Supplementary material for Quasi-convex Feasibility Problems: Subgradient Methods and Convergence Rates

Yaohua Hu[a], Gongnong Li[a], Carisa Kwok Wai Yu[b,*], Tsz Leung Yip[c]

[a] *Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China*

[b] *Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Shatin, Hong Kong*

[c] *Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

A. Preliminary lemmas

We first recall an averaging scheme from (Kiwiel, 2004, Lemma 2.1) and a supermartingale convergence theorem from (Bertsekas & Tsitsiklis, 1996, pp. 148), which are useful in convergence analysis of subgradient methods.

Lemma A.1. *Let $\{a_k\}$ be a sequence of scalars, and let $\{v_k\}$ be a sequence of nonnegative scalars. Suppose that $\lim_{k\to\infty}\sum_{i=1}^k v_i = \infty$. Then it holds that*

$$\liminf_{k\to\infty} a_k \le \liminf_{k\to\infty} \frac{\sum_{i=1}^k v_i a_i}{\sum_{i=1}^k v_i} \le \limsup_{k\to\infty} \frac{\sum_{i=1}^k v_i a_i}{\sum_{i=1}^k v_i} \le \limsup_{k\to\infty} a_k.$$

Theorem A.1. *Let $\{Y_k\}$, $\{Z_k\}$ and $\{W_k\}$ be three sequences of random variables, and let $\{\mathcal{F}_k\}$ be a sequence of sets of random variables such that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for each $k \in \mathbb{N}$. Suppose for each $k \in \mathbb{N}$ that*

(a) *Y_k, Z_k and W_k are functions of nonnegative random variables in $\mathcal{F}_k$;*

(b) *$\mathbb{E}\{Y_{k+1} \mid \mathcal{F}_k\} \le Y_k - Z_k + W_k$;*

(c) *$\sum_{k=1}^\infty W_k < \infty$.*

Then $\sum_{k=1}^\infty Z_k < \infty$ and $\{Y_k\}$ converges to a random variable with probability 1.

*Corresponding author

Email addresses: mayhhu@szu.edu.cn (Yaohua Hu), gnli@szu.edu.cn (Gongnong Li), carisayu@hsu.edu.hk (Carisa Kwok Wai Yu), t.l.yip@polyu.edu.hk (Tsz Leung Yip)

B. Proof of Theorem 3.1

Proof. (i) By the basic inequality (3.9) in Lemma 3.2, Lemma A.1 is applicable (with $(F^+(x_j))^{\frac{1}{\beta}}$ and 1 in place of a_j and v_j) to concluding that

$$\begin{aligned}\liminf_{k\to\infty} \left(F^+(x_k)\right)^{\frac{1}{\beta}} &\le \liminf_{k\to\infty} \frac{\sum_{j=1}^k (F^+(x_j))^{\frac{1}{\beta}}}{k} \\ &\le \liminf_{k\to\infty} \frac{1}{2v\mu}\left(\frac{L}{\alpha}\right)^{\frac{1}{\beta}}\left(\frac{\|x_1-x\|^2}{k}+v^2\right) \\ &= \frac{v}{2\mu}\left(\frac{L}{\alpha}\right)^{\frac{1}{\beta}}.\end{aligned}$$

Consequently, the conclusion (i) is obtained.

(ii) Proving by contradiction, we assume that $F^+(x_k) > \frac{L}{\alpha}\left(\frac{v}{2\mu}+\delta\right)^{\beta}$ for each $1 \le k \le K_{\mathrm{m}}^{\mathrm{c}}$. Then it follows from (3.9) in Lemma 3.2 (taking $x := \mathrm{P}_S(x_k)$) that

$$\mathrm{d}^2(x_{k+1}, S) < \mathrm{d}^2(x_k, S) - 2v\mu\delta \quad \text{for each } 1 \le k \le K_{\mathrm{m}}^{\mathrm{c}}.$$

Summing the above inequality over $k = 1, \ldots, K_{\mathrm{m}}^{\mathrm{c}}$, we derive that

$$0 \le \mathrm{d}^2(x_{K_{\mathrm{m}}^{\mathrm{c}}+1}, S) < \mathrm{d}^2(x_1, S) - 2K_{\mathrm{m}}^{\mathrm{c}} v\mu\delta,$$

which contradicts with the definition of $K_{\mathrm{m}}^{\mathrm{c}}$. The proof is complete. □

