

# Proof Script for WS2019/20 Convex Optimization\*

Last updated: January 20, 2020

## 1 Convex Analysis

**Theorem 1.1** (separation of convex sets). *Let  $C_1, C_2$  be nonempty convex subsets of  $\mathbb{E}$ .*

1. *Assume  $C_1$  is closed and  $C_2 = \{w\} \subset \mathbb{E} \setminus C_1$ . Then  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  s.t.*

$$\langle v, w \rangle > \alpha \geq \langle v, u \rangle, \quad \forall u \in C_1.$$

2. *Assume  $C_1$  is open and  $C_2 = \{w\} \subset \mathbb{E} \setminus C_1$ . Then  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  s.t.*

$$\langle v, w \rangle \geq \alpha \geq \langle v, u \rangle, \quad \forall u \in C_1.$$

3. *Assume  $C_1 \cap C_2 = \emptyset$  and  $C_1$  is open. Then  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  s.t.*

$$\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle, \quad \forall u^1 \in C_1, u^2 \in C_2.$$

4. *Assume  $\emptyset \neq \text{int } C_1 \subset \mathbb{E} \setminus C_2$ . Then  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  s.t.*

$$\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle, \quad \forall u^1 \in C_1, u^2 \in C_2.$$

*Proof.* (1) Consider the projection of  $w$  onto  $C_1$ , i.e., set  $u^* := \arg \min_{u \in C_1} \frac{1}{2} \|u - w\|^2$  or, equivalently via the variational inequality:  $\langle u - u^*, u^* - w \rangle \geq 0 \forall u \in C_1$ . Now let  $v := w - u^* \neq 0$ . Then  $\forall u \in C_1$ , we have  $\langle v, w \rangle = \langle w - u^*, w \rangle = \|w - u^*\|^2 + \langle w - u^*, u^* \rangle \geq \|w - u^*\|^2 + \langle w - u^*, u \rangle = \|v\|^2 + \langle v, u \rangle$ . Set  $\alpha := \sup\{\langle v, u \rangle : u \in C_1\}$ . Note  $\alpha < \infty$  since  $\langle v, u \rangle \leq \langle v, u^* \rangle \forall u \in C$ . Thus  $\langle v, w \rangle > \langle v, w \rangle - \|v\|^2 \geq \alpha \geq \langle v, u \rangle \forall u \in C_1$ .

(2) We have either  $w \in \mathbb{E} \setminus \text{cl } C_1$  or  $w \in \text{bdry } C_1$ . In either case,  $\exists w^k \in \mathbb{E} \setminus \text{cl } C_1$  s.t.  $w^k \rightarrow w$ . For each  $w^k$ , by (1),  $\exists v^k \in \mathbb{E}$  with  $\|v^k\| \equiv 1$  s.t.  $\langle v^k, w^k \rangle \leq \langle v^k, u \rangle \forall u \in \text{cl } C_1$ . Hence  $v^k \rightarrow v \in \mathbb{E}$  along a subsequence s.t.  $\|v\| = 1$  and  $\alpha := \langle v, w \rangle \leq \langle v, u \rangle \forall u \in C_1 \subset \text{cl } C_1$ .

(3) Let  $C := C_2 - C_1 = \{u^2 - u^1 : u^1 \in C_1, u^2 \in C_2\}$ . Note that  $C$  is a convex, open set, and  $0 \notin C$ . By (2),  $\exists v \in \mathbb{E}$  with  $\|v\| = 1$  s.t.  $\langle -v, u^2 - u^1 \rangle \geq \langle -v, 0 \rangle = 0$  or, equivalently,  $\langle v, u^1 \rangle \geq \langle v, u^2 \rangle \forall u^1 \in C_1, u^2 \in C_2$ . Set  $\alpha := \sup\{\langle v, u^2 \rangle : u^2 \in C_2\}$ , then we conclude that  $\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle \forall u^1 \in C_1, u^2 \in C_2$ .

(4) By applying (3) to  $\text{int } C_1$  and  $C_2$ , we have  $\langle v, u^1 \rangle \geq \alpha \geq \langle v, u^2 \rangle \forall u^1 \in \text{int } C_1, u^2 \in C_2$ . The inequality remains true for all  $u_1 \in C_1$ .  $\square$

---

\*Please report typos to: tao.wu@tum.de

**Theorem 1.2.** *A proper convex function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  is locally Lipschitz at any  $u \in \text{rint dom } J$ .*

*Proof.* Without loss of generality, we restrict  $J : \text{aff dom } J \rightarrow \overline{\mathbb{R}}$ .

(i) Claim: If  $M = \sup\{J(v) : v \in B_\epsilon(u)\} < \infty$  with  $\epsilon > 0$ , then  $J$  is locally Lipschitz at  $u$ .

First, by convexity of  $J$  we have  $\forall v \in B_\epsilon(u) : J(v) \geq 2J(u) - J(2u - v) \geq 2J(u) - M$ . Thus,  $\sup\{|J(v)| : v \in B_\epsilon(u)\} \leq M + 2|J(u)|$ .

Next, we show  $J$  is Lipschitz on  $B_{\epsilon/2}(u)$ . Let  $v, w \in B_{\epsilon/2}(u)$  be given. Take  $z \in B_\epsilon(u)$  s.t.  $w = (1 - t)v + tz$  for some  $t \in [0, 1]$  and  $\|z - v\| \geq \epsilon/2$ . By convexity,  $J(w) - J(v) \leq t(J(z) - J(v)) \leq 2t(M - J(u))$ . Since  $t(z - v) = w - v$ , we have  $t = \|w - v\|/\|z - v\| \leq 2\|w - v\|/\epsilon$  and  $J(w) - J(v) \leq (4(M - J(u))/\epsilon)\|w - v\|$ . Analogously, one can show  $J(v) - J(w) \leq (4(M - J(u))/\epsilon)\|w - v\|$ . Hence,  $J$  is Lipschitz on  $B_{\epsilon/2}(u)$  with modulus  $4(M - J(u))/\epsilon$ .

