Errata of Accelerated Optimization for Machine Learning

Zhouchen Lin[1], Huan Li[2], and Cong Fang[3]

[1]**Peking University**
[2]**Nankai University**
[3]**University of Pennsylvania**

ABSTRACT

The corrections for the book Accelerated Optimization for Machine Learning are listed below. Fortunately, they are all non-critical and some of them are actually for making the book better, rather than errors. Most of them are corrected in the Chinese version of the book. If you detect other errors, please send your correction information to: zclin2000@hotmail.com.

Keywords: None

1. Page 2, line 29, "

$$\min_{\mathbf{U}\in\mathbb{R}^{m\times r},\mathbf{V}\in\mathbb{R}^{n\times r}} \frac{1}{2}\sum_{(i,j)\in\Omega}\left\|\mathbf{U}_i\mathbf{V}_j^T-\mathbf{D}_{ij}\right\|_F^2+\frac{\lambda}{2}\left(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2\right).$$

" should be "

$$\min_{\mathbf{U}\in\mathbb{R}^{m\times r},\mathbf{V}\in\mathbb{R}^{n\times r}} \frac{1}{2}\sum_{(i,j)\in\Omega}\left\|\mathbf{U}_{i:}\mathbf{V}_{j:}^T-\mathbf{D}_{ij}\right\|_F^2+\frac{\lambda}{2}\left(\|\mathbf{U}\|_F^2+\|\mathbf{V}\|_F^2\right),$$

where $\mathbf{U}_{i:}$ and $\mathbf{V}_{j:}$ are the i-th row and the j-th row of $\mathbf{U}$ and $\mathbf{V}$, respectively."

2. Page 17, line 3, "$\theta_0 \le 1$" should be "$0 < \theta_0 \le 1$".
3. Page 19, line 7, "$\mathbf{w}^{k+1}$ and $\mathbf{y}^k$" should be "$\mathbf{z}_{k+1}$ and $\mathbf{y}_k$".
4. Page 19, line 9, "For Algorithm 2.1" should be "For Algorithm 2.2".
5. Page 20, line 15, "By reorganizing the terms in $\mathbf{x}^* - \frac{1}{\theta_k}\mathbf{y}^k + \frac{1-\theta_k}{\theta_k}\mathbf{x}^k$" should be "By reorganizing the terms in $\mathbf{x}^* - \frac{1}{\theta_k}\mathbf{y}_k + \frac{1-\theta_k}{\theta_k}\mathbf{x}_k$".
6. Page 33, line 15, "

$$F(\mathbf{x}_{K+1})-F(\mathbf{x}^*)\le\frac{4}{(K+2)^2}\left(L\|\mathbf{x}_0-\mathbf{x}^*\|^2+\frac{2}{1+2\delta}+\frac{18}{\delta^2}\right).$$

" should be "

$$F(\mathbf{x}_{K+1})-F(\mathbf{x}^*)\le\frac{4}{(K+2)^2}\left[L\|\mathbf{x}_0-\mathbf{x}^*\|^2+\frac{4(1+\delta)}{1+2\delta}+18\left(\frac{1+\delta}{\delta}\right)^2\right].$$

".

7. Page 34, line 6, "

$$\le L\|\mathbf{x}_0-\mathbf{x}^*\|^2+\frac{2}{1+2\delta}+\frac{18}{\delta^2}$$

" should be
"

$$\leq L\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + 2\left(1 + \frac{1}{1+2\delta}\right) + 18\left(1 + \frac{1}{\delta}\right)^2$$

".

8. Page 35, line 10, "where in $\overset{a}{\leq}$ we use $\epsilon_k \leq [1-(1-\delta)\theta]^{k+1}$, $\|\mathbf{e}_k\| \leq [1-(1-\delta)\theta]^{\frac{k+1}{2}}$" should be "where in the second inequality we use $\epsilon_k \leq [1-(1-\delta)\theta]^{k+1}$ and $\|\mathbf{e}_k\| \leq [1-(1-\delta)\theta]^{\frac{k+1}{2}}$, and in $\overset{a}{\leq}$ we use".

9. Page 35, line 15, "

$$\sum_{k=1}^{K}\left[\frac{1-(1-\delta)\theta}{1-\theta}\right]^{k/2} = \frac{q^K - 1}{q-1}$$

" should be
"

$$\sum_{k=1}^{K}\left[\frac{1-(1-\delta)\theta}{1-\theta}\right]^{k/2} = q\frac{q^K - 1}{q-1}$$

".

10. Page 36, line 12, "

$$F(\mathbf{x}_{K+1}) - F(\mathbf{x}^*) \leq \frac{4}{(K+2)^2}\left(L\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{2}{1+2\delta} + \frac{18}{\delta^2}\right).$$

" should be "

$$F(\mathbf{x}_{K+1}) - F(\mathbf{x}^*) \leq \frac{4}{(K+2)^2}\left[L\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{4(1+\delta)}{1+2\delta} + 18\left(\frac{1+\delta}{\delta}\right)^2\right].$$

".

11. Page 37, line 11, "

$$F(\mathbf{x}_{K+1}) - F(\mathbf{x}^*) \leq \frac{4}{(K+2)^2}\left(\tau\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{2}{1+2\delta} + \frac{18}{\delta^2}\right).$$

" should be "

$$F(\mathbf{x}_{K+1}) - F(\mathbf{x}^*) \leq \frac{4}{(K+2)^2}\left[\tau\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \frac{4(1+\delta)}{1+2\delta} + 18\left(\frac{1+\delta}{\delta}\right)^2\right].$$

".

12. Page 38, Algorithm 2.8 line 1, "Initialize $\mathbf{x}_{K_{-1}} = \mathbf{x}_0$." should be "Initialize $\mathbf{x}_{K_0} = \mathbf{x}_0$.".

13. Page 38, line 16, "Suppose that $F(\mathbf{x})$ is convex and L-smooth and satisfies the Hölderian error bound condition." should be "Suppose that $f(\mathbf{x})$ is convex and L-smooth, $h(\mathbf{x})$ is convex, and $F(\mathbf{x})$ satisfies the Hölderian error bound condition.".

14. Page 39, line 5, "

$$\leq \frac{2L}{(K_t+1)^2}\nu^2\left(f(\mathbf{x}_{K_{t-1}}) - f^*\right)^{2\vartheta}$$

" should be
"

$$\leq \frac{2L}{(K_t+1)^2}\nu^2\left(F(\mathbf{x}_{K_{t-1}}) - F^*\right)^{2\vartheta}$$

".

15. Page 48, line 8, “right-hand side” should be “left-hand side”.

16. Page 57, line 11, “and the primal-dual method.” should be “the primal-dual method, and the Frank-Wolfe algorithm.”.

17. Page 60, line 13, “So $\partial d_\beta(\mathbf{u})$ is a singleton. By Danskin’s theorem (Theorem A.1), we know” should be “So by Danskin’s theorem (Theorem A.1), $\partial d_\beta(\mathbf{u})$ is a singleton. Then by Proposition A.9 we know that”.

18. Page 61, line 12, “
$$f(\mathbf{x}) - f(\mathbf{x}^*) \le \epsilon \text{ and } -\|\boldsymbol{\lambda}^*\|\|\mathbf{A}\mathbf{x} - \mathbf{b}\| \le -\langle \boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle \le f(\mathbf{x}) - f(\mathbf{x}^*).$$
” should be “
$$f(\mathbf{x}) - f(\mathbf{x}^*) \le \epsilon - \frac{\beta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \text{ and}$$
$$-\|\boldsymbol{\lambda}^*\|\|\mathbf{A}\mathbf{x} - \mathbf{b}\| \le -\langle \boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x} - \mathbf{b}\rangle \le f(\mathbf{x}) - f(\mathbf{x}^*).$$
”

19. Page 67, line 20, “due to Danskin’s theorem [Bertsekas, 1999] (Theorem A.1) we know” should be “due to Danskin’s theorem [Bertsekas, 1999] (Theorem A.1) and Proposition A.9 we know”

20. Page 67, after (3.11), add “where $\boldsymbol{\lambda} = (\mathbf{u}, \mathbf{v})$. ” and accordingly, remove “$\boldsymbol{\lambda} = (\mathbf{u}, \mathbf{v})$ and ” from line 11 of page 68.

21. Page 68, line 21, “
$$-d(\boldsymbol{\lambda}_{K+1}) + d(\boldsymbol{\lambda}^*) \le \frac{L}{(K+2)^2}\|\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}^*\|^2$$
” should be “
$$-d(\boldsymbol{\lambda}_{K+1}) + d(\boldsymbol{\lambda}^*) \le \frac{2L}{(K+2)^2}\|\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}^*\|^2$$
”.

22. Page 69, line 8, “
$$f(\mathbf{x}^*) - f(\mathbf{x}) \le \|\boldsymbol{\lambda}^*\|\sqrt{\frac{2(p+m)}{L}\left[d(\boldsymbol{\lambda}^*) - d(\boldsymbol{\lambda})\right]}$$
” should be “
$$f(\mathbf{x}^*) - f(\mathbf{x}) \le \|\boldsymbol{\lambda}^*\|_\infty\sqrt{\frac{2(p+m)}{L}\left[d(\boldsymbol{\lambda}^*) - d(\boldsymbol{\lambda})\right]}$$
”.

23. Page 71, line 2, “
$$\le f(\mathbf{x}) + \sqrt{p+m}\|\boldsymbol{\lambda}^*\|\sqrt{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \sum_{i=1}^{p}(\max\{0, g_i(\mathbf{x})\})^2}$$
” should be “
$$\le f(\mathbf{x}) + \sqrt{p+m}\|\boldsymbol{\lambda}^*\|_\infty\sqrt{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \sum_{i=1}^{p}(\max\{0, g_i(\mathbf{x})\})^2}$$
”.

24. Page 74, line 15, “The following lemma can be used to prove the monotonicity of ADMM.” should be deleted.

25. Page 74, line 23, “Based on Lemma 3.9, we can give the monotonicity in the following lemma.” should be “The following lemma gives the monotonicity of ADMM.”.

26. Page 75, line 7, "$(\mathbf{x},\mathbf{y},\boldsymbol{\lambda}) = (\mathbf{x}_{k+1},\mathbf{y}_{k+1},\boldsymbol{\lambda}_{k_1})$" should be "$(\mathbf{x},\mathbf{y},\boldsymbol{\lambda}) = (\mathbf{x}_{k+1},\mathbf{y}_{k+1},\boldsymbol{\lambda}_{k+1})$".

27. Page 76, proof of lemma 3.11, "

$$\begin{aligned}
&\left\langle \hat{\nabla} f(\mathbf{x}_{k+1}), \mathbf{x}_{k+1}-\mathbf{x}^*\right\rangle + \left\langle \hat{\nabla} g(\mathbf{y}_{k+1}), \mathbf{y}_{k+1}-\mathbf{y}^*\right\rangle + \langle \boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b}\rangle \\
=& -\langle \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b}\rangle + \beta_k \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \mathbf{A}\mathbf{x}_{k+1}-\mathbf{A}\mathbf{x}^*\rangle \\
=& -\frac{1}{\beta_k}\langle \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle + \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle \\
& -\beta_k \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}^*\rangle \\
\overset{a}{=}& \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_k-\boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\|^2 \\
& +\frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_k-\mathbf{B}\mathbf{y}^*\|^2 - \frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}^*\|^2 - \frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k\|^2 \\
& +\langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle, \qquad (3.24)
\end{aligned}$$

" should be "

$$\begin{aligned}
&\left\langle \hat{\nabla} f(\mathbf{x}_{k+1}), \mathbf{x}_{k+1}-\mathbf{x}^*\right\rangle + \left\langle \hat{\nabla} g(\mathbf{y}_{k+1}), \mathbf{y}_{k+1}-\mathbf{y}^*\right\rangle + \langle \boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b}\rangle \\
=& -\langle \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*, \mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b}\rangle + \beta_k \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \mathbf{A}\mathbf{x}_{k+1}-\mathbf{A}\mathbf{x}^*\rangle \\
=& -\frac{1}{\beta_k}\langle \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle + \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle \\
& -\beta_k \langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}^*\rangle \qquad (3.24) \\
\overset{a}{=}& \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_k-\boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta_k}\|\boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\|^2 \\
& +\frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_k-\mathbf{B}\mathbf{y}^*\|^2 - \frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}^*\|^2 - \frac{\beta_k}{2}\|\mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k\|^2 \\
& +\langle \mathbf{B}\mathbf{y}_{k+1}-\mathbf{B}\mathbf{y}_k, \boldsymbol{\lambda}_{k+1}-\boldsymbol{\lambda}_k\rangle,
\end{aligned}$$

"

28. Page 81, line 15, "

$$\frac{1}{2L}\|\mathbf{B}^T(\boldsymbol{\lambda}_k-\boldsymbol{\lambda}^*)+\beta\mathbf{B}^T(\mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b})\|^2$$

" should be "

$$\frac{1}{2L}\|\mathbf{B}^T(\boldsymbol{\lambda}_k-\boldsymbol{\lambda}^*)+\beta_k\mathbf{B}^T(\mathbf{A}\mathbf{x}_{k+1}+\mathbf{B}\mathbf{y}_{k+1}-\mathbf{b})\|^2$$

".

