Journal of Scientific Computing (2023) 96:48
https://doi.org/10.1007/s10915-023-02267-6

CORRECTION

Correction to: A Line Search Based Proximal Stochastic Gradient Algorithm with Dynamical Variance Reduction

Giorgia Franchini[1] · Federica Porta[1] · Valeria Ruggiero[2] · Ilaria Trombini[2,3]

Published online: 23 June 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
This file contains a revised version of the proofs of Theorems 3 and 4 of the paper [1]. In particular, a more correct argument is employed to obtain the inequality (A11) from (A10), provided that a stronger hypothesis on the sequence $\{\varepsilon_k\}$ is included. The practical implementation of the algorithm (Section 3) remains as it is and all the numerical experiments (Section 4) are still valid since the stronger hypothesis on $\{\varepsilon_k\}$ was already satisfied by the selected setting of the hyperparameters.

Correction to:
Journal of Scientific Computing 94:23
https://doi.org/10.1007/s10915-022-02084-3

We restated both the statement and the proof of Theorem 3. We stress that the proof only changes in obtaining the inequality (A11) from (A10), but for a better readability we report all the arguments of the proof.

Theorem 3 *Under the Assumptions 1 and 2, let* $\{x^{(k)}\}$ *be the sequence generated by the iteration (7) with* $\mathbb{E}(\|e_g^{(k)}\|^2|\mathcal{F}_k) \leq \varepsilon_k$ *where* $\{\varepsilon_k\}$ *is a nonnegative non-increasing sequence*

The original article can be found online at https://doi.org/10.1007/s10915-022-02084-3.

✉ Giorgia Franchini
giorgia.franchini@unimore.it

Federica Porta
federica.porta@unimore.it

Valeria Ruggiero
valeria.ruggiero@unife.it

Ilaria Trombini
ilaria.trombini@unife.it

1 Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, Via Campi, 213/B, Modena 41125, Italy

2 Department of Mathematics and Computer Science, University of Ferrara, Via Machiavelli, 30, Ferrara 44121, Italy

3 Department of Mathematical, Physical and Computer Sciences, University of Parma, Parco Area delle Scienze, 7/A, Parma 43124, Italy

such that $\sum_{k=0}^{+\infty}\sqrt{\varepsilon_k} < +\infty$ *and* $\alpha_k \in [\alpha_{min}, \alpha_{max}]$. *Moreover, assume that condition (9) holds and the function* F *is convex. Then the sequence* $\{x^{(k)}\}$ *converges to a solution of (1) a.s.*

Proof Let $x^* \in X^*$. Since $\dfrac{x^{(k)} - x^{(k+1)}}{\alpha_k} - g^{(k)} \in \partial R(x^{(k+1)})$, it holds that

$$R(y) \geq R(x^{(k+1)}) + \frac{1}{\alpha_k}(x^{(k)} - x^{(k+1)} - \alpha_k g^{(k)})^T(y - x^{(k+1)}), \quad \forall y \in \mathbb{R}^d.$$

It follows that, $\forall y \in \mathbb{R}^d$,

$$\begin{aligned}\alpha_k R(y) &\geq \alpha_k R(x^{(k+1)}) + (x^{(k)} - x^{(k+1)} - \alpha_k g^{(k)})^T(y - x^{(k+1)})\\ &= \alpha_k R(x^{(k+1)}) + (x^{(k)} - x^{(k+1)})^T(y - x^{(k+1)}) - \alpha_k g^{(k)^T}(y - x^{(k+1)}),\end{aligned}$$

and, hence, the following inequality holds

$$(x^{(k+1)} - x^{(k)})^T(y - x^{(k+1)}) \geq \alpha_k \left(R(x^{(k+1)}) - R(y) + g^{(k)^T}(x^{(k+1)} - y)\right). \tag{A8}$$

For $y = x^*$ the previous inequality gives

$$\begin{aligned}&(x^{(k+1)} - x^{(k)})^T(x^* - x^{(k)} + x^{(k)} - x^{(k+1)}) \geq\\ &\geq \alpha_k \left(R(x^{(k+1)}) - R(x^*) + g^{(k)^T}(x^{(k+1)} - x^{(k)} + x^{(k)} - x^*)\right).\end{aligned}$$

