In the following appendices, we provide detailed proofs of theorems stated in the main paper. In Section A we first prove a basic inequality which is useful throughout the rest of the convergence analysis. Section B contains general analysis of the batch primal-dual algorithm that are common for proving both Theorem 1 and Theorem 3. Sections C, D, E and F give proofs for Theorem 1, Theorem 3, Theorem 2 and Theorem 4, respectively.

A. A basic lemma

Lemma 2. *Let h be a strictly convex function and $\mathcal{D}_h$ be its Bregman divergence. Suppose ψ is ν-strongly convex with respect to $\mathcal{D}_h$ and $1/\delta$-smooth (with respect to the Euclidean norm), and*

$$\hat{y} = \arg\min_{y\in C} \big\{\psi(y) + \eta\mathcal{D}_h(y,\bar{y})\big\},$$

where C is a compact convex set that lies within the relative interior of the domains of h and ψ (i.e., both h and ψ are differentiable over C). Then for any $y \in C$ and $\rho \in [0,1]$, we have

$$\psi(y) + \eta\mathcal{D}_h(y,\bar{y}) \geq \psi(\hat{y}) + \eta\mathcal{D}_h(\hat{y},\bar{y}) + \big(\eta + (1-\rho)\nu\big)\mathcal{D}_h(y,\hat{y}) + \frac{\rho\delta}{2}\left\|\nabla\psi(y) - \nabla\psi(\hat{y})\right\|^2.$$

Proof. The minimizer $\hat{y}$ satisfies the following first-order optimality condition:

$$\langle\nabla\psi(\hat{y}) + \eta\nabla\mathcal{D}_h(\hat{y},\bar{y}),\ y - \hat{y}\rangle \geq 0, \quad \forall\, y \in C.$$

Here $\nabla\mathcal{D}$ denotes partial gradient of the Bregman divergence with respect to its first argument, i.e., $\nabla\mathcal{D}(\hat{y},\bar{y}) = \nabla h(\hat{y}) - \nabla h(\bar{y})$. So the above optimality condition is the same as

$$\big\langle\nabla\psi(\hat{y}) + \eta(\nabla h(\hat{y}) - \nabla h(\bar{y})),\ y - \hat{y}\big\rangle \geq 0, \quad \forall\, y \in C. \tag{17}$$

Since ψ is ν-strongly convex with respect to $\mathcal{D}_h$ and $1/\delta$-smooth, we have

$$\begin{aligned}\psi(y) &\geq \psi(\hat{y}) + \langle\nabla\psi(\hat{y}), y - \hat{y}\rangle + \nu\mathcal{D}_h(y,\hat{y}),\\ \psi(y) &\geq \psi(\hat{y}) + \langle\nabla\psi(\hat{y}), y - \hat{y}\rangle + \frac{\delta}{2}\big\|\nabla\psi(y) - \nabla\psi(\hat{y})\big\|^2.\end{aligned}$$

For the second inequality, see, e.g., Theorem 2.1.5 in Nesterov (2004). Multiplying the two inequalities above by $(1-\rho)$ and ρ respectively and adding them together, we have

$$\psi(y) \geq \psi(\hat{y}) + \langle\nabla\psi(\hat{y}), y - \hat{y}\rangle + (1-\rho)\nu\mathcal{D}_h(y,\hat{y}) + \frac{\rho\delta}{2}\big\|\nabla\psi(y) - \nabla\psi(\hat{y})\big\|^2.$$

The Bregman divergence $\mathcal{D}_h$ satisfies the following equality:

$$\mathcal{D}_h(y,\bar{y}) = \mathcal{D}_h(y,\hat{y}) + \mathcal{D}_h(\hat{y},\bar{y}) + \big\langle\nabla h(\hat{y}) - \nabla h(\bar{y}),\ y - \hat{y}\big\rangle.$$

We multiply this equality by η and add it to the last inequality to obtain

$$\begin{aligned}\psi(y) + \eta\mathcal{D}_h(y,\bar{y}) \quad\geq\quad & \psi(\hat{y}) + \eta\mathcal{D}_h(y,\hat{y}) + \big(\eta + (1-\rho)\nu\big)\mathcal{D}_h(\hat{y},\bar{y}) + \frac{\rho\delta}{2}\big\|\nabla\psi(y) - \nabla\psi(\hat{y})\big\|^2\\ & + \big\langle\nabla\psi(\hat{y}) + \eta(\nabla h(\hat{y}) - \nabla h(\bar{y})),\ y - \hat{y}\big\rangle.\end{aligned}$$

Using the optimality condition in (17), the last term of inner product is nonnegative and thus can be dropped, which gives the desired inequality. □

B. Common Analysis of Batch Primal-Dual Algorithms

We consider the general primal-dual update rule as:

Iteration: $(\hat{x}, \hat{y}) = \mathrm{PD}_{\tau,\sigma}(\bar{x}, \bar{y}, \tilde{x}, \tilde{y})$

$$\hat{x} = \arg\min_{x\in\mathbb{R}^d} \left\{ g(x) + \tilde{y}^T A x + \frac{1}{2\tau}\|x - \bar{x}\|^2 \right\}, \tag{18}$$

$$\hat{y} = \arg\min_{y\in\mathbb{R}^n} \left\{ f^*(y) - y^T A\tilde{x} + \frac{1}{\sigma}\mathcal{D}(y, \bar{y}) \right\}. \tag{19}$$

Each iteration of Algorithm 1 is equivalent to the following specification of $\mathrm{PD}_{\tau,\sigma}$:

$$\begin{aligned} &\hat{x} = x^{(t+1)}, \qquad \bar{x} = x^{(t)}, \qquad \tilde{x} = x^{(t)} + \theta(x^{(t)} - x^{(t-1)}), \\ &\hat{y} = y^{(t+1)}, \qquad \bar{y} = y^{(t)}, \qquad \tilde{y} = y^{(t+1)}. \end{aligned} \tag{20}$$

Besides Assumption 2, we also assume that f^* is ν-strongly convex with respect to a kernel function h, i.e.,

$$f^*(y') - f^*(y) - \langle \nabla f^*(y), y' - y\rangle \geq \nu \mathcal{D}_h(y', y),$$

where $\mathcal{D}_h$ is the Bregman divergence defined as

$$\mathcal{D}_h(y', y) = h(y') - h(y) - \langle \nabla h(y), y' - y\rangle.$$

We assume that h is γ'-strongly convex and $1/\delta'$-smooth. Depending on the kernel function h, this assumption on f^* may impose additional restrictions on f. In this paper, we are mostly interested in two special cases: $h(y) = (1/2)\|y\|^2$ and $h(y) = f^*(y)$ (for the latter we always have $\nu = 1$). From now on, we will omit the subscript h and use $\mathcal{D}$ denote the Bregman divergence.

Under the above assumptions, any solution $(x^\star, y^\star)$ to the saddle-point problem (6) satisfies the optimality condition:

$$-A^T y^\star \in \partial g(x^\star), \tag{21}$$

$$Ax^\star = \nabla f^*(y^\star). \tag{22}$$

The optimality conditions for the updates described in equations (18) and (19) are

$$-A^T\tilde{y} + \frac{1}{\tau}(\bar{x} - \hat{x}) \in \partial g(\hat{x}), \tag{23}$$

$$A\tilde{x} - \frac{1}{\sigma}\big(\nabla h(\hat{y}) - \nabla h(\bar{y})\big) = \nabla f^*(\hat{y}). \tag{24}$$

Applying Lemma 2 to the dual minimization step in (19) with $\psi(y) = f^*(y) - y^T A\tilde{x}$, $\eta = 1/\sigma$, $y = y^\star$ and $\rho = 1/2$, we obtain

$$\begin{aligned} f^*(y^\star) - {y^\star}^T A\tilde{x} + \frac{1}{\sigma}\mathcal{D}(y^\star, \bar{y}) \geq\ & f^*(\hat{y}) - \hat{y}^T A\tilde{x} + \frac{1}{\sigma}\mathcal{D}(\hat{y}, \bar{y}) \\ & + \Big(\frac{1}{\sigma} + \frac{\nu}{2}\Big)\mathcal{D}(y^\star, \hat{y}) + \frac{\delta}{4}\big\|\nabla f^*(y^\star) - \nabla f^*(\hat{y})\big\|^2. \end{aligned} \tag{25}$$

Similarly, for the primal minimization step in (18), we have (setting $\rho = 0$)

$$g(x^\star) + \tilde{y}^T A x^\star + \frac{1}{2\tau}\|x^\star - \bar{x}\|^2 \geq g(\hat{x}) + \tilde{y}^T A\hat{x} + \frac{1}{2\tau}\|\hat{x} - \bar{x}\|^2 + \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big)\|x^\star - \hat{x}\|^2. \tag{26}$$