29. Page 85, line 18, "Summing over $k = 0, 1, \cdots, K-1$, we have" should be "Summing over $k = 0, 1, \cdots, K$, we have".

30. Page 88, line 6, "

$$\|\hat{\boldsymbol{\lambda}}_{k+1}-\overline{\boldsymbol{\lambda}}_{k+1}\|_2 = \frac{\beta}{\theta_k}\|\mathbf{B}(\mathbf{y}_{k+1}-\mathbf{v}_k)\|_2.$$

" should be "

$$\|\hat{\boldsymbol{\lambda}}_{k+1}-\overline{\boldsymbol{\lambda}}_{k+1}\| = \frac{\beta}{\theta_k}\|\mathbf{B}(\mathbf{y}_{k+1}-\mathbf{v}_k)\|.$$

"

31. Page 90, line 4, "Letting $\mathbf{x} = \mathbf{x}^*$ and $\mathbf{x} = \mathbf{x}_k$ in (3.33)" should be "Letting $\mathbf{x} = \mathbf{x}^*$ and $\mathbf{y} = \mathbf{y}^*$ and $\mathbf{x} = \mathbf{x}_k$ and $\mathbf{y} = \mathbf{y}_k$ in (3.33)".

32. Page 93, lines 7, to Page 94, line 4 "
Further define

$$R_K = \sum_{k=1}^{K} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{K} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)} = \sum_{k=1}^{K} r_k,$$

then

$$R_1 = \textstyle\sum_{k=1}^{1} \frac{1}{1/\tau+(k-1)(1/\tau-1)} \prod_{j=k+1}^{1} \frac{j(1/\tau-1)}{1/\tau+(j-1)(1/\tau-1)} = \tau,$$

and

$$\begin{aligned}
&R_{K+1}\\
=&\frac{1}{1/\tau + K(1/\tau - 1)} + \sum_{k=1}^{K} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{K+1} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)}\\
=&\frac{1}{1/\tau + K(1/\tau - 1)}\\
&+ \frac{(K+1)(1/\tau - 1)}{1/\tau + K(1/\tau - 1)} \sum_{k=1}^{K} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{K} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)}\\
=&\frac{1}{1/\tau + K(1/\tau - 1)} + \frac{(K+1)(1/\tau - 1)}{1/\tau + K(1/\tau - 1)} R_K.
\end{aligned}$$

Next, we prove $R_K < 1, \forall K \geq 1$, by induction. It can be easily checked that $R_1 = \tau < 1$. Assume that $R_K < 1$ holds, then

$$R_{K+1} < \frac{1}{1/\tau + K(1/\tau - 1)} + \frac{(K+1)(1/\tau - 1)}{1/\tau + K(1/\tau - 1)} = 1.$$

So by induction we can have $R_K < 1, \forall K \geq 1$. " should be "
Further define

$$R_T = \sum_{k=1}^{T} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{T} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)}.$$

Then we have $\sum_{k=1}^{K+1} r_k = R_{K+1}$,

$$R_1 = \sum_{k=1}^{1} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{1} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)} = \tau,$$

and

$$\begin{aligned}
&R_{T+1}\\
=&\frac{1}{1/\tau + T(1/\tau - 1)} + \sum_{k=1}^{T} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{T+1} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)}\\
=&\frac{1}{1/\tau + T(1/\tau - 1)}\\
&+ \frac{(T+1)(1/\tau - 1)}{1/\tau + T(1/\tau - 1)} \sum_{k=1}^{T} \frac{1}{1/\tau + (k-1)(1/\tau - 1)} \prod_{j=k+1}^{T} \frac{j(1/\tau - 1)}{1/\tau + (j-1)(1/\tau - 1)}\\
=&\frac{1}{1/\tau + T(1/\tau - 1)} + \frac{(T+1)(1/\tau - 1)}{1/\tau + T(1/\tau - 1)} R_T.
\end{aligned}$$

Next, we prove $R_T < 1, \forall T \geq 1$, by induction. It can be easily checked that $R_1 = \tau < 1$. Assume that $R_T < 1$ holds, then

$$R_{T+1} < \frac{1}{1/\tau + T(1/\tau - 1)} + \frac{(T+1)(1/\tau - 1)}{1/\tau + T(1/\tau - 1)} = 1.$$

So by induction we can have $R_T < 1, \forall T \geq 1$. Specially, we have $R_{K+1} < 1$. ". (Change the index K of R to T)

33. Page 99, line 12, "$\tau_k = \frac{1}{(2k+1)\mu_g}$" should be "$\tau_k = \frac{3}{(k+1)\mu_g}$".

34. Page 99, line 15, "

$$\leq \frac{1}{2\mu_g(K+1)^2}\left(\mu_g^2\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \|\mathbf{K}\|_2^2\|\mathbf{y}_0 - \mathbf{y}^*\|^2\right)$$

" should be "

$$\leq \frac{3}{\mu_g(K+1)^2}\left(\frac{\mu_g^2}{9}\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + \|\mathbf{K}\|_2^2\|\mathbf{y}_0 - \mathbf{y}^*\|^2\right)$$

".

35. Page 111, line 20, "

$$\alpha\sum_{k=0}^{\infty}\|\mathbf{x}_k - \mathbf{y}_k\|^2 \leq F(\mathbf{y}_k) - \inf F < \infty.$$

" should be "

$$\alpha\sum_{k=0}^{\infty}\|\mathbf{x}_k - \mathbf{y}_k\|^2 \leq F(\mathbf{y}_0) - \inf F < \infty.$$

".

36. Page 112, line 19, "$\limsup_{k\to\infty} g(\mathbf{x}_k) \geq g(\mathbf{z})$" should be "$\liminf_{k\to\infty} g(\mathbf{x}_k) \geq g(\mathbf{z})$".

37. Page 113, line 14, add "Without loss of generality, we assume that $F(\mathbf{x}_k) \neq F^*$ for all k." After "Proof".

38. Page 113, line 19-20, "K_0" should be "k_0".

39. Page 114, line 7-9, " Summing over $k = 1, 2, \cdots, \infty$, we have

$$\sum_{k=2}^{\infty}\|\mathbf{x}_k - \mathbf{y}_k\| \leq \|\mathbf{x}_1 - \mathbf{y}_1\| + c\Psi_1.$$

" should be " Summing over $k = k_0, k_0 + 1, \cdots, \infty$, we have

$$\sum_{k=k_0+1}^{\infty}\|\mathbf{x}_k - \mathbf{y}_k\| \leq \|\mathbf{x}_{k_0} - \mathbf{y}_{k_0}\| + c\Psi_{k_0}.$$

"

40. Page 114, line 13-15, "

$$\sum_{k=2}^{\infty}\|\mathbf{x}_k - \mathbf{x}_{k-1}\| - \sum_{k=1}^{\infty}\beta\|\mathbf{x}_k - \mathbf{x}_{k-1}\| \leq \|\mathbf{x}_1 - \mathbf{y}_1\| + c\Psi_1.$$

So

$$(1-\beta)\sum_{k=2}^{\infty}\|\mathbf{x}_k - \mathbf{x}_{k-1}\| \leq \beta\|\mathbf{x}_1 - \mathbf{x}_0\| + \|\mathbf{x}_1 - \mathbf{y}_1\| + c\Psi_1.$$

" should be "

$$\sum_{k=k_0+1}^{\infty} \|\mathbf{x}_k - \mathbf{x}_{k-1}\| - \sum_{k=k_0}^{\infty} \beta\|\mathbf{x}_k - \mathbf{x}_{k-1}\| \leq \|\mathbf{x}_{k_0} - \mathbf{y}_{k_0}\| + c\Psi_{k_0}.$$

So

$$(1-\beta)\sum_{k=k_0+1}^{\infty} \|\mathbf{x}_k - \mathbf{x}_{k-1}\| \leq \beta\|\mathbf{x}_{k_0} - \mathbf{x}_{k_0-1}\| + \|\mathbf{x}_{k_0} - \mathbf{y}_{k_0}\| + c\Psi_{k_0}.$$

"

41. Page 115, line 19, "$1 \leq d_1 C^2(r_k - r_{k+1})$." should be "$1 \leq d_1 C^2(r_{k-1} - r_k)$."
42. Page 115, line 23-24, and Page 116, line 3, "$\hat{k}_3$" better be "$\hat{k}_2$".
43. Page 116, line 13, delete "for all $k > k_0$,"
44. Page 118, equation (4.7), "

$$\begin{aligned} f(\mathbf{x}_j) &\geq f(\mathbf{y}_j) + \langle \nabla f(\mathbf{y}_j), \mathbf{x}_i - \mathbf{y}_j\rangle + \frac{\sigma}{2}\|\mathbf{x}_i - \mathbf{y}_i\|^2, \\ f(\mathbf{w}) &\geq f(\mathbf{y}_j) + \langle \nabla f(\mathbf{y}_j), \mathbf{w} - \mathbf{y}_j\rangle + \frac{\sigma}{2}\|\mathbf{w} - \mathbf{y}_i\|^2, \end{aligned} \quad j = 0, 1, \cdots, t-1,$$

" should be "

$$\begin{aligned} f(\mathbf{x}_j) &\geq f(\mathbf{y}_j) + \langle \nabla f(\mathbf{y}_j), \mathbf{x}_j - \mathbf{y}_j\rangle + \frac{\sigma}{2}\|\mathbf{x}_j - \mathbf{y}_j\|^2, \\ f(\mathbf{w}) &\geq f(\mathbf{y}_j) + \langle \nabla f(\mathbf{y}_j), \mathbf{w} - \mathbf{y}_j\rangle + \frac{\sigma}{2}\|\mathbf{w} - \mathbf{y}_j\|^2, \end{aligned} \quad j = 0, 1, \cdots, t-1,$$

".

45. Page 122, line 1, "Then $c \leq -\sigma$." should be "Then the above inequality gives $c \leq -\sigma$.".
46. Page 122, line 23, "Expoint-NC-Pair" should be "Exploit-NC-Pair".
47. Page 123, line 7, "Expoint-NC-Pair" should be "Exploit-NC-Pair".
48. Page 126, line 15, "

$$\left[\sqrt{\frac{L}{\sigma}} + \frac{5\sqrt{L\sigma}}{\epsilon^2}(f(\mathbf{z}_1) - f(\mathbf{z}_J))\right]\log 1/\epsilon.$$

" should be "

$$\left[1 + \frac{5\sigma}{\epsilon^2}(f(\mathbf{z}_1) - f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L}{\sigma}}\log 1/\epsilon\right).$$

".

49. Page 127, equation (4.17), "

$$\left[1 + \frac{5\sigma}{\epsilon^2}(f(\mathbf{z}_1) - f(\mathbf{z}_J))\right]\sqrt{\frac{L}{\sigma}}\log 1/\epsilon = \left[\sqrt{\frac{L}{\sigma}} + \frac{5\sqrt{L\sigma}}{\epsilon^2}(f(\mathbf{z}_1) - f(\mathbf{z}_J))\right]\log 1/\epsilon.$$

" should be "

$$\left[1 + \frac{5\sigma}{\epsilon^2}(f(\mathbf{z}_1) - f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L}{\sigma}}\log 1/\epsilon\right).$$

".

50. Page 128, line 4, "$\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v} \leq -O(\alpha)$ " should be "$\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v} \leq -\frac{1}{2}\alpha$".

