A New Randomized Primal-Dual Algorithm for Convex Optimization with Optimal Last Iterate Rates

Deyi Liu and **Quoc Tran-Dinh**

Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill (UNC)

March 15, 2022

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Outline of the talk

Introduction

Our Algorithm

Main theorems

Numerical results

Introduction

We consider the following nonsmooth constrained convex optimization problem:

$$F^\star := \min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \left\{ F(x,w) := h(x) + \sum_{i=1}^n f_i(x_i) + g(w) \quad \text{s.t.} \quad Kx + Bw = b \right\}, \text{ (P)}$$

where

- $h : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ is smooth and convex
- $f_i : \mathbb{R}^{p_i} \to \mathbb{R} \cup \{+\infty\}$ is possibly nonsmooth and convex, $i = 1, \cdots, n$
- $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is possibly nonsmooth, and convex
- $K \in \mathbb{R}^{d\times p}$, $B \in \mathbb{R}^{d\times m}$, and $b \in \mathbb{R}^d$

Introduction

We consider the following nonsmooth constrained convex optimization problem:

$$F^\star := \min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \Big\{ F(x,w) := h(x) + \sum_{i=1}^n f_i(x_i) + g(w) \quad \text{s.t.} \quad Kx + Bw = b \Big\}, \quad \text{(P)}$$

where

- $h : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ is smooth and convex
- $f_i : \mathbb{R}^{p_i} \to \mathbb{R} \cup \{+\infty\}$ is possibly nonsmooth and convex, $i = 1, \cdots, n$
- $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is possibly nonsmooth, and convex
- $K \in \mathbb{R}^{d\times p}$, $B \in \mathbb{R}^{d\times m}$, and $b \in \mathbb{R}^d$

The corresponding dual problem of (P) can be written as

$$D^\star := \max_{y\in\mathbb{R}^d} D(y), \quad \text{where } D(y) := \min_{x,w} \Big\{ F(x,w) + \langle Kx + Bw - b, y\rangle \Big\}. \quad \text{(D)}$$

Two Special Cases of Our Problem

- If $b = 0$ and $B = -\mathbb{I}$, then problem (P) reduces to:

$$\min_{x \in \mathbb{R}^p} \Big\{ F(x) := \textstyle\sum_{i=1}^n f_i(x_i) + h(x) + g(Kx) \Big\}. \tag{1}$$

- If $b = 0$ and $K = -\mathbb{I}$, then problem (P) reduces to:

$$\min_{w \in \mathbb{R}^p} \Big\{ F(w) := \textstyle\sum_{i=1}^n f_i(B_i w) + h(Bw) + g(w) \Big\}, \tag{2}$$

  where B_i is the i-th row block of B.

Two Special Cases of Our Problem

- If $b = 0$ and $B = -\mathbb{I}$, then problem (P) reduces to:

$$\min_{x\in\mathbb{R}^p}\Big\{F(x) := \sum_{i=1}^n f_i(x_i) + h(x) + g(Kx)\Big\}. \tag{1}$$

- If $b = 0$ and $K = -\mathbb{I}$, then problem (P) reduces to:

$$\min_{w\in\mathbb{R}^p}\Big\{F(w) := \sum_{i=1}^n f_i(B_i w) + h(Bw) + g(w)\Big\}, \tag{2}$$

  where B_i is the i-th row block of B.

The above two problems are both structurally complex and require carefully designed algorithms. Therefore, it is more convenient to consider the general problem (P).

Introduction

Assumptions

- The primal solution set $\mathcal{X}^\star \times \mathcal{W}^\star$ of (P) is nonempty and the Slater condition holds:

$$\mathrm{ri}\,(\mathsf{dom}(f+h) \times \mathsf{dom}(g) \cap \{(x,w) \mid Kx + Bw = b\}) \neq \emptyset.$$

- The function h is convex and partially $L_{h,i}$-smooth for all $i \in [n]$, i.e., for any $x \in \mathbb{R}^p$ and $d_i \in \mathbb{R}^{p_i}$ with $i \in [n]$, we have

$$\|\nabla_{x_i} h(x + U_i d_i) - \nabla_{x_i} h(x)\| \leq L_{h,i}\|d_i\|, \tag{3}$$

  where $U_i \in \mathbb{R}^{p \times p_i}$ has p_i unit vectors such that $[U_1, U_2, \cdots, U_n]$ forms the identity matrix $\mathbb{I}$ in $\mathbb{R}^{p\times p}$.

