Supplementary material for the paper: "Linear Convergence of Randomized Primal-Dual Coordinate Method for Large-scale Linear Constrained Convex Programming"

First of all, we have the following observations:

In algorithm RPDC, the indices $i(k)$, $k = 0, 1, 2, \dots$ are random variables. After k iterations, RPDC method generates a random output (u^{k+1}, p^{k+1}). Recall the definition of filtration $\mathcal{F}_k$ which is generated by the random variable $i(0), i(1), \dots, i(k)$, i.e.,

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

Additionally, $\mathcal{F} = (\mathcal{F}_k)_{k\in\mathbb{N}}$, $\mathbb{E}_{\mathcal{F}_{k+1}} = \mathbb{E}(\cdot|\mathcal{F}_k)$ is the conditional expectation w.r.t. $\mathcal{F}_k$ and the conditional expectation in term of $i(k)$ given $i(0), i(1), \dots, i(k-1)$ as $\mathbb{E}_{i(k)}$.

Knowing $\mathcal{F}_{k-1} = \{i(0), i(1), \dots, i(k-1)\}$, we have:

$$\mathbb{E}_{i(k)}\langle \nabla_{i(k)} G(u^k), (u^k - u)_{i(k)}\rangle = \frac{1}{N}\langle \nabla G(u^k), u^k - u\rangle \geq \frac{1}{N}\big[G(u^k) - G(u)\big], \tag{A.1}$$

$$\mathbb{E}_{i(k)}\big[J_{i(k)}(u^k_{i(k)}) - J_{i(k)}(u_{i(k)})\big] = \frac{1}{N}\big[J(u^k) - J(u)\big], \tag{A.2}$$

and

$$\mathbb{E}_{i(k)}\langle q^k, A_{i(k)}(u^k - u)_{i(k)}\rangle = \frac{1}{N}\langle q^k, A(u^k - u)\rangle. \tag{A.3}$$

Secondly, reconsidering the point $T(w^k) = \big(T_u(w^k), T_p(w^k)\big)$ generated by one deterministic iteration of APP-AL (Cohen & Zhu, 1984) for given w^k,

APP-AL

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

with $q^k = p^k + \gamma(Au^k - b)$, we have the following observations. The convex combination of u^k and $T_u(w^k)$ provides the expected value of u^{k+1} as following.

$$\mathbb{E}_{i(k)}u^{k+1} = \frac{1}{N}T_u(w^k) + (1 - \frac{1}{N})u^k, \tag{A.4}$$

or

$$T_u(w^k) = N\mathbb{E}_{i(k)}u^{k+1} - (N-1)u^k. \tag{A.5}$$

Moreover, the point $T(w^k)$ satisfies that: for any $(u, p) \in \mathbf{U} \times \mathbf{R}^m$,

$$\begin{cases} \langle \nabla G(u^k), u - T_u(w^k)\rangle + J(u) - J(T_u(w^k)) + \langle q^k, A(u - T_u(w^k))\rangle \\ \qquad + \frac{1}{\epsilon}\langle \nabla K(T_u(w^k)) - \nabla K(u^k), u - T_u(w^k)\rangle \geq 0, \\ \gamma\left[AT_u(w^k) - b\right] = T_p(w^k) - p^k. \end{cases} \tag{A.6}$$

1. Proof of Lemma 1

Proof. Take $w' = w^*$ in (9), we have that

$$\begin{aligned}\Lambda(w,w^*) &= D(u^*,u) + \frac{\epsilon}{2N\rho}\|p-p^*\|^2 + \frac{\epsilon(N-1)}{N}[L(u,p)-L(u^*,p^*)] + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2 \\ &= D(u^*,u) + \frac{\epsilon}{2N\rho}\|p-p^*\|^2 + \frac{\epsilon(N-1)}{N}[L(u,p^*)-L(u^*,p^*)] + \frac{\epsilon(N-1)}{N}\langle p-p^*, Au-b\rangle \\ &\quad + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2. \qquad \text{(A.7)}\end{aligned}$$

(i) Since $L(u,p^*) - L(u^*,p^*) \geq 0$ and $\frac{1}{2\gamma}\|p-p^*\|^2 + \frac{\gamma}{2}\|Au-b\|^2 + \langle p-p^*, Au-b\rangle \geq 0$, (A.7) follows that

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

From Assumption 2, we have $D(u^*,u) \geq \frac{\beta}{2}\|u-u^*\|^2$. Together with the fact $Au^* = b$ and $\rho < \frac{2\gamma}{2N-1}$, above inequality follows that

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

with $d_1 = \min\left\{\frac{1}{2N}[N\beta - \epsilon\gamma\lambda_{\max}(A^\top A)], \frac{\epsilon}{4N\gamma}\right\}$.

