Linear Convergence of Randomized Primal-Dual Coordinate Method for Large-scale Linear Constrained Convex Programming

Daoli Zhu and **Lei Zhao**

Shanghai Jiao Tong University
ICML 2020

July 16, 2020

Outline

1. Research Problem

Linear Constrained Convex Programming (LCCP):

$$(\text{P}): \quad \begin{array}{ll} \min & F(u) = G(u) + J(u) \\ \text{s.t} & Au - b = 0 \\ & u \in \mathbf{U} \end{array} \tag{1.1}$$

Assumption 1

(H_1) *J is a convex, lower semi-continuous function (not necessarily differentiable) such that* $\mathbf{dom} J \cap \mathbf{U} \neq \emptyset$.

(H_2) *G is convex and differentiable, and its derivative is Lipschitz with constant* B_G.

(H_3) *There exists at least one saddle point for the Lagrangian of (P).*

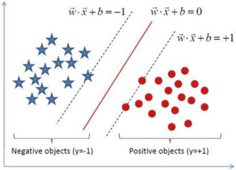
Decomposition for partial structured problem:

- Space decomposition of $\mathbf{U}$: $\mathbf{U} = \mathbf{U}_1 \times \mathbf{U}_2 \cdots \times \mathbf{U}_N, \mathbf{U}_i \subset \mathbf{R}^{n_i}, \sum_{i=1}^{N} n_i = n$.
- $J(u) = \sum_{i=1}^{N} J_i(u_i)$ and $A = (A_1, A_2, \cdots, A_N) \in \mathbf{R}^{m \times n}$ is an appropriate partition of A, where A_i is an $m \times n_i$ matrix.

1.1 Motivation

Support vector machine (SVM) problem:

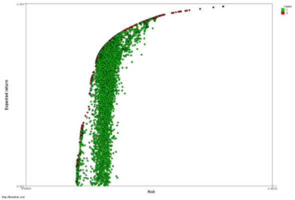
$$\text{(SVM)} \quad \min_{u \in [0,c]^n} \quad \frac{1}{2} u^\top Q u - \mathbf{1}_n^\top u \qquad \text{s.t.} \quad y^\top u = 0$$



$Q \in \mathbf{R}^{n \times n}$ is symmetric and positive-definite.
$c > 0$, $y \in \{-1, 1\}^n$.

Machine learning portfolio (MLP) problem:

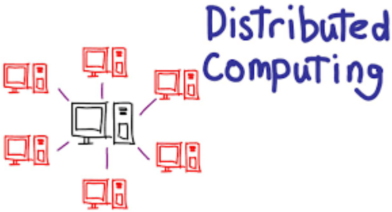
$$\text{(MLP)} \quad \min_{u \in \mathbf{R}^n} \quad \frac{1}{2} u^\top \Sigma u + \lambda \|u\|_1 \qquad \text{s.t.} \quad \mu^\top u = \rho, \quad \mathbf{1}_n^\top u = 1$$



$\Sigma \in \mathbf{R}^{n \times n}$ is the estimated covariance matrix of asset returns.
$\mu \in \mathbf{R}^n$ is the expectation of asset returns.
ρ is a predefined prospective growth rate.

1.1 Motivation

- In the big data era, the datasets used for computation are very big and are often distributed in different locations.
- It is often impractical to assume that optimization algorithms can traverse an entire dataset once in each iteration, because doing so is either time consuming or unreliable.
- Coordinate-type methods can make progress by using distributed information and thus, provide much flexibility for their implementation in the distributed environments.
- Therefore, we adopt randomized coordinate methods for the constrained optimization problem with emphasis on the convergence and rate of convergence properties.



1.2 Related works: augmented Lagrangian decomposition method

The augmented Lagrangian of (P) is $L_\gamma(u,p) = F(u) + \langle p, Au - b\rangle + \frac{\gamma}{2}\|Au - b\|^2$.