Introduction

Assumptions

- The primal solution set $\mathcal{X}^\star \times \mathcal{W}^\star$ of (P) is nonempty and the Slater condition holds:

$$\mathrm{ri}\,(\mathsf{dom}(f+h) \times \mathsf{dom}(g) \cap \{(x,w) \mid Kx + Bw = b\}) \neq \emptyset.$$

- The function h is convex and partially $L_{h,i}$-smooth for all $i \in [n]$, i.e., for any $x \in \mathbb{R}^p$ and $d_i \in \mathbb{R}^{p_i}$ with $i \in [n]$, we have

$$\|\nabla_{x_i} h(x + U_i d_i) - \nabla_{x_i} h(x)\| \leq L_{h,i}\|d_i\|, \tag{3}$$

where $U_i \in \mathbb{R}^{p \times p_i}$ has p_i unit vectors such that $[U_1, U_2, \cdots, U_n]$ forms the identity matrix $\mathbb{I}$ in $\mathbb{R}^{p \times p}$.

The above assumption implies strong duality, i.e., $F^\star = D^\star$, and the solution set $\mathcal{Y}^\star$ of (D) is also nonempty and bounded.

Introduction

Assumptions

- The primal solution set $\mathcal{X}^\star \times \mathcal{W}^\star$ of (P) is nonempty and the Slater condition holds:
$$\mathrm{ri}\left(\mathsf{dom}(f+h) \times \mathsf{dom}(g) \cap \{(x,w) \mid Kx + Bw = b\}\right) \neq \emptyset.$$
- The function h is convex and partially $L_{h,i}$-smooth for all $i \in [n]$, i.e., for any $x \in \mathbb{R}^p$ and $d_i \in \mathbb{R}^{p_i}$ with $i \in [n]$, we have
$$\|\nabla_{x_i} h(x + U_i d_i) - \nabla_{x_i} h(x)\| \leq L_{h,i}\|d_i\|, \tag{3}$$
where $U_i \in \mathbb{R}^{p \times p_i}$ has p_i unit vectors such that $[U_1, U_2, \cdots, U_n]$ forms the identity matrix $\mathbb{I}$ in $\mathbb{R}^{p\times p}$.

The above assumption implies strong duality, i.e., $F^\star = D^\star$, and the solution set $\mathcal{Y}^\star$ of (D) is also nonempty and bounded.

Our Contributions

- We develop a randomized primal-dual algorithm to solve (P). We obtain optimal $\mathcal{O}(1/k)$ and $\mathcal{O}(1/k^2)$ convergence rates in the convex and strongly-convex cases, respectively.
- Our rates are on the last primal iterate (x^k, w^k).

Introduction

We write (P) into the minimax form:

$$\min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \max_{y\in\mathbb{R}^d} \Big\{ \mathcal{L}(x, w, y) := F(x, w) + \langle Kx + Bw - b, y\rangle \Big\}, \tag{4}$$

For any point $(x^\star, w^\star, y^\star) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star$, we have

$$\mathcal{L}(x^\star, w^\star, y) \le \mathcal{L}(x^\star, w^\star, y^\star) \le \mathcal{L}(x, w, y^\star), \quad \forall x \in \mathbb{R}^p, w \in \mathbb{R}^m, y \in \mathbb{R}^d. \tag{5}$$

Introduction

We write (P) into the minimax form:

$$\min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \max_{y\in\mathbb{R}^d} \Big\{ \mathcal{L}(x,w,y) := F(x,w) + \langle Kx + Bw - b, y\rangle \Big\}, \tag{4}$$

For any point $(x^\star, w^\star, y^\star) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star$, we have

$$\mathcal{L}(x^\star, w^\star, y) \le \mathcal{L}(x^\star, w^\star, y^\star) \le \mathcal{L}(x, w, y^\star), \quad \forall x \in \mathbb{R}^p, w \in \mathbb{R}^m, y \in \mathbb{R}^d. \tag{5}$$

Suppose (x, w, y) are random variables. Given a compact set $\mathcal{Z}$ such that $\mathcal{Z}\cap \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star \neq \emptyset$, we define the following primal-dual expected gap for (x, w, y).