Combining the two inequalities above with the definition $\mathcal{L}(x, y) = g(x) + y^T Ax - f^*(y)$, we get

$$\begin{aligned} \mathcal{L}(\hat{x}, y^\star) - \mathcal{L}(x^\star, \hat{y}) =\ & g(\hat{x}) + {y^\star}^T A\hat{x} - f^*(y^\star) - g(x^\star) - \hat{y}^T A x^\star + f^*(\hat{y}) \\ \leq\ & \frac{1}{2\tau}\|x^\star - \bar{x}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star, \bar{y}) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big)\|x^\star - \hat{x}\|^2 - \Big(\frac{1}{\sigma} + \frac{\nu}{2}\Big)\mathcal{D}(y^\star, \hat{y}) \\ & - \frac{1}{2\tau}\|\hat{x} - \bar{x}\|^2 - \frac{1}{\sigma}\mathcal{D}(\hat{y}, \bar{y}) - \frac{\delta}{4}\big\|\nabla f^*(y^\star) - \nabla f^*(\hat{y})\big\|^2 \\ & + {y^\star}^T A\hat{x} - \hat{y}^T A x^\star + \tilde{y}^T A x^\star - \tilde{y}^T A\hat{x} - {y^\star}^T A\tilde{x} + \hat{y}^T A\tilde{x}. \end{aligned}$$

We can simplify the inner product terms as

$$y^{\star T}A\hat{x} - \hat{y}^TAx^\star + \tilde{y}^TAx^\star - \tilde{y}^TA\hat{x} - y^{\star T}A\tilde{x} + \hat{y}^TA\tilde{x} = (\hat{y}-\tilde{y})^TA(\hat{x}-x^\star) - (\hat{y}-y^\star)^TA(\hat{x}-\tilde{x}).$$

Rearranging terms on the two sides of the inequality, we have

$$\begin{aligned}\frac{1}{2\tau}\|x^\star-\bar{x}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star,\bar{y}) \quad\geq\quad & \mathcal{L}(\hat{x},y^\star) - \mathcal{L}(x^\star,\hat{y}) \\ & + \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-\hat{x}\|^2 + \Big(\frac{1}{\sigma}+\frac{\nu}{2}\Big)\mathcal{D}(y^\star,\hat{y}) \\ & + \frac{1}{2\tau}\|\hat{x}-\bar{x}\|^2 + \frac{1}{\sigma}\mathcal{D}(\hat{y},\bar{y}) + \frac{\delta}{4}\big\|\nabla f^*(y^\star)-\nabla f^*(\hat{y})\big\|^2 \\ & + (\hat{y}-y^\star)^TA(\hat{x}-\tilde{x}) - (\hat{y}-\tilde{y})^TA(\hat{x}-x^\star).\end{aligned}$$

Applying the substitutions in (20) yields

$$\begin{aligned}\frac{1}{2\tau}\|x^\star-x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star,y^{(t)}) \quad\geq\quad & \mathcal{L}(x^{(t+1)},y^\star) - \mathcal{L}(x^\star,y^{(t+1)}) \\ & + \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t+1)}\|^2 + \Big(\frac{1}{\sigma}+\frac{\nu}{2}\Big)\mathcal{D}(y^\star,y^{(t+1)}) \\ & + \frac{1}{2\tau}\|x^{(t+1)}-x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^{(t+1)},y^{(t)}) + \frac{\delta}{4}\big\|\nabla f^*(y^\star)-\nabla f^*(y^{(t+1)})\big\|^2 \\ & + (y^{(t+1)}-y^\star)^TA\big(x^{(t+1)}-(x^{(t)}+\theta(x^{(t)}-x^{(t-1)})\big). \end{aligned} \tag{27}$$

We can rearrange the inner product term in (27) as

$$\begin{aligned} & (y^{(t+1)}-y^\star)^TA\big(x^{(t+1)}-(x^{(t)}+\theta(x^{(t)}-x^{(t-1)}))\big) \\ = \quad & (y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)}) - \theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)}) - \theta(y^{(t+1)}-y^{(t)})^TA(x^{(t)}-x^{(t-1)}).\end{aligned}$$

Using the optimality conditions in (22) and (24), we can also bound $\|\nabla f^*(y^\star)-\nabla f^*(y^{(t+1)})\|^2$:

$$\begin{aligned} & \big\|\nabla f^*(y^\star)-\nabla f^*(y^{(t+1)})\big\|^2 \\ = \quad & \Big\|Ax^\star - A\big(x^{(t)}+\theta(x^{(t)}-x^{(t-1)})\big) + \frac{1}{\sigma}\big(\nabla h(y^{(t+1)})-\nabla h(y^{(t)})\big)\Big\|^2 \\ \geq \quad & \Big(1-\frac{1}{\alpha}\Big)\big\|A(x^\star-x^{(t)})\big\|^2 - (\alpha-1)\Big\|\theta A(x^{(t)}-x^{(t-1)})\big) - \frac{1}{\sigma}\big(\nabla h(y^{(t+1)})-\nabla h(y^{(t)})\big)\Big\|^2,\end{aligned}$$

where $\alpha>1$. With the definition $\mu=\sqrt{\lambda_{\min}(A^TA)}$, we also have $\|A(x^\star-x^{(t)})\|^2\geq\mu^2\|x^\star-x^{(t)}\|^2$. Combining them with the inequality (27) leads to

$$\begin{aligned} & \frac{1}{2\tau}\|x^\star-x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star,y^{(t)}) + \theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)}) \\ \geq\quad & \mathcal{L}(x^{(t+1)},y^\star) - \mathcal{L}(x^\star,y^{(t+1)}) \\ & + \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t+1)}\|^2 + \Big(\frac{1}{\sigma}+\frac{\nu}{2}\Big)\mathcal{D}(y^\star,y^{(t+1)}) + (y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)}) \\ & + \frac{1}{2\tau}\|x^{(t+1)}-x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^{(t+1)},y^{(t)}) - \theta(y^{(t+1)}-y^{(t)})^TA(x^{(t)}-x^{(t-1)}) \\ & + \Big(1-\frac{1}{\alpha}\Big)\frac{\delta\mu^2}{4}\|x^\star-x^{(t)}\|^2 - (\alpha-1)\frac{\delta}{4}\Big\|\theta A(x^{(t)}-x^{(t-1)})\big) - \frac{1}{\sigma}\big(\nabla h(y^{(t+1)})-\nabla h(y^{(t)})\big)\Big\|^2. \end{aligned} \tag{28}$$

C. Proof of Theorem 1

Let the kernel function be $h(y) = (1/2)\|y\|^2$. In this case, we have $\mathcal{D}(y', y) = (1/2)\|y' - y\|^2$ and $\nabla h(y) = y$. Moreover, $\gamma' = \delta' = 1$ and $\nu = \gamma$. Therefore, the inequality (28) becomes

$$\begin{aligned}
& \frac{1}{2}\left(\frac{1}{\tau} - \left(1 - \frac{1}{\alpha}\right)\frac{\delta\mu^2}{2}\right)\|x^\star - x^{(t)}\|^2 + \frac{1}{2\sigma}\|y^\star - y^{(t)}\|^2 + \theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)}) \\
\geq\ & \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \\
& + \frac{1}{2}\left(\frac{1}{\tau} + \lambda\right)\|x^\star - x^{(t+1)}\|^2 + \frac{1}{2}\left(\frac{1}{\sigma} + \frac{\gamma}{2}\right)\|y^\star - y^{(t+1)}\|^2 + (y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)}) \\
& + \frac{1}{2\tau}\|x^{(t+1)} - x^{(t)}\|^2 + \underline{\frac{1}{2\sigma}\|y^{(t+1)} - y^{(t)}\|^2 - \theta(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)})} \\
& - (\alpha - 1)\frac{\delta}{4}\left\|\theta A(x^{(t)} - x^{(t-1)})) - \frac{1}{\sigma}(y^{(t+1)} - y^{(t)})\right\|^2.
\end{aligned} \tag{29}$$