C. Proof of Theorem 3.3

Proof. (i) By the basic inequality (3.16) in Lemma 3.3, Lemma A.1 is applicable (with $(F^+(x_{sj}))^{\frac{1}{\beta}}$ and 1 in place of a_j and v_j) to concluding that

$$\begin{aligned}\liminf_{k\to\infty} \left(F^+(x_{sk})\right)^{\frac{1}{\beta}} &\le \liminf_{k\to\infty} \frac{\sum_{j=1}^k (F^+(x_{sj}))^{\frac{1}{\beta}}}{k} \\ &\le \liminf_{k\to\infty} \frac{1}{4v\mu}(2L)^{\frac{1}{\beta}}\left(\frac{\|x_s-x\|^2}{k}+sv^2(1+2\mu)\right) \\ &= \frac{sv(1+2\mu)}{4\mu}(2L)^{\frac{1}{\beta}}.\end{aligned}$$

Consequently, Theorem 3.3(i) is obtained.

(ii) Proving by contradiction, we assume that $F^+(x_k) > 2L\left(\frac{sv(1+2\mu)}{4\mu}+\delta\right)^{\beta}$ for each $1 \le k \le K_{\mathrm{c}}^{\mathrm{c}}$. Then it follows from (3.16) (taking $x := \mathrm{P}_S(x_{sk})$) that

$$\mathrm{d}^2(x_{s(k+1)}, S) < \mathrm{d}^2(x_{sk}, S) - 4v\mu\delta \quad \text{for each } 1 \le k \le \frac{K_{\mathrm{c}}^{\mathrm{c}}}{s}.$$

Summing the above inequality over $k = 1, \ldots, \frac{K_{\mathrm{c}}^{\mathrm{c}}}{s}$, we derive that

$$0 \le \mathrm{d}^2(x_{K_{\mathrm{c}}^{\mathrm{c}}+s}, S) < \mathrm{d}^2(x_s, S) - 4\frac{K_{\mathrm{c}}^{\mathrm{c}}}{s} v\mu\delta,$$

which contradicts with the definition of $K_{\mathrm{c}}^{\mathrm{c}}$. The proof is complete. □

D. Proof of Theorem 3.4

Proof. Firstly, we claim that

$$\mathrm{d}^2(x_{s(k+1)}, S) \leq \mathrm{d}^2(x_{sk}, S) - 4v\mu(2\kappa L)^{-\frac{1}{\beta}}\mathrm{d}^{\frac{q}{\beta}}(x_{sk}, S) + sv^2(1+2\mu) \tag{D.1}$$

for each $k \in \mathbb{N}$. To this end, recall from (3.16) (taking $x := \mathrm{P}_S(x_{sk})$) that

$$\mathrm{d}^2(x_{s(k+1)}, S) \leq \mathrm{d}^2(x_{sk}, S) - 4v\mu(2L)^{-\frac{1}{\beta}}\left(F^+(x_{sk})\right)^{\frac{1}{\beta}} + sv^2(1+2\mu). \tag{D.2}$$

By the assumption of the Hölder-type error bound property, (3.10) holds for each x_{sk}. This, together with (D.2), yields (D.1), as desired.

(i) Suppose that $q = 2\beta$. Setting $\tau := (1 - 4v\mu(2\kappa L)^{-\frac{1}{\beta}})_+ \in [0, 1)$ and by (D.1), we achieve that

$$\mathrm{d}^2(x_{s(k+1)}, S) \leq \tau \mathrm{d}^2(x_{sk}, S) + sv^2(1+2\mu) \quad \text{for each } k \in \mathbb{N}.$$

Then we inductively obtain the conclusion (i).