Primal-dual expected gap

$$\mathcal{G}_{\mathcal{Z}}(x,w,y) := \sup_{(\hat{x},\hat{w},\hat{y})\in\mathcal{Z}} \mathbb{E}\left[\mathcal{L}(x,w,\hat{y}) - \mathcal{L}(\hat{x},\hat{w},y)\right], \tag{6}$$

Introduction

We write (P) into the minimax form:

$$\min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \max_{y\in\mathbb{R}^d} \Big\{ \mathcal{L}(x,w,y) := F(x,w) + \langle Kx + Bw - b, y\rangle \Big\}, \tag{4}$$

For any point $(x^\star, w^\star, y^\star) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star$, we have

$$\mathcal{L}(x^\star, w^\star, y) \le \mathcal{L}(x^\star, w^\star, y^\star) \le \mathcal{L}(x, w, y^\star), \quad \forall x \in \mathbb{R}^p, w \in \mathbb{R}^m, y \in \mathbb{R}^d. \tag{5}$$

Suppose (x, w, y) are random variables. Given a compact set $\mathcal{Z}$ such that $\mathcal{Z} \cap \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star \neq \emptyset$, we define the following primal-dual expected gap for (x, w, y).

Primal-dual expected gap

$$\mathcal{G}_{\mathcal{Z}}(x,w,y) := \sup_{(\hat{x},\hat{w},\hat{y})\in\mathcal{Z}} \mathbb{E}\left[\mathcal{L}(x,w,\hat{y}) - \mathcal{L}(\hat{x},\hat{w},y)\right], \tag{6}$$

- $\mathcal{G}_{\mathcal{Z}}(x,w,y) \ge 0$ and $(x,w,y) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star \implies \mathcal{G}_{\mathcal{Z}}(x,w,y) = 0$.

Introduction

We write (P) into the minimax form:

$$\min_{x\in\mathbb{R}^p, w\in\mathbb{R}^m} \max_{y\in\mathbb{R}^d} \Big\{ \mathcal{L}(x,w,y) := F(x,w) + \langle Kx + Bw - b, y\rangle \Big\}, \tag{4}$$

For any point $(x^\star, w^\star, y^\star) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star$, we have

$$\mathcal{L}(x^\star, w^\star, y) \le \mathcal{L}(x^\star, w^\star, y^\star) \le \mathcal{L}(x, w, y^\star), \quad \forall x \in \mathbb{R}^p, w \in \mathbb{R}^m, y \in \mathbb{R}^d. \tag{5}$$

Suppose (x, w, y) are random variables. Given a compact set $\mathcal{Z}$ such that $\mathcal{Z} \cap \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star \neq \emptyset$, we define the following primal-dual expected gap for (x, w, y).

Primal-dual expected gap

$$\mathcal{G}_{\mathcal{Z}}(x,w,y) := \sup_{(\hat{x},\hat{w},\hat{y})\in\mathcal{Z}} \mathbb{E}\left[\mathcal{L}(x,w,\hat{y}) - \mathcal{L}(\hat{x},\hat{w},y)\right], \tag{6}$$

- $\mathcal{G}_{\mathcal{Z}}(x,w,y) \ge 0$ and $(x,w,y) \in \mathcal{X}^\star \times \mathcal{W}^\star \times \mathcal{Y}^\star \implies \mathcal{G}_{\mathcal{Z}}(x,w,y) = 0$.
- $\mathcal{G}_{\mathcal{Z}}(x,w,y) \le \mathbb{E}\left[\sup_{(\hat{x},\hat{w},\hat{y})\in\mathcal{Z}} \{\mathcal{L}(x,w,\hat{y}) - \mathcal{L}(\hat{x},\hat{w},y)\}\right]$

Outline of the talk

Introduction

Our Algorithm

Main theorems

Numerical results

Our Algorithm

Our method relies on the following augmented Lagrangian function:

$$\mathcal{L}_\rho(x, w, y) := f(x) + h(x) + g(w) + \underbrace{\langle Kx + Bw - b, y\rangle + \frac{\rho}{2}\|Kx + Bw - b\|^2}_{\psi_\rho(x,w,y)}. \quad (7)$$

Our main idea is presented as follows:

- Minimizing $\mathcal{L}_\rho(x, w, y)$ w.r.t. w given $(\hat{x}^k, \hat{w}^k, \hat{y}^k)$:

$$w^{k+1} \in \arg\min_{w \in \mathbb{R}^m} \left\{ g(w) + \langle \hat{y}^k, Bw\rangle + \frac{\rho_k}{2}\|Bw + K\hat{x}^k - b\|^2 \right\}. \quad (8)$$