Next we derive another form of the underlined items above:

$$\begin{aligned}
& \frac{1}{2\sigma}\|y^{(t+1)} - y^{(t)}\|^2 - \theta(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)}) \\
=\ & \frac{\sigma}{2}\left(\frac{1}{\sigma^2}\|y^{(t+1)} - y^{(t)}\|^2 - \frac{\theta}{\sigma}(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)})\right) \\
=\ & \frac{\sigma}{2}\left(\left\|\theta A(x^{(t)} - x^{(t-1)}) - \frac{1}{\sigma}(y^{(t+1)} - y^{(t)})\right\|^2 - \theta^2\|A(x^{(t)} - x^{(t-1)})\|^2\right) \\
\geq\ & \frac{\sigma}{2}\left\|\theta A(x^{(t)} - x^{(t-1)}) - \frac{1}{\sigma}(y^{(t+1)} - y^{(t)})\right\|^2 - \frac{\sigma\theta^2 L^2}{2}\|x^{(t)} - x^{(t-1)}\|^2,
\end{aligned}$$

where in the last inequality we used $\|A\| \leq L$ and hence $\|A(x^{(t)} - x^{(t-1)})\|^2 \leq L^2\|x^{(t)} - x^{(t-1)}\|^2$. Combining with inequality (29), we have

$$\begin{aligned}
& \frac{1}{2}\left(\frac{1}{\tau} - \left(1 - \frac{1}{\alpha}\right)\frac{\delta\mu^2}{2}\right)\|x^{(t)} - x^\star\|^2 + \frac{1}{2\sigma}\|y^{(t)} - y^\star\|^2 + \theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)}) + \frac{\sigma\theta^2 L^2}{2}\|x^{(t)} - x^{(t-1)}\|^2 \\
\geq\ & \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \\
& + \frac{1}{2}\left(\frac{1}{\tau} + \lambda\right)\|x^{(t+1)} - x^\star\|^2 + \frac{1}{2}\left(\frac{1}{\sigma} + \frac{\gamma}{2}\right)\|y^{(t+1)} - y^\star\|^2 + (y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)}) + \frac{1}{2\tau}\|x^{(t+1)} - x^{(t)}\|^2 \\
& + \left(\frac{\sigma}{2} - (\alpha - 1)\frac{\delta}{4}\right)\left\|\theta A(x^{(t)} - x^{(t-1)}) - \frac{1}{\sigma}(y^{(t+1)} - y^{(t)})\right\|^2.
\end{aligned} \tag{30}$$

We can remove the last term in the above inequality as long as its coefficient is nonnegative, i.e.,

$$\frac{\sigma}{2} - (\alpha - 1)\frac{\delta}{4} \geq 0.$$

In order to maximize $1 - 1/\alpha$, we take the equality and solve for the largest value of α allowed, which results in

$$\alpha = 1 + \frac{2\sigma}{\delta}, \qquad 1 - \frac{1}{\alpha} = \frac{2\sigma}{2\sigma + \delta}.$$

Applying these values in (30) gives

$$\begin{aligned}
& \frac{1}{2}\left(\frac{1}{\tau} - \frac{\sigma\delta\mu^2}{2\sigma + \delta}\right)\|x^{(t)} - x^\star\|^2 + \frac{1}{2\sigma}\|y^{(t)} - y^\star\|^2 + \theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)}) + \frac{\sigma\theta^2 L^2}{2}\|x^{(t)} - x^{(t-1)}\|^2 \\
\geq\ & \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \\
& + \frac{1}{2}\left(\frac{1}{\tau} + \lambda\right)\|x^{(t+1)} - x^\star\|^2 + \frac{1}{2}\left(\frac{1}{\sigma} + \frac{\gamma}{2}\right)\|y^{(t+1)} - y^\star\|^2 + (y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)}) + \frac{1}{2\tau}\|x^{(t+1)} - x^{(t)}\|^2.
\end{aligned} \tag{31}$$

We use $\Delta^{(t+1)}$ to denote the last row in (31). Equivalently, we define

$$\begin{aligned}\Delta^{(t)} &= \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star - x^{(t)}\|^2 + \frac{1}{2}\Big(\frac{1}{\sigma}+\frac{\gamma}{2}\Big)\|y^\star - y^{(t)}\|^2 + (y^{(t)}-y^\star)^T A(x^{(t)}-x^{(t-1)}) + \frac{1}{2\tau}\|x^{(t)}-x^{(t-1)}\|^2 \\ &= \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star - x^{(t)}\|^2 + \frac{\gamma}{4}\|y^\star - y^{(t)}\|^2 + \frac{1}{2}\begin{bmatrix} x^{(t)}-x^{(t-1)} \\ y^\star - y^{(t)} \end{bmatrix}^T \begin{bmatrix} \frac{1}{\tau}I & -A^T \\ -A & \frac{1}{\sigma}\end{bmatrix}\begin{bmatrix} x^{(t)}-x^{(t-1)} \\ y^\star - y^{(t)} \end{bmatrix}.\end{aligned}$$

The quadratic form in the last term is nonnegative if the matrix

$$M = \begin{bmatrix} \frac{1}{\tau}I & -A^T \\ -A & \frac{1}{\sigma}\end{bmatrix}$$

is positive semidefinite, for which a sufficient condition is $\tau\sigma \le 1/L^2$. Under this condition,

$$\Delta^{(t)} \;\ge\; \frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star - x^{(t)}\|^2 + \frac{\gamma}{4}\|y^\star - y^{(t)}\|^2 \;\ge\; 0. \tag{32}$$

If we can to choose τ and σ so that

$$\frac{1}{\tau} - \frac{\sigma\delta\mu^2}{2\sigma+\delta} \le \theta\Big(\frac{1}{\tau}+\lambda\Big), \qquad \frac{1}{\sigma} \le \theta\Big(\frac{1}{\sigma}+\frac{\gamma}{2}\Big), \qquad \frac{\sigma\theta^2 L^2}{2} \le \theta\frac{1}{2\tau}, \tag{33}$$

then, according to (31), we have

$$\Delta^{(t+1)} + \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \;\le\; \theta\Delta^{(t)}.$$

Because $\Delta^{(t)} \ge 0$ and $\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^{(t)}) \ge 0$ for any $t \ge 0$, we have

$$\Delta^{(t+1)} \;\le\; \theta\Delta^{(t)},$$

which implies

$$\Delta^{(t)} \;\le\; \theta^t\Delta^{(0)}$$

and

$$\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^{(t)}) \;\le\; \theta^t\Delta^{(0)}.$$

Let θ_x and θ_y be two contraction factors determined by the first two inequalities in (33), i.e.,

$$\begin{aligned}\theta_x &= \Big(\frac{1}{\tau} - \frac{\sigma\delta\mu^2}{2\sigma+\delta}\Big)\Big/\Big(\frac{1}{\tau}+\lambda\Big) = \Big(1 - \frac{\tau\sigma\delta\mu^2}{2\sigma+\delta}\Big)\frac{1}{1+\tau\lambda}, \\ \theta_y &= \frac{1}{\sigma}\Big/\Big(\frac{1}{\sigma}+\frac{\gamma}{2}\Big) = \frac{1}{1+\sigma\gamma/2}.\end{aligned}$$

Then we can let $\theta = \max\{\theta_x, \theta_y\}$. We note that any $\theta < 1$ would satisfy the last condition in (33) provided that

$$\tau\sigma \;=\; \frac{1}{L^2},$$

which also makes the matrix M positive semidefinite and thus ensures the inequality (32).

Among all possible pairs τ, σ that satisfy $\tau\sigma = 1/L^2$, we choose

$$\tau = \frac{1}{L}\sqrt{\frac{\gamma}{\lambda+\delta\mu^2}}, \qquad \sigma = \frac{1}{L}\sqrt{\frac{\lambda+\delta\mu^2}{\gamma}}, \tag{34}$$

which give the desired results of Theorem 1.

D. Proof of Theorem 3

If we choose $h = f^*$, then

- h is γ-strongly convex and $1/\delta$-smooth, i.e., $\gamma' = \gamma$ and $\delta' = \delta$;
- f^* is 1-strongly convex with respect to h, i.e., $\nu = 1$.

For convenience, we repeat inequality (28) here:

$$\begin{aligned}
&\frac{1}{2\tau}\|x^\star - x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star, y^{(t)}) + \theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)}) \\
\geq\;& \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \\
&+ \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big)\|x^\star - x^{(t+1)}\|^2 + \Big(\frac{1}{\sigma} + \frac{\nu}{2}\Big)\mathcal{D}(y^\star, y^{(t+1)}) + (y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)}) \\
&+ \frac{1}{2\tau}\|x^{(t+1)} - x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^{(t+1)}, y^{(t)}) - \theta(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)}) \\
&+ \Big(1 - \frac{1}{\alpha}\Big)\frac{\delta\mu^2}{4}\|x^\star - x^{(t)}\|^2 - (\alpha - 1)\frac{\delta}{4}\Big\|\theta A(x^{(t)} - x^{(t-1)})) - \frac{1}{\sigma}\big(\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\big)\Big\|^2. \qquad (35)
\end{aligned}$$

We first bound the Bregman divergence $\mathcal{D}(y^{(t+1)}, y^{(t)})$ using the assumption that the kernel h is γ-strongly convex and $1/\delta$-smooth. Using similar arguments as in the proof of Lemma 2, we have for any $\rho \in [0, 1]$,

$$\begin{aligned}
D(y^{(t+1)}, y^{(t)}) &= h(y^{(t+1)}) - h(y^{(t)}) - \langle \nabla h(y^{(t)}), y^{(t+1)} - y^{(t)} \rangle \\
&\geq (1 - \rho)\frac{\gamma}{2}\|y^{(t+1)} - y^{(t)}\|^2 + \rho\frac{\delta}{2}\big\|\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\big\|^2. \qquad (36)
\end{aligned}$$