(ii) Let  $u \in \text{rint dom } J$  and  $n = \dim(\text{aff dom } J)$ . Then by Carathéodory's theorem,  $\exists\{\alpha^i\}_{i=1}^{n+1} \subset (0, 1)$ ,  $\{u^i\}_{i=1}^{n+1} \subset \text{dom } J$  s.t.  $u = \sum_{i=1}^{n+1} \alpha^i u^i$ ,  $\sum_{i=1}^{n+1} \alpha^i = 1$ , i.e.,  $u$  belongs to the relative interior of the convex hull of  $\{u^i\}_{i=1}^{n+1}$ . Thus one can apply (i) to assert that  $J$  is locally Lipschitz at  $u$ .  $\square$

**Theorem 1.3.** *For any proper convex function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , if  $u^* \in \text{dom } J$  is a local minimizer of  $J$ , then it is also a global minimizer.*

*Proof.* By the definition of a local minimizer,  $\exists\epsilon > 0$  s.t.  $J(u^*) \leq J(u) \forall u \in B_\epsilon(u^*)$ . For the sake of contradiction, assume  $\exists\bar{u} \in \mathbb{E}$  s.t.  $J(\bar{u}) < J(u^*)$ . By convexity of  $J$ , we have  $J(\alpha\bar{u} + (1 - \alpha)u^*) \leq J(u^*) - \alpha(J(u^*) - J(\bar{u})) < J(u^*) \forall \alpha \in (0, 1]$ . This violates the local optimality of  $u^*$  as  $\alpha \rightarrow 0^+$ .  $\square$

**Theorem 1.4.** *Any proper function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , which is bounded from below, coercive, and lsc, has a (global) minimizer.*

*Proof.* Let  $\{u^k\}$  be an infimizing sequence for  $J$ , i.e.,  $\lim_{k \rightarrow \infty} J(u^k) = \inf_{u \in \mathbb{E}} J(u) \in \mathbb{R}$ . Since  $\{J(u^k)\}$  is uniformly bounded from above, by coercivity of  $J$  we have  $\{u^k\}$  uniformly bounded. By compactness,  $u^k \rightarrow u^*$  along a subsequence. Since  $J$  is lsc, we have  $J(u^*) \leq \liminf_{k \rightarrow \infty} J(u^k) = \inf_{u \in \mathbb{E}} J(u)$ , which implies  $J(u^*) = \inf_{u \in \mathbb{E}} J(u)$  or  $u^*$  is a minimizer of  $J$ .  $\square$

**Theorem 1.5.** *The minimizer of a strictly convex function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  is unique.*

*Proof.* Let  $u, v \in \mathbb{E}$  be two (global) minimizers s.t.  $u \neq v$  and  $J(u) = J(v) = J^*$ . By strict convexity of  $J$ ,  $J(\alpha u + (1 - \alpha)v) < \alpha J(u) + (1 - \alpha)J(v) = J^*$  for all  $\alpha \in (0, 1)$ , which contradicts the global optimality of  $u$  and  $v$ .  $\square$

**Theorem 1.6.** *Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be a convex function. Then for any  $u \in \text{int dom } J$ ,  $\partial J(u)$  is a nonempty, compact, and convex subset.*

*Proof.* (i) nonemptiness. Since  $(u, J(u)) \notin \text{int epi } J$ , by Theorem 1.1 we have  $\exists(p, -\alpha) \in \mathbb{E} \times \mathbb{R}$  s.t.  $(p, -\alpha) \neq (0, 0)$ ,  $\alpha \geq 0$  by our choice, and  $\langle (p, -\alpha), (u - v, J(u) - J(v)) \rangle \geq 0 \forall v \in \text{dom } J$ . In fact, we must have  $\alpha > 0$  since otherwise  $p = 0$ . Thus, we conclude that  $p/\alpha \in \partial J(u)$ .

(ii) boundedness. By Theorem 1.2,  $J$  is locally Lipschitz at  $u$  with modulus  $L_u$ . Let  $p \in \partial J(u)$  be fixed. For any  $h \in (\text{dom } J) - u$  whenever  $\|h\|$  is sufficiently small, we have  $\langle p, h \rangle \leq J(u + h) - J(u) \leq L_u\|h\|$ . This holds true only if  $\|p\| \leq L_u$ , which implies boundedness of  $\partial J(u)$ .

(iii) closedness. Let  $v \in \mathbb{E}$  be arbitrarily fixed and  $p^k \rightarrow p^*$  where each  $p^k \in \partial J(u)$ . Then  $\forall k : J(v) - J(u) \geq \langle p^k, v - u \rangle$ . By continuity,  $J(v) - J(u) \geq \langle p^*, v - u \rangle$  when passing  $k \rightarrow \infty$ . Since  $v$  can be arbitrary, we assert  $p^* \in \partial J(u)$ .

(iv) convexity. Let  $v \in \mathbb{E}$  be arbitrarily fixed, and  $p, q \in \partial J(u)$ . Then we have

$$\begin{aligned} J(v) &\geq J(u) + \langle p, v - u \rangle, \\ J(v) &\geq J(u) + \langle q, v - u \rangle. \end{aligned}$$

Hence,  $\forall 0 \leq \alpha \leq 1 : J(v) \geq J(u) + \langle \alpha p + (1 - \alpha)q, v - u \rangle$ , i.e.,  $\alpha p + (1 - \alpha)q \in \partial J(u)$ .  $\square$

**Theorem 1.7.** *Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be a convex function. Then  $\partial J$  is a monotone operator, i.e.  $\forall u^1, u^2 \in \text{dom } J$ ,  $p^1 \in \partial J(u^1)$ ,  $p^2 \in \partial J(u^2)$  :*

$$\langle p^1 - p^2, u^1 - u^2 \rangle \geq 0.$$

*Proof.* By applying the definition of subdifferential at arbitrarily given  $u^1, u^2 \in \text{dom } J$ , we have

$$\begin{aligned} J(u^2) &\geq J(u^1) + \langle p^1, u^2 - u^1 \rangle, \\ J(u^1) &\geq J(u^2) + \langle p^2, u^1 - u^2 \rangle. \end{aligned}$$

Adding the two inequalities yields  $\langle p^1 - p^2, u^1 - u^2 \rangle \geq 0$ .  $\square$

**Theorem 1.8.** *Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be a proper, convex, lsc function. Then  $\partial J$  is a closed set-valued map, i.e.,  $p^* \in \partial J(u^*)$  whenever*

$$\exists (u^k, p^k) \rightarrow (u^*, p^*) \in (\text{dom } J) \times \mathbb{E} \text{ s.t. } p^k \in \partial J(u^k) \forall k.$$

*Proof.* Let  $v \in \mathbb{E}$  be arbitrarily fixed. For each  $k$ ,  $p^k \in \partial J(u^k) \Rightarrow J(v) \geq J(u^k) + \langle p^k, v - u^k \rangle$ . Passing  $k \rightarrow \infty$ , we have  $\langle p^k, v - u^k \rangle \rightarrow \langle p^*, v - u^* \rangle$  and  $J(u^*) \leq \liminf_{k \rightarrow \infty} J(u^k)$ . Hence,  $J(u^*) + \langle p^*, v - u^* \rangle \leq \liminf_{k \rightarrow \infty} \{J(u^k) + \langle p^k, v - u^k \rangle\} \leq J(v)$ . Since  $v$  can be arbitrary,  $p^* \in \partial J(u^*)$ .  $\square$

**Theorem 1.9.** *Given any proper convex function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , the sufficient and necessary condition for  $u^*$  being a (global) minimizer for  $J$  is:  $0 \in \partial J(u^*)$ .*

*Proof.* (i) sufficiency.  $0 \in \partial J(u^*) \Rightarrow J(u) \geq J(u^*) + \langle 0, u - u^* \rangle = J(u^*) \forall u \in \mathbb{E}$ .

