Errata of Alternating Direction Method of Multipliers for Machine Learning

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

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

ABSTRACT

The corrections for the book Alternating Direction Method of Multipliers for Machine Learning are listed below. Fortunately, they are all non-critical. Actually, some corrections are just for making the book better. They are all 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. Change all "Cauchy-Schwartz" to "Cauchy-Schwarz"
2. Page xii, change "all the algorithms introduced in this book use a fixed penalty." to "all the algorithms introduced in this book use a fixed penalty (some of the accelerated algorithms can be regarded as using a variable penalty)."
3. Pages 3-4, before and after Eqn. (1.7), three "$\mathbf{Y} \geq 0$" should be "$\mathbf{Y} \geq \mathbf{0}$"
4. Page 5, line 18, "[6,23]" should be "[23]"
5. Page 17, line 5, delete "$k \leftarrow k + 1$."
6. Page 29, lines 18 and 19, "When $g(\mathbf{y})$ is strongly convex, ... When $g(\mathbf{y})$ is L-smooth," should be "When $g(\mathbf{y})$ is μ-strongly convex, ... When $g(\mathbf{y})$ is L-smooth and convex,"
7. Page 36, line 8, "$\nabla g(y^{k+1}) = -\mathbf{B}^T\boldsymbol{\lambda}^{k+1} \quad \text{and} \quad \nabla g(y^*) = -\mathbf{B}^T\boldsymbol{\lambda}^*$." should be "$\nabla g(\mathbf{y}^{k+1}) = -\mathbf{B}^T\boldsymbol{\lambda}^{k+1} \quad \text{and} \quad \nabla g(\mathbf{y}^*) = -\mathbf{B}^T\boldsymbol{\lambda}^*$."
8. Page 55, 3 lines below Algorithm 3.4, "keep L_g from $\psi(\mathbf{y}^*, \mathbf{y}^0)$ defined in (3.34))" should be "keep L_g from $D_\psi(\mathbf{y}^*, \mathbf{y}^0)$ defined in (3.34))"
9. Page 72, line 17, "
$$\eta_K \mathbf{s}^{k+1} + (1 - \eta_K)\mathbf{s}^K.$$
" should be "
$$\eta_K \mathbf{s}^{K+1} + (1 - \eta_K)\mathbf{s}^K.$$
".
10. Page 83, last line, "$2D_\psi(\mathbf{y}^*, y^0)$" should be "$2D_\psi(\mathbf{y}^*, \mathbf{y}^0)$".
11. Page 86, line 13, "$\widetilde{\mathbf{x}}_i^{k+1} = \operatorname{argmin}_{\mathbf{x}_i} L(\widetilde{\mathbf{x}}_1^{k+1}, \cdots, \widetilde{\mathbf{x}}_{i-1}^{k+1}, \mathbf{x}_i, \mathbf{x}_{i+1}^k \cdots, \mathbf{x}_m^k, \boldsymbol{\lambda}^k)$," should be "$\widetilde{\mathbf{x}}_i^{k+1} = \operatorname{argmin}_{\mathbf{x}_i} L(\widetilde{\mathbf{x}}_1^{k+1}, \cdots, \widetilde{\mathbf{x}}_{i-1}^{k+1}, \mathbf{x}_i, \mathbf{x}_{i+1}^k, \cdots, \mathbf{x}_m^k, \boldsymbol{\lambda}^k)$,"
12. Page 93, line 6, "
$$\tilde{L}_i(\widetilde{\mathbf{x}}_1^{k+1}, \cdots, \widetilde{\mathbf{x}}_{i-1}^{k+1}, \mathbf{x}_i, \boldsymbol{\xi}_{i+1}^k \cdots, \boldsymbol{\xi}_m^k, \boldsymbol{\lambda}^k),$$
" should be "
$$\tilde{L}_i(\widetilde{\mathbf{x}}_1^{k+1}, \cdots, \widetilde{\mathbf{x}}_{i-1}^{k+1}, \mathbf{x}_i, \boldsymbol{\xi}_{i+1}^k, \cdots, \boldsymbol{\xi}_m^k, \boldsymbol{\lambda}^k),$$
"

13. Page 93, line 14, add "We present the above iterations in Algorithm 3.10." after "$\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T, \cdots, \boldsymbol{\xi}_m^T\right)^T$."

