C_h^2\mu^t/(2\underline{\mathrm{d}})}_{\triangleq \mathbb{U}^t}$$

  - **FADMM-Q:**

$$\mathbb{P}^t \triangleq \underbrace{\mathcal{K}(\alpha^t, \mathbf{x}^t, \mathbf{y}^t; \mathbf{z}^t; \beta^t, \mu^t)}_{\triangleq \mathbb{K}^t} + \underbrace{12\overline{\alpha}^2(1+\xi)C_h^2/(\beta^0 t)}_{\triangleq \mathbb{T}^t} + \underbrace{\tfrac{1}{2}\overline{\alpha}^2 C_h^2\mu^t}_{\triangleq \mathbb{U}^t}$$

- **A Summerable Property**

$$\min(\varepsilon_x, \varepsilon_y, \varepsilon_z)\mathcal{E}^t \leq \mathbb{P}^t - \mathbb{P}^{t+1}$$

$$\mathcal{E}^t \triangleq \beta^t\{\|\mathbf{x}^{t+1} - \mathbf{x}^t\|_2^2 + \|\mathbf{y}^{t+1} - \mathbf{y}^t\|_2^2 + \|\mathbf{A}\mathbf{x}^{t+1} - \mathbf{y}^{t+1}\|_2^2\}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathcal{E}_+^t \leq \mathcal{O}(T^{(p-1)/2})$$

$$\mathcal{E}_+^t \triangleq \beta^t\{\|\mathbf{x}^{t+1} - \mathbf{x}^t\| + \|\mathbf{y}^{t+1} - \mathbf{y}^t\| + \|\mathbf{A}\mathbf{x}^{t+1} - \mathbf{y}^{t+1}\|\}$$

- **Approximate Critical Point**

Definition 6.1. *(ϵ-Critical Point) We define* $\mathrm{Crit}(\mathbf{x}^+, \mathbf{x}, \mathbf{y}^+, \mathbf{y}, \mathbf{z}^+, \mathbf{z}) \triangleq \|\mathbf{x}^+ - \mathbf{x}\| + \|\mathbf{y}^+ - \mathbf{y}\| + \|\mathbf{z}^+ - \mathbf{z}\| + \|\mathbf{A}\mathbf{x}^+ - \mathbf{y}^+\| + \|\partial h(\mathbf{y}^+) - \mathbf{z}^+\| + \|\partial\delta(\mathbf{x}^+) + \nabla f(\mathbf{x}^+) - \partial g(\mathbf{x}) + \mathbf{A}^\top\mathbf{z}^+ - \varphi(\mathbf{x}, \mathbf{y})\partial d(\mathbf{x})\|$, *and* $\varphi(\mathbf{x}, \mathbf{y}) = \{f(\mathbf{x}) + \delta(\mathbf{x}) - g(\mathbf{x}) + h(\mathbf{y})\}/d(\mathbf{x})$. *A solution* $(\bar{\mathbf{x}}^+, \bar{\mathbf{x}}, \bar{\mathbf{y}}^+, \bar{\mathbf{y}}, \bar{\mathbf{z}}^+, \bar{\mathbf{z}})$ *is a critical point of Problem (1) if:*

$$\mathrm{Crit}(\bar{\mathbf{x}}^+, \bar{\mathbf{x}}, \bar{\mathbf{y}}^+, \bar{\mathbf{y}}, \bar{\mathbf{z}}^+, \bar{\mathbf{z}}) \leq \epsilon.$$

- **The Complexity Result**

$$\mathrm{Crit}(\mathcal{W}^t) \leq \mathcal{O}(T^{-p}) + \mathcal{O}(T^{(p-1)/2})$$

with the choice $p = 1/3$, *we have* $\mathrm{Crit}(\mathcal{W}^t) \leq \mathcal{O}(T^{-1/3})$

7. Applications and Experiment Result

- **Sparse Fisher Discriminant Analysis**

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \frac{\mathrm{tr}(\mathbf{X}^\top\mathbf{C}\mathbf{X}) + \rho(\|\mathbf{X}\|_1 - \|\mathbf{X}\|_{[k]})}{\mathrm{tr}(\mathbf{X}^\top\mathbf{D}\mathbf{X})}, \;\mathrm{s.t.}\; \mathbf{X} \in \Omega \qquad \Omega \triangleq \{\mathbf{X} \,|\, \mathbf{X}^\top\mathbf{X} = \mathbf{I}_r\}$$