51. Page 128, line 6, "The above is actually a direct consequece of Corollary 2.7 of [8]." should be added at the beginning of the paragraph.

52. Page 128, Algorithm 4.7 line 1, "Let $\delta' = \frac{\delta}{1+\frac{3L_2^2}{\alpha^3}[f(\mathbf{z}_1)-\min_{\mathbf{z}} f(\mathbf{z})]}$." should be "Let $\delta' = \frac{\delta}{1+\frac{12L_2^2}{\alpha^3}[f(\mathbf{z}_1)-\min_{\mathbf{z}} f(\mathbf{z})]}$.".

53. Page 128, Algorithm 4.7 line 4, "$\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j \leq -O(\alpha)$" should be "$\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j \leq -\frac{1}{2}\alpha$".

54. Page 128, Algorithm 4.7 line 5, "$\eta_j = \frac{\left|\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j\right|}{L_2}\text{sign}\left(\mathbf{v}_j^T\nabla f(\mathbf{z}_j)\right)$" should be "$\eta_j = \frac{2\left|\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j\right|}{L_2}\text{sign}\left(\mathbf{v}_j^T\nabla f(\mathbf{z}_j)\right)$".

55. Page 128, line 15, "

$$\left[1+\frac{3L_2^2}{\alpha^3}(f(\mathbf{z}_1)-f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L_1}{\alpha}}\log 1/\delta'\right).$$

" should be "

$$\left[1+\frac{12L_2^2}{\alpha^3}(f(\mathbf{z}_1)-f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L_1}{\alpha}}\log 1/\delta'\right).$$

".

56. Page 128, line 20, "

$$= -\frac{\left|\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j\right|^3}{3L_2^2} \leq -\frac{\alpha^3}{3L_2^2}.$$

" should be
"

$$= -\frac{2\left|\mathbf{v}_j^T\nabla^2 f(\mathbf{z}_j)\mathbf{v}_j\right|^3}{3L_2^2} \leq -\frac{\alpha^3}{12L_2^2}.$$

".

57. Page 129, equation (4.19), "

$$(J-1)\frac{\alpha^3}{3L_2^2} \leq f(\mathbf{z}_1)-f(\mathbf{z}_J).$$

" should be "

$$(J-1)\frac{\alpha^3}{12L_2^2} \leq f(\mathbf{z}_1)-f(\mathbf{z}_J).$$

".

58. Page 129, line 4, "

$$J \leq 1+\frac{3L_2^2}{\alpha^3}[f(\mathbf{z}_1)-f(\mathbf{z}_J)].$$

" should be
"

$$J \leq 1+\frac{12L_2^2}{\alpha^3}[f(\mathbf{z}_1)-f(\mathbf{z}_J)].$$

".

59. Page 129, line 6, "

$$\left[1+\frac{3L_2^2}{\alpha^3}(f(\mathbf{z}_1)-f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L_1}{\alpha}}\log 1/\delta'\right),$$

” should be “

$$\left[1+\frac{12L_2^2}{\alpha^3}(f(\mathbf{z}_1)-f(\mathbf{z}_J))\right] O\left(\sqrt{\frac{L_1}{\alpha}}\log 1/\delta'\right),$$

”.

60. Page 129, Algorithm 4.8 line 1, “$K = \frac{3L_2^2\Delta_f}{\alpha^3}$” should be “$K = 1 + \frac{15\sqrt{L_2}\Delta_f}{\epsilon^{3/2}}$”.

61. Page 129, line 19, “2. in $\left(\frac{L_2^{1/4}L_1^{1/2}\Delta_f}{\epsilon^{7/4}}\right)\log\frac{1}{\delta\epsilon}$ running time,” should be “2. in $O\left(\frac{L_2^{1/4}L_1^{1/2}\Delta_f}{\epsilon^{7/4}}\log\frac{1}{\delta^{1/3}\epsilon}\right)$ running time,”.

62. Page 130, line 1, “where $\Delta_f = f(\hat{\mathbf{x}}_0) - \min_{\mathbf{x}} f(\mathbf{x})$” should be “where $\Delta_f = f(\hat{\mathbf{x}}_1) - \min_{\mathbf{x}} f(\mathbf{x})$”.

63. Page 131, line 8, “

$$\frac{\alpha^3}{3L_2^2} \le f(\mathbf{z}_1) - f(\mathbf{z}_2) \le f(\mathbf{z}_1) - f(\mathbf{z}_J),$$

” should be “

$$\frac{\alpha^3}{12L_2^2} \le f(\mathbf{z}_1) - f(\mathbf{z}_2) \le f(\mathbf{z}_1) - f(\mathbf{z}_J),$$

”.

64. Page 131, equation (4.20), “

$$\frac{\alpha^3}{3L_2^2} \le f(\mathbf{x}_k) - f(\hat{\mathbf{x}}_k).$$

” should be “

$$\frac{\alpha^3}{12L_2^2} \le f(\mathbf{x}_k) - f(\hat{\mathbf{x}}_k).$$

”.

65. Page 131, line 13, “

$$\frac{\alpha^3}{3L_2^2} \le f(\hat{\mathbf{x}}_{k-1}) - f(\hat{\mathbf{x}}_k),$$

” should be
“

$$\frac{\alpha^3}{12L_2^2} \le f(\hat{\mathbf{x}}_{k-1}) - f(\hat{\mathbf{x}}_k),$$

”.

66. Page 131, line 14, “Summing over $k = 1, \cdots, K$” should be “Summing over $k = 2, \cdots, K$”.

67. Page 131, line 15, “

$$K\frac{\alpha^3}{3L_2^2} \le f(\hat{\mathbf{x}}_0) - f(\hat{\mathbf{x}}_K).$$

” should be
“

$$(K-1)\frac{\alpha^3}{12L_2^2} \le f(\hat{\mathbf{x}}_1) - f(\hat{\mathbf{x}}_K).$$

”.

68. Page 131, line 17, "

$$K \leq \frac{3L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_0) - f(\hat{\mathbf{x}}_K)) \leq \frac{3L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_0) - \min_{\mathbf{x}} f(\mathbf{x})) = \frac{3L_2^2\Delta_f}{\alpha^3}.$$

" should be "

$$\begin{aligned} K \leq 1 + \frac{12L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_1) - f(\hat{\mathbf{x}}_K)) &\leq 1 + \frac{12L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_1) - \min_{\mathbf{x}} f(\mathbf{x})) \\ &= 1 + \frac{12L_2^2\Delta_f}{\alpha^3} = 1 + \frac{12\sqrt{L_2}\Delta_f}{\epsilon^{3/2}}. \end{aligned}$$

".

69. Page 132, line 4, "The right hand side of (4.21) is greater than the previous upper bound for K. So the number of outer iterations given in (4.21) is sufficient for terminating Algorithm 4.8." should be added below (4.21).

70. Page 132, line 14, "

$$\begin{aligned} \sum_{k=1}^{K}(j_k - 1) &\leq \sum_{k=1}^{K} \frac{3L_2^2}{\alpha^3}(f(\mathbf{x}_k) - f(\hat{\mathbf{x}}_k)) \\ &\leq \sum_{k=1}^{K} \frac{3L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_{k-1}) - f(\hat{\mathbf{x}}_k)) \\ &= \frac{3L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_0) - f(\hat{\mathbf{x}}_K)) \leq \frac{3L_2^2\Delta_f}{\alpha^3}. \end{aligned}$$

" should be "

$$\begin{aligned} \sum_{k=1}^{K}(j_k - 1) &\leq \sum_{k=1}^{K} \frac{12L_2^2}{\alpha^3}(f(\mathbf{x}_k) - f(\hat{\mathbf{x}}_k)) \\ &\leq \sum_{k=1}^{K} \frac{12L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_{k-1}) - f(\hat{\mathbf{x}}_k)) \\ &= \frac{12L_2^2}{\alpha^3}(f(\hat{\mathbf{x}}_0) - f(\hat{\mathbf{x}}_K)) \leq \frac{12L_2^2\Delta_f}{\alpha^3}. \end{aligned}$$

".

71. Page 132, line 18, "

$$\sum_{k=1}^{K} j_k \leq \frac{3L_2^2\Delta_f}{\alpha^3} + K \leq \frac{18\sqrt{L_2}\Delta_f}{\epsilon^{3/2}}$$

" should be "

$$\sum_{k=1}^{K} j_k \leq \frac{12L_2^2\Delta_f}{\alpha^3} + K \leq \frac{27\sqrt{L_2}\Delta_f}{\epsilon^{3/2}} + 1$$

".

72. Page 133, equation (4.22), "

$$\frac{18\sqrt{L_2}\Delta_f}{\epsilon^{3/2}} O\left(\sqrt{\frac{L_1}{\alpha}} \log 1/\delta'\right) = O\left(\frac{L_1^{1/2}L_2^{1/4}\Delta_f}{\epsilon^{7/4}} \log \frac{1}{\delta\alpha}\right), \tag{4.22}$$

" should be "

$$\frac{27\sqrt{L_2}\Delta_f}{\epsilon^{3/2}} O\left(\sqrt{\frac{L_1}{\alpha}} \log 1/\delta''\right) = O\left(\frac{L_1^{1/2}L_2^{1/4}\Delta_f}{\epsilon^{7/4}} \log \frac{1}{\delta^{1/3}\epsilon}\right), \tag{4.23}$$

".

73. Page 133, line 3, "where we use the fact that δ' defined in Algorithm 4.7 is at the same order of $\delta\alpha^3$." should be "where we use the fact that δ' defined in Algorithm 4.7 is actually the δ'' in Algorithm 4.8, $K = O(\epsilon^{-3/2})$, $\delta' = O(\delta\alpha^3)$ and $\alpha = \sqrt{L_2\epsilon}$. So $\delta'' = \delta'/K = O(\delta\epsilon^3)$.".

74. Page 133, equation (4.23), "

$$\begin{aligned}
&\sum_{k=1}^{K-1}\left[\sqrt{\frac{L_1}{\alpha}} + \frac{\sqrt{L_1\alpha}}{\epsilon^2}(f(\hat{\mathbf{x}}_k) - f(\mathbf{x}_{k+1}))\right]\log 1/\epsilon \\
\leq &\sum_{k=1}^{K-1}\left[\sqrt{\frac{L_1}{\alpha}} + \frac{\sqrt{L_1\alpha}}{\epsilon^2}(f(\hat{\mathbf{x}}_k) - f(\hat{\mathbf{x}}_{k+1}))\right]\log 1/\epsilon \\
\leq &\left(K\sqrt{\frac{L_1}{\alpha}} + \frac{\Delta_f\sqrt{L_1\alpha}}{\epsilon^2}\right)\log 1/\epsilon \\
\overset{a}{\leq} &\frac{16L_1^{1/2}L_2^{1/4}\Delta_f}{\epsilon^{7/4}}\log 1/\epsilon,
\end{aligned} \tag{4.23}$$

" should be "

$$\begin{aligned}
&\sum_{k=1}^{K-1}\left[\sqrt{\frac{L_1}{\alpha}} + \frac{5\sqrt{L_1\alpha}}{\epsilon^2}(f(\hat{\mathbf{x}}_k) - f(\mathbf{x}_{k+1}))\right]\log 1/\epsilon \\
\leq &\sum_{k=1}^{K-1}\left[\sqrt{\frac{L_1}{\alpha}} + \frac{5\sqrt{L_1\alpha}}{\epsilon^2}(f(\hat{\mathbf{x}}_k) - f(\hat{\mathbf{x}}_{k+1}))\right]\log 1/\epsilon \\
\leq &\left(K\sqrt{\frac{L_1}{\alpha}} + \frac{5\Delta_f\sqrt{L_1\alpha}}{\epsilon^2}\right)\log 1/\epsilon \\
\overset{a}{\leq} &\frac{20L_1^{1/2}L_2^{1/4}\Delta_f}{\epsilon^{7/4}}\log 1/\epsilon,
\end{aligned} \tag{4.24}$$

".

75. Page 139, equation (5.4), "

$$\min_{\mathbf{x}\in\mathcal{R}^n} F(\mathbf{x}) \equiv h(\mathbf{x}) + f(\mathbf{x}),$$

" should be "

$$\min_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x}) \equiv h(\mathbf{x}) + f(\mathbf{x}),$$

".