For any $\beta > 0$, we can lower bound the inner product term

$$-\theta(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)}) \geq -\frac{\beta}{2}\|y^{(t+1)} - y^{(t)}\|^2 - \frac{\theta^2 L^2}{2\beta}\|x^{(t)} - x^{(t-1)}\|^2.$$

In addition, we have

$$\Big\|\theta A(x^{(t)} - x^{(t-1)})) - \frac{1}{\sigma}\big(\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\big)\Big\|^2 \;\leq\; 2\theta^2 L^2\|x^{(t)} - x^{(t-1)}\|^2 + \frac{2}{\sigma^2}\big\|\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\big\|^2.$$

Combining these bounds with (35) and (36) with $\rho = 1/2$, we arrive at

$$\begin{aligned}
&\frac{1}{2}\Big(\frac{1}{\tau} - \Big(1 - \frac{1}{\alpha}\Big)\frac{\delta\mu^2}{2}\Big)\|x^\star - x^{(t)}\|^2 + \frac{1}{\sigma}\mathcal{D}(y^\star, y^{(t)}) + \theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)}) \\
&+ \Big(\frac{\theta^2 L^2}{2\beta} + (\alpha - 1)\frac{\delta\theta^2 L^2}{2}\Big)\|x^{(t)} - x^{(t-1)}\|^2 \\
\geq\;& \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}) \\
&+ \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big)\|x^\star - x^{(t+1)}\|^2 + \Big(\frac{1}{\sigma} + \frac{1}{2}\Big)\mathcal{D}(y^\star, y^{(t+1)}) + (y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)}) \\
&+ \Big(\frac{\gamma}{4\sigma} - \frac{\beta}{2}\Big)\|y^{(t+1)} - y^{(t)}\|^2 + \Big(\frac{\delta}{4\sigma} - \frac{(\alpha - 1)\delta}{2\sigma^2}\Big)\big\|\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\big\|^2 \\
&+ \frac{1}{2\tau}\|x^{(t+1)} - x^{(t)}\|^2. \qquad (37)
\end{aligned}$$

We choose α and β in (37) to zero out the coefficients of $\|y^{(t+1)} - y^{(t)}\|^2$ and $\|\nabla h(y^{(t+1)}) - \nabla h(y^{(t)})\|^2$:

$$\alpha = 1 + \frac{\sigma}{2}, \qquad \beta = \frac{\gamma}{2\sigma}.$$

Then the inequality (37) becomes

$$\begin{aligned}
&\frac{1}{2}\left(\frac{1}{\tau}-\frac{\sigma\delta\mu^2}{4+2\sigma}\right)\|x^\star-x^{(t)}\|^2+\frac{1}{\sigma}\mathcal{D}(y^\star,y^{(t)})+\theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})\\
&+\left(\frac{\sigma\theta^2L^2}{\gamma}+\frac{\delta\sigma\theta^2L^2}{4}\right)\|x^{(t)}-x^{(t-1)}\|^2\\
\geq\;&\mathcal{L}(x^{(t+1)},y^\star)-\mathcal{L}(x^\star,y^{(t+1)})\\
&+\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t+1)}\|^2+\left(\frac{1}{\sigma}+\frac{1}{2}\right)\mathcal{D}(y^\star,y^{(t+1)})+(y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)})\\
&+\frac{1}{2\tau}\|x^{(t+1)}-x^{(t)}\|^2.
\end{aligned}$$

The coefficient of $\|x^{(t)}-x^{(t-1)}\|^2$ can be bounded as

$$\frac{\sigma\theta^2L^2}{\gamma}+\frac{\delta\sigma\theta^2L^2}{4}=\Big(\frac{1}{\gamma}+\frac{\delta}{4}\Big)\sigma\theta^2L^2=\frac{4+\gamma\delta}{4\gamma}\sigma\theta^2L^2<\frac{2\sigma\theta^2L^2}{\gamma},$$

where in the inequality we used $\gamma\delta\leq 1$. Therefore we have

$$\begin{aligned}
&\frac{1}{2}\left(\frac{1}{\tau}-\frac{\sigma\delta\mu^2}{4+2\sigma}\right)\|x^\star-x^{(t)}\|^2+\frac{1}{\sigma}\mathcal{D}(y^\star,y^{(t)})+\theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})+\frac{2\sigma\theta^2L^2}{\gamma}\|x^{(t)}-x^{(t-1)}\|^2\\
\geq\;&\mathcal{L}(x^{(t+1)},y^\star)-\mathcal{L}(x^\star,y^{(t+1)})\\
&+\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t+1)}\|^2+\left(\frac{1}{\sigma}+\frac{1}{2}\right)\mathcal{D}(y^\star,y^{(t+1)})+(y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)})+\frac{1}{2\tau}\|x^{(t+1)}-x^{(t)}\|^2.
\end{aligned}$$

We use $\Delta^{(t+1)}$ to denote the last row of the above inequality. Equivalently, we define

$$\Delta^{(t)}=\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t)}\|^2+\left(\frac{1}{\sigma}+\frac{1}{2}\right)\mathcal{D}(y^\star,y^{(t)})+(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})+\frac{1}{2\tau}\|x^{(t)}-x^{(t-1)}\|^2.$$

Since h is γ-strongly convex, we have $\mathcal{D}(y^\star,y^{(t)})\geq\frac{\gamma}{2}\|y^\star-y^{(t)}\|^2$, and thus

$$\begin{aligned}
\Delta^{(t)}\;&\geq\;\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t)}\|^2+\frac{1}{2}\mathcal{D}(y^\star,y^{(t)})+\frac{\gamma}{2\sigma}\|y^{(t)}-y^\star\|^2+(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})+\frac{1}{2\tau}\|x^{(t)}-x^{(t-1)}\|^2\\
&=\;\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t)}\|^2+\frac{1}{2}\mathcal{D}(y^\star,y^{(t)})+\frac{1}{2}\begin{bmatrix}x^{(t)}-x^{(t-1)}\\ y^\star-y^{(t)}\end{bmatrix}^T\begin{bmatrix}\frac{1}{\tau}I & -A^T\\ -A & \frac{\gamma}{\sigma}\end{bmatrix}\begin{bmatrix}x^{(t)}-x^{(t-1)}\\ y^\star-y^{(t)}\end{bmatrix}.
\end{aligned}$$

The quadratic form in the last term is nonnegative if $\tau\sigma\leq\gamma/L^2$. Under this condition,

$$\Delta^{(t)}\;\geq\;\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)\|x^\star-x^{(t)}\|^2+\frac{1}{2}\mathcal{D}(y^\star,y^{(t)})\;\geq\;0. \tag{38}$$

If we can to choose τ and σ so that

$$\frac{1}{\tau}-\frac{\sigma\delta\mu^2}{4+2\sigma}\leq\theta\Big(\frac{1}{\tau}+\lambda\Big),\qquad \frac{1}{\sigma}\leq\theta\Big(\frac{1}{\sigma}+\frac{1}{2}\Big),\qquad \frac{2\sigma\theta^2L^2}{\gamma}\leq\theta\frac{1}{2\tau}, \tag{39}$$

then we have

$$\Delta^{(t+1)}+\mathcal{L}(x^{(t+1)},y^\star)-\mathcal{L}(x^\star,y^{(t+1)})\;\leq\;\theta\Delta^{(t)}.$$

Because $\Delta^{(t)}\geq 0$ and $\mathcal{L}(x^{(t)},y^\star)-\mathcal{L}(x^\star,y^{(t)})\geq 0$ for any $t\geq 0$, we have

$$\Delta^{(t+1)}\;\leq\;\theta\Delta^{(t)},$$

which implies

$$\Delta^{(t)}\;\leq\;\theta^t\Delta^{(0)}$$

and

$$\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^{(t)}) \;\leq\; \theta^t \Delta^{(0)}.$$

To satisfy the last condition in (39) and also ensure the inequality (38), it suffices to have

$$\tau\sigma \;\leq\; \frac{\gamma}{4L^2}.$$

We choose

$$\tau = \frac{1}{2L}\sqrt{\frac{\gamma}{\lambda+\delta\mu^2}}, \qquad \sigma = \frac{1}{2L}\sqrt{\gamma(\lambda+\delta\mu^2)}.$$

With the above choice and assuming $\gamma(\lambda+\delta\mu^2) \ll L^2$, we have

$$\theta_y = \frac{\frac{1}{\sigma}}{\frac{1}{\sigma}+\frac{1}{2}} = \frac{1}{1+\sigma/2} = \frac{1}{1+\sqrt{\gamma(\lambda+\delta\mu^2)}/(4L)} \approx 1 - \frac{\sqrt{\gamma(\lambda+\delta\mu^2)}}{4L}.$$

For the contraction factor over the primal variables, we have

$$\theta_x = \frac{\frac{1}{\tau} - \frac{\sigma\delta\mu^2}{4+2\sigma}}{\frac{1}{\tau}+\lambda} = \frac{1-\frac{\tau\sigma\delta\mu^2}{4+2\sigma}}{1+\tau\lambda} = \frac{1-\frac{\gamma\delta\mu^2}{4(4+2\sigma)L^2}}{1+\tau\lambda} \approx 1 - \frac{\gamma\delta\mu^2}{16L^2} - \frac{\lambda}{2L}\sqrt{\frac{\gamma}{\lambda+\delta\mu^2}}.$$

This finishes the proof of Theorem 3.