(ii) By Young's inequality, (A.7) follows that

$$\begin{aligned}\Lambda(w,w^*) &\leq D(u^*,u) + \frac{\epsilon}{2N\rho}\|p-p^*\|^2 + \frac{\epsilon(N-1)}{N}[L(u,p^*)-L(u^*,p^*)] \\ &\quad + \frac{\epsilon(N-1)}{N}[\frac{1}{2\gamma}\|p-p^*\|^2 + \frac{\gamma}{2}\|Au-b\|^2] + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2.\end{aligned}$$

From Assumption 2, we have $D(u^*,u) \leq \frac{B}{2}\|u-u^*\|^2$. Together with the fact $Au^* = b$ and $2\gamma > (2N-1)\rho$, above inequality follows that

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

with $d_2 = \max\left\{\frac{(4N-3)\epsilon}{(4N-2)N\rho}, \frac{NB + \epsilon(2N-3)\gamma\lambda_{\max}(A^\top A)}{2N}\right\}$.

(iii) By the definition of $\Lambda(w,w')$, we have

$$\begin{aligned}\Lambda(w,w') &\geq \frac{\epsilon(N-1)}{N}[L(u,p)-L(u^*,p^*)] + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2 \\ &= \frac{\epsilon(N-1)}{N}[L(u,p)-L(u,p^*)] + \frac{\epsilon(N-1)}{N}[L(u,p^*)-L(u^*,p^*)] + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2 \\ &\geq \frac{\epsilon(N-1)}{N}[L(u,p)-L(u,p^*)] + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2 \\ &= \frac{\epsilon(N-1)}{N}\langle p-p^*, Au-b\rangle + \frac{\epsilon(N-2)\gamma}{2N}\|Au-b\|^2 \\ &\geq -d_3\|p-p^*\|^2, \qquad \text{(A.8)}\end{aligned}$$

with $d_3 = \frac{\epsilon(N-1)^2}{2\gamma N(N-2)}$.

□

2. Proof of Lemma 2

Proof. *Step 1: Estimate* $\frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1},q^k)-L(u,q^k)\big]$;
For all $u\in\mathbf{U}$, the unique solution u^{k+1} of the primal problem of RPDC is characterized by the following variational inequality:

$$\langle\nabla_{i(k)}G(u^k),(u^{k+1}-u)_{i(k)}\rangle+J_{i(k)}(u^{k+1}_{i(k)})-J_{i(k)}(u_{i(k)})+\langle q^k,A_{i(k)}(u^{k+1}-u)_{i(k)}\rangle$$
$$+\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u^{k+1}-u\rangle\leq 0,$$

which follows that

$$\langle\nabla_{i(k)}G(u^k),\big(u^k-u-(u^k-u^{k+1})\big)_{i(k)}\rangle+J_{i(k)}(u^k_{i(k)})-J_{i(k)}(u_{i(k)})-\big(J_{i(k)}(u^k_{i(k)})-J_{i(k)}(u^{k+1}_{i(k)})\big)$$
$$+\langle q^k,A_{i(k)}\big(u^k-u-(u^k-u^{k+1})\big)_{i(k)}\rangle+\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u^{k+1}-u\rangle\leq 0. \tag{A.9}$$

Observing that for any separable mapping $\psi(u)=\sum_{i=1}^{N}\psi_i(u_i)$, we have $\psi_{i(k)}(u^k_{i(k)})-\psi_{i(k)}(u^{k+1}_{i(k)})=\psi(u^k)-\psi(u^{k+1})$. Therefore, (A.9) follows that

$$\begin{aligned}&\langle\nabla_{i(k)}G(u^k),(u^k-u)_{i(k)}\rangle+J_{i(k)}(u^k_{i(k)})-J_{i(k)}(u_{i(k)})+\langle q^k,A_{i(k)}(u^k-u)_{i(k)}\rangle\\ \leq\;&\langle\nabla G(u^k),u^k-u^{k+1}\rangle+J(u^k)-J(u^{k+1})+\langle q^k,A(u^k-u^{k+1})\rangle\\ &+\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u-u^{k+1}\rangle.\end{aligned} \tag{A.10}$$

Taking expectation with respect to $i(k)$ on both side of (A.10), together the condition expectation (A.1)-(A.3), we get

$$\begin{aligned}\frac{1}{N}\big[L(u^k,q^k)-L(u,q^k)\big]\;\leq\;&\mathbb{E}_{i(k)}\Big\{\langle\nabla G(u^k),u^k-u^{k+1}\rangle+J(u^k)-J(u^{k+1})\\ &+\langle q^k,A(u^k-u^{k+1})\rangle+\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u-u^{k+1}\rangle\Big\}.\end{aligned} \tag{A.11}$$

or

$$\begin{aligned}\frac{1}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1},q^k)-L(u,q^k)\big]\;\leq\;&\mathbb{E}_{i(k)}\Big\{\underbrace{\langle\nabla G(u^k),u^k-u^{k+1}\rangle}_{\mathfrak{a}_1}+J(u^k)-J(u^{k+1})\\ &+\langle q^k,A(u^k-u^{k+1})\rangle+\frac{1}{N}\big[L(u^{k+1},q^k)-L(u^k,q^k)\big]\\ &\underbrace{+\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u-u^{k+1}\rangle}_{\mathfrak{a}_2}\Big\}.\end{aligned} \tag{A.12}$$