76. Page 140, line 9, "

$$\boldsymbol{\delta} = \operatorname*{argmin}_{\boldsymbol{\delta}}\left(h_{i_k}(\mathbf{x}_{i_k}^k + \boldsymbol{\delta}) + \left\langle\nabla_{i_k}f(\mathbf{x}^k), \boldsymbol{\delta}\right\rangle + \frac{\theta_k}{2\gamma}\boldsymbol{\delta}^2\right),$$

" should be "

$$\delta = \operatorname*{argmin}_{\delta}\left(h_{i_k}(\mathbf{x}_{i_k}^k + \delta) + \left\langle\nabla_{i_k}f(\mathbf{x}^k), \delta\right\rangle + \frac{n\theta_k}{2\gamma}\delta^2\right),$$

".

77. Page 144, equation (5.16), "

$$\theta_k\left\langle\boldsymbol{\xi}_{i_k}^k, \mathbf{x}_{i_k}^* - \mathbf{z}_{i_k}^{k+1}\right\rangle \leq \theta_k h_{i_k}(\mathbf{x}^*) - \theta_k h_{i_k}(\mathbf{z}_{i_k}^{k+1}) - \frac{\mu\theta_k}{2}\left(\mathbf{z}_{i_k}^{k+1} - \mathbf{x}_{i_k}^*\right)^2.$$

” should be “

$$\theta_k \left\langle \boldsymbol{\xi}_{i_k}^k, \mathbf{x}_{i_k}^* - \mathbf{z}_{i_k}^{k+1} \right\rangle \leq \theta_k h_{i_k}(\mathbf{x}_{i_k}^*) - \theta_k h_{i_k}(\mathbf{z}_{i_k}^{k+1}) - \frac{\mu\theta_k}{2} \left(\mathbf{z}_{i_k}^{k+1} - \mathbf{x}_{i_k}^*\right)^2.$$

”.

78. Page 146, line 10, “

$$\begin{aligned} & \mathbb{E}_{i_k} f(\mathbf{x}^{k+1}) + \mathbb{E}\hat{h}_{k+1} - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \mathbb{E}_{i_k} \left\|\mathbf{z}^{k+1} - \mathbf{x}^*\right\|^2 \\ \leq \quad & (1-\theta)\left(f(\mathbf{x}^k) + \mathbb{E}\hat{h}_k - F(\mathbf{x}^*)\right) + \frac{n^2\theta^2 + (n-1)\theta\mu\gamma}{2\gamma} \left\|\mathbf{z}^k - \mathbf{x}^*\right\|^2. \end{aligned}$$

” should be “

$$\begin{aligned} & \mathbb{E}_{i_k} f(\mathbf{x}^{k+1}) + \mathbb{E}_{i_k}\hat{h}_{k+1} - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \mathbb{E}_{i_k} \left\|\mathbf{z}^{k+1} - \mathbf{x}^*\right\|^2 \\ \leq \quad & (1-\theta)\left(f(\mathbf{x}^k) + \hat{h}_k - F(\mathbf{x}^*)\right) + \frac{n^2\theta^2 + (n-1)\theta\mu\gamma}{2\gamma} \left\|\mathbf{z}^k - \mathbf{x}^*\right\|^2. \end{aligned}$$

”.

79. Page 146, line 13, “

$$\theta^2 + (n-1)\theta\mu\gamma = (1-\theta)(n^2\theta^2 + n\theta\mu\gamma).$$

” should be “

$$n^2\theta^2 + (n-1)\theta\mu\gamma = (1-\theta)(n^2\theta^2 + n\theta\mu\gamma).$$

”.

80. Page 146, line 15, “

$$\begin{aligned} & \mathbb{E}_{i_k} f(\mathbf{x}^{k+1}) + \mathbb{E}\hat{h}_{k+1} - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \mathbb{E}_{i_k} \left\|\mathbf{z}^{k+1} - \mathbf{x}^*\right\|^2 \\ \leq \quad & (1-\theta)\left(f(\mathbf{x}^k) + \mathbb{E}\hat{h}_k - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \left\|\mathbf{z}^k - \mathbf{x}^*\right\|^2\right). \end{aligned}$$

” should be “

$$\begin{aligned} & \mathbb{E}_{i_k} f(\mathbf{x}^{k+1}) + \mathbb{E}_{i_k}\hat{h}_{k+1} - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \mathbb{E}_{i_k} \left\|\mathbf{z}^{k+1} - \mathbf{x}^*\right\|^2 \\ \leq \quad & (1-\theta)\left(f(\mathbf{x}^k) + \hat{h}_k - F(\mathbf{x}^*) + \frac{n^2\theta^2 + n\theta\mu\gamma}{2\gamma} \left\|\mathbf{z}^k - \mathbf{x}^*\right\|^2\right). \end{aligned}$$

”.

81. Page 147, equation (5.21), “

$$\min_{\mathbf{a}\in\mathbb{R}^n} D(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^n \phi_i^*(-\mathbf{a}_i) + \frac{\lambda}{2}\left\|\frac{1}{\lambda n}\mathbf{A}\mathbf{a}\right\|^2,$$

” should be “

$$\min_{\mathbf{a}\in\mathbb{R}^n} D(\mathbf{a}) \equiv \frac{1}{n}\sum_{i=1}^n \phi_i^*(-\mathbf{a}_i) + \frac{\lambda}{2}\left\|\frac{1}{\lambda n}\mathbf{A}\mathbf{a}\right\|^2,$$

”.

82. Page 148, Algorithm 5.2 line 1, "step size γ." should be "step size η.".

83. Page 148, Algorithm 5.2 line 6, "$\mathbf{x}_{k+1}^s = \mathbf{x}_k^s - \gamma\tilde{\nabla} f(\mathbf{x}_k^s)$," should be "$\mathbf{x}_{k+1}^s = \mathbf{x}_k^s - \eta\tilde{\nabla} f(\mathbf{x}_k^s)$,".

84. Page 149, equation (5.24), "

$$\begin{aligned}\mathbb{E}_k\left(\tilde{\nabla} f_{i_{k,s}}(\mathbf{x}_k^s)\right) &= \mathbb{E}_k\nabla f_{i_{k,s}}(\mathbf{x}_k^s) - \mathbb{E}_k\left(\nabla f_{i_{k,s}}(\tilde{\mathbf{x}}^s) - \frac{1}{n}\sum_{i=1}^n \nabla f_i(\tilde{\mathbf{x}}^s)\right)\\ &= \nabla f(\mathbf{x}_k^s).\end{aligned}$$

So $\tilde{\nabla} f_{i_{k,s}}(\mathbf{x}_k^s)$ is an unbiased estimator of $\nabla f(\mathbf{x}_k^s)$." should be "

$$\begin{aligned}\mathbb{E}_k\left(\tilde{\nabla} f(\mathbf{x}_k^s)\right) &= \mathbb{E}_k\nabla f_{i_{k,s}}(\mathbf{x}_k^s) - \mathbb{E}_k\left(\nabla f_{i_{k,s}}(\tilde{\mathbf{x}}^s) - \frac{1}{n}\sum_{i=1}^n \nabla f_i(\tilde{\mathbf{x}}^s)\right)\\ &= \nabla f(\mathbf{x}_k^s).\end{aligned}$$

So $\tilde{\nabla} f(\mathbf{x}_k^s)$ is an unbiased estimator of $\nabla f(\mathbf{x}_k^s)$.".

85. Page 150, line 9, "result with $k=1$ to m" should be "result with $k=0$ to $m-1$".

86. Page 150, line 13, "

$$\leq 2\left(\mu^{-1} + 2Lm\eta^2\right)\mathbb{E}\left(f(\tilde{\mathbf{x}}^s - f(\mathbf{x}^*)\right),$$

" should be "

$$\leq 2\left(\mu^{-1} + 2Lm\eta^2\right)\mathbb{E}\left(f(\tilde{\mathbf{x}}^s) - f(\mathbf{x}^*)\right),$$

".

87. Page 152, line 27, "The algorithm is shown in Algorithm 5.3." should be "Consider problem (5.4). The algorithm is shown in Algorithm 5.3.".

88. Page 153, Algorithm 5.3 line 12, " $\tilde{\mathbf{x}}^{s+1} = \left(\sum_{k=0}^{m-1}\theta_3^k\right)^{-1}\sum_{k=0}^{m-1}\theta_3^k\mathbf{x}_k^s$." should be " $\tilde{\mathbf{x}}^{s+1} = \left(\sum_{k=0}^{m-1}\theta_3^k\right)^{-1}\sum_{k=0}^{m-1}\theta_3^k\mathbf{x}_{k+1}^s$.".

89. Page 154, line 6, "

$$F(\tilde{\mathbf{x}}^{S+1}) - F(\mathbf{x}^*) \leq \theta_3^{-Sn}\left[\frac{1}{4n\gamma}\|\mathbf{z}_0^0 - \mathbf{x}^*\|^2 + \left(1+\frac{1}{n}\right)\left(F(\mathbf{x}_0^0) - F(\mathbf{x}^*)\right)\right].$$

" should be "

$$\mathbb{E}\left(F(\tilde{\mathbf{x}}^{S+1}) - F(\mathbf{x}^*)\right) \leq \theta_3^{-(S+1)n}\left[\left(1+\frac{1}{n}\right)\left(F(\mathbf{x}_0^0) - F(\mathbf{x}^*)\right) + \frac{1}{4n\gamma}\|\mathbf{z}_0^0 - \mathbf{x}^*\|^2\right].$$

".

90. Page 157, equation (5.50), "

$$\mathbb{E}_k F(\mathbf{x}_{k+1}^s) + \frac{1+\frac{\mu\gamma}{\theta_1}}{2\gamma}\left\|\theta_1\mathbf{z}_{k+1}^s - \theta_1\mathbf{x}^*\right\|^2$$

" should be "

$$\mathbb{E}_k F(\mathbf{x}_{k+1}^s) + \frac{1+\frac{\mu\gamma}{\theta_1}}{2\gamma}\mathbb{E}_k\left\|\theta_1\mathbf{z}_{k+1}^s - \theta_1\mathbf{x}^*\right\|^2$$

".

91. Page 157, line 15, "$\theta_3 = \frac{\mu\gamma}{\theta_1} + 1 \leq \frac{L\gamma}{\theta_1} + 1 \leq \frac{1}{3}\sqrt{\frac{\mu}{Ln}} + 1$. Taking expectation on (5.50) for the first $k-1$ iterations," should be "$\theta_3 = \frac{\mu\gamma}{\theta_1} + 1 \leq \frac{1}{3}\sqrt{\frac{\mu}{Ln}} + 1$. Taking full expectation on (5.50),".

92. Page 157, equation (5.51), "

$$\sum_{k=1}^{m}\theta_3^{k-1}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)-a\sum_{k=0}^{m-1}\theta_3^{k}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)-\theta_2\sum_{k=0}^{m-1}\left[\theta_3^{k}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)\right]$$
$$+\frac{\theta_3^m}{2\gamma}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\le\frac{1}{2\gamma}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.$$

" should be "

$$\sum_{k=1}^{m}\theta_3^{k-1}\mathbb{E}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)-a\sum_{k=0}^{m-1}\theta_3^{k}\mathbb{E}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)-\theta_2\sum_{k=0}^{m-1}\left[\theta_3^{k}\mathbb{E}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)\right]$$
$$+\frac{\theta_3^m}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\le\frac{1}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.$$

".

93. Page 158, line 2, "

$$\begin{aligned}&[\theta_1+\theta_2-(1-1/\theta_3)]\sum_{k=1}^{m}\theta_3^{k}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)+\theta_3^m a\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)\\&+\frac{\theta_3^m}{2\gamma}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\\le\ &\theta_2\sum_{k=0}^{m-1}\left[\theta_3^k\mathbb{E}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)\right]+a\mathbb{E}\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.\end{aligned}$$

" should be "

$$\begin{aligned}&[\theta_1+\theta_2-(1-1/\theta_3)]\sum_{k=1}^{m}\theta_3^{k}\mathbb{E}\left(F(\mathbf{x}_k^s)-F(\mathbf{x}^*)\right)+\theta_3^m a\mathbb{E}\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)\\&+\frac{\theta_3^m}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\\le\ &\theta_2\sum_{k=0}^{m-1}\left[\theta_3^k\mathbb{E}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)\right]+a\mathbb{E}\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.\end{aligned}$$

".