(ii) necessity.  $J(u^*) \leq J(u) \forall u \in \mathbb{E} \Rightarrow J(u^*) + \langle 0, u - u^* \rangle \leq J(u) \forall u \Rightarrow 0 \in \partial J(u^*)$ .  $\square$

**Theorem 1.10** (Fenchel-Young inequality). *For any function  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  and  $(u, p) \in \mathbb{E} \times \mathbb{E}$ , we have*

$$J(u) + J^*(p) \geq \langle u, p \rangle.$$

*If  $J$  is also convex, the equality holds iff  $p \in \partial J(u)$  with  $(u, p) \in \text{dom } J \times \text{dom } J^*$ .*

*Proof.* (i)  $J(u) + J^*(p) \geq \langle u, p \rangle$  follows directly from the definition of convex conjugate. (ii) The equality holds only if  $(u, p) \in \text{dom } J \times \text{dom } J^*$ . Moreover,  $p \in \partial J(u)$  is the sufficient and necessary condition for  $\min_{u \in \mathbb{E}} \{J(u) - \langle u, p \rangle\}$ .  $\square$

**Theorem 1.11** (order reversing). *For any  $J_1, J_2 : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , we have  $J_1^*(\cdot) \leq J_2^*(\cdot)$  whenever  $J_1(\cdot) \geq J_2(\cdot)$ .*

*Proof.* Given any  $(u, p)$ , we have  $\langle u, p \rangle - J_1(u) \leq \langle u, p \rangle - J_2(u)$ . Taking supremum over  $u$  on both sides yields  $J_1^*(p) \leq J_2^*(p)$ .  $\square$

**Theorem 1.12.** *Let  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ , and  $J^{**} = (J^*)^*$  be the biconjugate of  $J$ . In general:*

1.  $J^{**}(\cdot) \leq J(\cdot)$ .

2.  $J^*$  is convex and lsc.

If  $J$  is proper, convex, and lsc, then:

3.  $J^{**}(\cdot) = J(\cdot)$ .

4.  $p \in \partial J(u)$  iff  $u \in \partial J^*(p)$ .

*Proof.* (1) Since  $J^{**}(u) = \sup_p \{ \langle p, u \rangle - J^*(p) \}$  and, by Theorem 1.10,  $\langle p, u \rangle - J^*(p) \leq J(u) \forall p$ , we have  $J^{**}(u) \leq J(u)$  for arbitrarily given  $u$ .

(2) (i) convexity. Let  $p, q \in \mathbb{E}$ ,  $0 \leq \alpha \leq 1$ . Then  $J^*(\alpha p + (1-\alpha)q) = \sup_u \{ \langle u, \alpha p + (1-\alpha)q \rangle - J(u) \} \leq \sup_u \{ \langle \alpha u, p \rangle - \alpha J(u) \} + \sup_u \{ \langle (1-\alpha)u, q \rangle - (1-\alpha)J(u) \} = \alpha J^*(p) + (1-\alpha)J^*(q)$ .

(ii) lsc. Note  $\text{epi } J^* = \{ (p, \alpha) \in \mathbb{E} \times \mathbb{R} : \langle u, p \rangle - J(u) \leq \alpha \forall u \} = \bigcap_u \text{epi } \Phi_u$  where  $\Phi_u(\cdot) = \langle u, \cdot \rangle - J(u)$ . Since each  $\text{epi } \Phi_u$  and any arbitrary intersection of closed sets is closed,  $\text{epi } J^*$  is closed and hence  $J^*$  is lsc.

(3) For the sake of contradiction, assume  $\exists \bar{u} \in \text{dom } J^{**}$  s.t.  $J(\bar{u}) > J^{**}(\bar{u})$ . Let  $0 < d < J(\bar{u}) - J^{**}(\bar{u})$  be fixed. Since  $(\bar{u}, J(\bar{u}) - d) \notin \text{epi } J$  and  $\text{epi } J$  is convex and closed, by Theorem 1.1,  $\exists (\bar{p}, -1) \in \mathbb{E} \times \mathbb{R}$  s.t.  $\langle (\bar{p}, -1), (\bar{u}, J(\bar{u}) - d) \rangle \geq \langle (\bar{p}, -1), (u, \alpha) \rangle \forall (u, \alpha) \in \text{epi } J$ . In particular,  $\langle \bar{p}, \bar{u} \rangle - J(\bar{u}) + d \geq \langle \bar{p}, u \rangle - J(u) \forall u \in \text{dom } J$ . Hence,  $\langle \bar{p}, \bar{u} \rangle - J(\bar{u}) + d \geq J^*(\bar{p}) \geq \langle \bar{p}, \bar{u} \rangle - J^{**}(\bar{u})$  by Theorem 1.10. Thus we have  $J^{**}(\bar{u}) + d \geq J(\bar{u})$  as a contradiction to our assumption.

(4)  $p \in \partial J(u) \Leftrightarrow J(u) + J^*(p) = \langle u, p \rangle \Leftrightarrow J^{**}(u) + J^*(p) = \langle u, p \rangle \Leftrightarrow u \in \partial J^*(p)$ .  $\square$

**Theorem 1.13.** Assume that  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  is proper, convex, and lsc. Then  $J$  is  $\mu$ -strongly convex iff  $J^*$  is  $\frac{1}{\mu}$ -Lipschitz differentiable.

*Proof.* (only if) Let  $p \in \partial J(u)$  be arbitrarily given. By  $\mu$ -strong convexity of  $J$ , we have

$$J(v) \geq J(u) + \langle p, v - u \rangle + \frac{\mu}{2} \|v - u\|^2 \quad \forall v. \quad (1)$$

Then  $\forall q : J^*(q) = \sup_v \{ \langle q, v \rangle - J(v) \} \leq \sup_v \{ \langle q, v \rangle - J(u) - \langle p, v - u \rangle - \frac{\mu}{2} \|v - u\|^2 \} = \langle q, u \rangle - J(u) + \sup_v \{ \langle q - p, v - u \rangle - \frac{\mu}{2} \|v - u\|^2 \} = \langle q, u \rangle - J(u) + \frac{1}{2\mu} \|q - p\|^2 = \langle p, u \rangle - J(u) + \langle q - p, u \rangle + \frac{1}{2\mu} \|q - p\|^2 = J^*(p) + \langle q - p, u \rangle + \frac{1}{2\mu} \|q - p\|^2$ . Here we have used the identity  $\langle p, u \rangle - J(u) = J^*(p)$ . We have actually derived  $\lim_{q \rightarrow p} \|J^*(q) - J^*(p) - \langle q - p, u \rangle\| / \|q - p\| = 0$ , which asserts that  $J^*$  is (Fréchet-)differentiable at  $p$  with  $\nabla J^*(p) = u$ .