Augmented Lagrangian method (ALM) (Hestenes, 1969; Powell, 1969)

$$\begin{cases} u^{k+1} = \arg\min_{u \in \mathbf{U}} L_\gamma(u, p^k); \\ p^{k+1} = p^k + \gamma(Au^{k+1} - b). \end{cases}$$

does not preserve separability

Augmented Lagrangian decomposition method (I)

Alternating Direction Method of Multipliers (ADMM) (Fortin & Glowinski, 1983)

$$\begin{cases} u_1^{k+1} = \arg\min_{u_1 \in \mathbf{U}_1} L_\gamma(u_1, u_2^k, u_3^k, \dots, u_{N-1}^k, u_N^k, p^k); \\ u_2^{k+1} = \arg\min_{u_1 \in \mathbf{U}_1} L_\gamma(u_1^{k+1}, u_2, u_3^k, \dots, u_{N-1}^k, u_N^k, p^k); \\ \quad\vdots \\ u_N^{k+1} = \arg\min_{u_1 \in \mathbf{U}_1} L_\gamma(u_1^{k+1}, u_2^{k+1}, u_3^{k+1}, \dots, u_{N-1}^{k+1}, u_N, p^k); \\ p^{k+1} = p^k + \gamma(Au^{k+1} - b). \end{cases}$$

Gauss-Seidel method for ALM

1.2 Related works: augmented Lagrangian decomposition method

Augmented Lagrangian decomposition method (II)

Auxiliary Problem Principle of Augmented Lagrangian (APP-AL) (Cohen & Zhu, 1983)

$$\begin{cases} u^{k+1} = \arg\min_{u \in \mathbf{U}} \langle \nabla G(u^k), u \rangle + J(u) \\ \qquad\qquad + \langle p^k + \gamma(Au^k - b), Au \rangle \\ \qquad\qquad\qquad + \frac{1}{\epsilon} D(u, u^k); \\ p^{k+1} = p^k + \gamma(Au^{k+1} - b). \end{cases}$$

linearize the smooth term in primal problem of ALM and add a regularization term

where $D(u,v) = K(u) - K(v) - \langle \nabla K(v), u - v \rangle$ is a Bregman like function.

Randomized Primal-Dual Coordinate method (RPDC) (This paper)

$$\begin{cases} \text{Choose } i(k) \text{ from } \{1, \ldots, N\} \text{ with equal probability;} \\ u^{k+1} = \arg\min_{u \in \mathbf{U}} \langle \nabla_{i(k)} G(u^k), u_{i(k)} \rangle + J_{i(k)}(u_{i(k)}) \\ \qquad\qquad + \langle p^k + \gamma(Au^k - b), A_{i(k)} u_{i(k)} \rangle \\ \qquad\qquad\qquad + \frac{1}{\epsilon} D(u, u^k); \\ p^{k+1} = p^k + \rho(Au^{k+1} - b). \end{cases}$$

randomly updates one block of variables in primal subproblem of APP-AL

1.2 Related works: comparison between RPDC and Randomized Coordinate Descent algorithm (RCD) by Necoara & Patrascu, 2014

Randomized Primal-Dual Coordinate method (RPDC) (This paper)

$$\begin{cases} \text{Choose } i(k) \text{ from } \{1, ..., N\} \text{ with equal probability;} \\ u^{k+1} = \arg\min_{u \in \mathbf{U}} \langle \nabla_{i(k)} G(u^k), u_{i(k)} \rangle + J_{i(k)}(u_{i(k)}) + \langle p^k + \gamma(Au^k - b), A_{i(k)} u_{i(k)} \rangle + \frac{1}{\epsilon} D(u, u^k); \\ p^{k+1} = p^k + \rho(Au^{k+1} - b). \end{cases}$$

Necoara & Patrascu, 2014 consider problem (P) with $A \in \mathbf{R}^{1 \times n}$, $b = 0$, and $\mathbf{U} = \mathbf{R}^n$:

$$\text{(P'):} \quad \min_{u \in \mathbf{R}^n} \quad G(u) + J(u), \quad \text{s.t} \quad a^\top u = 0.$$

where $a = (a_1, ..., a_n)^\top \in \mathbf{R}^n$. And the randomized coordinate descent algorithm (RCD) by Necoara & Patrascu, 2014 for (P') is

$$\begin{cases} \text{Choose } i(k) \text{ and } j(k) \text{ from } \{1, ..., n\} \text{ with equal probability;} \\ u^{k+1} = \arg\min_{a_{i(k)} u_{i(k)} + a_{j(k)} u_{j(k)} = 0} \langle \nabla_{i(k)} G(u^k), u_{i(k)} \rangle + \langle \nabla_{j(k)} G(u^k), u_{j(k)} \rangle + J_{i(k)}(u_{i(k)}) \\ \qquad + J_{j(k)}(u_{j(k)}) + \frac{1}{2\epsilon} \|u - u^k\|^2. \end{cases}$$

The RPDC method can deal with more complex problem than RCD.