E. Proof of Theorem 2

We consider the SPDC algorithm in the Euclidean case with $h(x) = (1/2)\|x\|^2$. The corresponding batch case analysis is given in Section C. For each $i = 1, \ldots, n$, let $\tilde{y}_i$ be

$$\tilde{y}_i = \arg\min_y \left\{ \phi_i^*(y) + \frac{(y - y_i^{(t)})^2}{2\sigma} - y\langle a_i, \tilde{x}^{(t)}\rangle \right\}.$$

Based on the first-order optimality condition, we have

$$\langle a_i, \tilde{x}^{(t)}\rangle - \frac{(\tilde{y}_i - y_i^{(t)})}{\sigma} \in \phi_i^{*'}(\tilde{y}_i).$$

Also, since $y_i^\star$ minimizes $\phi_i^*(y) - y\langle a_i, x^\star\rangle$, we have

$$\langle a_i, x^\star\rangle \in \phi_i^{*'}(y_i^\star).$$

By Lemma 2 with $\rho = 1/2$, we have

$$\begin{aligned} -y_i^\star\langle a_i, \tilde{x}^{(t)}\rangle + \phi_i^*(y_i^\star) + \frac{(y_i^{(t)} - y_i^\star)^2}{2\sigma} \geq \left(\frac{1}{\sigma}+\frac{\gamma}{2}\right)\frac{(\tilde{y}_i - y_i^\star)^2}{2} + \phi_i^*(\tilde{y}_i) - \tilde{y}_i\langle a_i, \tilde{x}^{(t)}\rangle \\ + \frac{(\tilde{y}_i - y_i^{(t)})^2}{2\sigma} + \frac{\delta}{4}(\phi_i^{*'}(\tilde{y}_i) - \phi_i^{*'}(y_i^\star))^2, \end{aligned}$$

and re-arranging terms, we get

$$\begin{aligned} \frac{(y_i^{(t)} - y_i^\star)^2}{2\sigma} \geq \left(\frac{1}{\sigma}+\frac{\gamma}{2}\right)\frac{(\tilde{y}_i - y_i^\star)^2}{2} + \frac{(\tilde{y}_i - y_i^{(t)})^2}{2\sigma} - (\tilde{y}_i - y_i^\star)\langle a_i, \tilde{x}^{(t)}\rangle + (\phi_i^*(\tilde{y}_i) - \phi_i^*(y_i^\star)) \\ + \frac{\delta}{4}(\phi_i^{*'}(\tilde{y}_i) - \phi_i^{*'}(y_i^\star))^2. \end{aligned} \tag{40}$$

Notice that

$$
\begin{aligned}
\mathbb{E}[y_i^{(t+1)}] &= \frac{1}{n} \cdot \tilde{y}_i + \frac{n-1}{n} \cdot y_i^{(t)}, \\
\mathbb{E}[(y_i^{(t+1)} - y_i^\star)^2] &= \frac{(\tilde{y}_i - y_i^\star)^2}{n} + \frac{(n-1)(y_i^{(t)} - y_i^\star)^2}{n}, \\
\mathbb{E}[(y_i^{(t+1)} - y_i^{(t)})^2] &= \frac{(\tilde{y}_i - y_i^{(t)})^2}{n}, \\
\mathbb{E}[\phi_i^*(y_i^{(t+1)})] &= \frac{1}{n} \cdot \phi_i^*(\tilde{y}_i) + \frac{n-1}{n} \cdot \phi_i^*(y_i^{(t)}).
\end{aligned}
$$

Plug the above relations into (40) and divide both sides by n, we have

$$
\begin{aligned}
\left(\frac{1}{2\sigma} + \frac{(n-1)\gamma}{4n}\right)(y_i^{(t)} - y_i^\star)^2 \geq{}& \left(\frac{1}{2\sigma} + \frac{\gamma}{4}\right)\mathbb{E}[(y_i^{(t+1)} - y_i^\star)^2] + \frac{1}{2\sigma}\mathbb{E}[(y_i^{(t+1)} - y_i^{(t)})^2] \\
&- \left(\mathbb{E}[(y_i^{(t+1)} - y_i^{(t)})] + \frac{1}{n}(y_i^{(t)} - y_i^\star)\right)\langle a_i, \tilde{x}^{(t)}\rangle \\
&+ \mathbb{E}[\phi_i^*(y_i^{(t+1)})] - \phi_i^*(y_i^{(t)}) + \frac{1}{n}(\phi_i^*(y_i^{(t)}) - \phi_i^*(y_i^\star)) \\
&+ \frac{\delta}{4n}\left(\langle a_i, \tilde{x}^{(t)} - x^\star\rangle - \frac{(\tilde{y}_i - y_i^{(t)})}{\sigma}\right)^2,
\end{aligned}
$$

and summing over $i = 1, \ldots, n$, we get

$$
\begin{aligned}
\left(\frac{1}{2\sigma} + \frac{(n-1)\gamma}{4n}\right)\|y^{(t)} - y^\star\|^2 \geq{}& \left(\frac{1}{2\sigma} + \frac{\gamma}{4}\right)\mathbb{E}[\|y^{(t+1)} - y^\star\|^2] + \frac{\mathbb{E}[\|y^{(t+1)} - y^{(t)}\|^2]}{2\sigma} \\
&+ \phi_k^*(y_k^{(t+1)}) - \phi_k^*(y_k^{(t)}) + \frac{1}{n}\sum_{i=1}^n(\phi_i^*(y_i^{(t)}) - \phi_i^*(y_i^\star)) \\
&- \left\langle n(u^{(t+1)} - u^{(t)}) + (u^{(t)} - u^*), \tilde{x}^{(t)}\right\rangle \\
&+ \frac{\delta}{4n}\left\|A(x^* - \tilde{x}^{(t)}) + \frac{(\tilde{y} - y^{(t)})}{\sigma}\right\|^2,
\end{aligned}
$$

where

$$
u^{(t)} = \frac{1}{n}\sum_{i=1}^n y_i^{(t)} a_i, \quad u^{(t+1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(t+1)} a_i, \quad \text{and} \quad u^\star = \frac{1}{n}\sum_{i=1}^n y_i^\star a_i.
$$

On the other hand, since $x^{(t+1)}$ minimizes the $\frac{1}{\tau} + \lambda$-strongly convex objective

$$
g(x) + \left\langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x\right\rangle + \frac{\|x - x^{(t)}\|^2}{2\tau},
$$

we can apply Lemma 2 with $\rho = 0$ to obtain

$$
\begin{aligned}
& g(x^\star) + \langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^\star\rangle + \frac{\|x^{(t)} - x^\star\|^2}{2\tau} \\
& \geq g(x^{(t+1)}) + \langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^{(t+1)}\rangle + \frac{\|x^{(t+1)} - x^{(t)}\|^2}{2\tau} + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|x^{(t+1)} - x^\star\|^2,
\end{aligned}
$$

and re-arranging terms we get

$$
\begin{aligned}
\frac{\|x^{(t)} - x^\star\|^2}{2\tau} \geq{}& \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \frac{\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2]}{2\tau} + \mathbb{E}[g(x^{(t+1)}) - g(x^\star)] \\
&+ \mathbb{E}[\langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - x^\star\rangle].
\end{aligned}
$$

Also notice that

$$\begin{aligned}
&\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)})) - (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) \\
&= \frac{1}{n}\sum_{i=1}^{n}(\phi_i^*(y_i^{(t)}) - \phi_i^*(y^\star)) + (\phi_k^*(y_k^{(t+1)}) - \phi_k^*(y_k^{(t)})) + g(x^{(t+1)}) - g(x^\star) \\
&+ \langle u^\star, x^{(t+1)}\rangle - \langle u^{(t)}, x^\star\rangle + n\langle u^{(t)} - u^{(t+1)}, x^\star\rangle.
\end{aligned}$$

Combining everything together, we have

$$\begin{aligned}
&\frac{\|x^{(t)} - x^\star\|^2}{2\tau} + \left(\frac{1}{2\sigma} + \frac{(n-1)\gamma}{4n}\right)\|y^{(t)} - y^\star\|^2 + (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) \\
&\geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{2\sigma} + \frac{\gamma}{4}\right)\mathbb{E}[\|y^{(t+1)} - y^\star\|^2] + \frac{\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2]}{2\tau} + \frac{\mathbb{E}[\|y^{(t+1)} - y^{(t)}\|^2]}{2\sigma} \\
&+ \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))] \\
&+ \mathbb{E}[\langle u^{(t)} - u^\star + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - \bar{x}^t\rangle] + \frac{\delta}{4n}\left\|A(x^\star - \tilde{x}^{(t)}) + \frac{(\tilde{y} - y^{(t)})}{\sigma}\right\|^2.
\end{aligned}$$