Finally we show  $\nabla J^*$  is  $\frac{1}{\mu}$ -Lipschitz. Let  $u = \nabla J^*(p)$ ,  $v = \nabla J^*(q)$ , or equivalently  $p \in \partial J(u)$ ,  $q \in \partial J(v)$ . Then by (1) we have

$$\begin{aligned} J(v) &\geq J(u) + \langle p, v - u \rangle + \frac{\mu}{2} \|v - u\|^2, \\ J(u) &\geq J(v) + \langle q, u - v \rangle + \frac{\mu}{2} \|u - v\|^2. \end{aligned}$$

Adding the above two inequalities, we obtain  $\mu \|u - v\|^2 \leq \langle p - q, u - v \rangle \leq \|p - q\| \|u - v\|$  and thus  $\|u - v\| \leq \frac{1}{\mu} \|p - q\|$ .

(if) Note that  $J^*(q) = J^*(p) + \int_0^1 \langle \nabla J^*(p + s(q - p)), q - p \rangle ds = J^*(p) + \langle \nabla J^*(p), q - p \rangle + \int_0^1 \langle \nabla J^*(p + s(q - p)) - \nabla J^*(p), q - p \rangle ds \leq J^*(p) + \langle \nabla J^*(p), q - p \rangle + \frac{1}{2\mu} \|q - p\|^2$ . Let  $p \in \partial J(u) \Leftrightarrow u = \nabla J^*(p)$ . Then  $J^*(q) \leq J^*(p) + \langle q - p, u \rangle + \frac{1}{2\mu} \|q - p\|^2$ . Taking the convex conjugate on both sides, we deduce  $J(v) = J^{**}(v) \geq \sup_q \{ \langle q, v \rangle - (J^*(p) + \langle q - p, u \rangle + \frac{1}{2\mu} \|q - p\|^2) \} = -J^*(p) + \langle p, v \rangle + \frac{\mu}{2} \|v - u\|^2 = J(u) + \langle p, v - u \rangle + \frac{\mu}{2} \|v - u\|^2$ .  $\square$

**Theorem 1.14** (weak duality). Let  $K \in \mathbb{R}^{m \times n}$ , and  $F : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ ,  $G : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  are proper, convex, and lsc. Then it holds that  $\inf_u \{ F(Ku) + G(u) \} \geq \sup_p \{ -G^*(-K^\top p) - F^*(p) \}$ .

*Proof.* Let  $\mathcal{L}(u, p) = \langle p, Ku \rangle - F^*(p) + G(u)$ , then  $\inf_u \{F(Ku) + G(u)\} = \inf_u \sup_p \mathcal{L}(u, p)$  and  $\sup_p \{-G^*(-K^\top p) - F^*(p)\} = \sup_p \inf_u \mathcal{L}(u, p)$ . It remains to verify  $\inf_u \sup_p \mathcal{L}(u, p) \geq \sup_p \inf_u \mathcal{L}(u, p)$ . For an arbitrarily fixed  $(u, p)$ , we have  $\sup_{p'} \mathcal{L}(u, p') \geq \mathcal{L}(u, p) \geq \inf_{u'} \mathcal{L}(u', p)$ . Taking infimum over  $u$  yields  $\inf_u \sup_{p'} \mathcal{L}(u, p') \geq \inf_{u'} \mathcal{L}(u', p)$ ; Taking supremum over  $p$  yields  $\sup_{p'} \mathcal{L}(u, p') \geq \sup_p \inf_{u'} \mathcal{L}(u', p)$ . Hence, the conclusion follows.  $\square$

**Theorem 1.15** (Fenchel-Rockafellar duality). *Assume  $\exists \bar{u} \in \text{dom } G$  s.t.  $F$  is continuous at  $K\bar{u}$ . Then the strong duality holds:  $\mathcal{P}^* = \mathcal{D}^*$ . Moreover,  $(u^*, p^*)$  is the optimal primal-dual pair iff*

$$\begin{cases} Ku^* \in \partial F^*(p^*), \\ -K^\top p^* \in \partial G(u^*). \end{cases}$$

*Proof.* Define  $\Phi(v) := \inf_u \{F(Ku + v) + G(u)\}$ . Since  $\forall v^1, v^2 \in \mathbb{R}^m$ ,  $\alpha \in [0, 1] : \alpha\Phi(v^1) + (1 - \alpha)\Phi(v^2) = \inf_{u^1} \{\alpha F(Ku^1 + v^1) + \alpha G(u^1)\} + \inf_{u^2} \{(1 - \alpha)F(Ku^2 + v^2) + (1 - \alpha)G(u^2)\} = \inf_{u^1, u^2} \{\alpha F(Ku^1 + v^1) + (1 - \alpha)F(Ku^2 + v^2) + \alpha G(u^1) + (1 - \alpha)G(u^2)\} \geq \inf \{F(Ku + \alpha v^1 + (1 - \alpha)v^2) + G(u) : u = \alpha u^1 + (1 - \alpha)u^2\} \geq \Phi(\alpha v^1 + (1 - \alpha)v^2)$ , we prove that  $\Phi$  is a convex function.

Without loss of generality, assume  $\Phi(0) > -\infty$ . By our assumption,  $\exists \epsilon > 0$  s.t.  $\forall \|v\| < \epsilon : \Phi(v) \leq F(K\bar{u} + v) + G(\bar{u}) \leq M$  for some  $M < \infty$ , i.e.,  $v \in \text{dom } \Phi$ . By Theorem 1.2,  $\Phi$  is locally Lipschitz at 0, and  $\Phi(0) = \Phi^{**}(0) = \sup_p -\Phi^*(p)$ , where  $\Phi^*(p) = \sup_v \{\langle p, v \rangle - \inf_u \{F(Ku + v) + G(u)\}\} = \sup_{v, u} \{\langle p, v + Ku \rangle + \langle -K^\top p, u \rangle - F(Ku + v) - G(u)\} = F^*(p) + G^*(-K^\top p)$ . Thus,  $\mathcal{P}^* = \mathcal{D}^*$  is proven.