14. Page 98, line 2,"

$$L(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}^*), \forall \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}.$$

" should be "

$$L(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}^*), \quad \forall \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}.$$

"

15. Page 98, line 2,"

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \cdots \\ \mathbf{x}_m \end{pmatrix}$$

" should be "

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{pmatrix}$$

"

16. Page 102, equation (3.94), "

$$\mathbf{M}^T\mathbf{H}^T + \mathbf{H}\mathbf{M} - \mathbf{M}^T\mathbf{H}\mathbf{M} \succeq \mathbf{0}, \quad \mathbf{P}^T\mathbf{H}\mathbf{P} \succeq \mathbf{0}, \quad \text{and} \quad \mathbf{P}^T\mathbf{H}\mathbf{P} \neq \mathbf{0}.$$

should be "

$$\mathbf{M}^T\mathbf{H}^T + \mathbf{H}\mathbf{M} - \mathbf{M}^T\mathbf{H}\mathbf{M} \succeq \mathbf{0} \quad \text{and} \quad \mathbf{H} \succeq \mathbf{0}.$$

"

17. Page 135, line 19, "[2,5]" should be "[5]"

18. Page 144, lines 19 & 21, "$F(\mathbf{x}_1, \xi)$" should be "$F(\mathbf{x}_1; \xi)$"

19. Page 148, line 13, "$F_1(\mathbf{x}_1, \xi)$" should be "$F_1(\mathbf{x}_1; \xi)$"

20. Page 167, line 2-3, "the convexity of $J_1(\cdot)$" should be "and the convexity of $h_1(\cdot)$ and $f_1(\cdot)$".

21. Page 168, line 16-17, "and the convexity of $J_2(\cdot)$" should be "and the convexity of $h_2(\cdot)$ and $f_2(\cdot)$".

22. Page 191, line 13, "$F(\mathbf{x}, \xi)$" should be "$F(\mathbf{x}; \xi)$"

23. Page 191, line 19, "$F(\mathbf{x}^k, \xi)$" should be "$F(\mathbf{x}^k; \xi)$"

24. Page 196, line 23, "$F_i(\mathbf{x}_i, \xi_i)$" should be "$F_i(\mathbf{x}_i; \xi_i)$"

25. Page 197, line 15, "$F_i(\mathbf{x}_i^k, \xi_i)$" should be "$F_i(\mathbf{x}_i^k; \xi_i)$"

26. Page 197, line 18, and Page 198, lines 5, 7, 9 , 11 and 15, "$F_i(\mathbf{x}_i^k, \xi_i)$" should be "$F_i(\mathbf{x}_i^k; \xi_i)$", "$F_i(\mathbf{x}_i^{k-1}, \xi_i)$" should be "$F_i(\mathbf{x}_i^{k-1}; \xi_i)$"

27. Page 198, line 17,"equality" should be "inequality".

28. Page 212, line 2 of Algorithm 6.6, "Initialize: $\mathbf{x}_i^0$, $\boldsymbol{\lambda}_i^0$, $i \in [m]$, and $\widetilde{\mathbf{x}}_i^0$." should be "Initialize: $\mathbf{x}_i^0$, $\boldsymbol{\lambda}_i^0$ and $\widetilde{\mathbf{x}}_i^0$, $i \in [m]$."

29. Page 218, above Theorem 6.5, "From Theorem 3.14 or 3.8" should be "Following the proof of Theorem 3.14 or 3.8"

30. Page 218, last line, "where $\sigma_{\mathbf{L}}$ is the smallest positive eigenvalue of $\mathbf{L}$." should be "where $\sigma_{\mathbf{L}} > 0$ is the second smallest eigenvalue of $\mathbf{L}$."

31. Page 223, line 11, "
$$\ell_{k+1} \leq O\left(1 - \sqrt{\frac{\mu\sigma_{\mathbf{L}}}{2Ld_{\max}}}\right)\ell_k.$$
" should be "
$$\ell_{k+1} \leq \left(1 - \sqrt{\frac{\mu\sigma_{\mathbf{L}}}{2Ld_{\max}}}\right)\ell_k.$$
"

32. Page 225, lines 8-9, "The squared Euclidean distance is obtained when $\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$, in which case $D_\phi(\mathbf{y}, \mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2$." should be "The squared Euclidean distance is obtained when $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$, in which case $D_\phi(\mathbf{y}, \mathbf{x}) = \|\mathbf{x} - \mathbf{y}\|^2$."

33. Page 242, line 19, change "a constant penalty parameter β has been used." to "a constant penalty parameter β has been used (some of the accelerated algorithms can be regarded as using a variable penalty)."

34. Page 251, line 6, "rank$\mathbf{A} = r$" should be "rank$(\mathbf{A}) = r$"

35. Page 256, line 11, "where C is the intersection of the domains of f and g." should be "where C is the intersection of the domains of f and $\mathbf{g}$."

36. Page 257, line 6, "$\mathbf{v}_i \geq \mathbf{0}$" should be "$\mathbf{v}_i \geq 0$"

37. Page 257, lines 24 and 26, "$\nabla_x \phi$" should be "$\nabla_{\mathbf{x}} \phi$"