1.2 Related works: similar schemes

Paper	**Problem**	**Algorithm**	**Theoretical Results**
Xu & Zhang, 2018	(P)	similar to RPDC	F is strongly convex: $O(1/t^2)$ *rate*.
Gao, Xu & Zhang, 2019	(P)	similar to RPDC	F is convex: $O(1/t)$ *rate*.
This paper	(P)	RPDC	F is convex: (i) *Almost surely convergence*; (ii) $O(1/t)$ *rate*; Global strong metric subregularity: (iii) *Linear convergence*.

1.3 Contribution

We propose the randomized primal-dual coordinate (RPDC) method based on the first-order primal-dual method Cohen & Zhu, 1984; Zhao & Zhu, 2019.

(i) We show that the sequence generated by RPDC converges to an optimal solution with probability 1.

(ii) We show RPDC has expected $O(1/t)$ rate for general LCCP.

(iii) We establish the expected linear convergence of RPDC under global strong metric subregularity.

(iv) We show that SVM and MLP problems satisfy global strong metric subregularity under some reasonable conditions.

2. Preliminaries

Lagrangian of (P):

$$L(u,p) = F(u) + \langle p, Au - b\rangle,$$

Saddle point inequality:

$$\forall u \in \mathbf{U}, p \in \mathbf{R}^m : L(u^*, p) \leq L(u^*, p^*) \leq L(u, p^*). \tag{2.2}$$

Karush-Kuhn-Tucker (KKT) system of (P):

Let $w = (u,p)$ and $\mathbf{U}^* \times \mathbf{P}^*$ be the set of saddle points. $\forall w \in \mathbf{U}^* \times \mathbf{P}^*$,

$$0 \in H(w) = \begin{pmatrix} \partial_u L(u,p) + \mathcal{N}_{\mathbf{U}}(u) \\ -\nabla_p L(u,p) \end{pmatrix} = \begin{pmatrix} \nabla G(u) + \partial J(u) + A^\top p + \mathcal{N}_{\mathbf{U}}(u) \\ b - Au \end{pmatrix},$$

with $\mathcal{N}_{\mathbf{U}}(u) = \{\xi : \langle \xi, \zeta - u\rangle \leq 0, \forall \zeta \in \mathbf{U}\}$ is the normal cone at u to $\mathbf{U}$.

3. Convergence and Convergence Rate Analysis of RPDC: RPDC Algorithm

Algorithm 1: Randomized Primal-Dual Coordinate method (RPDC)

for $k = 1$ **to** t

Choose $i(k)$ *from* $\{1, \dots, N\}$ *with equal probability;*

$$u^{k+1} = \arg\min_{u \in \mathbf{U}} \langle \nabla_{i(k)} G(u^k), u_{i(k)} \rangle + J_{i(k)}(u_{i(k)}) + \langle q^k, A_{i(k)} u_{i(k)} \rangle + \frac{1}{\epsilon} D(u, u^k);$$

$$p^{k+1} = p^k + \rho(Au^{k+1} - b).$$

end for

where $q^k = p^k + \gamma(Au^k + b)$ and $D(u, v) = K(u) - K(v) - \langle \nabla K(v), u - v \rangle$ is a Bregman like function with K is strongly convex and gradient Lipschitz.

Assumption 2

(i) *K is strongly convex with parameter β and gradient Lipschitz continuous with parameter B.*

(ii) *The parameters ϵ and ρ satisfy:* $0 < \epsilon < \beta / [B_G + \gamma \lambda_{\max}(A^\top A)]$ *and* $0 < \rho < \frac{2\gamma}{2N-1}$.

3. Convergence and Convergence Rate Analysis of RPDC: Preparation

Filtration:

$$\mathcal{F}_k \overset{def}{=} \{i(0), i(1), \ldots, i(k)\}, \mathcal{F}_k \subset \mathcal{F}_{k+1}.$$

The conditional expectation with respect to $\mathcal{F}_k$: $\mathbb{E}_{\mathcal{F}_{k+1}} = \mathbb{E}(\cdot|\mathcal{F}_k)$.
The conditional expectation in the $i(k)$ term for given $i(0), i(1), \ldots, i(k-1)$: $\mathbb{E}_{i(k)}$.