By the gradient Lipschitz of G, term $\mathfrak{a}_1$ in (A.12) is bounded by

$$\mathfrak{a}_1=\langle\nabla G(u^k),u^k-u^{k+1}\rangle\leq G(u^k)-G(u^{k+1})+\frac{B_G}{2}\|u^k-u^{k+1}\|^2. \tag{A.13}$$

The simple algebraic operation and Assumption 2 follows that

$$\begin{aligned}\mathfrak{a}_2=\frac{1}{\epsilon}\langle\nabla K(u^{k+1})-\nabla K(u^k),u-u^{k+1}\rangle\;&=\;\frac{1}{\epsilon}\big[D(u,u^k)-D(u,u^{k+1})-D(u^{k+1},u^k)\big]\\ &\leq\;\frac{1}{\epsilon}\big[D(u,u^k)-D(u,u^{k+1})\big]-\frac{\beta}{2\epsilon}\|u^k-u^{k+1}\|^2.\end{aligned} \tag{A.14}$$

Combining (A.12)-(A.14), we obtain that

$$\begin{aligned}\frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1},q^k)-L(u,q^k)\big]\;\leq\;&\big[D(u,u^k)-\mathbb{E}_{i(k)}D(u,u^{k+1})\big]+\mathbb{E}_{i(k)}\Big\{\frac{\epsilon(N-1)}{N}\underbrace{\big[L(u^k,q^k)-L(u^{k+1},q^k)\big]}_{\mathfrak{a}_3}\\ &-\frac{\beta-\epsilon B_G}{2}\|u^k-u^{k+1}\|^2\Big\}\end{aligned} \tag{A.15}$$

Since $p^{k+1} = p^k + \rho(Au^{k+1} - b)$ and $q^k = p^k + \gamma(Au^k - b)$, term $\mathfrak{a}_3$ in (A.15) follows that

$$\begin{aligned}
\mathfrak{a}_3 &= L(u^k, q^k) - L(u^{k+1}, q^k) \\
&= L(u^k, p^k) - L(u^{k+1}, p^{k+1}) + \langle q^k - p^k, Au^k - b\rangle + \langle p^{k+1} - q^k, Au^{k+1} - b\rangle \\
&= L(u^k, p^k) - L(u^{k+1}, p^{k+1}) + \gamma\|Au^k - b\|^2 + \rho\|Au^{k+1} - b\|^2 - \gamma\langle Au^k - b, Au^{k+1} - b\rangle \\
&= L(u^k, p^k) - L(u^{k+1}, p^{k+1}) + \frac{\gamma}{2}\|Au^k - b\|^2 + (\rho - \frac{\gamma}{2})\|Au^{k+1} - b\|^2 + \frac{\gamma}{2}\|A(u^k - u^{k+1})\|^2 \\
&\leq L(u^k, p^k) - L(u^{k+1}, p^{k+1}) + \frac{\gamma}{2}\|Au^k - b\|^2 + (\rho - \frac{\gamma}{2})\|Au^{k+1} - b\|^2 \\
&\quad + \frac{\gamma\lambda_{\max}(A^\top A)}{2}\|u^k - u^{k+1}\|^2.
\end{aligned} \tag{A.16}$$

Combining (A.15)-(A.16), we have that

$$\begin{aligned}
\frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1}, q^k) - L(u, q^k)\big] \leq\ & \big[D(u, u^k) - \mathbb{E}_{i(k)}D(u, u^{k+1})\big] + \mathbb{E}_{i(k)}\bigg\{\frac{\epsilon(N-1)}{N}\big[L(u^k, p^k) - L(u^{k+1}, p^{k+1})\big] \\
& - \frac{\beta - \epsilon[B_G + \frac{N-1}{N}\gamma\lambda_{\max}(A^\top A)]}{2}\|u^k - u^{k+1}\|^2 + \frac{\epsilon\gamma(N-1)}{2N}\|Au^k - b\|^2 \\
& + \frac{\epsilon(2\rho - \gamma)(N-1)}{2N}\|Au^{k+1} - b\|^2\bigg\}
\end{aligned} \tag{A.17}$$