Our Algorithm

Our method relies on the following augmented Lagrangian function:

$$\mathcal{L}_\rho(x, w, y) := f(x) + h(x) + g(w) + \underbrace{\langle Kx + Bw - b, y\rangle + \frac{\rho}{2}\|Kx + Bw - b\|^2}_{\psi_\rho(x,w,y)}. \quad (7)$$

Our main idea is presented as follows:

- Minimizing $\mathcal{L}_\rho(x, w, y)$ w.r.t. w given $(\hat{x}^k, \hat{w}^k, \hat{y}^k)$:

$$w^{k+1} \in \arg\min_{w\in\mathbb{R}^m} \left\{ g(w) + \langle \hat{y}^k, Bw\rangle + \frac{\rho_k}{2}\|Bw + K\hat{x}^k - b\|^2 \right\}. \quad (8)$$

- Minimizing $\mathcal{L}_\rho(x, w, y)$ w.r.t. x given $(\hat{x}^k, \hat{w}^k, \hat{y}^k)$:

$$\begin{aligned} \tilde{x}_i^{k+1} \quad = \quad & \arg\min_{x_i\in\mathbb{R}^{p_i}} \Big\{ f_i(x_i) + \langle \nabla_{x_i} h(\hat{x}^k) + \nabla_{x_i}\psi_{\rho_k}(\hat{x}^k, w^{k+1}, \hat{y}^k), x_i - \hat{x}_i^k\rangle \\ & + \tfrac{\tau_k\sigma_i}{2\tau_0\beta_k}\|x_i - \tilde{x}_i^k\|^2 \Big\}, \end{aligned} \quad (9)$$

where σ_i is the scaling parameter for each block i.

Our Algorithm

- We apply the Tseng's accelerated framework to get the optimal convergence rate
- We use the following novel dual update to make the last primal iterate (x^k, w^k) converge:

$$\hat{y}^{k+1} := \hat{y}^k + \eta_k \left[(Kx^{k+1} + Bw^{k+1} - b) - (1-\tau_k)(Kx^k + Bw^k - b) \right]$$

Algorithm 1 (Randomized Block-Coordinate Alternating Primal-Dual Algorithm)

1: **For** $k := 1, \cdots, K$

2: Update $\hat{x}^k := (1-\tau_k)x^k + \tau_k \tilde{x}^k$.

3: Update w^{k+1} by solving (8).

4: Sample a block-coordinate i_k with distribution that **Prob**$(i_k = i) = q_i$.

5: Maintain $\tilde{x}_i^{k+1} := \tilde{x}_i^k$ for all $i \neq i_k$, and for $i = i_k$, update $\tilde{x}_i^{k+1}$ by solving (9).

6: Update $x^{k+1} := \hat{x}^k + \frac{\tau_k}{\tau_0}(\tilde{x}^{k+1} - \tilde{x}^k)$.

7: Update $\hat{y}^{k+1} := \hat{y}^k + \eta_k[(Kx^{k+1} + Bw^{k+1} - b) - (1-\tau_k)(Kx^k + Bw^k - b)]$.

8: Update $\bar{y}^{k+1} := (1-\tau_k)\bar{y}^k + \tau_k \left[\hat{y}^k + \rho_k (K\hat{x}^k + Bw^{k+1} - b) \right]$ (if necessary).

9: **End**

Outline of the talk

Introduction

Our Algorithm

Main theorems

Numerical results

Main theorems

Parameter update (general convex case)

$$\tau_k := \frac{\tau_0}{k+1}, \quad \rho_k := \frac{\rho_0 \tau_0}{\tau_k}, \quad \beta_k := \frac{1}{L_\sigma^h + 2\bar{L}_\sigma \rho_k}, \quad \text{and} \quad \eta_k := \frac{\rho_k}{2}, \tag{10}$$

where $\bar{L}_\sigma := \max_{i\in[n]} \left\{ \frac{\|K_i\|^2}{\sigma_i} \right\}$, $\quad L_\sigma^h := \max_{i\in[n]} \left\{ \frac{L_{h,i}}{\sigma_i} \right\}$, $\quad$ and $\quad \tau_0 := \min_{i\in[n]} \{q_i\}$.