As a consequence, we obtain the following relations:

$$\begin{aligned}(x^{(k+1)} - x^{(k)})^T(x^* - x^{(k)}) &\geq \alpha_k \left(R(x^{(k+1)}) - R(x^*) + g^{(k)^T}(x^{(k)} - x^*)\right) +\\ &\quad - (x^{(k+1)} - x^{(k)})^T(x^{(k)} - x^{(k+1)}) + \alpha_k g^{(k)^T}(x^{(k+1)} - x^{(k)})\\ &= \alpha_k \left(R(x^{(k+1)}) - R(x^*) + (\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k)} - x^*)\right) +\\ &\quad + (x^{(k+1)} - x^{(k)})^T(x^{(k+1)} - x^{(k)}) + \alpha_k(\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k+1)} - x^{(k)})\\ &\geq \alpha_k \left(R(x^{(k+1)}) - R(x^*) + F(x^{(k)}) - F(x^*)\right) + \alpha_k e_g^{(k)^T}(x^{(k)} - x^*) +\\ &\quad + \|x^{(k+1)} - x^{(k)}\|^2 + \alpha_k(\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k+1)} - x^{(k)})\\ &= \alpha_k \left(R(x^{(k+1)}) + R(x^{(k)}) - R(x^{(k)}) + F(x^{(k)}) - P(x^*)\right) +\\ &\quad + \|x^{(k+1)} - x^{(k)}\|^2 + \alpha_k e_g^{(k)^T}(x^{(k)} - x^*) +\\ &\quad + \alpha_k(\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k+1)} - x^{(k)})\\ &= \alpha_k \left(R(x^{(k+1)}) - R(x^{(k)}) + P(x^{(k)}) - P(x^*)\right) + \|x^{(k+1)} - x^{(k)}\|^2 +\\ &\quad + \alpha_k e_g^{(k)^T}(x^{(k)} - x^*) + \alpha_k(\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k+1)} - x^{(k)})\\ &\geq \alpha_k \left(R(x^{(k+1)}) - R(x^{(k)})\right) + \|x^{(k+1)} - x^{(k)}\|^2 + \alpha_k e_g^{(k)^T}(x^{(k)} - x^*) +\\ &\quad + \alpha_k(\nabla F(x^{(k)}) + e_g^{(k)})^T(x^{(k+1)} - x^{(k)}),\end{aligned} \tag{A9}$$

where the second inequality follows from the convexity of F and the last inequality follows from the fact that $P(x^{(k)}) - P(x^*) \geq 0$. From a basic property of the Euclidean norm[1] we

[1] $\|a - b\|^2 + \|b - c\|^2 - \|a - c\|^2 = 2(a - b)^T(c - b), \quad \forall a, b, c \in \mathbb{R}^d.$