Reference point:

$w^k = (u^k, p^k)$

APP-AL:

$$\begin{cases} T_u(w^k) = \arg\min_{u \in \mathbf{U}} \langle \nabla G(u^k), u \rangle + J(u) + \langle q^k, Au \rangle \\ \qquad\qquad + \frac{1}{\epsilon} D(u, u^k); \\ T_p(w^k) = p^k + \gamma \left[A T_u(w^k) - b \right]. \end{cases}$$

$T(w^k) = (T_u(w^k), T_p(w^k))$

$$\mathbb{E}_{i(k)} u^{k+1} = \frac{1}{N} T_u(w^k) + (1 - \frac{1}{N}) u^k$$

3. Convergence and Convergence Rate Analysis of RPDC: Preparation

For any $w, w' \in \mathbf{U} \times \mathbf{R}^m$, we construct the function

$$\Lambda(w, w') = \frac{\epsilon(N-1)}{N}[L(u,p) - L(u^*,p^*)] + D(u',u) + \frac{\epsilon}{2N\rho}\|p - p'\|^2 + \frac{\epsilon(N-2)\gamma}{2N}\|Au - b\|^2.$$

Let $w' = w^*$,

$$\Lambda(w, w^*) = \underbrace{\frac{\epsilon(N-1)}{N}[L(u,p) - L(u^*,p^*)]}_{\text{Lagrangian residual}} + \underbrace{D(u^*,u) + \frac{\epsilon}{2N\rho}\|p - p^*\|^2}_{\text{primal, dual residual}} + \underbrace{\frac{\epsilon(N-2)\gamma}{2N}\|Au - b\|^2}_{\text{feasibility residual}}.$$

Lemma 1 (Boundness of $\Lambda(w, w^*)$ and $\Lambda(w, w')$)

There exist $d_1 > 0$, $d_2 > 0$ and $d_3 > 0$, such that

(i) *Lower bound of $\Lambda(w, w^*)$: $\Lambda(w, w^*) \geq d_1\|w - w^*\|^2$;*

(ii) *Upper bound of $\Lambda(w, w^*)$: $\Lambda(w, w^*) \leq d_2\|w - w^*\|^2 + \frac{\epsilon(N-1)}{N}[L(u,p^*) - L(u^*,p^*)]$;*

(iii) *Lower bound of $\Lambda(w, w')$: $\Lambda(w, w') \geq -d_3\|p - p^*\|^2$.*

3. Convergence and Convergence Rate Analysis of RPDC: Preparation

From Assumption 1, 2 and RPDC scheme

$$\begin{aligned}&\tfrac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1},q^k)-L(u,q^k)\right]\\ \leq\ &\left[D(u,u^k)-\mathbb{E}_{i(k)}D(u,u^{k+1})\right]\\ &+\mathbb{E}_{i(k)}\Big\{\tfrac{\epsilon(N-1)}{N}\left[L(u^k,p^k)-L(u^{k+1},p^{k+1})\right]\\ &-\tfrac{\beta-\epsilon[B_G+\frac{N-1}{N}\gamma\lambda_{\max}(A^\top A)]}{2}\|u^k-u^{k+1}\|^2\\ &+\tfrac{\epsilon\gamma(N-1)}{2N}\|Au^k-b\|^2\\ &+\tfrac{\epsilon(2\rho-\gamma)(N-1)}{2N}\|Au^{k+1}-b\|^2\Big\}\end{aligned}$$

From RPDC scheme

$$\begin{aligned}&\tfrac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1},p)-L(u^{k+1},q^k)\right]\\ \leq\ &\tfrac{\epsilon}{2N\rho}\left[\|p-p^k\|^2-\mathbb{E}_{i(k)}\|p-p^{k+1}\|^2\right]\\ &+\tfrac{\epsilon\frac{1}{N}\gamma\lambda_{\max}(A^\top A)}{2}\mathbb{E}_{i(k)}\|u^k-u^{k+1}\|^2\\ &-\tfrac{\epsilon\gamma}{2N}\|Au^k-b\|^2+\tfrac{\epsilon(\rho-\gamma)}{2N}\mathbb{E}_{i(k)}\|Au^{k+1}-b\|^2.\end{aligned}$$

Assumption 2 and $\mathbb{E}_{i(k)}u^{k+1}=\frac{1}{N}T_u(w^k)+(1-\frac{1}{N})u^k$