Step 2: Estimate $\frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1}, p) - L(u^{k+1}, q^k)\big]$

$$\begin{aligned}
L(u^{k+1}, p) - L(u^{k+1}, q^k) &= \langle p - q^k, Au^{k+1} - b\rangle \\
&= \frac{1}{\rho}\langle p - p^k, p^{k+1} - p^k\rangle - \gamma\langle Au^k - b, Au^{k+1} - b\rangle \\
&= \frac{1}{2\rho}\big[\|p - p^k\|^2 - \|p - p^{k+1}\|^2 + \|p^k - p^{k+1}\|^2\big] - \gamma\langle Au^k - b, Au^{k+1} - b\rangle \\
&= \frac{1}{2\rho}\big[\|p - p^k\|^2 - \|p - p^{k+1}\|^2 + \|p^k - p^{k+1}\|^2\big] + \frac{\gamma}{2}\|A(u^k - u^{k+1})\|^2 \\
&\quad - \frac{\gamma}{2}\|Au^k - b\|^2 - \frac{\gamma}{2}\|Au^{k+1} - b\|^2 \\
&= \frac{1}{2\rho}\big[\|p - p^k\|^2 - \|p - p^{k+1}\|^2\big] + \frac{\gamma}{2}\|A(u^k - u^{k+1})\|^2 \\
&\quad - \frac{\gamma}{2}\|Au^k - b\|^2 + \frac{\rho - \gamma}{2}\|Au^{k+1} - b\|^2 \qquad \text{(since } p^{k+1} = p^k + \rho(Au^{k+1} - b).\text{)} \\
&\leq \frac{1}{2\rho}\big[\|p - p^k\|^2 - \|p - p^{k+1}\|^2\big] + \frac{\gamma\lambda_{\max}(A^\top A)}{2}\|u^k - u^{k+1}\|^2 \\
&\quad - \frac{\gamma}{2}\|Au^k - b\|^2 + \frac{\rho - \gamma}{2}\|Au^{k+1} - b\|^2
\end{aligned} \tag{A.18}$$

Multiply $\frac{\epsilon}{N}$ on both side of above inequality, we obtain that: $\forall p \in \mathbf{R}^m$

$$\begin{aligned}
\frac{\epsilon}{N}\big[L(u^{k+1}, p) - L(u^{k+1}, q^k)\big] \leq\ & \frac{\epsilon}{2N\rho}\big[\|p - p^k\|^2 - \|p - p^{k+1}\|^2\big] + \frac{\epsilon\frac{1}{N}\gamma\lambda_{\max}(A^\top A)}{2}\|u^k - u^{k+1}\|^2 \\
& - \frac{\epsilon\gamma}{2N}\|Au^k - b\|^2 + \frac{\epsilon(\rho - \gamma)}{2N}\|Au^{k+1} - b\|^2.
\end{aligned} \tag{A.19}$$

Taking expectation with respect to $i(k)$ on both side of inequality (A.19), we have

$$\begin{aligned}
\frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1}, p) - L(u^{k+1}, q^k)\big] \leq\ & \frac{\epsilon}{2N\rho}\big[\|p - p^k\|^2 - \mathbb{E}_{i(k)}\|p - p^{k+1}\|^2\big] + \frac{\epsilon\frac{1}{N}\gamma\lambda_{\max}(A^\top A)}{2}\mathbb{E}_{i(k)}\|u^k - u^{k+1}\|^2 \\
& - \frac{\epsilon\gamma}{2N}\|Au^k - b\|^2 + \frac{\epsilon(\rho - \gamma)}{2N}\mathbb{E}_{i(k)}\|Au^{k+1} - b\|^2.
\end{aligned} \tag{A.20}$$

Step 3: Estimate the variance of $\Lambda(w^k, w)$.

Summing inequalities (A.17) and (A.20), with $d_4 = \frac{\min\left\{\frac{\beta-\epsilon[B_G+\gamma\lambda_{\max}(A^\top A)]}{2}, \frac{\epsilon[2\gamma-(2N-1)\rho]}{2N}\right\}}{\max\{N^2+2\gamma^2(N^2+2)\lambda_{\max}(A^\top A), 4\gamma^2\}}$, we have that