can write

$$
\begin{aligned}
\|x^{(k+1)}-x^*\|^2 &= \|x^{(k+1)}-x^{(k)}\|^2+\|x^{(k)}-x^*\|^2-2(x^{(k+1)}-x^{(k)})^T(x^*-x^{(k)}) \\
&\overset{(A9)}{\leq} \|x^{(k+1)}-x^{(k)}\|^2+\|x^{(k)}-x^*\|^2-2\alpha_k\left(R(x^{(k+1)})-R(x^{(k)})\right)+ \\
&\quad -2\|x^{(k+1)}-x^{(k)}\|^2+ \\
&\quad -2\alpha_k e_g^{(k)^T}(x^{(k)}-x^*)-2\alpha_k(\nabla F(x^{(k)})+e_g^{(k)})^T(x^{(k+1)}-x^{(k)}) \\
&= \|x^{(k)}-x^*\|^2-\|x^{(k+1)}-x^{(k)}\|^2-2\alpha_k\left(R(x^{(k+1)})-R(x^{(k)})\right)+ \\
&\quad -2\alpha_k\nabla F(x^{(k)})^T(x^{(k+1)}-x^{(k)})-2\alpha_k e_g^{(k)^T}(x^{(k)}-x^*)+ \\
&\quad -2\alpha_k e_g^{(k)^T}(x^{(k+1)}-x^{(k)}) \\
&= \|x^{(k)}-x^*\|^2-2\alpha_k e_g^{(k)^T}(x^{(k)}-x^*)-2\alpha_k e_g^{(k)^T}(x^{(k+1)}-x^{(k)})+ \\
&\quad -2\alpha_k\Big(R(x^{(k+1)})-R(x^{(k)})+\nabla F(x^{(k)})^T(x^{(k+1)}-x^{(k)})+ \\
&\quad +\frac{1}{2\alpha_k}\|x^{(k+1)}-x^{(k)}\|^2\Big) \\
&= \|x^{(k)}-x^*\|^2-2\alpha_k\left(h_{\alpha_k}(x^{(k+1)};x^{(k)})+e_g^{(k)^T}(x^{(k+1)}-x^{(k)})\right)+ \\
&\quad -2\alpha_k e_g^{(k)^T}(x^{(k)}-x^*) \\
&\leq \|x^{(k)}-x^*\|^2-2\alpha_{max}\left(h_{\alpha_k}(x^{(k+1)};x^{(k)})+e_g^{(k)^T}(x^{(k+1)}-x^{(k)})\right)+ \\
&\quad -2\alpha_k e_g^{(k)^T}(x^{(k)}-x^*).
\end{aligned}
$$

In view of the hypotheses on the sequence $\{\varepsilon_k\}$, it follows that

$$
\begin{aligned}
\|x^{(k+1)}-x^*\|^2 &\leq \|x^{(k)}-x^*\|^2-2\alpha_{max}\left(h_{\alpha_k}(x^{(k+1)};x^{(k)})+e_g^{(k)^T}(x^{(k+1)}-x^{(k)})\right)+ \\
&\quad +\alpha_k\frac{\|e_g^{(k)}\|^2}{\sqrt{\varepsilon_k}}+\alpha_k\sqrt{\varepsilon_k}\|x^{(k)}-x^*\|^2 \\
&\leq \|x^{(k)}-x^*\|^2-2\alpha_{max}\left(h_{\alpha_k}(x^{(k+1)};x^{(k)})+e_g^{(k)^T}(x^{(k+1)}-x^{(k)})\right)+ \\
&\quad +\alpha_{max}\frac{\|e_g^{(k)}\|^2}{\sqrt{\varepsilon_k}}+\alpha_{max}\sqrt{\varepsilon_k}\|x^{(k)}-x^*\|^2. \qquad \text{(A10)}
\end{aligned}
$$

Taking the conditional expectation in (A10) with respect to the σ-algebra $\mathcal{F}_k$, we obtain

$$
\begin{aligned}
\mathbb{E}\left(\|x^{(k+1)}-x^*\|^2|\mathcal{F}_k\right) &\leq (1+\alpha_{max}\sqrt{\varepsilon_k})\|x^{(k)}-x^*\|^2+ \\
&-2\alpha_{max}\mathbb{E}\left(h_{\alpha_k}(x^{(k+1)};x^{(k)})+e_g^{(k)^T}(x^{(k+1)}-x^{(k)})|\mathcal{F}_k\right)+\alpha_{max}\sqrt{\varepsilon_k} \qquad \text{(A11)}
\end{aligned}
$$

By combining (A11) and part *i)* of Theorem 1 together with Lemma 3, we can state that the sequence $\{\|x^{(k)}-x^*\|\}_{k\in\mathbb{N}}$ converges a.s.

Next we prove the almost sure convergence of the sequence $\{x^{(k)}\}$ by following a strategy similar to the one employed in [2, Theorem 2.1]. Let $\{x_i^*\}_i$ be a countable subset of the relative interior $\mathrm{ri}(X^*)$ that is dense in X^*. From the almost sure convergence of $\|x^{(k)}-x^*\|$, $x^*\in X^*$, we have that for each i, the probability $\mathrm{Prob}(\{\|x^{(k)}-x_i^*\|\}$ is not convergent$)=0$.