(ii) Suppose that $q > 2\beta$ and $v < (s(1+2\mu))^{\frac{q-2\beta}{2(\beta-q)}}(4\mu)^{\frac{\beta}{\beta-q}}(2\kappa L)^{\frac{1}{q-\beta}}\left(\frac{2\beta}{q}\right)^{\frac{q}{2(q-\beta)}}$. Then by (D.1), Lemma 2.3(ii) is applicable (with $\mathrm{d}^2(x_{sk}, S)$, $\frac{q}{2\beta} - 1$, $4v\mu(2\kappa L)^{-\frac{1}{\beta}}$, $sv^2(1+2\mu)$ in place of u_k, r, a, b) to obtaining the conclusion (ii). The proof is complete. □

E. Proof of Theorem 3.5

Proof. (i) Fix $\delta > 0$, and define a set $S_\delta \subseteq \mathbb{R}^n$ by

$$S_\delta := \{x \in X : F^+(x) < L\left(\frac{mv}{2} + \delta\right)^\beta, \forall i \in I\}.$$

Let $y_\delta \in X$ be such that $F^+(y_\delta) = \delta^\beta$ (y_δ is well-defined by the consistency of the QFP and the continuity of each f_i); hence $y_\delta \in S_\delta$ by construction. We define a new process $\{\hat{x}_k\}$ by letting $\hat{x}_0 := x_0$ and

$$\hat{x}_{k+1} := \begin{cases} \mathrm{P}_X\left(\hat{x}_k - v\sum_{i\in\{\hat{\omega}_k\}\cap I(\hat{x}_k)} \hat{g}_{k,i}\right), & \text{if } \hat{x}_k \notin S_\delta, \\ y_\delta, & \text{otherwise,} \end{cases}$$

where $\hat{g}_{k,i} \in \partial^* f_i(\hat{x}_k) \cap \mathbb{S}$. By comparing the above process with Algorithm 3.1, $\{\hat{x}_k\}$ is identical to $\{x_k\}$, except that $\hat{x}_k$ enters S_δ and then the process terminates with $\hat{x}_k = y_\delta \in S_\delta$.

Assume that $\hat{x}_k \notin S_\delta$ for each $k \in \mathbb{N}$, and let $\hat{\mathcal{F}}_k := \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_k\}$ for each $k \in \mathbb{N}$. It says that $F^+(\hat{x}_k) \geq L\left(\frac{mv}{2} + \delta\right)^\beta$, and then follows from (3.21) in Lemma 3.4 that the following relation holds for any $x \in S$ and $k \in \mathbb{N}$:

$$\mathbb{E}\left\{\|\hat{x}_{k+1} - x\|^2 \mid \hat{\mathcal{F}}_k\right\} \leq \|\hat{x}_k - x\|^2 - \frac{2v}{m}\delta.$$

Then it follows from Theorem A.1 (applied to $\|\hat{x}_k - x\|^2$, $\frac{2v}{m}\delta$, 0 in place of Y_k, Z_k, W_k) that $\sum_{k=0}^\infty \frac{2v}{m}\delta < \infty$ with probability 1, that is impossible. Hence $\hat{x}_k \in S_\delta$ must occur for infinitely many times; consequently, in the original process, it holds with probability 1 that $\liminf_{k\to\infty} F^+(x_k) \leq L\left(\frac{mv}{2} + \delta\right)^\beta$. Then the conclusion (i) is achieved by letting δ tend to 0.

(ii) Proving by contradiction, we assume that $\mathbb{E}\{F^+(x_k)\} > L\left(\frac{mv}{2}+\delta\right)^\beta$ for each $1 \le k \le K_{\rm s}^{\rm c}$. Consequently, $\mathbb{E}\left\{(F^+(x_k))^{\frac{1}{\beta}}\right\} > L^{\frac{1}{\beta}}\left(\frac{mv}{2}+\delta\right)$ by the convexity of $t^{\frac{1}{\beta}}$ on $\mathbb{R}_+$ (as $\beta \le 1$). Taking the expectation on (3.21) (with $x := {\rm P}_S(x_k)$), one has for each $k = 1, \dots, K_{\rm s}^{\rm c}$ that

$$\begin{aligned}\mathbb{E}\left\{{\rm d}^2(x_{k+1}, S)\right\} &\le \mathbb{E}\left\{{\rm d}^2(x_k, S)\right\} - \frac{2v}{m}L^{-\frac{1}{\beta}}\mathbb{E}\left\{\left(F^+(x_k)\right)^{\frac{1}{\beta}}\right\} + v^2 \\ &\le \mathbb{E}\left\{{\rm d}^2(x_k, S)\right\} - \frac{2v}{m}\delta.\end{aligned}$$