As for the optimality condition, note that  $\forall (u, p) : \mathcal{G}(u, p) = F(Ku) + G(u) + G^*(-K^\top p) + F^*(p) = F(Ku) + F^*(p) - \langle Ku, p \rangle + G(u) + G^*(-K^\top p) - \langle -K^\top p, u \rangle \geq 0$ . The equality holds, i.e.,  $\mathcal{G}(u^*, p^*) = 0$ , iff  $Ku^* \in \partial F^*(p^*)$  and  $-K^\top p^* \in \partial G(u^*)$  according to Theorem 1.10.  $\square$

**Theorem 1.16** (Moreau identity). *Let  $\tau > 0$  and  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be proper, convex, and lsc. Then the following identity holds:*

$$\text{id}(\cdot) = \text{prox}_{\tau J}(\cdot) + \tau \text{prox}_{\frac{1}{\tau} J^*}(\cdot/\tau).$$

*Proof.*  $v = \tau \text{prox}_{\frac{1}{\tau} J^*}(u/\tau) \Leftrightarrow (I + \frac{1}{\tau} \partial J^*)^{-1}(u/\tau) \ni v/\tau \Leftrightarrow \partial J^*(v/\tau) \ni u - v \Leftrightarrow v/\tau \in \partial J(u - v) \Leftrightarrow u - v = (I + \tau \partial J)^{-1}(u) = \text{prox}_{\tau J}(u)$ .  $\square$

**Theorem 1.17.** *Let  $F, G : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be proper, convex, and lsc. Then*

$$(F \square G)^* = F^* + G^*.$$

*Proof.*  $\forall p \in \mathbb{E} : (F \square G)^*(p) = \sup_{u, v} \{\langle p, u \rangle - F(v) - G(u - v)\} = \sup_{u, v} \{\langle p, v \rangle - F(v) + \langle p, u - v \rangle - G(u - v)\} = F^*(p) + G^*(p)$ .  $\square$

## 2 Optimization Algorithms

**Theorem 2.1.** *If  $\langle \nabla J(u^k), d^k \rangle < 0$ , then  $J(u^k + \tau d^k) < J(u^k)$  for all sufficiently small  $\tau > 0$ .*

*Proof.* The conclusion follows from Taylor expansion:  $J(u^k + \tau d^k) = J(u^k) + \tau \langle \nabla J(u^k), d^k \rangle + o(\tau) = J(u^k) + \tau (\langle \nabla J(u^k), d^k \rangle + o(1)) < J(u^k)$ , for all  $\tau > 0$  sufficiently small.  $\square$

**Lemma 2.2** (feasibility of line search). *Assume that  $J : \mathbb{E} \rightarrow \mathbb{R}$  is continuously differentiable,  $\langle \nabla J(u^k), d^k \rangle < 0 \forall k$ , and  $0 < c_1 < c_2 < 1$ . Then there exists an open interval in which the step size  $\tau$  satisfies the Armijo- and the curvature conditions.*

*Proof.* Consider  $\phi(\tau) := J(u^k + \tau d^k)$  and  $\psi(\tau) := J(u^k) + \tau c_1 \langle \nabla J(u^k), d^k \rangle$  for  $\tau \geq 0$ . Since  $\phi'(0) = \langle \nabla J(u^k), d^k \rangle < \psi'(0)$ ,  $\phi(\tau) < \psi(\tau)$  for all  $\tau > 0$  sufficiently close to 0. On the other hand,  $\phi(\cdot)$  is bounded from below but  $\psi(\cdot)$  is not. Hence,  $\phi$  and  $\psi$  intersect at  $\tau = \tau' > 0$  (for the first time as  $\tau$  increases from 0). Thus,  $0 < \tau < \tau'$  fulfills the Armijo condition.

By the mean value theorem,  $\exists \tau'' \in (0, \tau')$  s.t.  $J(u^k + \tau'' d^k) - J(u^k) = \tau'' \langle \nabla J(u^k + \tau'' d^k), d^k \rangle$ . This implies  $\langle \nabla J(u^k + \tau'' d^k), d^k \rangle = c_1 \langle \nabla J(u^k), d^k \rangle > c_2 \langle \nabla J(u^k), d^k \rangle$  since  $c_1 < c_2$  and  $\langle \nabla J(u^k), d^k \rangle < 0$ . By continuity, this inequality holds in a neighborhood of  $\tau''$ .  $\square$

**Theorem 2.3** (Zoutendijk). *Assume that  $J : \mathbb{E} \rightarrow \mathbb{R}$  is continuously differentiable, and the Armijo- and curvature-conditions are both satisfied with  $0 < c_1 < c_2 < 1$  for each  $k$ . In addition,  $J$  is  $\mu$ -Lipschitz differentiable on  $\{u \in \mathbb{E} : J(u) \leq J(u^0)\}$ . Then  $\sum_{k=0}^{\infty} \frac{|\langle \nabla J(u^k), d^k \rangle|^2}{\|d^k\|^2} < \infty$ .*

*Proof.* From the curvature condition, we have  $\langle \nabla J(u^{k+1}) - \nabla J(u^k), d^k \rangle \geq (c_2 - 1) \langle \nabla J(u^k), d^k \rangle$ . Since  $\nabla J$  is  $\mu$ -Lipschitz,  $\langle \nabla J(u^{k+1}) - \nabla J(u^k), d^k \rangle \leq \tau^k \mu \|d^k\|^2$ . Altogether we have  $\tau^k \geq \frac{(c_2 - 1) \langle \nabla J(u^k), d^k \rangle}{\mu \|d^k\|^2}$ . Using the Armijo condition, we have  $J(u^{k+1}) \leq J(u^k) - \frac{c_1(1 - c_2) |\langle \nabla J(u^k), d^k \rangle|^2}{\mu \|d^k\|^2}$ .

Summing up this inequality from  $k = 0$  to  $\infty$ , we have  $\sum_{k=0}^{\infty} \frac{|\langle \nabla J(u^k), d^k \rangle|^2}{\|d^k\|^2} < \infty$ .  $\square$

**Lemma 2.4.** *Assume that  $J : \mathbb{E} \rightarrow \mathbb{R}$  is  $\mu$ -Lipschitz differentiable. Then  $\forall u, v \in \mathbb{E}$ :*

$$|J(v) - J(u) - \langle \nabla J(u), v - u \rangle| \leq \frac{\mu}{2} \|v - u\|^2.$$

*Proof.* Since  $J(v) = J(u) + \int_0^1 \langle \nabla J(u + t(v - u)), v - u \rangle dt = J(u) + \langle \nabla J(u), v - u \rangle + \int_0^1 \langle \nabla J(u + t(v - u)) - \nabla J(u), v - u \rangle dt$ , we have  $|J(v) - J(u) - \langle \nabla J(u), v - u \rangle| = \left| \int_0^1 \langle \nabla J(u + t(v - u)) - \nabla J(u), v - u \rangle dt \right| \leq \int_0^1 |\langle \nabla J(u + t(v - u)) - \nabla J(u), v - u \rangle| dt \leq \int_0^1 \|\nabla J(u + t(v - u)) - \nabla J(u)\| \|v - u\| dt \leq \int_0^1 t \mu \|v - u\|^2 dt = \frac{\mu}{2} \|v - u\|^2$ .  $\square$

**Theorem 2.5** (convergence of gradient descent). *Assume that  $J : \mathbb{E} \rightarrow \mathbb{R}$  is  $\mu$ -Lipschitz differentiable. Then the gradient descent iteration  $u^{k+1} = u^k - \tau \nabla J(u^k)$  with  $\tau \in (0, 1/\mu]$  yields  $\lim_{k \rightarrow \infty} \nabla J(u^k) = 0$ .*