Therefore, we observe that

$$\text{Prob}(\forall i\ \exists b_i \text{ s.t. } \lim_{k\to+\infty}\|x^{(k)}-x_i^*\|=b_i)=1-\text{Prob}(\{\|x^{(k)}-x_i^*\|\} \text{ is not convergent})$$
$$\geq 1-\sum_i \text{Prob}(\{\|x^{(k)}-x_i^*\|\} \text{ is not convergent})=1,$$

where the inequality follows from the union bound, i.e. for each i, $\{\|x^{(k)}-x_i^*\|\}$ is a convergent sequence a.s. For a contradiction, suppose that there are convergent subsequences $\{u_{k_j}\}_{k_j}$ and $\{v_{k_j}\}_{k_j}$ of $\{x^{(k)}\}$ which converge to their limiting points u^* and v^* respectively, with $\|u^*-v^*\|=r>0$. By Theorem 2, u^* and v^* are stationary; in particular, since P is convex, they are minimum points, i.e. $u^*, v^*\in X^*$. Since $\{x_i^*\}_i$ is dense in X^*, we may assume that for all $\epsilon>0$, we have $x_{i_1}^*$ and $x_{i_2}^*$ are such that $\|x_{i_1}^*-u^*\|<\epsilon$ and $\|x_{i_2}^*-v^*\|<\epsilon$. Therefore, for all k_j sufficiently large,

$$\|u_{k_j}-x_{i_1}^*\|\leq\|u_{k_j}-u^*\|+\|u^*-x_{i_1}^*\|<\|u_{k_j}-u^*\|+\epsilon.$$

On the other hand, for sufficiently large j, we have

$$\|v_{k_j}-x_{i_1}^*\|\geq\|v^*-u^*\|-\|u^*-x_{i_1}^*\|-\|v_{k_j}-v^*\|>r-\epsilon-\|v_{k_j}-v^*\|>r-2\epsilon.$$

This contradicts with the fact that $x^{(k)}-x_{i_1}^*$ is convergent. Therefore, we must have $u^*=v^*$, hence there exists $\bar{x}\in X^*$ such that $x^{(k)}\longrightarrow\bar{x}$. □

The same stronger assumption on the sequence $\{\varepsilon_k\}$ included in Theorem 3 must be added also in Theorem 4. The proof of this theorem is revised since its starting point is inequality (A11) which has been modified. We remark that the main arguments of the proof do not change: indeed it is only different how to obtain (A13) from (A12).

Theorem 4 *Under the Assumptions 1 and 2, let $\{x^{(k)}\}$ be the sequence generated by the iteration (7) with $\mathbb{E}(\|e_g^{(k)}\|^2|\mathcal{F}_k)\leq\varepsilon_k$ where $\{\varepsilon_k\}$ is a nonnegative non-increasing sequence such that $\sum_{k=0}^{+\infty}\sqrt{\varepsilon_k}<+\infty$ and $\alpha_k\in[\alpha_{min},\alpha_{max}]$. Moreover, assume that condition (9) holds, the function F is convex. Then, by denoting $\overline{x}^{(K)}=\frac{1}{K+1}\sum_{k=0}^{K}x^{(k)}$, we have*

$$\mathbb{E}(P(\overline{x}^{(K)})-P(x^*))=\mathcal{O}\left(\frac{1}{K}\right). \tag{10}$$

Furthermore, when $\mathbb{E}\left(\sum_{k=0}^{\infty}k\eta_k\right)<\infty$, we have

$$\mathbb{E}(P(x^{(k)})-P(x^*))=\mathcal{O}\left(\frac{1}{k}\right). \tag{11}$$

Proof If we do not neglect the term $P(x^{(k)})-P(x^*)$ in (A9) and in all the subsequent inequalities, instead of (A11) we obtain

$$\begin{aligned}\mathbb{E}\left(\|x^{(k+1)}-x^*\|^2|\mathcal{F}_k\right)\leq&(1+\alpha_{max}\sqrt{\varepsilon_k})\|x^{(k)}-x^*\|^2+\\&+2\alpha_{max}\mathbb{E}\left(-h_{\alpha_k}(x^{(k+1)};x^{(k)})-{e_g^{(k)}}^T(x^{(k+1)}-x^{(k)})|\mathcal{F}_k\right)+\\&-2\alpha_{min}\mathbb{E}\left(P(x^{(k)})-P(x^*)|\mathcal{F}_k\right)+\alpha_{max}\sqrt{\varepsilon_k}.\end{aligned} \tag{A12}$$