Lemma 2 (Estimation on the variance of $\Lambda(w^k,w)$)

Lemma 2 (Estimation on the variance of $\Lambda(w^k,w)$)

There exists $d_4>0$, such that

$$\Lambda(w^k,w)-\mathbb{E}_{i(k)}\Lambda(w^{k+1},w)\geq\tfrac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1},p)-L(u,q^k)\right]+d_4\|w^k-T(w^k)\|^2.$$

3. Convergence and Convergence Rate Analysis of RPDC: Convergence Analysis

Lemma 1
$\Lambda(w, w^*) \geq d_1 \|w - w^*\|^2 \geq 0$

Take $w = w^*$ *in Lemma 2,*
$\Lambda(w^k, w^*) - \mathbb{E}_{i(k)} \Lambda(w^{k+1}, w^*) \geq d_4 \|w^k - T(w^k)\|^2.$

↓

Robbins-Siegmund's Lemma (Robbins & Siegmund, 1971).

↓

Theorem 1 (Almost surely convergence)

Theorem 1 (Almost surely convergence)

(i) $\sum_{k=0}^{+\infty} \|w^k - T(w^k)\|^2 < +\infty$ *a.s.;*

(ii) *The sequence* $\{w^k\}$ *generated by RPDC is almost surely bounded;*

(iii) *Every cluster point of* $\{w^k\}$ *almost surely is a saddle point of Lagrangian for (P).*

3. Convergence and Convergence Rate Analysis of RPDC: Convergence Rate Analysis

From Lemma 1

$$h(w,w') = \Lambda(w,w') + \frac{d_3}{d_1}\Lambda(w,w^*) \geq 0$$

From Lemma 2

$$\mathbb{E}_{\mathcal{F}_t}[\Lambda(w^k,w) - \Lambda(w^{k+1},w)] \geq \frac{\epsilon}{N}\mathbb{E}_{\mathcal{F}_t}\left[L(u^{k+1},p) - L(u,q^k)\right]$$

and $\mathbb{E}_{\mathcal{F}_t}[\Lambda(w^k,w^*) - \Lambda(w^{k+1},w^*)] \geq 0$

$$\mathbb{E}_{\mathcal{F}_t}[h(w^k,w) - h(w^{k+1},w)] \geq \frac{\epsilon}{N}\mathbb{E}_{\mathcal{F}_t}\left[L(u^{k+1},p) - L(u,q^k)\right]$$

$$\bar{u}_t = \frac{\sum_{k=0}^{t} u^{k+1}}{t+1} \text{ and } \bar{p}_t = \frac{\sum_{k=0}^{t} q^k}{t+1}.$$

Theorem 2 ($O(1/t)$ convergence rate)

Theorem 2 ($O(1/t)$ convergence rate)

- (i) *Global estimate of expected bifunction values:* $\mathbb{E}_{\mathcal{F}_t}\left[L(\bar{u}_t,p) - L(u,\bar{p}_t)\right] \leq \frac{Nh(w^0,w)}{\epsilon(t+1)}$, $\forall u \in \mathbf{U}, p \in \mathbf{R}^m$, *$(u,p)$ could possibly be random;*
- (ii) *Expected feasibility:* $\mathbb{E}_{\mathcal{F}_t}\|A\bar{u}_t - b\| \leq O(1/t)$;
- (iii) *Expected suboptimality:* $-O(1/t) \leq \mathbb{E}_{\mathcal{F}_t}\left[F(\bar{u}_t) - F(u^*)\right] \leq O(1/t)$.

4. Linear Convergence of RPDC under Global Strong Metric Subregularity

Lemma 1
$\Lambda(w,w^*) \geq d_1\|w-w^*\|^2$;
$\Lambda(w,w^*) \leq d_2\|w-w^*\|^2 + \frac{\epsilon(N-1)}{N}[L(u,p^*)-L(u^*,p^*)]$.

Lemma 2
$\Lambda(w^k,w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1},w) \geq \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1},p)-L(u,q^k)\right] + d_4\|w^k - T(w^k)\|^2$

Lemma 3
(Boundness of $\phi(w,w^*)$ and descent inequality of $\phi(w^k,w^*)$)

where $\phi(w,w^*) = \Lambda(w,w^*) + \frac{\epsilon}{N}[L(u,p^*)-L(u^*,p^*)]$.