Next we notice that

$$\begin{aligned}
\frac{\delta}{4n}\left\|A(x^\star - \tilde{x}^{(t)}) + \frac{n(\mathbb{E}[y^{(t+1)}] - y^{(t)})}{\sigma}\right\|^2 =& \frac{\delta}{4n}\left\|A(x^\star - x^{(t)}) - \theta A(x^{(t)} - x^{(t-1)}) + \frac{(\tilde{y} - y^{(t)})}{\sigma}\right\|^2 \\
\geq& \left(1 - \frac{1}{\alpha}\right)\frac{\delta}{4n}\left\|A(x^\star - x^{(t)})\right\|^2 \\
&- (\alpha - 1)\frac{\delta}{4n}\left\|\theta A(x^{(t)} - x^{(t-1)}) + \frac{(\tilde{y} - y^{(t)})}{\sigma}\right\|^2,
\end{aligned}$$

for some $\alpha > 1$ and

$$\left\|A(x^\star - x^{(t)})\right\|^2 \geq \mu^2\|x^\star - x^{(t)}\|^2,$$

and

$$\begin{aligned}
\left\|\theta A(x^{(t)} - x^{(t-1)}) + \frac{(\tilde{y} - y^{(t)})}{\sigma}\right\|^2 \geq& -2\theta^2\|A(x^{(t)} - x^{(t-1)})\|^2 - \frac{2}{\sigma^2}\|\tilde{y} - y^{(t)}\|^2 \\
\geq& -2\theta^2 L^2\|x^{(t)} - x^{(t-1)}\|^2 - \frac{2n}{\sigma^2}\mathbb{E}[\|y^{(t+1)} - y^{(t)}\|^2].
\end{aligned}$$

We follow the same reasoning as in the standard SPDC analysis,

$$\begin{aligned}
&\langle u^{(t)} - u^\star + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - \tilde{x}^{(t)}\rangle = \frac{(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})}{n} - \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n} \\
&+ \frac{(n-1)}{n}(y^{(t+1)} - y^{(t)})^T A(x^{(t+1)} - x^{(t)}) - \theta(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)}),
\end{aligned}$$

and using Cauchy-Schwartz inequality, we have

$$\begin{aligned}
|(y^{(t+1)} - y^{(t)})^T A(x^{(t)} - x^{(t-1)})| \leq& \frac{\|(y^{(t+1)} - y^{(t)})^T A\|^2}{1/(2\tau)} + \frac{\|x^{(t)} - x^{(t-1)}\|^2}{8\tau} \\
\leq& \frac{\|y^{(t+1)} - y^{(t)}\|^2}{1/(2\tau R^2)},
\end{aligned}$$

and

$$\begin{aligned}
|(y^{(t+1)} - y^{(t)})^T A(x^{(t+1)} - x^{(t)})| \leq& \frac{\|(y^{(t+1)} - y^{(t)})^T A\|^2}{1/(2\tau)} + \frac{\|x^{(t+1)} - x^{(t)}\|^2}{8\tau} \\
\leq& \frac{\|y^{(t+1)} - y^{(t)}\|^2}{1/(2\tau R^2)}.
\end{aligned}$$

Thus we get

$$\langle u^{(t)} - u^\star + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - \tilde{x}^{(t)} \rangle \geq \frac{(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})}{n} - \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n}$$
$$- \frac{\|y^{(t+1)} - y^{(t)}\|^2}{1/(4\tau R^2)} - \frac{\|x^{(t+1)} - x^{(t)}\|^2}{8\tau} - \frac{\theta\|x^{(t)} - x^{(t-1)}\|^2}{8\tau}.$$

Putting everything together, we have

$$\left(\frac{1}{2\tau} - \frac{(1-1/\alpha)\delta\mu^2}{4n}\right)\|x^{(t)} - x^\star\|^2 + \left(\frac{1}{2\sigma} + \frac{(n-1)\gamma}{4n}\right)\|y^{(t)} - y^\star\|^2 + \theta(\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^\star))$$
$$+ (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) + \theta\left(\frac{1}{8\tau} + \frac{(\alpha-1)\theta\delta L^2}{2n}\right)\|x^{(t)} - x^{t-1}\|^2 + \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n}$$
$$\geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{2\sigma} + \frac{\gamma}{4}\right)\mathbb{E}[\|y^{(t+1)} - y^\star\|^2] + \frac{\mathbb{E}[(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})]}{n}$$
$$+ \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))]$$
$$+ \left(\frac{1}{2\tau} - \frac{1}{8\tau}\right)\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2]$$
$$+ \left(\frac{1}{2\sigma} - 4R^2\tau - \frac{(\alpha-1)\delta}{2\sigma^2}\right)\mathbb{E}[\|y^{(t+1)} - y^{(t)}\|^2].$$

If we choose the parameters as

$$\alpha = \frac{\sigma}{4\delta} + 1, \quad \sigma\tau = \frac{1}{16R^2},$$

then we know

$$\frac{1}{2\sigma} - 4R^2\tau - \frac{(\alpha-1)\delta}{2\sigma^2} = \frac{1}{2\sigma} - \frac{1}{4\sigma} - \frac{1}{8\sigma} > 0,$$

and

$$\frac{(\alpha-1)\theta\delta L^2}{2n} \leq \frac{\sigma L^2}{8n^2} \leq \frac{\sigma R^2}{8} \leq \frac{1}{256\tau},$$

thus

$$\frac{1}{8\tau} + \frac{(\alpha-1)\theta\delta L^2}{2n} \leq \frac{3}{8\tau}.$$

In addition, we have

$$1 - \frac{1}{\alpha} = \frac{\sigma}{\sigma + 4\delta}.$$

Finally we obtain

$$\left(\frac{1}{2\tau} - \frac{\sigma\delta\mu^2}{4n(\sigma+4\delta)}\right)\|x^{(t)} - x^\star\|^2 + \left(\frac{1}{2\sigma} + \frac{(n-1)\gamma}{4n}\right)\|y^{(t)} - y^\star\|^2 + \theta(\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^\star))$$
$$+ (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) + \theta \cdot \frac{3}{8\tau}\|x^{(t)} - x^{(t-1)}\|^2 + \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n}$$
$$\geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{2\sigma} + \frac{\gamma}{4}\right)\mathbb{E}[\|y^{(t+1)} - y^\star\|^2] + \frac{\mathbb{E}[(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})]}{n}$$
$$+ \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))] + \frac{3}{8\tau}\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2].$$

Now we can define θ_x and θ_y as the ratios between the coefficients in the x-distance and y-distance terms, and let $\theta = \max\{\theta_x, \theta_y\}$ as before. Choosing the step-size parameters as

$$\tau = \frac{1}{4R}\sqrt{\frac{\gamma}{n\lambda + \delta\mu^2}}, \quad \sigma = \frac{1}{4R}\sqrt{\frac{n\lambda + \delta\mu^2}{\gamma}}$$

gives the desired result.

F. Proof of Theorem 4

In this setting, for i-th coordinate of the dual variables y we choose $h = \phi_i^*$, let

$$\mathcal{D}_i(y_i, y_i') = \phi_i^*(y_i) - \phi_i^*(y_i') + \langle (\phi_i^*)'(y_i'), y_i - y_i' \rangle,$$

and define

$$\mathcal{D}(y, y') = \sum_{i=1}^{n} \mathcal{D}_i(y_i, y_i').$$

For $i = 1, \ldots, n$, let $\tilde{y}_i$ be

$$\tilde{y}_i = \arg\min_y \left\{ \phi_i^*(y) + \frac{\mathcal{D}_i(y, y_i^{(t)})}{\sigma} - y\langle a_i, \tilde{x}^{(t)} \rangle \right\}.$$

Based on the first-order optimality condition, we have

$$\langle a_i, \tilde{x}^{(t)} \rangle - \frac{(\phi_i^*)'(\tilde{y}_i) - (\phi_i^*)'(y_i^{(t)})}{\sigma} \in (\phi_i^*)'(\tilde{y}_i).$$