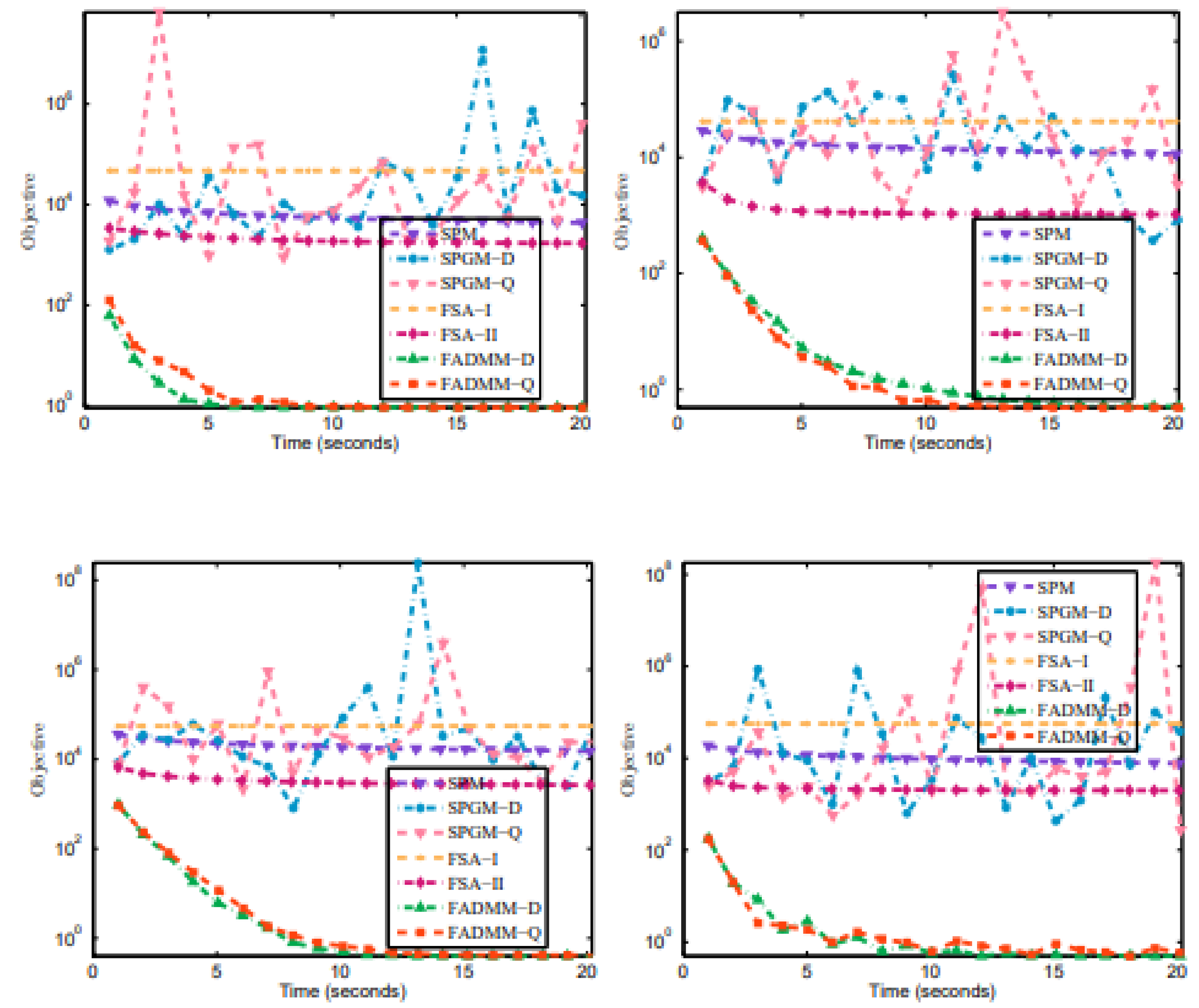
- **Robust Sharpe Ratio Maximization**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\max(0, \max(\mathbf{b} - \mathbf{D}\mathbf{x}))}{\max_{i=1}^{p} \mathbf{x}^\top\mathbf{C}_{(i)}\mathbf{x}}, \;\mathrm{s.t.}\; \mathbf{x} \in \Omega \qquad \Omega \triangleq \{\mathbf{x} \,|\, \mathbf{x} \geq \mathbf{0}, \mathbf{x}^\top\mathbf{1} = 1\}$$

- **Robust Sparse Recovery**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{\rho_1\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 + \rho_2\|\mathbf{x}\|_1}{\|\mathbf{x}\|_{[k]}}, \;\mathrm{s.t.}\; \mathbf{x} \in \Omega \qquad \Omega \triangleq \{\mathbf{x} \,|\, \|\mathbf{x}\|_\infty \leq \rho_0\}$$

- **Experiment Results**



Conclusions:

(a) SPM tends to be less efficient in comparison to other methods.

(b) SPGM-D and SPGM-Q, utilizing a variable smoothing strategy, generally demonstrates better performance than SPM.

(c) The proposed FADMM-D and FADMMQ generally exhibit similar performance, both achieving the lowest objective function values among all the methods examined. This supports the widely accepted view that primal-dual methods are generally more robust and faster than primal-only methods.

(d) The proposed FADMM-D and FADMM-Q still outperform FSA, which uses a sufficiently small step size to ensure convergence.