Lemma 3 (Boundness of $\phi(w,w^*)$ and descent inequality of $\phi(w^k,w^*)$)

(i) *Lower bound of $\phi(w,w^*)$:* $\phi(w,w^*) \geq d_1\|w-w^*\|^2$.

(ii) *Upper bound of $\phi(w,w^*)$:* $\phi(w,w^*) \leq d_2\|w-w^*\|^2 + \epsilon[L(u,p^*)-L(u^*,p^*)]$.

(iii) *Descent inequality of $\phi(w^k,w^*)$:*
$\phi(w^k,w^*) - \mathbb{E}_{i(k)}\phi(w^{k+1},w^*) \geq d_4\|w^k - T(w^k)\|^2 + \frac{\epsilon}{N}[L(u^k,p^*)-L(u^*,p^*)]$.

4. Linear Convergence of RPDC under Global Strong Metric Subregularity

Definition (Global strong metric subregularity (GS-MS))

Let $\mathcal{H}(x)$ be a set-valued mapping between real spaces $\mathbf{X}$ and $\mathbf{Y}$. Then $\mathcal{H}(x)$ is called global strong metric subregular at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in \mathcal{H}(\bar{x})$ if there exists positive number $\mathfrak{c}$ such that

$$dist(x, \bar{x}) \leq \mathfrak{c} dist\left(\bar{y}, \mathcal{H}(x)\right), \quad \text{for all } x \in \mathbf{X}.$$

$H(w)$ is global strong metric subregularity at w^* for 0

$$\|T(w^k) - w^*\| \leq \mathfrak{c} dist(0, H(T(w^k)))$$

From APP-AL scheme

$$v(T(w^k)) \in H(T(w^k))$$

and $\|v(T(w^k))\|^2 \leq \delta \|w^k - T(w^k)\|^2$

since $\|w^k - w^*\| \leq \|T(w^k) - w^*\| + \|w^k - T(w^k)\|$

$$\|w^k - w^*\| \leq (\mathfrak{c}\sqrt{\delta} + 1)\|w^k - T(w^k)\|$$

where
$$v(T(w^k)) = \begin{pmatrix} \nabla G(T_u(w^k)) - \nabla G(u^k) + A^\top (T_p(w^k) - q^k) \\ \qquad + \frac{1}{\epsilon}\left[\nabla K(u^k) - \nabla K(T_u(w^k))\right] \\ \frac{1}{\gamma}\left[p^k - T_p(w^k)\right] \end{pmatrix} \in H(T(w^k)).$$

4. Linear Convergence of RPDC under Global Strong Metric Subregularity

Global strong metric subregularity

$$\|w^k - w^*\| \leq (\mathfrak{c}\sqrt{\delta}+1)\|w^k - T(w^k)\|.$$

Lemma 3

$$\phi(w^k, w^*) - \mathbb{E}_{i(k)}\phi(w^{k+1}, w^*) \geq d_4\|w^k - T(w^k)\|^2 + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)]$$

$$\phi(w^k, w^*) - \mathbb{E}_{i(k)}\phi(w^{k+1}, w^*) \geq \delta'\{d_2\|w^k - w^*\|^2 + \epsilon[L(u^k, p^*) - L(u^*, p^*)]\}$$

with $\delta' = \min\{\frac{d_4}{\max\{d_2(\mathfrak{c}\sqrt{\delta}+1)^2, d_4+1\}}, \frac{1}{N+1}\} < 1.$

Lemma 3

$$\phi(w, w^*) \leq d_2\|w - w^*\|^2 + \epsilon[L(u, p^*) - L(u^*, p^*)]$$

Theorem 3 (linear convergence of RPDC)

Theorem 3 (Global strong metric subregularity of $H(w)$ implies linear convergence of RPDC)

For given saddle point w^*, if $H(w)$ is global strong metric subregular at w^* for 0, then there exists $\alpha = 1 - \delta' \in (0,1)$ such that $\mathbb{E}_{\mathcal{F}_{k+1}}\phi(w^{k+1}, w^*) \leq \alpha^{k+1}\phi(w^0, w^*), \quad \forall k.$

4. Linear Convergence of RPDC under Global Strong Metric Subregularity

Theorem 3 (linear convergence of RPDC)
$\mathbb{E}_{\mathcal{F}_k}\phi(w^k, w^*) \leq \alpha^k \phi(w^0, w^*), \ \forall k$