94. Page 158, line 5, "$\tilde{\mathbf{x}}^{s+1}=\left(\sum_{j=0}^{m-1}\theta_3^j\right)^{-1}\sum_{j=0}^{m-1}\theta_3^j\mathbf{x}_j^s$" should be "$\tilde{\mathbf{x}}^{s+1}=\left(\sum_{j=0}^{m-1}\theta_3^j\right)^{-1}\sum_{j=0}^{m-1}\theta_3^j\mathbf{x}_{j+1}^s$".

95. Page 158, equation (5.52), "

$$\begin{aligned}&[\theta_1+\theta_2-(1-1/\theta_3)]\theta_3\left(\sum_{k=0}^{m-1}\theta_3^k\right)\left(F(\tilde{\mathbf{x}}^{s+1})-F(\mathbf{x}^*)\right)+\theta_3^m a\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)\\&+\frac{\theta_3^m}{2\gamma}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\\le\ &\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)+a\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.\end{aligned}$$

" should be "

$$\begin{aligned}
&[\theta_1+\theta_2-(1-1/\theta_3)]\theta_3\left(\sum_{k=0}^{m-1}\theta_3^k\right)\mathbb{E}\left(F(\tilde{\mathbf{x}}^{s+1})-F(\mathbf{x}^*)\right)+\theta_3^m a\mathbb{E}\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)\\
&+\frac{\theta_3^m}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\
\le\quad&\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\mathbb{E}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)+a\mathbb{E}\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.
\end{aligned}$$

".

96. Page 159, line 7, "

$$\begin{aligned}
&\theta_2\theta_3^m\left(\sum_{k=0}^{m-1}\theta_3^k\right)\left(F(\tilde{\mathbf{x}}^{s+1})-F(\mathbf{x}^*)\right)+\theta_3^m a\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)+\frac{\theta_3^m}{2\gamma}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\
\le\quad&\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)+a\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.
\end{aligned}$$

" should be "

$$\begin{aligned}
&\theta_2\theta_3^m\left(\sum_{k=0}^{m-1}\theta_3^k\right)\mathbb{E}\left(F(\tilde{\mathbf{x}}^{s+1})-F(\mathbf{x}^*)\right)+\theta_3^m a\mathbb{E}\left(F(\mathbf{x}_m^s)-F(\mathbf{x}^*)\right)+\frac{\theta_3^m}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_m^s-\theta_1\mathbf{x}^*\right\|^2\\
\le\quad&\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\mathbb{E}\left(F(\tilde{\mathbf{x}}^s)-F(\mathbf{x}^*)\right)+a\mathbb{E}\left(F(\mathbf{x}_0^s)-F(\mathbf{x}^*)\right)+\frac{1}{2\gamma}\mathbb{E}\left\|\theta_1\mathbf{z}_0^s-\theta_1\mathbf{x}^*\right\|^2.
\end{aligned}$$

".

97. Page 159, line 12, "

$$\begin{aligned}
&\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\left(F(\tilde{\mathbf{x}}^{S+1})-F(\mathbf{x}^*)\right)+(1-\theta_1-\theta_2)\left(F(\mathbf{x}_m^S)-F(\mathbf{x}^*)\right)\\
&+\frac{\theta_1^2}{2\gamma}\left\|\mathbf{z}_0^{S+1}-\mathbf{x}^*\right\|^2\\
\le\quad&\theta_3^{-Sm}\left\{\left[\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)+(1-\theta_1-\theta_2)\right]\left(F(\mathbf{x}_0^0)-F(\mathbf{x}^*)\right)+\frac{\theta_1^2}{2\gamma}\left\|\mathbf{z}_0^0-\mathbf{x}^*\right\|^2\right\}.
\end{aligned}$$

" should be "

$$\begin{aligned}
&\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)\mathbb{E}\left(F(\tilde{\mathbf{x}}^{S+1})-F(\mathbf{x}^*)\right)+(1-\theta_1-\theta_2)\mathbb{E}\left(F(\mathbf{x}_m^S)-F(\mathbf{x}^*)\right)\\
&+\frac{\theta_1^2}{2\gamma}\mathbb{E}\left\|\mathbf{z}_0^{S+1}-\mathbf{x}^*\right\|^2\\
\le\quad&\theta_3^{-(S+1)m}\left\{\left[\theta_2\left(\sum_{k=0}^{m-1}\theta_3^k\right)+(1-\theta_1-\theta_2)\right]\left(F(\mathbf{x}_0^0)-F(\mathbf{x}^*)\right)+\frac{\theta_1^2}{2\gamma}\left\|\mathbf{z}_0^0-\mathbf{x}^*\right\|^2\right\}.
\end{aligned}$$

".

98. Page 159, line 16, "$\sum_{k=0}^{m-1}\theta_3^k\ge n$" should be "$\sum_{k=0}^{m-1}\theta_3^k\ge m=n$".

99. Page 159, line 18, "

$$F(\tilde{\mathbf{x}}^{S+1}) - F(\mathbf{x}^*) \le \theta_3^{-Sn}\left[\left(1+\frac{1}{n}\right)\left(F(\mathbf{x}_0^0) - F(\mathbf{x}^*)\right) + \frac{1}{4n\gamma}\left\|\mathbf{z}_0^0 - \mathbf{x}^*\right\|^2\right].$$

" should be "

$$\mathbb{E}\left(F(\tilde{\mathbf{x}}^{S+1}) - F(\mathbf{x}^*)\right) \le \theta_3^{-(S+1)n}\left[\left(1+\frac{1}{n}\right)\left(F(\mathbf{x}_0^0) - F(\mathbf{x}^*)\right) + \frac{1}{4n\gamma}\left\|\mathbf{z}_0^0 - \mathbf{x}^*\right\|^2\right].$$

".

100. Page 163, line 20, "

$$\stackrel{a}{=} G_k(\mathbf{x}_k) - \epsilon_k + \frac{\kappa+\mu}{2}\|(\mathbf{x}-\mathbf{x}_k) + (\mathbf{x}_k - \mathbf{x}_k^*)\|^2 - \frac{\kappa}{2}\|\mathbf{x} - \mathbf{y}_{k-1}\|^2$$

" should be "

$$\stackrel{a}{\ge} G_k(\mathbf{x}_k) - \epsilon_k + \frac{\kappa+\mu}{2}\|(\mathbf{x}-\mathbf{x}_k) + (\mathbf{x}_k - \mathbf{x}_k^*)\|^2 - \frac{\kappa}{2}\|\mathbf{x} - \mathbf{y}_{k-1}\|^2$$

".

101. Page 164, line 1, "where in $\stackrel{a}{=}$" should be "where in $\stackrel{a}{\ge}$".

102. Page 168, line 14, "$h(\mathbf{x}) = \mathbf{0}$ " should be "$h(\mathbf{x}) \equiv 0$".

103. Page 171, line 1, "$24L\Delta\sigma\cdot\epsilon^{-3} + 2\sigma^2\epsilon^{-2} + 4\sigma n_0^{-1}\epsilon^{-1}$" should be "$24L\Delta\sigma\cdot\epsilon^{-3} + 2\sigma^2\epsilon^{-2} + 6\sigma n_0^{-1}\epsilon^{-1}$".

104. Page 174, line 9, "Line 17" should be "Line 18".

105. Page 174, line (5.94), "

$$\le \left[3\left(\frac{4Ln_0\Delta}{\epsilon^2}\right) + 2\right]\frac{2\sigma}{\epsilon n_0} + \frac{2\sigma^2}{\epsilon^2} = \frac{24L\sigma\Delta}{\epsilon^3} + \frac{4\sigma}{n_0\epsilon} + \frac{2\sigma^2}{\epsilon^2},$$

" should be "

$$\le \left[3\left(\frac{4Ln_0\Delta}{\epsilon^2}\right) + 3\right]\frac{2\sigma}{\epsilon n_0} + \frac{2\sigma^2}{\epsilon^2} = \frac{24L\sigma\Delta}{\epsilon^3} + \frac{6\sigma}{n_0\epsilon} + \frac{2\sigma^2}{\epsilon^2},$$

".

106. Page 174, line 18, "$24L\Delta\sigma\epsilon^{-3} + 2\sigma^2\epsilon^{-2} + 4\sigma n_0^{-1}\epsilon^{-1}$" should be "$24L\Delta\sigma\epsilon^{-3} + 2\sigma^2\epsilon^{-2} + 6\sigma n_0^{-1}\epsilon^{-1}$".

107. Page 175, line 5, "$n + 12(L\Delta)\cdot n^{1/2}\epsilon^{-2} + 2n_0^{-1}n^{1/2}$" should be "$n + 12(L\Delta)\cdot n^{1/2}\epsilon^{-2} + 3n_0^{-1}n^{1/2}$".

108. Page 175, line 6, "Treating Δ, L, and σ" should be "Treating Δ and L".

109. Page 175, line 10, "$S_2 = \frac{n^{1/2}}{\epsilon n_0}$" should be "$S_2 = \frac{n^{1/2}}{n_0}$".

110. Page 175, equation (5.97), "

$$\le \left[3\left(\frac{4Ln_0\Delta}{\epsilon^2}\right) + 2\right]\frac{n^{1/2}}{n_0} + n = \frac{12(L\Delta)\cdot n^{1/2}}{\epsilon^2} + \frac{2n^{1/2}}{n_0} + n,$$

" should be "

$$\le \left[3\left(\frac{4Ln_0\Delta}{\epsilon^2}\right) + 3\right]\frac{n^{1/2}}{n_0} + n = \frac{12(L\Delta)\cdot n^{1/2}}{\epsilon^2} + \frac{3n^{1/2}}{n_0} + n,$$

".

111. Page 175, line 20, "$n + 12(L\Delta) \cdot n^{1/2}\epsilon^{-2} + 2n_0^{-1}n^{1/2}$" should be "$n + 12(L\Delta) \cdot n^{1/2}\epsilon^{-2} + 3n_0^{-1}n^{1/2}$".

112. Page 176, line 23, "$\tilde{O}(\min(n^{3/4}\epsilon^{-1.75}, n^{1/2}\varepsilon^{-2}, \varepsilon^{-3}))$" should be "$\tilde{O}(\min(n^{3/4}\epsilon^{-1.75}, n^{1/2}\epsilon^{-2}, \epsilon^{-3}))$".

113. Page 177, line 17-18, "We use L_1 to denote the Lipschitz constant of $f_1(\mathbf{x}_1)$, and L_2 to denote the Lipschitz constant of $f_{2,i}(\mathbf{x}_2)$" should be "We use L_1 to denote the Lipschitz constant of $\nabla f_1(\mathbf{x}_1)$, and L_2 to denote the Lipschitz constant of $\nabla f_{2,i}(\mathbf{x}_2)$".

114. Page 178, line 18, "

$$\mathbb{E}_{i_k} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_2 - \theta_1) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

" should be "

$$\mathbb{E}_{i_k} \tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_2 - \theta_1) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

".

115. Page 179, Algorithm 5.7 line 1, "$\tilde{\mathbf{b}}_0 = \mathbf{0}$" should be added.

116. Page 179, line 4, "$L(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda}) = F_1(\mathbf{x}_1) + F_2(\mathbf{x}_2) + \langle \boldsymbol{\lambda}, \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 - \mathbf{b} \rangle$ is the Lagrangian function" should be "$\tilde{L}(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda}) = L(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda}) - L(\mathbf{x}_1^*, \mathbf{x}_2^*, \boldsymbol{\lambda}^*)$ is the shifted Lagrangian function in which $L(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda}) = F_1(\mathbf{x}_1) + F_2(\mathbf{x}_2) + \langle \boldsymbol{\lambda}, \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 - \mathbf{b} \rangle$ is the Lagrangian function".

117. Page 179, line 6, "$\mathbf{G}_2 = \left(\left(1 + \frac{1}{b\theta_2}\right) L_2 + \frac{\beta \|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1} \right) \mathbf{I}$" should be "$\mathbf{G}_2 = \left[\left(1 + \frac{1}{b\theta_2}\right) L_2 + \frac{\beta \|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1} \right] \mathbf{I}$".