Also since $y_i^\star$ minimizes $\phi_i^*(y) - y\langle a_i, x^* \rangle$, we have

$$\langle a_i, x^* \rangle \in (\phi_i^*)'(y_i^\star).$$

Using Lemma 2 with $\rho = 1/2$, we obtain

$$\begin{aligned} -y_i^\star \langle a_i, \tilde{x}^{(t)} \rangle + \phi_i^*(y_i^\star) + \frac{\mathcal{D}_i(y_i^\star, y_i^{(t)})}{\sigma} \geq & \left(\frac{1}{\sigma} + \frac{1}{2}\right) \mathcal{D}_i(y_i^\star, \tilde{y}_i) + \phi_i^*(\tilde{y}_i) - \tilde{y}_i \langle a_i, \tilde{x}^{(t)} \rangle \\ & + \frac{\mathcal{D}_i(\tilde{y}_i, y_i^{(t)})}{\sigma} + \frac{\delta}{4}((\phi_i^*)'(\tilde{y}_i) - (\phi_i^*)'(y_i^\star))^2, \end{aligned}$$

and rearranging terms, we get

$$\begin{aligned} \frac{\mathcal{D}_i(y_i^\star, y_i^{(t)})}{\sigma} \geq & \left(\frac{1}{\sigma} + \frac{1}{2}\right) \mathcal{D}_i(y_i^\star, \tilde{y}_i) + \frac{\mathcal{D}_i(\tilde{y}_i, y_i^{(t)})}{\sigma} - (\tilde{y}_i - y_i^\star)\langle a_i, \tilde{x}^{(t)} \rangle + (\phi_i^*(\tilde{y}_i) - \phi_i^*(y_i^\star)) \\ & + \frac{\delta}{4}((\phi_i^*)'(\tilde{y}_i) - (\phi_i^*)'(y_i^\star))^2. \end{aligned} \tag{41}$$

With i.i.d. random sampling at each iteration, we have the following relations:

$$\begin{aligned} \mathbb{E}[y_i^{(t+1)}] &= \frac{1}{n} \cdot \tilde{y}_i + \frac{n-1}{n} \cdot y_i^{(t)}, \\ \mathbb{E}[\mathcal{D}_i(y_i^{(t+1)}, y_i^\star)] &= \frac{\mathcal{D}_i(\tilde{y}_i, y_i^\star)}{n} + \frac{(n-1)\mathcal{D}_i(y_i^{(t)}, y_i^\star)}{n}, \\ \mathbb{E}[\mathcal{D}_i(y_i^{(t+1)}, y_i^{(t)})] &= \frac{\mathcal{D}_i(\tilde{y}_i, y_i^{(t)})}{n}, \\ \mathbb{E}[\phi_i^*(y_i^{(t+1)})] &= \frac{1}{n} \cdot \phi_i^*(\tilde{y}_i) + \frac{n-1}{n} \cdot \phi_i^*(y_i^{(t)}). \end{aligned}$$

Plugging the above relations into (41) and dividing both sides by n, we have

$$\begin{aligned} \left(\frac{1}{\sigma} + \frac{(n-1)}{2n}\right) \mathcal{D}_i(y_i^{(t)}, y_i^\star) \geq & \left(\frac{1}{\sigma} + \frac{1}{2}\right) \mathcal{D}_i(y_i^{(t+1)}, y_i^\star) + \frac{1}{\sigma}\mathbb{E}[\mathcal{D}_i(y_i^{(t+1)}, y_i^{(t)})] \\ & - \left(\mathbb{E}[(y_i^{(t+1)} - y_i^{(t)})] + \frac{1}{n}(y_i^{(t)} - y_i^\star)\right) \langle a_i, \tilde{x}^{(t)} \rangle \\ & + \mathbb{E}[\phi_i^*(y_i^{(t+1)})] - \phi_i^*(y_i^{(t)}) + \frac{1}{n}(\phi_i^*(y_i^{(t)}) - \phi_i^*(y_i^\star)) \\ & + \frac{\delta}{4n}\left(\langle a_i, \tilde{x}^{(t)} - x^\star \rangle - \frac{((\phi_i^*)'(\tilde{y}_i) - (\phi_i^*)'(y_i^{(t)}))}{\sigma}\right)^2, \end{aligned}$$

and summing over $i = 1, \ldots, n$, we get

$$\begin{aligned}\left(\frac{1}{\sigma} + \frac{(n-1)}{2n}\right)\mathcal{D}(y^{(t)}, y^\star) \geq \left(\frac{1}{\sigma} + \frac{1}{2}\right)\mathbb{E}[\mathcal{D}(y^{(t+1)}, y^\star)] + \frac{\mathbb{E}[\mathcal{D}(y^{(t+1)}, y^{(t)})]}{\sigma} \\ + \phi_k^*(y_k^{(t+1)}) - \phi_k^*(y_k^{(t)}) + \frac{1}{n}\sum_{i=1}^n(\phi_i^*(y_i^{(t)}) - \phi_i^*(y_i^\star)) \\ - \left\langle n(u^{(t+1)} - u^{(t)}) + (u^{(t)} - u^*), \tilde{x}^{(t)}\right\rangle \\ + \frac{\delta}{4n}\left\|A(x^\star - \tilde{x}^{(t)}) + \frac{(\phi^{*'}(\tilde{y}) - \phi^{*'}(y^{(t)}))}{\sigma}\right\|^2,\end{aligned}$$

where $\phi^{*'}(y^{(t)})$ is a n-dimensional vector such that the i-th coordinate is

$$[\phi^{*'}(y^{(t)})]_i = (\phi_i^*)'(y_i^{(t)}),$$

and

$$u^{(t)} = \frac{1}{n}\sum_{i=1}^n y_i^{(t)} a_i, \quad u^{(t+1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(t+1)} a_i, \quad \text{and} \quad u^\star = \frac{1}{n}\sum_{i=1}^n y_i^\star a_i.$$

On the other hand, since $x^{(t+1)}$ minimizes a $\frac{1}{\tau} + \lambda$-strongly convex objective

$$g(x) + \left\langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x\right\rangle + \frac{\|x - x^{(t)}\|^2}{2\tau},$$

we can apply Lemma 2 with $\rho = 0$ to obtain

$$\begin{aligned}& g(x^\star) + \langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^\star\rangle + \frac{\|x^{(t)} - x^\star\|^2}{2\tau} \\ & \geq g(x^{(t+1)}) + \langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^{(t+1)}\rangle + \frac{\|x^{(t+1)} - x^{(t)}\|^2}{2\tau} + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|x^{(t+1)} - x^\star\|^2,\end{aligned}$$

and rearranging terms, we get

$$\begin{aligned}\frac{\|x^{(t)} - x^\star\|^2}{2\tau} \geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \frac{\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2]}{2\tau} + \mathbb{E}[g(x^{(t+1)}) - g(x^\star)] \\ + \mathbb{E}[\langle u^{(t)} + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - x^\star\rangle].\end{aligned}$$

Notice that

$$\begin{aligned}& \mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)})) - (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) \\ & = \frac{1}{n}\sum_{i=1}^n(\phi_i^*(y_i^{(t)}) - \phi_i^*(y^\star)) + (\phi_k^*(y_k^{(t+1)}) - \phi_k^*(y_k^{(t)})) + g(x^{(t+1)}) - g(x^\star) \\ & + \langle u^\star, x^{(t+1)}\rangle - \langle u^{(t)}, x^\star\rangle + n\langle u^{(t)} - u^{(t+1)}, x^\star\rangle,\end{aligned}$$

so

$$\begin{aligned}& \frac{\|x^{(t)} - x^\star\|^2}{2\tau} + \left(\frac{1}{\sigma} + \frac{(n-1)}{2n}\right)\mathcal{D}(y^{(t)}, y^\star) + (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) \\ & \geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{\sigma} + \frac{1}{2}\right)\mathbb{E}[\mathcal{D}(y^{(t+1)}, y^\star)] + \frac{\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2]}{2\tau} + \frac{\mathbb{E}[\mathcal{D}(y^{(t+1)}, y^{(t)})]}{\sigma} \\ & + \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))] \\ & + \mathbb{E}[\langle u^{(t)} - u^\star + n(u^{(t+1)} - u^{(t)}), x^{(t+1)} - \bar{x}^t\rangle] + \frac{\delta}{4n}\left\|A(x^\star - \tilde{x}^{(t)}) + \frac{(\phi^{*'}(\tilde{y}) - \phi^{*'}(y^{(t)}))}{\sigma}\right\|^2.\end{aligned}$$