$$\begin{aligned}
& \Lambda(w^k, w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w) \\
\geq\ & \mathbb{E}_{i(k)}\left\{\frac{\epsilon}{N}\left[L(u^{k+1}, p) - L(u, q^k)\right] + \frac{\beta-\epsilon[B_G+\gamma\lambda_{\max}(A^\top A)]}{2}\|u^k - u^{k+1}\|^2 + \frac{\epsilon[2\gamma-(2N-1)\rho]}{2N}\|Au^{k+1} - b\|^2\right\} \\
\geq\ & \mathbb{E}_{i(k)}\left\{\frac{\epsilon}{N}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\left[\left(N^2 + 2\gamma^2(N^2+2)\lambda_{\max}(A^\top A)\right)\|u^k - u^{k+1}\|^2 + 4\gamma^2\|Au^{k+1} - b\|^2\right]\right\} \\
\geq\ & \mathbb{E}_{i(k)}\left\{\frac{\epsilon}{N}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\left[\left(1 + 2\gamma^2\lambda_{\max}(A^\top A)\right)N^2\|u^k - u^{k+1}\|^2 + 4\gamma^2[\|A(u^k - u^{k+1})\|^2 + \|Au^{k+1} - b\|^2]\right]\right\} \\
\geq\ & \mathbb{E}_{i(k)}\left\{\frac{\epsilon}{N}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\left[\left(1 + 2\gamma^2\lambda_{\max}(A^\top A)\right)N^2\|u^k - u^{k+1}\|^2 + 2\gamma^2\|Au^k - b\|^2\right]\right\} \\
=\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\left[\left(1 + 2\gamma^2\lambda_{\max}(A^\top A)\right)N^2\mathbb{E}_{i(k)}\|u^k - u^{k+1}\|^2 + 2\gamma^2\|Au^k - b\|^2\right]. \qquad \text{(A.21)}
\end{aligned}$$

By Jensen's inequality, (A.21) follows that

$$\begin{aligned}
\Lambda(w^k, w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w) \geq\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] \\
& + d_4\left[\left(1 + 2\gamma^2\lambda_{\max}(A^\top A)\right)N^2\|u^k - \mathbb{E}_{i(k)}u^{k+1}\|^2 + 2\gamma^2\|Au^k - b\|^2\right] \qquad \text{(A.22)}
\end{aligned}$$

Since $\mathbb{E}_{i(k)}u^{k+1} - u^k = \frac{1}{N}[T_u(w^k) - u^k]$ in (A.4), (A.22) yields that

$$\begin{aligned}
\Lambda(w^k, w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w) \geq\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] \\
& + d_4\left[\left(1 + 2\gamma^2\lambda_{\max}(A^\top A)\right)\|u^k - T_u(w^k)\|^2 + 2\gamma^2\|Au^k - b\|^2\right]. \qquad \text{(A.23)}
\end{aligned}$$

Since $\lambda_{\max}(A^\top A)\|u^k - T_u(w^k)\|^2 \geq \|A[u^k - T_u(w^k)]\|^2$ and $T_p(w^k) - p^k = \gamma[AT_u(w^k) - b]$, (A.23) follows that

$$\begin{aligned}
\Lambda(w^k, w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w) \geq\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] \\
& + d_4\left[\|u^k - T_u(w^k)\|^2 + 2\gamma^2\|A[u^k - T_u(w^k)]\|^2 + 2\gamma^2\|Au^k - b\|^2\right] \\
\geq\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\left[\|u^k - T_u(w^k)\|^2 + \gamma^2\|AT_u(w^k) - b\|^2\right] \\
\geq\ & \frac{\epsilon}{N}\mathbb{E}_{i(k)}\left[L(u^{k+1}, p) - L(u, q^k)\right] + d_4\|w^k - T(w^k)\|^2.
\end{aligned}$$

Then we have the result of Lemma 2. □

3. Proof of Theorem 1 (Almost surely convergence)

Proof.

(i) Take $w = w^*$ in Lemma 2, we have

$$\Lambda(w^k, w^*) \geq \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w^*) + \frac{\epsilon}{N}\mathbb{E}_{i(k)}\big[L(u^{k+1}, p^*) - L(u^*, q^k)\big] + d_4\|w^k - T(w^k)\|^2. \tag{A.24}$$

Observe that $L(u^{k+1}, p^*) - L(u^*, q^k) \geq 0$. From statement (i) of Lemma 1, we have that $\Lambda(w^k, w^*)$ is nonnegative. By the Robbins-Siegmund Lemma (Robbins & Siegmund, 1971), we obtain that $\lim_{k\to+\infty} \Lambda(w^k, w^*)$ almost surely exists, $\sum_{k=0}^{+\infty} \|w^k - T(w^k)\|^2 < +\infty$ a.s..

(ii) Since $\lim_{k\to+\infty} \Lambda(w^k, w^*)$ almost surely exists, thus $\Lambda(w^k, w^*)$ is almost surely bounded. Thanks statement (i) of Lemma 1, it implies the sequences $\{w^k\}$ is almost surely bounded.

(iii) From statement (i) we have that

$$\lim_{k\to\infty} \|w^k - T(w^k)\| = 0 \quad \text{a.s..}$$

By variational inequality system (A.6), we have that any cluster point of a realization sequence generated by RPDC almost surely is a saddle point of Lagrangian for (P).

□

4. Proof of Theorem 2 (Expected primal suboptimality and expected feasibility)

Proof.