Summing the above inequality over $k = 1, \dots, K_{\rm s}^{\rm c}$, we derive that

$$0 \le \mathbb{E}\left\{{\rm d}^2(x_{k+1}, S)\right\} < {\rm d}^2(x_1, S) - K_{\rm s}^{\rm c}\frac{2v}{m}\delta,$$

which contradicts with the definition of $K_{\rm s}^{\rm c}$. The proof is complete. $\square$

F. Proof of Theorem 3.7

Proof. (i) Note by Lemma 3.5(ii) that $\{x_k\}$ is bounded, and thus, must have a cluster point, denoted by $\bar{x}$. It follows from (3.32) in Lemma 3.6 that

$$\sum_{k=1}^{\infty}\left(F^+(x_k)\right)^{\frac{2}{\beta}} \le \frac{1}{\underline{v}(2-\overline{v})\mu}\left(\frac{L}{\alpha}\right)^{\frac{2}{\beta}}{\rm d}^2(x_1, S) < \infty.$$

Consequently, $\lim_{k\to\infty} F^+(x_k) = 0$, which shows that its cluster point $\bar{x} \in S$ by the continuity of each f_i. Recall from Lemma 3.5(ii) that $\{\|x_k - \bar{x}\|\}$ is monotonically decreasing. Hence $\{x_k\}$ converges to this $\bar{x} \in S$, as desired.

(ii) Proving by contradiction, we assume that $F^+(x_k) > \delta$ for each $1 \le k \le K_{\rm m}^{\rm d}$. Then it follows from (3.32) that, for each $1 \le k \le K_{\rm m}^{\rm d}$,

$${\rm d}^2(x_{k+1}, S) < {\rm d}^2(x_k, S) - \underline{v}(2-\overline{v})\mu\left(\frac{\alpha\delta}{L}\right)^{\frac{2}{\beta}}.$$

Summing the above inequality over $k = 1, \dots, K_{\rm m}^{\rm d}$, we derive that

$$0 \le {\rm d}^2(x_{K_{\rm m}^{\rm d}+1}, S) < {\rm d}^2(x_1, S) - K_{\rm m}^{\rm d}\underline{v}(2-\overline{v})\mu\left(\frac{\alpha\delta}{L}\right)^{\frac{2}{\beta}},$$

which contradicts with the definition of $K_{\rm m}^{\rm d}$. The proof is complete. $\square$

G. Proof of Theorem 3.9

Proof. (i) Note by Lemma 3.5(ii) that $\{x_k\}$ is bounded, and thus, must have a cluster point, denoted by $\bar{x}$. It follows from (3.36) that

$$\sum_{k=1}^{\infty}\left(F^+(x_{sk})\right)^{\frac{2}{\beta}} \le \frac{1+4s}{2\underline{v}(2-\overline{v})\mu}(2L)^{\frac{2}{\beta}}{\rm d}^2(x_s, S) < \infty.$$

This shows that $\lim_{k\to\infty} F^+(x_{sk}) = 0$, and thus, the cluster point $\bar{x}$ of $\{x_{sk}\}$ falls in S by the continuity of each f_i. This, together with Lemma 3.5(ii), shows that $\{x_k\}$ converges to $\bar{x} \in S$.

(ii) Proving by contradiction, we assume that $F^+(x_k) > \delta$ for each $1 \le k \le K_{\rm c}^{\rm d}$. Then it follows from (3.36) that, for each $1 \le k \le \frac{K_{\rm c}^{\rm d}}{s}$,

$$\mathrm{d}^2(x_{s(k+1)}, S) < \mathrm{d}^2(x_{sk}, S) - \frac{2\underline{v}(2-\overline{v})\mu}{1+4s}\left(\frac{\delta}{2L}\right)^{\frac{2}{\beta}}.$$

Summing the above inequality over $k = 1, \ldots, \frac{K_{\rm c}^{\rm d}}{s}$, we derive that

$$0 \le \mathrm{d}^2(x_{K_{\rm c}^{\rm d}+s}, S) < \mathrm{d}^2(x_s, S) - \frac{K_{\rm c}^{\rm d}}{s}\frac{2\underline{v}(2-\overline{v})\mu}{1+4s}\left(\frac{\delta}{2L}\right)^{\frac{2}{\beta}},$$

which contradicts with the definition of $K_{\rm c}^{\rm d}$. The proof is complete. □