Lemma 3
$\phi(w, w^*) \geq d_1 \|w - w^*\|^2$

$$\|\mathbb{E}_{\mathcal{F}_k} w^k - w^*\| \leq \hat{M}(\sqrt{\alpha})^k, \quad \textit{with} \quad \hat{M} = \sqrt{\frac{\phi(w^0, w^*)}{d_1}}.$$

Corollary 1
(R-linear convergence of $\{\mathbb{E}_{\mathcal{F}_k} w^k\}$)

Corollary (R-linear rate of $\{\mathbb{E}_{\mathcal{F}_k} w^k\}$)

The sequence $\{\mathbb{E}_{\mathcal{F}_k} w^k\}$ converges to the desired saddle point w^* at R-linear rate; i.e.,

$$\lim_{k\to\infty} \sup \sqrt[k]{\|\mathbb{E}_{\mathcal{F}_k} w^k - w^*\|} = \sqrt{\alpha}.$$

5. Numerical Analysis: SVM

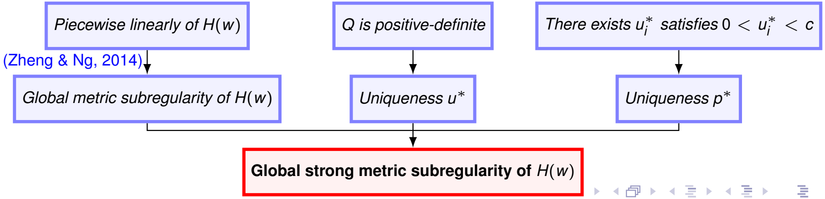
$$\text{(SVM)} \quad \min_{u\in[0,c]^n} \quad \frac{1}{2}u^\top Qu - \mathbf{1}_n^\top u \qquad \text{s.t.} \quad y^\top u = 0$$

KKT mapping of (SVM):

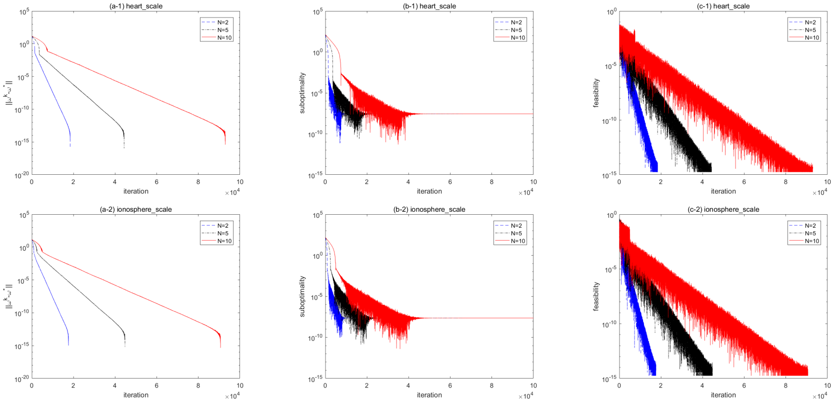
$$H(w) = \begin{pmatrix} Qu - \mathbf{1}_n + py + \mathcal{N}_{[0,c]^n}(u) \\ y^\top u \end{pmatrix}$$

Proposition

Assume there exists at least one component u_i^* of optimal solution u^* that satisfies $0 < u_i^* < c$. Then the KKT mapping for SVM is global strong metric subregular.



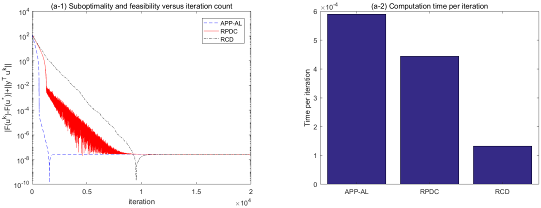
5. Numerical Analysis: SVM



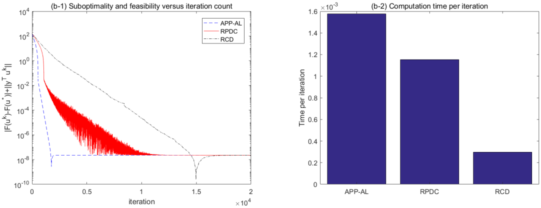
(c) $\|w^k - w^*\|$ value (d) Suboptimality (e) Feasibility