*Proof.* First, note that  $J(u^{k+1}) \leq J(u^k) \forall k$ . Since  $J$  has finite infimum by assumption,  $\lim_{k \rightarrow \infty} |J(u^{k+1}) - J(u^k)| = 0$ . Due to the majorization property and  $\mu \leq 1/\tau$ , we have  $J(u^{k+1}) \leq J(u^k) + \langle \nabla J(u^k), u^{k+1} - u^k \rangle + \frac{1}{2\tau} \|u^{k+1} - u^k\|^2 = J(u^k) - \frac{1}{2\tau} \|u^{k+1} - u^k\|^2$ . Hence, we conclude  $\|\nabla J(u^k)\| = \frac{1}{\tau} \|u^{k+1} - u^k\| \rightarrow 0$ .  $\square$

**Proposition 2.6.** *Let  $C$  be a nonempty, closed, convex subset of  $\mathbb{E}$ ,  $\Phi : C \rightarrow \mathbb{E}$ , and  $\alpha \in (0, 1)$ . Then the following statements are equivalent:*

1.  $\Phi$  is  $\alpha$ -averaged.
2.  $(1 - \frac{1}{\alpha})I + \frac{1}{\alpha}\Phi$  is nonexpansive.
3.  $\forall u, v \in C : \|\Phi(u) - \Phi(v)\|^2 \leq \|u - v\|^2 - \frac{1-\alpha}{\alpha}\|(I - \Phi)(u) - (I - \Phi)(v)\|^2$ .
4.  $\forall u, v \in C : \|\Phi(u) - \Phi(v)\|^2 + (1 - 2\alpha)\|u - v\|^2 \leq 2(1 - \alpha) \langle u - v, \Phi(u) - \Phi(v) \rangle$ .

*Proof.* By the definition of the averaged operator,  $\Phi = (1 - \alpha)I + \alpha\Psi$  for some nonexpansive operator  $\Psi : C \rightarrow \mathbb{E}$ , or  $\Psi = (1 - \frac{1}{\alpha})I + \frac{1}{\alpha}\Phi$ . Hence, (1)  $\Leftrightarrow$  (2) follows.

(2)  $\Leftrightarrow \Psi = (1 - \frac{1}{\alpha})I + \frac{1}{\alpha}\Phi$  is nonexpansive  $\Leftrightarrow \|\Psi(u) - \Psi(v)\| \leq \|u - v\| \Leftrightarrow \alpha^2\|u - v\|^2 \geq \|((\alpha - 1)I + \Phi)(u) - ((\alpha - 1)I + \Phi)(v)\|^2 = \|\Phi(u) - \Phi(v)\|^2 + (\alpha - 1)^2\|u - v\|^2 + 2(\alpha - 1) \langle u - v, \Phi(u) - \Phi(v) \rangle \Leftrightarrow$  (4).

Note that  $-2 \langle u - v, \Phi(u) - \Phi(v) \rangle = \|(I - \Phi)(u) - (I - \Phi)(v)\|^2 - \|u - v\|^2 - \|\Phi(u) - \Phi(v)\|^2$ . Hence, (4)  $\Leftrightarrow \|\Phi(u) - \Phi(v)\|^2 + (1 - 2\alpha)\|u - v\|^2 \leq (1 - \alpha)\|(I - \Phi)(u) - (I - \Phi)(v)\|^2 - (1 - \alpha)\|u - v\|^2 - (1 - \alpha)\|\Phi(u) - \Phi(v)\|^2 \Leftrightarrow$  (3).  $\square$

**Theorem 2.7** (Baillon-Haddad). *Let  $J : \mathbb{E} \rightarrow \mathbb{R}$  be a convex, continuously differentiable function. Then  $\nabla J$  is a nonexpansive operator iff  $\nabla J$  is firmly nonexpansive.*

*Proof.* (if) Obvious.

(only if) Define  $H(\cdot) := \frac{1}{2}\|\cdot\|^2 - J(\cdot)$ . Note that  $H$  is continuously differentiable and  $\nabla H = I - \nabla J$ . Since  $\nabla J$  is nonexpansive, we have  $\forall u, v : \langle \nabla H(v) - \nabla H(u), v - u \rangle \geq \|v - u\|(\|v - u\| - \|\nabla J(v) - \nabla J(u)\|) \geq 0$ .

This implies  $\forall u, v : H(v) - H(u) = \int_0^1 \langle \nabla H(u + t(v - u)), v - u \rangle dt \geq \int_0^1 \langle \nabla H(u), v - u \rangle dt = \langle \nabla H(u), v - u \rangle$ . Furthermore,  $H(v) \geq H(u) + \langle \nabla H(u), v - u \rangle \Rightarrow \frac{1}{2}\|v\|^2 - J(v) \geq \frac{1}{2}\|u\|^2 - J(u) + \langle u - \nabla J(u), v - u \rangle \Rightarrow J(v) - J(u) - \langle \nabla J(u), v - u \rangle \leq \frac{1}{2}\|v\|^2 - \frac{1}{2}\|u\|^2 + \langle u, u - v \rangle = \frac{1}{2}\|v - u\|^2$ .

Define  $D_J(w, u) := J(w) - J(u) - \langle \nabla J(u), w - u \rangle, \forall w, u \in \mathbb{E}$ . The above result says  $\frac{1}{2}\|w - u\|^2 \geq D_J(w, u), \forall w, u$ . Fix  $u$  temporarily and let  $d(\cdot) = D_J(\cdot, u)$ . Then  $d$  is convex,  $d(\cdot) \geq 0$ ,  $\nabla d(\cdot) = \nabla J(\cdot) - \nabla J(u)$ , and  $D_J(\cdot, u) = D_d(\cdot, u)$ . Therefore, we have  $\frac{1}{2}\|w - v\|^2 \geq D_d(w, v) = d(w) - d(v) - \langle \nabla d(v), w - v \rangle = d(w) - d(v) - \langle \nabla J(v) - \nabla J(u), w - v \rangle$ . Set  $w = v - \nabla J(v) + \nabla J(u)$ , then we have  $D_J(v, u) = d(v) \geq d(w) + \frac{1}{2}\|\nabla J(v) - \nabla J(u)\|^2 \geq \frac{1}{2}\|\nabla J(v) - \nabla J(u)\|^2$ .