Theorem 1

Set $\tilde{x}^0 := x^0$, $\bar{y}^0 := \hat{y}^0$. *Let* $\left\{(x^k, w^k, \bar{y}^k)\right\}$ *be generated by Algorithm 1, where* τ_k, β_k, ρ_k, *and* η_k *are updated by* (10). *Then,*

$$\begin{cases} \mathbb{E}\left[F(x^k, w^k) - F^\star\right] & \leq \dfrac{\bar{\mathcal{E}}_0 + \|y^\star\|(2\bar{\mathcal{E}}_0/\rho_0)^{1/2}}{\tau_0 k + 1 - \tau_0}, \\ \mathbb{E}\left[\|Kx^k + Bw^k - b\|^2\right] & \leq \dfrac{2\bar{\mathcal{E}}_0}{\rho_0(\tau_0 k + 1 - \tau_0)^2}, \\ \mathcal{G}_{\mathcal{Z}}(x^k, w^k, \bar{y}^k) & \leq \dfrac{F(x^0, w^0) - D(\hat{y}^0) + \bar{R}^2_{\mathcal{Z}}}{\tau_0 k + 1 - \tau_0}, \end{cases} \tag{11}$$

where $u^0 := Kx^0 + Bw^0 - b$, *and*

$$\begin{cases} \bar{\mathcal{E}}_0 & := F(x^0, w^0) - D(\hat{y}^0) + \frac{(L^h_\sigma + 2\rho_0\bar{L}_\sigma)\tau_0}{2}\|x^\star - x^0\|^2_{\sigma/q} + \frac{2}{\rho_0}\|y^\star - \hat{y}^0\|^2 \\ & \quad + \frac{1}{\rho_0}\|\hat{y}^0\|^2 + \frac{\rho_0(2-\tau_0)}{2}\|u^0\|^2, \\ \bar{R}^2_{\mathcal{Z}} & := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\left\{\frac{(L^h_\sigma + 2\rho_0\bar{L}_\sigma)\tau_0}{2}\|x - x^0\|^2_{\sigma/q} + \frac{2}{\rho_0}\|y - \hat{y}^0\|^2\right\} \\ & \quad + \frac{1}{\rho_0}\|\hat{y}^0\|^2 + \frac{\rho_0(2-\tau_0)}{2}\|u^0\|^2. \end{cases} \tag{12}$$

Main theorems

Parameter update (strongly convex case)

$$\tau_k := \frac{\tau_{k-1}\left[\sqrt{\tau_{k-1}^2+4}-\tau_{k-1}\right]}{2}, \rho_k := \frac{\rho_{k-1}}{1-\tau_k}, \beta_k := \frac{1}{L_\sigma^h + 2\bar{L}_\sigma\rho_k}, \text{and } \eta_k := \frac{\rho_k}{2}, \tag{13}$$

where $\bar{L}_\sigma := \max_{i\in[n]}\left\{\frac{\|K_i\|^2}{\sigma_i}\right\}$, $\quad L_\sigma^h := \max_{i\in[n]}\left\{\frac{L_{h,i}}{\sigma_i}\right\}$, $\quad$ and $\quad \tau_0 := \min_{i\in[n]}\{q_i\}$.

Main theorems

Theorem 2

Set $\tilde{x}^0 := x^0$, $\bar{y}^0 := \hat{y}^0$. Let $\left\{(x^k, w^k, \bar{y}^k)\right\}$ be generated by Algorithm 1, where τ_k, β_k, ρ_k, and η_k are updated by (13). Then,

$$\begin{cases} \mathbb{E}\left[F(x^k, w^k) - F^\star\right] & \leq & \dfrac{4\left[\tilde{\mathcal{E}}_0 + \|y^\star\|(2\tilde{\mathcal{E}}_0/\rho_0)^{1/2}\right]}{(\tau_0 k + 2)^2}, \\ \mathbb{E}\left[\|Kx^k + Bw^k - b\|^2\right] & \leq & \dfrac{8\tilde{\mathcal{E}}_0}{\rho_0(\tau_0 k + 2)^4}, \\ \mathcal{G}_{\mathcal{Z}}(x^k, w^k, \bar{y}^k) & \leq & \dfrac{F(x^0, w^0) - D(\hat{y}^0) + \tilde{R}^2_{\mathcal{Z}}}{(\tau_0 k + 2)^2}, \end{cases} \tag{14}$$