Figure: Number of blocks, $\|w^k - w^*\|$ value, suboptimality, and feasibility with respect to iteration count

5. Numerical Analysis: SVM



(a) Comparison on `heart_scale`



(b) Comparison on `ionosphere_scale`

Figure: Comparison among RPDC with $N = 2$, APP-AL and RCD (Necoara & Patrascu, 2014)

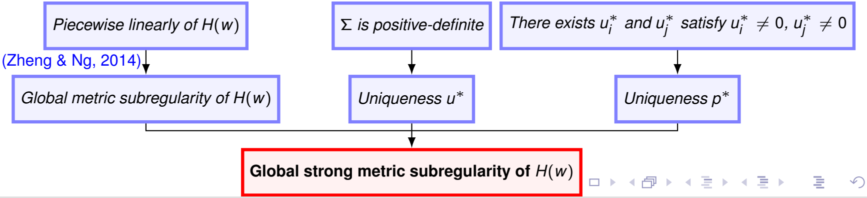
5. Numerical Analysis: MLP

$$
\text{(MLP)} \quad \begin{array}{ll} \min\limits_{u \in \mathbf{R}^n} & \frac{1}{2} u^\top \Sigma u + \lambda \|u\|_1 \\ \text{s.t.} & \mu^\top u = \rho \\ & \mathbf{1}_n^\top u = 1 \end{array}
$$

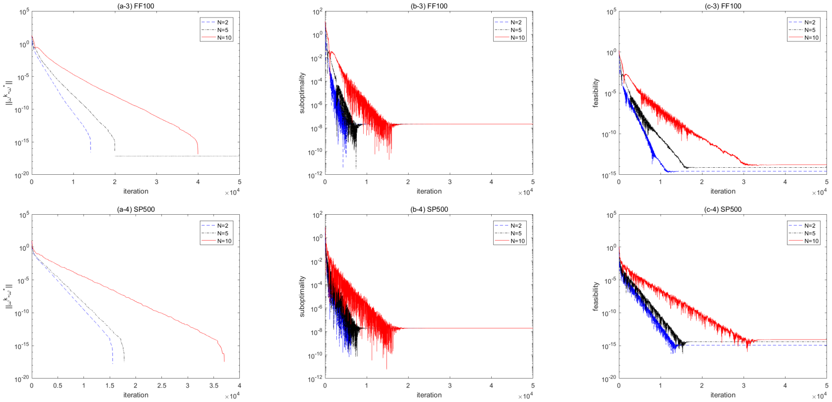
The KKT mapping of MLP: $H(w) = \begin{pmatrix} \Sigma u + \lambda \partial \|u\|_1 + p_1 \mathbf{1}_n + p_2 \mu \\ \mu^\top u - \rho \\ \mathbf{1}_n^\top u - 1 \end{pmatrix}$.

Proposition

Assume there exists at least two component u_i^* and u_j^* for optimal solution u^* that satisfy $u_i^* \neq 0$, $u_j^* \neq 0$; and $\mu_i \neq \mu_j$. Then the KKT mapping for MLP is global strong metric subregular.



5. Numerical Analysis: MLP



(a) $\|w^k - w^*\|$ value (b) Suboptimality (c) Feasibility

Figure: Number of blocks, $\|w^k - w^*\|$ value, suboptimality, and feasibility with respect to iteration count

6. Conclusions

This paper proposed a randomized coordinate extension of the first-order primal-dual method proposed by Cohen & Zhu, 1984 and Zhao & Zhu, 2019 to solve LCCP.

(i) We established almost surely convergence and expected $O(1/t)$ convergence rate for the general convex case.

(ii) Under global strong metric subregularity condition, we establish the expected linear convergence of RPDC.

(iii) SVM and MLP problems satisfy global strong metric subregularity under some reasonable conditions.

We also discussed the implementation details of RPDC and present numerical experiments on SVM and MLP problems to verify the linear convergence.

Future study will consider RPDC for nonlinearly constrained nonconvex and nonsmooth optimization.

For More Details and Results

Contact me by e-mail: l.zhao@sjtu.edu.cn

Download slides: https://drive.google.com/file/d/1SFt0tjV5yUx_r1fIGX0Rsa5FOgvff3Ft/view?usp=sharing

THANK YOU FOR YOUR ATTENTION!