Analogously, we can show  $D_J(u, v) \geq \frac{1}{2}\|\nabla J(u) - \nabla J(v)\|^2$ . Hence,  $\langle \nabla J(v) - \nabla J(u), v - u \rangle = D_J(u, v) + D_J(v, u) \geq \|\nabla J(v) - \nabla J(u)\|^2$ . Hence,  $\nabla J$  is firmly nonexpansive by Proposition 2.6.  $\square$

**Corollary 2.8.** *Assume  $G : \mathbb{E} \rightarrow \mathbb{R}$  is convex and  $\mu$ -Lipschitz differentiable, and  $\tau = 2\alpha/\mu$  with  $\alpha \in (0, 1)$ . Then  $I - \tau\nabla G$  is  $\alpha$ -averaged.*

*Proof.* Since  $\frac{1}{\mu}\nabla G$  is nonexpansive, by the Baillon-Haddad theorem,  $\frac{1}{\mu}\nabla G$  is firmly nonexpansive, i.e.,  $\exists \Psi : \mathbb{E} \rightarrow \mathbb{E}$  nonexpansive s.t.  $\frac{1}{\mu}\nabla G = \frac{1}{2}I + \frac{1}{2}\Psi$ . Hence,  $I - \tau\nabla G = (1 - \frac{\tau\mu}{2})I - \frac{\tau\mu}{2}\Psi = (1 - \alpha)I + \alpha(-\Psi)$ , i.e.,  $I - \tau\nabla G$  is  $\alpha$ -averaged.  $\square$

**Theorem 2.9** (composition of averaged operators). *Let  $C$  be a nonempty, closed, convex subset of  $\mathbb{E}$ . For each  $i \in \{1, \dots, m\}$ , let  $\alpha_i \in (0, 1)$  and  $\Phi_i : C \rightarrow C$  be an  $\alpha_i$ -averaged operator. Then*

$$\Phi = \Phi_m \circ \dots \circ \Phi_1$$

is  $\alpha$ -averaged with

$$\alpha = \frac{m}{m-1 + \frac{1}{\max_{1 \leq i \leq m} \alpha_i}}.$$

*Proof.* Let  $\kappa_i := \alpha_i/(1 - \alpha_i)$  for each  $i$ , and  $\kappa := \max_i \kappa_i$ . For arbitrarily fixed  $u, v \in C$ , we derive

$$\begin{aligned} & \|(I - \Phi)(u) - (I - \Phi)(v)\|^2/m \\ = & \|(I - \Phi_1)(u) - (I - \Phi_1)(v) + [(I - \Phi_2) \circ \Phi_1](u) - [(I - \Phi_2) \circ \Phi_1](v) + \dots \\ & + [(I - \Phi_m) \circ \Phi_{m-1} \circ \dots \circ \Phi_1](u) - [(I - \Phi_m) \circ \Phi_{m-1} \circ \dots \circ \Phi_1](v)\|^2/m \\ \leq & \|(I - \Phi_1)(u) - (I - \Phi_1)(v)\|^2 + \|[(I - \Phi_2) \circ \Phi_1](u) - [(I - \Phi_2) \circ \Phi_1](v)\|^2 + \dots \\ & + \|[(I - \Phi_m) \circ \Phi_{m-1} \circ \dots \circ \Phi_1](u) - [(I - \Phi_m) \circ \Phi_{m-1} \circ \dots \circ \Phi_1](v)\|^2 \\ \leq & \kappa_1(\|u - v\|^2 - \|\Phi_1(u) - \Phi_1(v)\|^2) + \kappa_2(\|\Phi_1(u) - \Phi_1(v)\|^2 - \|[\Phi_2 \circ \Phi_1](u) - [\Phi_2 \circ \Phi_1](v)\|^2) \\ & + \dots + \kappa_m(\|[\Phi_{m-1} \circ \dots \circ \Phi_1](u) - [\Phi_{m-1} \circ \dots \circ \Phi_1](v)\|^2 - \|[\Phi_m \circ \dots \circ \Phi_1](u) - [\Phi_m \circ \dots \circ \Phi_1](v)\|^2) \\ \leq & \kappa(\|u - v\|^2 - \|\Phi(u) - \Phi(v)\|^2). \end{aligned}$$

Since (1)  $\Leftrightarrow$  (3) in Proposition 2.6,  $\Phi$  is  $\alpha$ -averaged with  $\frac{1-\alpha}{\alpha} = \frac{1}{m\kappa}$ , or equivalently  $\alpha = \frac{m}{m+1/\kappa}$ .  $\square$

**Theorem 2.10** (Krasnoselskii). *Let  $C$  be a nonempty, closed, convex subset of  $\mathbb{E}$ , and  $u^{k+1} = \Phi(u^k)$  for  $k = 0, 1, 2, \dots$  where  $\Phi : C \rightarrow C$  satisfies:*

1.  $\Phi$  is  $\alpha$ -averaged for some  $\alpha \in (0, 1)$ .
2.  $\Phi$  has at least one fixed point.

Then  $\{u^k\}$  converges to a fixed point of  $\Phi$ .

*Proof.* Let  $\bar{u} \in C$  be an arbitrary fixed point of  $\Phi$ . Since  $\Phi$  is  $\alpha$ -averaged, we have  $\forall k : \|u^{k+1} - \bar{u}\|^2 = \|\Phi(u^k) - \Phi(\bar{u})\|^2 \leq \|u^k - \bar{u}\|^2 - \frac{1-\alpha}{\alpha} \|(I - \Phi)(u^k) - (I - \Phi)(\bar{u})\|^2 = \|u^k - \bar{u}\|^2 - \frac{1-\alpha}{\alpha} \|(I - \Phi)(u^k)\|^2$ . Summing up this inequality for all indices in  $l \in [0, k]$ , we have

$$\|u^{k+1} - \bar{u}\|^2 \leq \|u^0 - \bar{u}\|^2 - \frac{1-\alpha}{\alpha} \sum_{l=0}^k \|(I - \Phi)(u^l)\|^2.$$

This yields: (i)  $\|u^k - \bar{u}\| \searrow c \geq 0$ ; (ii)  $\sum_{k=0}^{\infty} \|(I - \Phi)(u^k)\|^2 < \infty$ .

By (i),  $\{u^k\}$  is uniformly bounded. Let  $\{u^{k'}\}$  be any convergent subsequence of  $\{u^k\}$  s.t.  $\lim_{k' \rightarrow \infty} u^{k'} = u^* \in C$ . By (ii),  $\|(I - \Phi)(u^*)\| = \lim_{k' \rightarrow \infty} \|(I - \Phi)(u^{k'})\| = 0$ , i.e.,  $u^*$  is a fixed point of  $\Phi$ .

Finally, we show the limit  $u^*$  is unique for any convergent subsequence of  $\{u^k\}$ . Assume that another subsequence of  $\{u^k\}$ , say  $\{u^{k''}\}$ , converges to  $u^{**} \in C$ . Then both  $\lim_{k \rightarrow \infty} \|u^k - u^*\|^2$  and  $\lim_{k \rightarrow \infty} \|u^k - u^{**}\|^2$  exist, and therefore  $2\langle u^k, u^{**} - u^* \rangle = \|u^k - u^*\|^2 - \|u^k - u^{**}\|^2 - \|u^*\|^2 + \|u^{**}\|^2 \rightarrow c' \in \mathbb{R}$ . Passing  $k \rightarrow \infty$  along subsequences  $\{k'\}$  and  $\{k''\}$  respectively, we have  $2\langle u^*, u^{**} - u^* \rangle = 2\langle u^{**}, u^{**} - u^* \rangle = c'$  and hence  $\|u^* - u^{**}\|^2 = 0$ . Thus, we have shown that  $\lim_{k \rightarrow \infty} u^k = u^*$ .  $\square$



**Theorem 2.11** (Krasnoselskii-Mann). *Let  $C$  be a nonempty, closed, convex subset of  $\mathbb{E}$ , and  $u^{k+1} = (1 - \tau^k)u^k + \tau^k\Psi(u^k)$  for  $k = 0, 1, 2, \dots$  where  $\{\tau^k\} \subset [0, 1]$  s.t.*

$$\sum_{k=0}^{\infty} \tau^k(1 - \tau^k) = \infty,$$

and  $\Psi : C \rightarrow C$  satisfies:

1.  $\Psi$  is nonexpansive.
2.  $\Psi$  has at least one fixed point.