118. Page 180, line 2-4, "By the same proof technique of Lemma 5.2 for Katyusha (Algorithm 5.3), we can bound the variance through:

$$\mathbb{E}_{i_k} \left\| \nabla f_2(\mathbf{y}_2^k) - \tilde{\nabla} f_2(\mathbf{y}_2^k) \right\|^2 \leq \frac{2L_2}{b} \left[f_2(\tilde{\mathbf{x}}_2) - f_2(\mathbf{y}_2^k) - \langle \nabla f_2(\mathbf{y}_2^k), \tilde{\mathbf{x}}_2 - \mathbf{y}_2^k \rangle \right]. \tag{5.104}$$

" should be deleted and moved to Step 2, i.e., the last line of page 185, after "and $\overset{e}{\leq}$ uses (5.104).".

119. Page 180, line 9, "we have" should be "we have (in order not to introduce more notations, in the sequel we also use $\partial h(\cdot)$ to denote an element in $\partial h(\cdot)$)".

120. Page 181, line 10, "$\overset{a}{\leq}$ " should be "$\overset{a}{=}$ ".

121. Page 189, line 5, "Define $L(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda}) = F_1(\mathbf{x}_1) - F_1(\mathbf{x}_1^*) + F_2(\mathbf{x}_2) - F_2(\mathbf{x}_2^*) + \langle \boldsymbol{\lambda}, \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 - \mathbf{b} \rangle$. We have" should be "By the definition of $\tilde{L}(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\lambda})$, we have".

122. Page 189, line 7, "

$$L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_1 - \theta_2) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

" should be "

$$\tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_1 - \theta_2) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

".

123. Page 189, line 13, "

$$\mathbb{E}_{i_k} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_2 - \theta_1) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

" should be "

$$\mathbb{E}_{i_k} \tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1 - \theta_2 - \theta_1) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

".

124. Page 190, line 7, “

$$+\left\langle \boldsymbol{\lambda}^* - \bar{\boldsymbol{\lambda}}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}), \mathbf{A}_2\left[\mathbf{x}_2^{k+1} - (1-\theta_1-\theta_2)\mathbf{x}_1^k - \theta_2\tilde{\mathbf{x}}_2 - \theta_1\mathbf{x}_2^*\right]\right\rangle$$

” should be “

$$+\left\langle \boldsymbol{\lambda}^* - \bar{\boldsymbol{\lambda}}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}), \mathbf{A}_2\left[\mathbf{x}_2^{k+1} - (1-\theta_1-\theta_2)\mathbf{x}_2^k - \theta_2\tilde{\mathbf{x}}_2 - \theta_1\mathbf{x}_2^*\right]\right\rangle$$

”.

125. Page 190, line 12, “

$$\mathbb{E}_{i_k} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

” should be “

$$\mathbb{E}_{i_k} \tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

”.

126. Page 191, line 3, “

$$\mathbb{E}_{i_k} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

” should be “

$$\mathbb{E}_{i_k} \tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

”.

127. Page 192, line 5, “

$$\mathbb{E}_{i_k} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

” should be “

$$\mathbb{E}_{i_k} \tilde{L}(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \theta_2 \tilde{L}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - (1-\theta_2-\theta_1) \tilde{L}(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*)$$

”.

128. Page 192, line 15, “If the conditions in Lemma 5.7 hold, then we have” should be “For Algorithm 5.7, we have”.

129. Page 193, line 5, “k” should be “$k+1$”.

130. Page 193, equation (5.134), “

$$\begin{aligned}
& \frac{1}{\theta_1}\mathbb{E}L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \boldsymbol{\lambda}^*) - \frac{\theta_2}{\theta_1}L(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \boldsymbol{\lambda}^*) - \frac{1-\theta_2-\theta_1}{\theta_1}L(\mathbf{x}_1^k, \mathbf{x}_2^k, \boldsymbol{\lambda}^*) \\
\leq\ & \frac{1}{2\beta}\left(\left\|\hat{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^*\right\|^2 - \mathbb{E}\left\|\hat{\boldsymbol{\lambda}}^{k+1} - \boldsymbol{\lambda}^*\right\|^2\right) \\
& + \frac{\theta_1}{2}\left\|\frac{1}{\theta_1}\left[\mathbf{y}_1^k - (1-\theta_1-\theta_2)\mathbf{x}_1^k - \theta_2\tilde{\mathbf{x}}_1\right] - \mathbf{x}_1^*\right\|^2_{\left(L_1 + \frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I} - \frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& - \frac{\theta_1}{2}\mathbb{E}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^{k+1} - (1-\theta_1-\theta_2)\mathbf{x}_1^k - \theta_2\tilde{\mathbf{x}}_1\right] - \mathbf{x}_1^*\right\|^2_{\left(L_1 + \frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I} - \frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& + \frac{\theta_1}{2}\left\|\frac{1}{\theta_1}\left[\mathbf{y}_2^k - (1-\theta_1-\theta_2)\mathbf{x}_2^k - \theta_2\tilde{\mathbf{x}}_2\right] - \mathbf{x}_2^*\right\|^2_{\left(\alpha L_2 + \frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}} \\
& - \frac{\theta_1}{2}\mathbb{E}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^{k+1} - (1-\theta_1-\theta_2)\mathbf{x}_2^k - \theta_2\tilde{\mathbf{x}}_2\right] - \mathbf{x}_2^*\right\|^2_{\left(\alpha L_2 + \frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}},
\end{aligned}$$

" should be "

$$\begin{aligned}
& \frac{1}{\theta_1}\mathbb{E}_{,s-1}\tilde{L}(\mathbf{x}_1^{k+1},\mathbf{x}_2^{k+1},\boldsymbol{\lambda}^*) - \frac{\theta_2}{\theta_1}\tilde{L}(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2,\boldsymbol{\lambda}^*) - \frac{1-\theta_2-\theta_1}{\theta_1}\mathbb{E}_{,s-1}\tilde{L}(\mathbf{x}_1^k,\mathbf{x}_2^k,\boldsymbol{\lambda}^*) \\
\leq\ & \frac{1}{2\beta}\left(\mathbb{E}_{,s-1}\left\|\hat{\boldsymbol{\lambda}}^k-\boldsymbol{\lambda}^*\right\|^2 - \mathbb{E}_{,s-1}\left\|\hat{\boldsymbol{\lambda}}^{k+1}-\boldsymbol{\lambda}^*\right\|^2\right) \\
& +\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{y}_1^k-(1-\theta_1-\theta_2)\mathbf{x}_1^k-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& -\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_1^k-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& +\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{y}_2^k-(1-\theta_1-\theta_2)\mathbf{x}_2^k-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}} \\
& -\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_2^k-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}},
\end{aligned}$$

".

131. Page 193, line 13, "where the expectation is taken under the condition that the randomness under the first s epochs are fixed. " should be "where $\mathbb{E}_{,s-1}$ denotes taking full expectation on $k+1$ inner iterations when fixing the first $s-1$ epochs.".

132. Page 193-194, equation (5.135), "

$$\begin{aligned}
& \frac{1}{\theta_1}\mathbb{E}L(\mathbf{x}_1^{k+1},\mathbf{x}_2^{k+1},\boldsymbol{\lambda}^*) - \frac{\theta_2}{\theta_1}L(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2,\boldsymbol{\lambda}^*) - \frac{1-\theta_2-\theta_1}{\theta_1}L(\mathbf{x}_1^k,\mathbf{x}_2^k,\boldsymbol{\lambda}^*) \\
\leq\ & \frac{1}{2\beta}\left(\left\|\hat{\boldsymbol{\lambda}}^k-\boldsymbol{\lambda}^*\right\|^2 - \mathbb{E}\left\|\hat{\boldsymbol{\lambda}}^{k+1}-\boldsymbol{\lambda}^*\right\|^2\right) \\
& +\frac{\theta_1}{2}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^k-(1-\theta_1-\theta_2)\mathbf{x}_1^{k-1}-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& -\frac{\theta_1}{2}\mathbb{E}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_1^k-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}} \\
& +\frac{\theta_1}{2}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^k-(1-\theta_1-\theta_2)\mathbf{x}_2^{k-1}-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}} \\
& -\frac{\theta_1}{2}\mathbb{E}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_2^k-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}}, \quad k\geq 1.
\end{aligned}$$

” should be “

$$
\begin{aligned}
& \frac{1}{\theta_1}\mathbb{E}_{,s-1}\tilde{L}(\mathbf{x}_1^{k+1},\mathbf{x}_2^{k+1},\boldsymbol{\lambda}^*)-\frac{\theta_2}{\theta_1}\tilde{L}(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2,\boldsymbol{\lambda}^*)\\
& -\frac{1-\theta_2-\theta_1}{\theta_1}\mathbb{E}_{,s-1}\tilde{L}(\mathbf{x}_1^{k},\mathbf{x}_2^{k},\boldsymbol{\lambda}^*)\\
\leq\ & \frac{1}{2\beta}\left(\mathbb{E}_{,s-1}\left\|\hat{\boldsymbol{\lambda}}^k-\boldsymbol{\lambda}^*\right\|^2-\mathbb{E}_{,s-1}\left\|\hat{\boldsymbol{\lambda}}^{k+1}-\boldsymbol{\lambda}^*\right\|^2\right)\\
& +\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^k-(1-\theta_1-\theta_2)\mathbf{x}_1^{k-1}-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}}\\
& -\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_1^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_1^{k}-\theta_2\tilde{\mathbf{x}}_1\right]-\mathbf{x}_1^*\right\|^2_{\left(L_1+\frac{\beta\|\mathbf{A}_1^T\mathbf{A}_1\|}{\theta_1}\right)\mathbf{I}-\frac{\beta\mathbf{A}_1^T\mathbf{A}_1}{\theta_1}}\\
& +\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^k-(1-\theta_1-\theta_2)\mathbf{x}_2^{k-1}-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}}\\
& -\frac{\theta_1}{2}\mathbb{E}_{,s-1}\left\|\frac{1}{\theta_1}\left[\mathbf{x}_2^{k+1}-(1-\theta_1-\theta_2)\mathbf{x}_2^{k}-\theta_2\tilde{\mathbf{x}}_2\right]-\mathbf{x}_2^*\right\|^2_{\left(\alpha L_2+\frac{\beta\|\mathbf{A}_2^T\mathbf{A}_2\|}{\theta_1}\right)\mathbf{I}},\quad k\geq 1.
\end{aligned}
$$

”.

133. Page 198, line 6, “For $\theta_{1,s-1}\geq\theta_{1,s}$, we have $\|\mathbf{x}\|^2_{\theta_{1,s-1}L}\geq\|\mathbf{x}\|^2_{\theta_{1,s}L}$, thus” should be “Since $\|\mathbf{x}\|^2_{\mathbf{M}_1}\geq\|\mathbf{x}\|^2_{\mathbf{M}_2}$ if $\mathbf{M}_1\succeq\mathbf{M}_2$, and $\theta_{1,s-1}\geq\theta_{1,s}$, we have ”.
134. Page 203, line 1, “We apply mini-batch AGD to solve” should be “We apply Stochastic Accelerated Gradient Descent (SAGD) to solve”.
135. Page 204, line 16, “Then using” should be “Then”.
136. Page 210, line 4, “Niu et al.[12]” should be “Recht et al.[12]”.
137. Page 210, last but one line, “$(k-\tau-1)$-th iteration” should be “$(k-\tau)$-th iteration”.
138. Page 211, line 9, “$j(k)\in\{k-\tau,k-\tau+1,\cdots,k\}$.” should be “$j(k)\in\{k-\tau+1,k-\tau+2,\cdots,k\}$.”.
139. Page 211, line 13, “(called Accelerated Proximal Gradient Method when $h(\mathbf{x})\neq 0$)” should be “(the special case of Accelerated Proximal Gradient Method when $h(\mathbf{x})\equiv 0$)”.
140. Page 213, line 14, “$\theta_k=\frac{-\gamma\mu+\sqrt{\gamma\mu^2+4\gamma\mu}}{2}$” should be “$\theta_k=\frac{-\gamma\mu+\sqrt{\gamma^2\mu^2+4\gamma\mu}}{2}$”.
141. Page 215, line 2, “

$$\left\langle\nabla f(\mathbf{w}^{j(k)})-\nabla f(\mathbf{y}^k),\mathbf{x}^{k+1}-\mathbf{y}^k\right\rangle$$

” should be “

$$-\left\langle\nabla f(\mathbf{w}^{j(k)})-\nabla f(\mathbf{y}^k),\mathbf{x}^{k+1}-\mathbf{y}^k\right\rangle$$

”.