H. Proof of Theorem 3.10

Proof. By the assumption of the Hölder-type error bound property, there exists $\kappa > 0$ such that (3.10) holds for each x_k. This, together with (3.36), yields

$$\mathrm{d}^2(x_{s(k+1)}, S) \le \mathrm{d}^2(x_{sk}, S) - \rho\, \mathrm{d}^{\frac{2q}{\beta}}(x_{sk}, S) \quad \text{for each } k \in \mathbb{N},$$

where $\rho := \frac{2\underline{v}(2-\overline{v})\mu}{1+4s}\left(\frac{1}{2\kappa L}\right)^{\frac{2}{\beta}}$. Consequently, there exists $c \ge 0$ such that

$$\mathrm{d}(x_k, S) \le \begin{cases} c\tau^k, & \text{if } q = \beta, \\ ck^{-\frac{\beta}{2(q-\beta)}}, & \text{if } q > \beta, \end{cases} \tag{H.1}$$

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3(i) (with $\mathrm{d}^2(x_k, S)$, ρ, $\frac{q}{\beta} - 1$ in place of u_k, a, r), for each $k \in \mathbb{N}$.

Fix $l > k \in \mathbb{N}$. It follows from Lemma 3.5(ii) (taking $x := \mathrm{P}_S(x_k)$) that

$$\|x_l - x_k\| \le \|x_l - \mathrm{P}_S(x_k)\| + \|x_k - \mathrm{P}_S(x_k)\| \le 2\|x_k - \mathrm{P}_S(x_k)\| = 2\mathrm{d}(x_k, S).$$

Hence, by the convergence of $\{x_l\}$ to $\bar{x} \in S$ as shown in Theorem 3.7, we obtain

$$\|x_k - \bar{x}\| = \lim_{l\to\infty} \|x_l - x_k\| \le 2\mathrm{d}(x_k, S).$$

This, together with (H.1), implies the conclusions. The proof is complete. □

I. Proof of Theorem 3.11

Proof. (i) By virtue of (3.41) in Lemma 3.8, Theorem A.1 is applicable to showing that $\{\|x_k - x\|\}$ is convergent and $\sum_{k=1}^{\infty} (F^+(x_k))^{\frac{2}{\beta}} < \infty$ with probability 1. Hence $\lim_{k\to\infty} F^+(x_k) = 0$, and the cluster point $\bar{x}$ of $\{x_k\}$ falls in S, with probability 1. This, together with Lemma 3.5(ii), shows that $\{x_k\}$ converges to $\bar{x} \in S$ with probability 1.

(ii) Proving by contradiction, we assume that $\mathbb{E}\{F^+(x_k)\} > \delta$ for each $1 \le k \le K_{\rm s}^{\rm d}$. Consequently, $\mathbb{E}\left\{(F^+(x_k))^{\frac{2}{\beta}}\right\} > \delta^{\frac{2}{\beta}}$ by the convexity of $t^{\frac{2}{\beta}}$ on $\mathbb{R}_+$ (as $\beta \le 1$). Taking the expectation on (3.41) (with $x := \mathrm{P}_S(x_k)$), one has for each $1 \le k \le K_{\rm s}^{\rm d}$ that

$$\begin{aligned} \mathbb{E}\left\{\mathrm{d}^2(x_{k+1}, S)\right\} &\le \mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\} - \frac{\underline{v}(2-\overline{v})}{m} L^{-\frac{2}{\beta}} \mathbb{E}\left\{(F^+(x_k))^{\frac{2}{\beta}}\right\} \\ &\le \mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\} - \frac{\underline{v}(2-\overline{v})}{m}\left(\frac{\delta}{L}\right)^{\frac{2}{\beta}}. \end{aligned}$$

Summing the above inequality over $k = 1, \ldots, K_{\mathrm{s}}^{\mathrm{d}}$, we derive that

$$0 \leq \mathbb{E}\left\{ \mathrm{d}^2(x_{k+1}, S) \right\} < \mathrm{d}^2(x_1, S) - K_{\mathrm{s}}^{\mathrm{d}} \frac{\underline{v}(2 - \overline{v})}{m} \left(\frac{\delta}{L} \right)^{\frac{2}{\beta}},$$

which contradicts with the definition of $K_{\mathrm{s}}^{\mathrm{d}}$. The proof is complete. $\square$

References

Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific.

Kiwiel, K. C. (2004). Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization*, *14*, 807–840.