(i) Let $h(w, w') = \Lambda(w, w') + \frac{d_3}{d_1}\Lambda(w, w^*)$. By statement (i) and (iii) in Lemma 1, we have $h(w, w') \geq 0$. From Lemma 2, we obtain that

$$\mathbb{E}_{i(k)}\frac{\epsilon}{N}\big[L(u^{k+1}, p) - L(u, q^k)\big] \leq \Lambda(w^k, w) - \mathbb{E}_{i(k)}\Lambda(w^{k+1}, w)$$

Taking expectation with respect to $\mathcal{F}_t$, $t > k$ for above inequality, we obtain that

$$\frac{\epsilon}{N}\mathbb{E}_{\mathcal{F}_t}\big[L(u^{k+1}, p) - L(u, q^k)\big] \leq \mathbb{E}_{\mathcal{F}_t}[\Lambda(w^k, w) - \Lambda(w^{k+1}, w)]. \tag{A.25}$$

Take $w = w^*$ in (A.25), we obtain

$$0 \leq \mathbb{E}_{\mathcal{F}_t}[\Lambda(w^k, w^*) - \Lambda(w^{k+1}, w^*)]. \tag{A.26}$$

By the combination of (A.25) and (A.26), it follows

$$\frac{\epsilon}{N}\mathbb{E}_{\mathcal{F}_t}\big[L(u^{k+1}, p) - L(u, q^k)\big] \leq \mathbb{E}_{\mathcal{F}_t}[h(w^k, w) - h(w^{k+1}, w)] \tag{A.27}$$

From the definition of $\bar{u}_t$ and $\bar{p}_t$, we have $\bar{u}_t \in \mathbf{U}$ and $\bar{p}_t \in \mathbf{R}^m$. From the convexity of set $\mathbf{U}$, $\mathbf{R}^m$ and the function $L(u', p) - L(u, p')$ is convex in u' and linear in p', for all $u \in \mathbf{U}$ and $p \in \mathbf{R}^m$, we have that

$$\mathbb{E}_{\mathcal{F}_t}\big[L(\bar{u}_t, p) - L(u, \bar{p}_t)\big] \leq \mathbb{E}_{\mathcal{F}_t}\frac{1}{t+1}\sum_{k=0}^{t}\big[L(u^{k+1}, p) - L(u, q^k)\big] \leq \frac{Nh(w^0, w)}{\epsilon(t+1)}. \tag{A.28}$$

(ii) If $\mathbb{E}_{\mathcal{F}_t}\|A\bar{u}_t - b\| = 0$, statement (ii) is obviously. Otherwise, $\mathbb{E}_{\mathcal{F}_t}\|A\bar{u}_t - b\| \neq 0$ i.e., there is set $\mathbb{W}$ such that $\mathbb{P}\{\omega \in \mathbb{W} | \|A\bar{u}_t - b\| \neq 0\} > 0$. Let $\hat{p}$ be a random vector:

$$\hat{p}(\omega) = \begin{cases} 0 & \omega \notin \mathbb{W} \\ \frac{M\left(A\bar{u}_t - b\right)}{\|A\bar{u}_t - b\|} & \omega \in \mathbb{W}. \end{cases} \tag{A.29}$$

Noted that for $\omega \notin \mathbb{W}$, we have $\hat{p}(\omega) = 0$ and $\|A\bar{u}_t - b\| = 0$. Thus

$$\langle \hat{p}(\omega), A\bar{u}_t - b\rangle = M\|A\bar{u}_t - b\| = 0. \tag{A.30}$$

Otherwise, for $\omega \in \mathbb{W}$, we have that

$$\langle \hat{p}(\omega), A\bar{u}_t - b\rangle = M\|A\bar{u}_t - b\|. \tag{A.31}$$

Together (A.30) and (A.31), we have

$$\langle \hat{p}, A\bar{u}_t - b\rangle = M\|A\bar{u}_t - b\| \tag{A.32}$$

Moreover, since $Au^* = b$, we have

$$L(\bar{u}_t, \hat{p}) - L(u^*, \bar{p}_t) = F(\bar{u}_t) + \langle \hat{p}, A\bar{u}_t - b\rangle - F(u^*) = F(\bar{u}_t) - F(u^*) + M\|A\bar{u}_t - b\|. \tag{A.33}$$

Moreover, by taking $u = \bar{u}_t$ in the right hand side of saddle point inequality, we have

$$F(\bar{u}_t) - F(u^*) \geq -\langle p^*, A\bar{u}_t - b\rangle \geq -\|p^*\|\|A\bar{u}_t - b\|. \tag{A.34}$$

Combine (A.33) and (A.34), we have that

$$\|A\bar{u}_t - b\| \leq \frac{L(\bar{u}_t, \hat{p}) - L(u^*, \bar{p}_t)}{(M - \|p^*\|)}.$$