142. Page 218, equation (6.23), “

$$
\begin{aligned}
&\overset{d}{\leq}\ \left[\sum_{ii=1}^{\min(\tau,k)}\left(1+\sum_{l=1}^{ii}1\right)\right]\sum_{ii=1}^{\tau}\left(1+\sum_{l=1}^{ii}1\right)\left\|\mathbf{x}^{k-ii+1}-\mathbf{y}^{k-ii}\right\|^2\\
&\leq\ \frac{\tau^2+3\tau}{2}\sum_{i=1}^{\min(\tau,k)}(i+1)\left\|\mathbf{x}^{k-i+1}-\mathbf{y}^{k-i}\right\|^2,
\end{aligned}
$$

" should be "

$$\begin{aligned} &\overset{d}{\leq} \left[\sum_{ii=1}^{\tau}\left(1+\sum_{l=1}^{ii}1\right)\right]\sum_{ii=1}^{k-j(k)}\left(1+\sum_{l=1}^{ii}1\right)\left\|\mathbf{x}^{k-ii+1}-\mathbf{y}^{k-ii}\right\|^2 \\ &= \frac{\tau^2+3\tau}{2}\sum_{i=1}^{k-j(k)}(i+1)\left\|\mathbf{x}^{k-i+1}-\mathbf{y}^{k-i}\right\|^2, \end{aligned}$$

".

143. Page 218, line 8, "(e.g., $k \geq 2(\tau-1)$)" should be "(e.g., $k \geq 2\tau$. In this case, $k - j(k) \leq \min(\tau, k-\tau)$.)".

144. Page 219, line 1, "for $\theta_k = \frac{2}{k+2}$, $k \geq 2(\tau-1)$, and $i \leq \min(\tau, k-\tau)$," should be "for $\theta_k = \frac{2}{k+2}$ and $i \leq \min(\tau, k-\tau)$,".

145. Page 221, line 9, "$\theta = \frac{-\gamma\mu+\sqrt{\gamma\mu^2+4\gamma\mu}}{2}$" should be "$\theta = \frac{-\gamma\mu+\sqrt{\gamma^2\mu^2+4\gamma\mu}}{2}$".

146. Page 221, line 14, "$\tau \geq 2$" should be "$\tau \geq 1$".

147. Page 222, line 4, "θ^{K-k}" should be "$(1-\theta)^{K-k}$".

148. Page 222, line 8, "

$$-\gamma\left(\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}\right)\sum_{i=0}^{K}(1-\theta)^{K-k}\left\|\mathbf{x}^{k+1}-\mathbf{y}^k\right\|^2$$

" should be "

$$-\gamma\left(\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}\right)\sum_{k=0}^{K}(1-\theta)^{K-k}\left\|\frac{1}{\gamma}\left(\mathbf{x}^{k+1}-\mathbf{y}^k\right)\right\|^2$$

".

149. Page 222, line 11, "

$$-\gamma\left[\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}-\left(\frac{\gamma^2L^2}{2C_1}+\gamma L\right)\frac{(\tau^2+3\tau)^2}{4(1-\theta)^\tau}\right]\sum_{i=0}^{K}(1-\theta)^{K-k}\left\|\mathbf{x}^{k+1}-\mathbf{y}^k\right\|^2$$

" should be "

$$-\gamma\left[\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}-\left(\frac{\gamma^2L^2}{2C_1}+\gamma L\right)\frac{(\tau^2+3\tau)^2}{4(1-\theta)^\tau}\right]\sum_{k=0}^{K}(1-\theta)^{K-k}\left\|\frac{1}{\gamma}\left(\mathbf{x}^{k+1}-\mathbf{y}^k\right)\right\|^2$$

".

150. Page 222, line 14, "

$$-\gamma\left[\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}-\left(\frac{\gamma^2L^2}{2C_1}+\gamma L\right)\frac{3(\tau^2+3\tau)^2}{8}\right]\sum_{i=0}^{K}(1-\theta)^{K-k}\left\|\mathbf{x}^{k+1}-\mathbf{y}^k\right\|^2,$$

" should be "

$$-\gamma\left[\frac{1}{2}-\frac{\gamma L}{2}-\frac{C_1}{2}-\left(\frac{\gamma^2L^2}{2C_1}+\gamma L\right)\frac{3(\tau^2+3\tau)^2}{8}\right]\sum_{k=0}^{K}(1-\theta)^{K-k}\left\|\frac{1}{\gamma}\left(\mathbf{x}^{k+1}-\mathbf{y}^k\right)\right\|^2,$$

".

151. Page 222, line 16, "$C_1 = \gamma L$" should be "$C_1 = 1.5\gamma L$".

152. Page 223, Algorithm 6.2 line 6, "3 $\quad \delta_k = \operatorname{argmin}_\delta h_{i_k}(\mathbf{z}_{i_k}^k + \delta) + \nabla_{i_k} f(\mathbf{w}^{j(k)}) \cdot \delta + \frac{\theta_k}{2\gamma}\delta^2$," should be "3 $\quad \delta_k = \operatorname{argmin}_\delta h_{i_k}(\mathbf{z}_{i_k}^k + \delta) + \nabla_{i_k} f(\mathbf{w}^{j(k)}) \cdot \delta + \frac{n\theta_k}{2\gamma}\delta^2$,".

153. Page 223, line 19, "$\tau \le \sqrt{n}$" should be "$\tau \le n$".

154. Page 226, line 17, "$\overset{a}{\le}$", "$\overset{b}{\le}$" should be "$\overset{a}{=}$", "$\overset{b}{=}$".

155. Page 228, line 3, "

$$\sum_{i_k=1}^{n} \langle \boldsymbol{\xi}_{i_k}^k, \theta_k \mathbf{z}_{i_k}^k - \theta_k \mathbf{x}_{i_k}^* \rangle.$$

" should be "

$$\sum_{i_k=1}^{n} \langle \boldsymbol{\xi}_{i_k}^k, \theta_k \mathbf{z}_{i_k}^k - \theta_k \mathbf{x}_{i_k}^* \rangle$$

".

156. Page 229, equation (6.46), "

$$\theta_k \langle \boldsymbol{\xi}_{i_k}^k, \mathbf{x}_{i_k}^* - \mathbf{z}_{i_k}^{k+1} \rangle \le \theta_k h_{i_k}(\mathbf{x}^*) - \theta_k h_{i_k}(\mathbf{z}_{i_k}^{k+1}) - \frac{\mu\theta_k}{2}\left(\mathbf{z}_{i_k}^{k+1} - \mathbf{x}_{i_k}^*\right)^2.$$

" should be "

$$\theta_k \langle \boldsymbol{\xi}_{i_k}^k, \mathbf{x}_{i_k}^* - \mathbf{z}_{i_k}^{k+1} \rangle \le \theta_k h_{i_k}(\mathbf{x}_{i_k}^*) - \theta_k h_{i_k}(\mathbf{z}_{i_k}^{k+1}) - \frac{\mu\theta_k}{2}\left(\mathbf{z}_{i_k}^{k+1} - \mathbf{x}_{i_k}^*\right)^2.$$

".

157. Page 231, equation (6.51), "

$$\begin{aligned}
&\overset{d}{\le} \left[\sum_{ii=1}^{\min(\tau,k)} \left(1 + \frac{1}{n}\sum_{l=1}^{ii} 1\right)\right] \sum_{ii=1}^{\tau} \left(1 + \frac{1}{n}\sum_{l=1}^{ii} 1\right) \left\|\mathbf{x}^{k-ii+1} - \mathbf{y}^{k-ii}\right\|^2 \\
&\le \left(\frac{\tau^2+\tau}{2n} + \tau\right) \sum_{i=1}^{\min(\tau,k)} \left(\frac{i}{n} + 1\right) \left\|\mathbf{x}^{k-i+1} - \mathbf{y}^{k-i}\right\|^2,
\end{aligned}$$

" should be "

$$\begin{aligned}
&\overset{d}{\le} \left[\sum_{ii=1}^{\tau} \left(1 + \frac{1}{n}\sum_{l=1}^{ii} 1\right)\right] \sum_{ii=1}^{\min(\tau,k)} \left(1 + \frac{1}{n}\sum_{l=1}^{ii} 1\right) \left\|\mathbf{x}^{k-ii+1} - \mathbf{y}^{k-ii}\right\|^2 \\
&= \left(\frac{\tau^2+\tau}{2n} + \tau\right) \sum_{i=1}^{\min(\tau,k)} \left(\frac{i}{n} + 1\right) \left\|\mathbf{x}^{k-i+1} - \mathbf{y}^{k-i}\right\|^2,
\end{aligned}$$

".

158. Page 231, line 5, "As in Theorem 6.1, we are more interested in the limiting case, when k is large (e.g., $k \ge 2(\tau - 1)$) we suppose that at the first τ steps, we run our algorithm in serial." should be deleted.

159. Page 235, line 1, "we assume $n \ge 2$." should be "we assume $n \ge 2$ and $\tau \ge 1$.".

160. Page 235, line 4, "

$$\le \frac{1}{1 - \frac{1}{2\sqrt{3}}} \le \frac{3}{2},$$

" should be "

$$\overset{d}{\le} \frac{1}{1 - \frac{1}{2\sqrt{3}}} \le \frac{3}{2},$$

".

161. Page 235, line 5, “and $\overset{c}{\leq}$ uses $\sqrt{\frac{\mu}{L_c}}/n \leq \frac{1}{n} \leq \frac{1}{2}$.” should be “$\overset{c}{\leq}$ uses $\sqrt{\frac{\mu}{L_c}}/n \leq \frac{1}{n} \leq \frac{1}{2}$, and $\overset{d}{\leq}$ uses the fact that $\left(1 - \frac{1}{2\sqrt{3}x}\right)^{-x}$ is a decreasing function when $x \geq 1$.”.

162. Page 235, line 6, “θ^{K-k}” should be “$(1-\theta)^{K-k}$”.

163. Page 237, line 21, “$\kappa = \frac{L}{\mu}$” should be “$\kappa = \frac{L}{\lambda}$”.

164. Page 238, line 18, “subset” should be “sub-vector”.

165. Page 239, line 3, “we have” should be “we have the following optimization problem”.

166. Page 239, line 16, “$L_c = \frac{1}{2mn^2\lambda} + \frac{1}{4mnL}$” should be “$L_c = \frac{1}{mn^2\lambda} + \frac{1}{2mnL}$”.

167. Page 239, line 24, “Moreover, $g_0(\mathbf{b})$ is $\frac{1}{6mn^2L}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ ($j \in [n]$, $i \in [m]$) are $\frac{3}{4nmL}$-strongly convex.” should be “Moreover, $g_0(\mathbf{b})$ is $\frac{1}{3mn^2L}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ ($j \in [n]$, $i \in [m]$) are $\frac{1}{2nmL}$-strongly convex.”.

168. Page 240, line 1, “Because $\frac{\mu_1}{6mn^2} = \frac{1}{6mn^2L} \leq \frac{1}{4mnL}$, (6.61) is right. Also, it is easy to prove that $g_0(\mathbf{b})$ is $\frac{\mu_1}{6mn^2}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ with $i \in [m]$ and $j \in [n]$ are $\frac{3\mu_1}{4nm}$-strongly convex.” should be
“We can also find that

$$\|\nabla_{\mathbf{b}} f(\mathbf{a}, \tilde{\mathbf{b}}) - \nabla_{\mathbf{b}} f(\mathbf{a}, \bar{\mathbf{b}})\| \leq \frac{1}{mn^2}\left|\frac{1}{\lambda} - \frac{3}{L}\right| \|\tilde{\mathbf{b}} - \bar{\mathbf{b}}\| < L_c\|\tilde{\mathbf{b}} - \bar{\mathbf{b}}\|.$$

So (6.61) is right. Also, it is easy to prove that $g_0(\mathbf{b})$ is $\frac{\mu_1}{3mn^2}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ with $i \in [m]$ and $j \in [n]$ are $\frac{1}{2nmL}$-strongly convex.”.