where $u^0 := Kx^0 + Bw^0 - b$, and

$$\begin{cases} \tilde{\mathcal{E}}_0 & := & F(x^0, w^0) - D(\hat{y}^0) + \sum_{i=1}^n \frac{\tau_0}{2q_i}\left[(L^h_\sigma + 2\rho_0\bar{L}_\sigma)\sigma_i + \mu_{f_i}\right]\|x^\star_i - x^0_i\|^2 \\ & & + \frac{2}{\rho_0}\|y^\star - \hat{y}^0\|^2 + \frac{1}{\rho_0}\|\hat{y}^0\|^2 + \frac{\rho_0(2-\tau_0)}{2}\|u^0\|^2, \\ \tilde{R}^2_{\mathcal{Z}} & := & \sup\limits_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left\{\sum_{i=1}^n \frac{\tau_0}{2q_i}\left[(L^h_\sigma + 2\rho_0\bar{L}_\sigma)\sigma_i + \mu_{f_i}\right]\|x_i - x^0_i\|^2 \right. \\ & & \left. + \ \frac{2}{\rho_0}\|y - \hat{y}^0\|^2\right\} + \frac{1}{\rho_0}\|\hat{y}^0\|^2 + \frac{\rho_0(2-\tau_0)}{2}\|u^0\|^2. \end{cases} \tag{15}$$

Outline of the talk

Introduction

Our Algorithm

Main theorems

Numerical results

Numerical results

Example 3 (Support vector machine)

Given a training set of m examples $\{(a_i, b_i)\}_{i=1}^m$, $a_i \in \mathbb{R}^p$ and class labels $b_i \in \{-1, +1\}$, the soft margin SVM problem is defined as

$$\min_{x \in \mathbb{R}^p} \left\{ F(x) := \frac{1}{m} \sum_{i=1}^{m} \max\left\{0, 1 - b_i \langle a_i, x \rangle\right\} + \frac{\lambda}{2} \|x\|^2 \right\}. \tag{16}$$

Let us define $g(w) := \frac{1}{m} \sum_{i=1}^m \max\{0, 1 - w_i\}$, $f(x) := \frac{\lambda}{2}\|x\|^2$, $h(x) := 0$, and using a linear constraint $Ax - w = 0$, where $b_i a_i$ is the i-th row of A. Then, (16) can be cast into (P).

Numerical results

- Verifying theoretical convergence rate

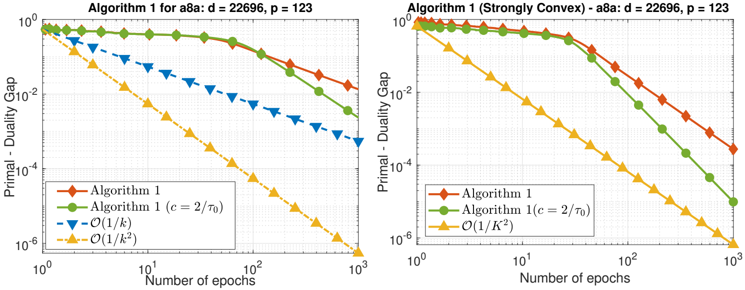

Figure: Convergence rates of Algorithm 1 and its variant (using a modified rule of $\tau_k = \frac{c\tau_0}{k+c}$) for solving (16).

Numerical results

- Comparing with PDHG and SPDHG

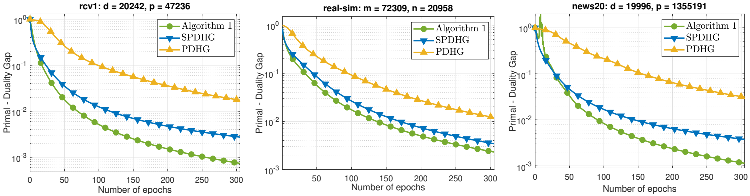

Figure: Comparison of Algorithm 1 with PDHG and SPDHG for solving (16).

Numerical results

Example 4 (Least absolute deviation (LAD) problem)

We consider the following well-studied least absolute deviations (LAD) problem:

$$\min_{x \in \mathbb{R}^p} \left\{ F(x) := \|Kx - b\|_1 + \lambda \|x\|_1 \right\}, \tag{17}$$

where $K \in \mathbb{R}^{d \times p}$, $b \in \mathbb{R}^d$ and $\lambda > 0$ is a regularization parameter.

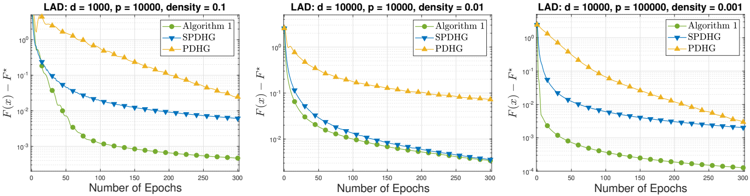

Figure: Comparison of Algorithm 1 with PDHG and SPDHG on (17) using synthetic data.

Thank you for your attention!