Next, we have

$$\begin{aligned}\frac{\delta}{4n}\left\|A(x^\star-\tilde{x}^{(t)})+\frac{(\phi^{*'}(\tilde{y})-\phi^{*'}(y^{(t)}))}{\sigma}\right\|^2 =& \frac{\delta}{4n}\left\|A(x^\star-x^{(t)})-\theta A(x^{(t)}-x^{(t-1)})+\frac{(\phi^{*'}(\tilde{y})-\phi^{*'}(y^{(t)}))}{\sigma}\right\|^2\\ \geq& \left(1-\frac{1}{\alpha}\right)\frac{\delta}{4n}\left\|A(x^\star-x^{(t)})\right\|^2\\ &-(\alpha-1)\frac{\delta}{4n}\left\|\theta A(x^{(t)}-x^{(t-1)})+\frac{(\phi^{*'}(\tilde{y})-\phi^{*'}(y^{(t)}))}{\sigma}\right\|^2,\end{aligned}$$

for any $\alpha > 1$ and

$$\left\|A(x^\star-x^{(t)})\right\|^2 \geq \mu^2\|x^\star-x^{(t)}\|^2,$$

and

$$\begin{aligned}\left\|\theta A(x^{(t)}-x^{(t-1)})+\frac{(\phi^{*'}(\tilde{y})-\phi^{*'}(y^{(t)}))}{\sigma}\right\|^2 \geq& -2\theta^2\|A(x^{(t)}-x^{(t-1)})\|^2-\frac{2}{\sigma^2}\|\phi^{*'}(\tilde{y})-\phi^{*'}(y^{(t)})\|^2]\\ \geq& -2\theta^2L^2\|x^{(t)}-x^{(t-1)}\|^2-\frac{2n}{\sigma^2}\mathbb{E}[\|\phi^{*'}(y^{(t+1)})-\phi^{*'}(y^{(t)})\|^2].\end{aligned}$$

Following the same reasoning as in the standard SPDC analysis, we have

$$\begin{aligned}\langle u^{(t)}-u^\star+n(u^{(t+1)}-u^{(t)}),x^{(t+1)}-\tilde{x}^{(t)}\rangle =& \frac{(y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)})}{n}-\frac{\theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})}{n}\\ &+\frac{(n-1)}{n}(y^{(t+1)}-y^{(t)})^TA(x^{(t+1)}-x^{(t)})-\theta(y^{(t+1)}-y^{(t)})^TA(x^{(t)}-x^{(t-1)}),\end{aligned}$$

and using Cauchy-Schwartz inequality, we have

$$\begin{aligned}|(y^{(t+1)}-y^{(t)})^TA(x^{(t)}-x^{(t-1)})| \leq& \frac{\|(y^{(t+1)}-y^{(t)})^TA\|^2}{1/(2\tau)}+\frac{\|x^{(t)}-x^{(t-1)}\|^2}{8\tau}\\ \leq& \frac{\|y^{(t+1)}-y^{(t)}\|^2}{1/(2\tau R^2)},\end{aligned}$$

and

$$\begin{aligned}|(y^{(t+1)}-y^{(t)})^TA(x^{(t+1)}-x^{(t)})| \leq& \frac{\|(y^{(t+1)}-y^{(t)})^TA\|^2}{1/(2\tau)}+\frac{\|x^{(t+1)}-x^{(t)}\|^2}{8\tau}\\ \leq& \frac{\|y^{(t+1)}-y^{(t)}\|^2}{1/(2\tau R^2)}.\end{aligned}$$

Thus we get

$$\begin{aligned}\langle u^{(t)}-u^*+n(u^{(t+1)}-u^{(t)}),x^{(t+1)}-\tilde{x}^{(t)}\rangle \geq& \frac{(y^{(t+1)}-y^\star)^TA(x^{(t+1)}-x^{(t)})}{n}-\frac{\theta(y^{(t)}-y^\star)^TA(x^{(t)}-x^{(t-1)})}{n}\\ &-\frac{\|y^{(t+1)}-y^{(t)}\|^2}{1/(4\tau R^2)}-\frac{\|x^{(t+1)}-x^{(t)}\|^2}{8\tau}-\frac{\theta\|x^{(t)}-x^{(t-1)}\|^2}{8\tau}.\end{aligned}$$

Also we can lower bound the term $\mathcal{D}(y^{(t+1)},y^{(t)})$ using Lemma 2 with $\rho=1/2$:

$$\begin{aligned}\mathcal{D}(y^{(t+1)},y^{(t)}) =& \sum_{i=1}^n\left(\phi_i^*(y_i^{(t+1)})-\phi_i^*(y_i^{(t)})-\langle(\phi_i^*)'(y_i^{(t)}),y_i^{(t+1)}-y_i^{(t)}\rangle\right)\\ \geq& \sum_{i=1}^n\left(\frac{\gamma}{2}(y_i^{(t+1)}-y_i^{(t)})^2+\frac{\delta}{2}((\phi_i^*)'(y_i^{(t+1)})-(\phi_i^*)'(y_i^{(t)}))^2\right)\\ =& \frac{\gamma}{2}\|y^{(t+1)}-y^{(t)}\|^2+\frac{\delta}{2}\|\phi^{*'}(y^{(t+1)})-\phi^{*'}(y^{(t)})\|^2.\end{aligned}$$

Combining everything above together, we have

$$\begin{aligned}
&\left(\frac{1}{2\tau} - \frac{(1-1/\alpha)\delta\mu^2}{4n}\right)\|x^{(t)} - x^\star\|^2 + \left(\frac{1}{\sigma} + \frac{(n-1)}{2n}\right)\mathcal{D}(y^{(t)}, y^\star) + \theta(\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^\star)) \\
&+ (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) + \theta\left(\frac{1}{8\tau} + \frac{(\alpha-1)\theta\delta L^2}{2n}\right)\|x^{(t)} - x^{t-1}\|^2 + \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n} \\
&\geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{\sigma} + \frac{1}{2}\right)\mathbb{E}[\mathcal{D}(y^{(t+1)}, y^\star)] + \frac{\mathbb{E}[(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})]}{n} \\
&+ \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))] \\
&+ \left(\frac{1}{2\tau} - \frac{1}{8\tau}\right)\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2] + \left(\frac{\gamma}{2\sigma} - 4R^2\tau\right)\mathbb{E}[\|y^{(t+1)} - y^{(t)}\|^2] \\
&+ \left(\frac{\delta}{2\sigma} - \frac{(\alpha-1)\delta}{2\sigma^2}\right)\mathbb{E}[\|\phi^{*'}(y^{(t+1)}) - \phi^{*'}(y^{(t)})\|^2].
\end{aligned}$$

If we choose the parameters as

$$\alpha = \frac{\sigma}{4} + 1, \quad \sigma\tau = \frac{\gamma}{16R^2},$$

then we know

$$\frac{\gamma}{2\sigma} - 4R^2\tau = \frac{\gamma}{2\sigma} - \frac{\gamma}{4\sigma} > 0,$$

and

$$\frac{\delta}{2\sigma} - \frac{(\alpha-1)\delta}{2\sigma^2} = \frac{\delta}{2\sigma} - \frac{\delta}{8\sigma} > 0$$

and

$$\frac{(\alpha-1)\theta\delta L^2}{2n} \leq \frac{\sigma\delta L^2}{8n^2} \leq \frac{\delta\sigma R^2}{8} \leq \frac{\delta\gamma}{256\tau} \leq \frac{1}{256\tau},$$

thus

$$\frac{1}{8\tau} + \frac{(\alpha-1)\theta\delta L^2}{2n} \leq \frac{3}{8\tau}.$$

In addition, we have

$$1 - \frac{1}{\alpha} = \frac{\sigma}{\sigma+4}.$$

Finally we obtain

$$\begin{aligned}
&\left(\frac{1}{2\tau} - \frac{\sigma\delta\mu^2}{4n(\sigma+4)}\right)\|x^{(t)} - x^\star\|^2 + \left(\frac{1}{\sigma} + \frac{(n-1)}{2n}\right)\mathcal{D}(y^{(t)}, y^\star) + \theta(\mathcal{L}(x^{(t)}, y^\star) - \mathcal{L}(x^\star, y^\star)) \\
&+ (n-1)(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t)})) + \theta \cdot \frac{3}{8\tau}\|x^{(t)} - x^{(t-1)}\|^2 + \frac{\theta(y^{(t)} - y^\star)^T A(x^{(t)} - x^{(t-1)})}{n} \\
&\geq \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\mathbb{E}[\|x^{(t+1)} - x^\star\|^2] + \left(\frac{1}{\sigma} + \frac{1}{2}\right)\mathbb{E}[\|y^{(t+1)} - y^\star\|^2] + \frac{\mathbb{E}[(y^{(t+1)} - y^\star)^T A(x^{(t+1)} - x^{(t)})]}{n} \\
&+ \mathbb{E}[\mathcal{L}(x^{(t+1)}, y^\star) - \mathcal{L}(x^\star, y^\star) + n(\mathcal{L}(x^\star, y^\star) - \mathcal{L}(x^\star, y^{(t+1)}))] + \frac{3}{8\tau}\mathbb{E}[\|x^{(t+1)} - x^{(t)}\|^2].
\end{aligned}$$

As before, we can define θ_x and θ_y as the ratios between the coefficients in the x-distance and y-distance terms, and let $\theta = \max\{\theta_x, \theta_y\}$. Then choosing the step-size parameters as

$$\tau = \frac{1}{4R}\sqrt{\frac{\gamma}{n\lambda + \delta\mu^2}}, \quad \sigma = \frac{1}{4R}\sqrt{\gamma(n\lambda + \delta\mu^2)}$$

gives the desired result.