Summing the previous inequality from 0 to K and taking the total expectation, we obtain

$$\begin{aligned}\sum_{k=0}^{K}\mathbb{E}\left(P(x^{(k)})-P(x^*)\right) &\le \frac{1}{2\alpha_{min}}\left(\|x^{(0)}-x^*\|^2-\mathbb{E}(\|x^{(K+1)}-x^*\|^2)\right)+\\&+\frac{\alpha_{max}}{\alpha_{min}}\mathbb{E}\left(\sum_{k=0}^{K}\mathbb{E}\left(-h_{\alpha_k}(x^{(k+1)};x^{(k)})-e_g^{(k)T}(x^{(k+1)}-x^{(k)})|\mathcal{F}_k\right)\right)\\&+\frac{\alpha_{max}}{2\alpha_{min}}\mathbb{E}\left(\sum_{k=0}^{K}\sqrt{\varepsilon_k}\|x^{(k)}-x^*\|^2\right)+\alpha_{max}\sum_{k=0}^{K}\sqrt{\varepsilon_k}\\&\le\frac{1}{2\alpha_{min}}\|x^{(0)}-x^*\|^2+\frac{\alpha_{max}}{\alpha_{min}}\sum_{k=0}^{K}\mathbb{E}\left(-h_{\alpha_k}(x^{(k+1)};x^{(k)})-e_g^{(k)T}(x^{(k+1)}-x^{(k)})\right)+\\&+\frac{\alpha_{max}}{2\alpha_{min}}\sum_{k=0}^{K}\sqrt{\varepsilon_k}\mathbb{E}(\|x^{(k)}-x^*\|^2)+\alpha_{max}\sum_{k=0}^{K}\sqrt{\varepsilon_k}\end{aligned}$$

where the second inequality follows by neglecting the term $-\mathbb{E}(\|x^{(K+1)}-x^*\|^2)$. Now, we observe that, at the end of the proof of Theorem 1, we prove that

$$\sum_{k}\mathbb{E}\left(-h_{\alpha_k}(x^{(k+1)};x^{(k)})-e_g^{(k)T}(x^{(k+1)}-x^{(k)})\right)<+\infty. \tag{*}$$

As a consequence,

$$\sum_{k=0}^{K}\mathbb{E}\left(-h_{\alpha_k}(x^{(k+1)};x^{(k)})-e_g^{(k)T}(x^{(k+1)}-x^{(k)})\right)\le S.$$

Moreover, by considering the total expectation in (A11), Lemma 2 in Section 2.2.1 of [3] together with (*) allows to state that $\mathbb{E}(\|x^{(k)}-x^*\|^2)$ converges and thus there exists M such that $\mathbb{E}(\|x^{(k)}-x^*\|^2)<M$. We can write

$$\begin{aligned}\sum_{k=0}^{K}\mathbb{E}\left(P(x^{(k)})-P(x^*)\right)\le\frac{1}{2\alpha_{min}}\|x^{(0)}-x^*\|^2+\frac{\alpha_{max}}{\alpha_{min}}S+\frac{\alpha_{max}}{2\alpha_{min}}\bar{\varepsilon}M+\\+\alpha_{max}\bar{\varepsilon}\end{aligned} \tag{A13}$$

where we set $\bar{\varepsilon}=\sum_{k=0}^{+\infty}\sqrt{\varepsilon_k}$. Setting $\overline{x}^{(K)}=\frac{1}{K+1}\sum_{k=0}^{K}x^{(k)}$, from the Jensen's inequality, we observe that $\mathbb{E}(P(\overline{x}^{(K)}))\le\frac{1}{K+1}\sum_{k=0}^{K}\mathbb{E}(P(x^{(k)}))$. Thus, by dividing (A13) by $K+1$, we can write

$$\begin{aligned}\mathbb{E}\left(P(\overline{x}^{(K)})-P(x^*)\right)\le\frac{1}{K+1}\left(\frac{1}{2\alpha_{min}}\|x^{(0)}-x^*\|^2+\frac{\alpha_{max}}{\alpha_{min}}S\right)+\\+\frac{1}{K+1}\left(\frac{\alpha_{max}}{2\alpha_{min}}\bar{\varepsilon}M+\alpha_{max}\bar{\varepsilon}\right)\end{aligned} \tag{A14}$$