Take expectation on both side of above inequality, we have that

$$\begin{aligned} \mathbb{E}_{\mathcal{F}_t}\|A\bar{u}_t - b\| \leq \frac{\mathbb{E}_{\mathcal{F}_t}[L(\bar{u}_t, \hat{p}) - L(u^*, \bar{p}_t)]}{(M - \|p^*\|)} \quad &\leq \quad \mathbb{E}_{\mathcal{F}_t}\frac{Nh(w^0, (u^*, \hat{p}))}{(M - \|p^*\|)\,\epsilon(t+1)} \qquad \text{(by (i))} \\ &\leq \quad \mathbb{E}_{\mathcal{F}_t}\frac{Nd_5}{(M - \|p^*\|)\,\epsilon(t+1)} \end{aligned} \tag{A.35}$$

where $d_5 = \sup\limits_{\|p\|<M} h(w^0, (u^*, p))$.

(iii) Again from (A.33), (A.34) and statement (ii), statement (iii) is coming.

□

5. Proof of Lemma 3

Proof.

(i) This statement directly follows from the definition of $\phi(w, w^*)$ and statement (i) in Lemma 1.

(ii) This statement directly follows from the definition of $\phi(w, w^*)$ and statement (ii) in Lemma 1.

(iii) By the definition of $\phi(w, w^*)$, we have that.

$$\begin{aligned} & \phi(w^k, w^*) - \mathbb{E}_{i(k)}\phi(w^{k+1}, w^*) \\ = \ & \Lambda(w^k, w^*) - \mathbb{E}_{i(k)}\left\{\Lambda(w^{k+1}, w^*) + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)] - \frac{\epsilon}{N}[L(u^{k+1}, p^*) - L(u^*, p^*)]\right\} \\ \geq \ & \Lambda(w^k, w^*) - \mathbb{E}_{i(k)}\left\{\Lambda(w^{k+1}, w^*) + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)] - \frac{\epsilon}{N}[L(u^{k+1}, p^*) - L(u^*, q^k)]\right\} \\ & \qquad \text{(by the definition of saddle point.)} \\ \geq \ & d_4[\|w^k - T(w^k)\|^2 + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)]. \qquad \text{(by Lemma 2)} \end{aligned}$$

□

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

Proof. Considering the reference point $T(w^k)$ associated with given point w^k, we have that

$$\begin{cases} 0 \in \nabla G(u^k) + \partial J(T_u(w^k)) + A^\top q^k + \frac{1}{\epsilon}\left[\nabla K(T_u(w^k)) - \nabla K(u^k)\right] + \mathcal{N}_{\mathbf{U}}(T_u(w^k)) \\ 0 = b - AT_u(w^k) + \frac{1}{\gamma}\left[T_p(w^k) - p^k\right] \end{cases} \tag{A.36}$$

Thus

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

From Assumption 1 and 2, there is $\delta > 0$ such that

$$\|v(T(w^k))\|^2 \leq \delta \|w^k - T(w^k)\|^2. \tag{A.37}$$

Since $H(w)$ is global strong metric subregular at w^* for 0, then

$$\|T(w^k) - w^*\| \leq \mathfrak{c}dist(0, H(T(w^k))) \leq \mathfrak{c}\|v(T(w^k))\| \leq \mathfrak{c}\sqrt{\delta}\|w^k - T(w^k)\|. \tag{A.38}$$

Since $\|w^k - w^*\| \leq \|T(w^k) - w^*\| + \|w^k - T(w^k)\|$, we have

$$\|w^k - w^*\| \leq (\mathfrak{c}\sqrt{\delta} + 1)\|w^k - T(w^k)\|. \tag{A.39}$$

From statement (iii) of Lemma 3, we have that

$$\begin{aligned} \phi(w^k, w^*) - \mathbb{E}_{i(k)}\phi(w^{k+1}, w^*) &\geq d_4\|w^k - T(w^k)\|^2 + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)] \\ &\geq \frac{d_4}{(\mathfrak{c}\sqrt{\delta}+1)^2}\|w^k - w^*\|^2 + \frac{\epsilon}{N}[L(u^k, p^*) - L(u^*, p^*)] \qquad \text{(by (A.39))} \\ &\geq \delta'\{d_2\|w^k - w^*\|^2 + \epsilon[L(u^k, p^*) - L(u^*, p^*)]\} \\ &\geq \delta'\phi(w^k, w^*). \qquad \text{(by (i) of Lemma 3)} \end{aligned} \tag{A.40}$$

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

$$\mathbb{E}_{i(k)}\phi(w^{k+1}, w^*) \leq \alpha\phi(w^k, w^*). \tag{A.41}$$

where $\alpha = 1 - \delta' \in (0, 1)$. Taking expectation with respect to $\mathcal{F}_{k+1}$ for above inequality, we obtain that

$$\mathbb{E}_{\mathcal{F}_{k+1}}\phi(w^{k+1}, w^*) \leq \alpha^{k+1}\phi(w^0, w^*). \tag{A.42}$$