169. Page 241, Algorithm 6.3 line 5, “If” should be “**if**”.

170. Page 241, Algorithm 6.3 line 8, “Else” should be “**else**”.

171. Page 241, Algorithm 6.3 line 11, “End If” should be “**end if**”.

172. Page 241, Algorithm 6.3, step numbers 7-14 should change to 6-13, respectively.

173. Page 243, line 6-7, “

$$f(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{m} f_i(\mathbf{a}_i, \mathbf{b}).$$

” should be added.

174. Page 243, line 9, “Suppose $n \leq \frac{L}{\lambda}$, and set $u_1 = \frac{1}{L}$ and $\mu_2 = 1/\kappa_g$.” should be “Suppose that l_{ij}’s are all convex and L-smooth and $n \leq \frac{L}{\lambda}$, and set $\mu_1 = \frac{1}{L}$ and $\mu_2 = 1/\kappa_g$.”.

175. Page 243, line 11, “$L_c = \frac{1}{2mn^2\lambda} + \frac{1}{4mnL}$” should be “$L_c = \frac{1}{mn^2\lambda} + \frac{1}{2mnL}$”.

176. Page 243, line 12, “$g_0(\mathbf{b})$ is $\frac{1}{6mn^2\kappa_g L}$-strongly convex. $g_{ij}(\mathbf{a}_{ij})$ with $i \in [m]$ and $j \in [n]$ are $\frac{1}{4nmL}$-strongly convex.” should be “$g_0(\mathbf{b})$ is $\frac{1}{3mn^2\kappa_g L}$-strongly convex. $g_{ij}(\mathbf{a}_{ij})$ with $i \in [m]$ and $j \in [n]$ are $\frac{1}{2nmL}$-strongly convex.”.

177. Page 243, line 16, “$\frac{1}{2mn^2\lambda} + \frac{1}{4mnL}$” should be “$\frac{1}{mn^2\lambda} + \frac{1}{2mnL}$”.

178. Page 243, line 16, “$g_0(\mathbf{b})$ is $\frac{\mu_1}{3mn^2}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ are $\frac{1}{4nmL}$-strongly convex, $i \in [m]$, $j \in [n]$.” should be “$g_0(\mathbf{b})$ is $\frac{\mu_1\mu_2}{3mn^2}$-strongly convex and $g_{ij}(\mathbf{a}_{ij})$ are $\frac{1}{2nmL}$-strongly convex, $i \in [m]$, $j \in [n]$.”.

179. Page 244, line 12, “u_3-strongly convex” should be “μ_3-strongly convex”.

180. Page 244, line 13, “u_4-strongly convex” should be “μ_4-strongly convex”.

181. Page 245, line 2, “$q \leq pb$” should be “$q \leq p_b$”.

182. Page 245, line 4, “Step 7” should be “Step 6”.

183. Page 245, equation (6.68), “

$$\partial g_0(\hat{\mathbf{b}}_c^{k+1}) + \nabla_{\mathbf{b}} f(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k) + \frac{\theta L_2}{p_b}(\hat{\mathbf{b}}_c^{k+1} - \hat{\mathbf{b}}^k) = \mathbf{0}.$$

” should be “

$$\mathbf{0} \in \partial g_0(\hat{\mathbf{b}}_c^{k+1}) + \nabla_{\mathbf{b}} f(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k) + \frac{\theta L_2}{p_b}(\hat{\mathbf{b}}_c^{k+1} - \hat{\mathbf{b}}^k).$$

”.

184. Page 245, line 6, “Step 13” should be “Step 12”.

185. Page 245, equation (6.69), “

$$\partial g_0(\hat{\mathbf{b}}_c^{k+1}) + \nabla_{\mathbf{b}} f(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k) + L_2(\mathbf{b}_c^{k+1} - \tilde{\mathbf{b}}^k) = \mathbf{0}.$$

” should be “

$$\mathbf{0} \in \partial g_0(\hat{\mathbf{b}}_c^{k+1}) + \nabla_{\mathbf{b}} f(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k) + L_2(\mathbf{b}_c^{k+1} - \tilde{\mathbf{b}}^k).$$

”.

186. Page 245, line 11, “we have” should be “we have (in order not to introduce more notations, in the sequel we also use $\partial g(\cdot)$ to denote an element in $\partial g(\cdot)$)”.

187. Page 245, equation (6.72), “

$$\partial g_{i,j(i)}\left(\hat{\mathbf{a}}_{i,j(i),c}^{k+1}\right) + \nabla_{\mathbf{a}_{i,j(i)}} f_i(\tilde{\mathbf{a}}_i^k, \tilde{\mathbf{b}}^k) + L_2\left(\mathbf{a}_{i,j(i),c}^{k+1} - \tilde{\mathbf{a}}_{i,j(i)}^k\right) = \mathbf{0}.$$

” should be “

$$\mathbf{0} \in \partial g_{i,j(i)}\left(\hat{\mathbf{a}}_{i,j(i),c}^{k+1}\right) + \nabla_{\mathbf{a}_{i,j(i)}} f_i(\tilde{\mathbf{a}}_i^k, \tilde{\mathbf{b}}^k) + L_2\left(\mathbf{a}_{i,j(i),c}^{k+1} - \tilde{\mathbf{a}}_{i,j(i)}^k\right).$$

”.

188. Page 246, line 1, “$\mathbf{a}$ is chosen at iteration k” should be “$\mathbf{a}$ is chosen to update at iteration k”.

189. Page 247, line 10, “Step 13” should be “Step 12”.

190. Page 248, line 4, “

$$-\left\langle \nabla_{\mathbf{a}_{i,j(i)}} f_i(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta \mathbf{a}_{ij}^* \right\rangle.$$

” should be “

$$-\left\langle \nabla_{\mathbf{a}_{i,j}} f_i(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta \mathbf{a}_{ij}^* \right\rangle.$$

”.

191. Page 248, line 8, “

$$+\frac{\theta^2 L_2}{2p_a^2}\mathbb{E}\left\|\hat{\mathbf{a}}^{k+1} - \mathbf{a}^*\right\|^2 - \frac{\theta^2 L_2}{2p_a^2}\left\|\hat{\mathbf{a}}^k - \mathbf{a}^*\right\|^2$$

” should be “

$$+\frac{\theta^2 L_2}{2p_a^2}\mathbb{E}_k\left\|\hat{\mathbf{a}}^{k+1} - \mathbf{a}^*\right\|^2 - \frac{\theta^2 L_2}{2p_a^2}\left\|\hat{\mathbf{a}}^k - \mathbf{a}^*\right\|^2$$

”.

192. Page 248, line 11, "

$$-\sum_{i=1}^{m}\sum_{j=1}^{n}\left\langle \nabla_{\mathbf{a}_{i,j(i)}} f_i(\tilde{\mathbf{a}}_i^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta\mathbf{a}_{ij}^*\right\rangle$$

" should be "

$$-\sum_{i=1}^{m}\sum_{j=1}^{n}\left\langle \nabla_{\mathbf{a}_{i,j}} f_i(\tilde{\mathbf{a}}_i^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta\mathbf{a}_{ij}^*\right\rangle$$

".

193. Page 249, line 8, "

$$-\sum_{i=1}^{m}\sum_{j=1}^{n}\left\langle \nabla_{\mathbf{a}_{i,j(i)}} f_i(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta\mathbf{a}_{ij}^*\right\rangle$$

" should be "

$$-\sum_{i=1}^{m}\sum_{j=1}^{n}\left\langle \nabla_{\mathbf{a}_{i,j}} f_i(\tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k), \tilde{\mathbf{a}}_{ij}^k - (1-\theta)\mathbf{a}_{ij}^k - \theta\mathbf{a}_{ij}^*\right\rangle$$

".

194. Page 250, line 6, "$\theta = \frac{p_b\sqrt{u_3/L_2}}{2} = \frac{p_a\sqrt{u_4/L_2}}{2}$." should be "$\theta = \frac{p_b\sqrt{\mu_3/L_2}}{2} = \frac{p_a\sqrt{\mu_4/L_2}}{2}$.".

195. Page 250, line 13-14, "u_3" and "u_4" should be "μ_3" and "μ_4".

196. Page 252, line 11, "From Steps 3 and 13" should be "From Steps 3 and 12".

197. Page 252, line 19 and line 22, "μ_4" should be "μ_3".

198. Page 253, line 17, "$L_2 = \frac{1}{2mn^2\lambda} + \frac{1}{4mnL}$, $\mu_3 = \frac{1}{6mn^2L}$, and $\mu_4 = \frac{3}{4mnL}$." should be "$L_2 = \frac{1}{mn^2\lambda} + \frac{1}{2mnL}$, $\mu_3 = \frac{1}{3mn^2L}$, and $\mu_4 = \frac{1}{2mnL}$.".

199. Page 253, last two lines, "The communication and the iteration costs" should be "The iteration and the communication costs"

200. Page 254, line 2, "from Steps 3, 7, and 13" should be "from Steps 3, 6, and 12".

201. Page 254, line 4, "$L_2 = \frac{1}{2mn^2\lambda} + \frac{1}{4mnL}$, $\mu_3 = \frac{1}{6mn^2\kappa_g L}$, and $\mu_4 = \frac{1}{4mnL}$." should be "$L_2 = \frac{1}{mn^2\lambda} + \frac{1}{2mnL}$, $\mu_3 = \frac{1}{3mn^2\kappa_g L}$, and $\mu_4 = \frac{1}{2mnL}$.".

202. Page 253, line 7, "The communication and the iteration costs" should be "The iteration and the communication costs"

203. Page 262, line 14-15, "

$\|\|\cdot\|\|$	Operator norm of an operator or a matrix.
$\|\|\cdot\|\|_2$ or $\|\cdot\|$	ℓ_2 norm of vectors, $\|\|\mathbf{v}\|\|_2 = \sqrt{\sum_i \mathbf{v}_i^2}$; $\|\cdot\|$ is also used for general norm of vectors.

" should be "

$\|\|\cdot\|\|_2$ or $\|\|\cdot\|\|$	Operator norm of an operator or a matrix.
$\|\cdot\|$	ℓ_2 norm of vectors, $\|\|\mathbf{v}\|\| = \sqrt{\sum_i \mathbf{v}_i^2}$; $\|\cdot\|$ is also used for general norm of vectors.

"

204. Page 262, last line, "$\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2$" should be "$\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2$".

205. Page 263, line 14, "factorization (1)" should be "factorization".

206. Page 266, equation (A.8), "$\frac{L}{6}\|\mathbf{y}-\mathbf{x}\|^2$" should be "$\frac{L}{6}\|\mathbf{y}-\mathbf{x}\|^3$".

207. Page 267, line 1, "if f is differentiable and μ-strongly convex" should be "if f is μ-strongly convex".

208. Page 267, line 4, "On the other hand, we can have" should be "On the other hand, if f is differentiable and μ-strongly convex, we can have".

209. Page 268, line 2, "strongly convex" should be "strictly convex".

210. Page 269, line 4, "C is the domain of f" should be "C is the intersection of the domains of f and $\mathbf{g}$".

211. Page 269, line 4, "$\mathcal{D} = \{(\mathbf{u}, \mathbf{v})|d(\mathbf{u}, \mathbf{v}) > -\infty\}$" should be "$\mathcal{D} = \{(\mathbf{u}, \mathbf{v})|\mathbf{v} \geq 0, d(\mathbf{u}, \mathbf{v}) > -\infty\}$".

212. Page 269, line 20, "When the Slater's condition holds" should be "When the Slater's condition holds for a convex problem".

213. Page 269, line 25, "Complementary slackness: $\mathbf{v}_i g_i(\mathbf{x}) = 0$" should be "Complementary slackness: $\mathbf{v}_i g_i(\mathbf{x}) = 0, i = 1, \cdots, p$".

214. Page 269, line 28-29, "when $f(\mathbf{x})$ and $g_i(\mathbf{x})$, $i = 1, \cdots, p$, are all convex" should be "when problem (A.12) is convex and satisfies the Slater's condition".

215. Page 270, line 16, "dom $g = \{\mathbf{x} \in \mathbb{R} : g(\mathbf{x}) < +\infty\}$" should be "dom $g = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) < +\infty\}$".

216. Page 271, line 26, "dom$\partial f := \{\mathbf{x} \in \mathbb{R}^n : \partial f(\mathbf{u}) \neq \emptyset\}$" should be "dom$\partial f := \{\mathbf{u} \in \mathbb{R}^n : \partial f(\mathbf{u}) \neq \emptyset\}$".