Thus, we obtain the $\mathcal{O}(1/K)$ ergodic convergence rate of $\mathbb{E}\left(P(\overline{x}^{(K)})-P(x^*)\right)$.
Now, we assume $\sum_{k=0}^{\infty}k\eta_k=\Sigma$. In (A13) the term $\sum_{k=0}^{K}\mathbb{E}\left(P(x^{(k)})-P(x^*)\right)$ is equal to $\mathbb{E}\left(\sum_{k=0}^{K}P(x^{(k)})\right)-(K+1)P(x^*)$. We observe that, since $0\le P(x^{(0)})-P(x^*)$, we can

write

$$\mathbb{E}\left(\sum_{k=1}^{K} P(x^{(k)})\right) - KP(x^*) \leq \mathbb{E}\left(\sum_{k=0}^{K} P(x^{(k)})\right) - (K+1)P(x^*)$$
$$\leq \frac{1}{2\alpha_{min}}\|x^{(0)} - x^*\|^2 + \frac{\alpha_{max}}{\alpha_{min}}S +$$
$$+ \frac{\alpha_{max}}{2\alpha_{min}}\bar{\varepsilon}M + \alpha_{max}\bar{\varepsilon}.$$

Now we determine a lower bound for $\mathbb{E}\left(\sum_{k=1}^{K} P(x^{(k)})\right)$. From the inequality (8), we have that $\mathbb{E}\left(P(x^{(k)}) - P(x^{(k+1)})|\mathcal{F}_k\right) + \eta_k \geq 0$ and, hence, by considering the total expectation we obtain $\mathbb{E}\left(P(x^{(k)}) - P(x^{(k+1)})\right) + \mathbb{E}(\eta_k) \geq 0$. Thus, we have

$$0 \leq \sum_{k=1}^{K} k\mathbb{E}\left(P(x^{(k)}) - P(x^{(k+1)})\right) + \sum_{k=1}^{K} k\mathbb{E}(\eta_k)$$
$$= \sum_{k=1}^{K} \mathbb{E}(P(x^{(k)})) - K\mathbb{E}(P(x^{(K+1)})) + \mathbb{E}\left(\sum_{k=1}^{K} k\eta_k\right). \tag{A15}$$

Then, we can write

$$K\mathbb{E}(P(x^{(K+1)})) - \Sigma \leq \sum_{k=1}^{K} \mathbb{E}\left(P(x^{(k)})\right). \tag{A16}$$

Consequently, we can conclude that

$$\mathbb{E}(P(x^{(K+1)}) - P(x^*)) \leq \frac{1}{K}\left(\frac{1}{2\alpha_{min}}\|x^{(0)} - x^*\|^2 + \frac{\alpha_{max}}{\alpha_{min}}S\right) +$$
$$+ \frac{1}{K}\left(\frac{\alpha_{max}}{2\alpha_{min}}\bar{\varepsilon}M + \alpha_{max}\bar{\varepsilon} + \Sigma\right).$$

□

References

1. Franchini, G., Porta, F., Ruggiero, V., Trombini, I.: A line search based proximal stochastic gradient algorithm with dynamical variance reduction. J. Sci. Comput. **94**, 23 (2023)
2. Poon, C., Liang, J., Schoenlieb, C.: Local convergence properties of SAGA/Prox-SVRG and acceleration, PMLR. In: Proceedings of the 35th International Conference on Machine Learning, Vol 80, pp. 4124–4132 (2018)
3. Polyak, B.T.: Introduction to Optimization. Optimization Software, New York (1987)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.