□

7. Proof of Corollary 1 (R-linear rate of the sequence $\{\mathbb{E}_{\mathcal{F}_k} w^k\}$)

Proof. By statement (i) in Lemma 3, we have that $\phi(w, w^*) \geq d_1\|w - w^*\|^2$. By Theorem 3, we have that

$$\mathbb{E}_{\mathcal{F}_k}\phi(w^k, w^*) \leq \alpha^k\phi(w^0, w^*).$$

Then we have that

$$\mathbb{E}_{\mathcal{F}_k}\|w^k - w^*\|^2 \leq \frac{\alpha^k\phi(w^0, w^*)}{d_1}.$$

By convexity of $\|\cdot\|^2$ and Jensen's inequality, we obtain that

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

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

$$\lim_{k\to\infty} \sup \sqrt[k]{\|\mathbb{E}_{\mathcal{F}_k} w^k - w^*\|} = \sqrt{\alpha} < 1.$$

□

8. Proof of Proposition 1

Proof. By the piecewise linear of $H(w)$ and Zheng and Ng (Zheng & Ng, 2014), we have that $H(w)$ is global metric subregular at w^* for 0. Since Q is positive-definite, then problem (SVM) has unique solution u^*. Hence, to show $H(w)$ is global strongly metric subregular, we need to prove uniqueness of the Lagrangian multiplier for (SVM). Suppose their are two multipliers p and p', thus we have

$$\begin{cases} 0 \in Qu^* - \mathbf{1}_n + py + \mathcal{N}_{[0,c]^n}(u^*) \\ 0 \in Qu^* - \mathbf{1}_n + p'y + \mathcal{N}_{[0,c]^n}(u^*) \end{cases}$$

Since there exists at least one component u_i^* of optimal solution u^* satisfies $0 < u_i^* < c$, then $\xi_i = \mathcal{N}_{[0,c]}(u_i^*) = 0$. Thus, we have that

$$\begin{cases} Q_i u^* - 1 + y_i p = 0 \\ Q_i u^* - 1 + y_i p' = 0 \end{cases} \tag{A.43}$$

We conclude that $p = p'$. Therefore $H(w)$ is global strongly metric subregular. □

9. Proof of Proposition 2

Proof. By the piecewise linear of $H(w)$ and Zheng and Ng (Zheng & Ng, 2014), we have that $H(w)$ is global metric subregular at w^* for 0. Since Σ is positive-definite, then problem (MLP) has unique solution u^*. Hence, to show $H(w)$ is global strongly metric subregular, we need to prove uniqueness of the Lagrangian multiplier for (MLP). Suppose their are two pare of multipliers (p_1, p_2) and (p_1', p_2'), thus we have

$$\begin{cases} 0 \in \Sigma u^* + \lambda \partial \|u^*\|_1 + p_1 \mu + p_2 \mathbf{1}_n \\ 0 \in \Sigma u^* + \lambda \partial \|u^*\|_1 + p_1' \mu + p_2' \mathbf{1}_n \end{cases}$$

Since $u_i^* \neq 0$, $u_j^* \neq 0$, thus $\xi_i = \partial |u_i^*|$ and $\xi_j = \partial |u_j^*|$ are single valued and we have

$$\begin{cases} \Sigma_i u^* + \lambda \xi_i + \mu_i p_1 + p_2 = 0 \\ \Sigma_i u^* + \lambda \xi_i + \mu_i p_1' + p_2' = 0 \end{cases} \tag{A.44}$$

$$\begin{cases} \Sigma_j u^* + \lambda \xi_j + \mu_j p_1 + p_2 = 0 \\ \Sigma_j u^* + \lambda \xi_j + \mu_j p_1' + p_2' = 0 \end{cases} \tag{A.45}$$

It follows that

$$\begin{cases} \mu_i (p_1 - p_1') + p_2 - p_2' = 0 \\ \mu_j (p_1 - p_1') + p_2 - p_2' = 0 \end{cases} \tag{A.46}$$

Since $\mu_i \neq \mu_j$, we conclude that $p_1 = p_1'$ and $p_2 = p_2'$. Therefore $H(w)$ is global strongly metric subregular. □

References

Cohen, G. and Zhu, D. Decomposition and coordination methods in large scale optimization problems: The nondifferentiable case and the use of augmented lagrangians. *Adv. in Large Scale Systems*, 1:203–266, 1984.

Robbins, H. and Siegmund, D. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics*, pp. 233–257. Elsevier, 1971.

Zheng, X. Y. and Ng, K. F. Metric subregularity of piecewise linear multifunctions and applications to piecewise linear multiobjective optimization. *SIAM Journal on Optimization*, 24(1):154–174, 2014.