Then  $\{u^k\}$  converges to a fixed point of  $\Psi$ .

*Proof.* Let  $\bar{u} \in C$  be an arbitrary fixed point of  $\Psi$ . Then  $\forall k : \|u^{k+1} - \bar{u}\|^2 = \|(1 - \tau^k)(u^k - \bar{u}) + \tau^k(\Psi(u^k) - \bar{u})\|^2 = (1 - \tau^k)\|u^k - \bar{u}\|^2 + \tau^k\|\Psi(u^k) - \bar{u}\|^2 - \tau^k(1 - \tau^k)\|\Psi(u^k) - u^k\|^2 \leq \|u^k - \bar{u}\|^2 - \tau^k(1 - \tau^k)\|\Psi(u^k) - u^k\|^2$ . Summing up this inequality for all indices in  $l \in [0, k]$ , we have

$$\|u^{k+1} - \bar{u}\|^2 \leq \|u^0 - \bar{u}\|^2 - \sum_{l=0}^k \tau^l(1 - \tau^l)\|(I - \Psi)(u^l)\|^2.$$

This yields: (i)  $\|u^k - \bar{u}\| \searrow c \geq 0$ ; (ii)  $\sum_{k=0}^{\infty} \tau^k(1 - \tau^k)\|(I - \Psi)(u^k)\|^2 < \infty$ .

(ii) further implies  $\liminf_{k \rightarrow \infty} \|(I - \Psi)(u^k)\| = 0$ . Otherwise  $\exists \bar{k} \in \mathbb{N}$ ,  $\epsilon > 0$ , s.t.  $\forall k \geq \bar{k} : \|(I - \Psi)(u^k)\| \geq \epsilon$ , and hence  $\infty > \sum_{k=0}^{\infty} \tau^k(1 - \tau^k)\|(I - \Psi)(u^k)\|^2 \geq \sum_{k=\bar{k}}^{\infty} \tau^k(1 - \tau^k)\|(I - \Psi)(u^k)\|^2 \geq \epsilon^2 \tau^k(1 - \tau^k) = \infty$  yields a contradiction. Moreover,  $\|(I - \Psi)(u^{k+1})\| = \|(1 - \tau^k)(u^k - \Psi(u^k)) + (\Psi(u^k) - \Psi(u^{k+1}))\| \leq (1 - \tau^k)\|u^k - \Psi(u^k)\| + \|u^{k+1} - u^k\| = \|(I - \Psi)(u^k)\|$ . Altogether, we obtain  $\lim_{k \rightarrow \infty} \|(I - \Psi)(u^k)\| = 0$ .

The remainder of the proof is identical to that for Theorem 2.10.  $\square$

**Lemma 2.12** (demiclosedness principle). *Let  $C$  be a nonempty, closed, convex subset of a real Hilbert space  $\mathbb{H}$ , and  $\Phi : C \rightarrow \mathbb{H}$  be nonexpansive. For any sequence  $\{u^k\} \subset C$  s.t.  $\{u^k\}$  weakly converges to  $u \in C$  and  $u^k - \Phi(u^k)$  strongly converges to  $v \in \mathbb{H}$ , we have  $u - \Phi(u) = v$ .*

*Proof.* Since  $\{u^k\}$  weakly converges to  $u^*$  and  $C$  is weakly closed (for being convex and strongly closed), we have  $u \in C$  and  $\Phi(u)$  is well defined. By the nonexpansiveness of  $\Phi$ , we derive

$$\begin{aligned} \|u - \Phi(u) - v\|^2 &= \|u^k - \Phi(u) - v\|^2 - \|u^k - u\|^2 - 2\langle u^k - u, u - \Phi(u) - v \rangle \\ &= \|u^k - \Phi(u^k) - v\|^2 + 2\langle u^k - \Phi(u^k) - v, \Phi(u^k) - \Phi(u) \rangle + \|\Phi(u^k) - \Phi(u)\|^2 - \|u^k - u\|^2 \\ &\quad - 2\langle u^k - u, u - \Phi(u) - v \rangle \\ &\leq \|u^k - \Phi(u^k) - v\|^2 + 2\langle u^k - \Phi(u^k) - v, \Phi(u^k) - \Phi(u) \rangle - 2\langle u^k - u, u - \Phi(u) - v \rangle \rightarrow 0. \end{aligned}$$

Note that, in the last inequality above,  $\Phi(u^k) - \Phi(u) = (\Phi(u^k) - u^k + v) + (u^k - \Phi(u) - v)$  weakly converges to  $u - \Phi(u) - v$  and is, therefore, bounded.  $\square$

**Theorem 2.13** (local quadratic convergence of proximal Newton). *The proximal Newton method converges locally quadratically to the (global) minimizer  $u^*$  if  $\nabla^2 G(u^*)$  is spd.*

*Proof.* Note that  $0 \in \partial F(u^*) + \nabla G(u^*)$ . Let  $u^k$  be in a small neighborhood of  $u^*$  where  $\nabla^2 G(\cdot) \succeq \mu I$  and  $\nabla^2 G$  is  $L$ -Lipschitz for some constants  $\mu, L > 0$ . Note that  $u^{k+1} = u^k + d^k = [\partial F + \nabla^2 G(u^k)]^{-1} \nabla^2 G(u^k)(u^k - [\nabla^2 G(u^k)]^{-1} \nabla G(u^k))$ , and  $[\partial F + \nabla^2 G(u^k)]^{-1} \nabla^2 G(u^k)$  is firmly nonexpansive under the scaled norm  $\|\cdot\|_{\nabla^2 G(u^k)}$ . Hence, the conclusion follows from

$$\begin{aligned}
& \sqrt{\mu} \|u^{k+1} - u^*\| \leq \|u^{k+1} - u^*\|_{\nabla^2 G(u^k)} \\
& = \|[\partial F + \nabla^2 G(u^k)]^{-1} \nabla^2 G(u^k)(u^k - [\nabla^2 G(u^k)]^{-1} \nabla G(u^k)) \\
& \quad - [\partial F + \nabla^2 G(u^k)]^{-1} \nabla^2 G(u^k)(u^* - [\nabla^2 G(u^k)]^{-1} \nabla G(u^*))\|_{\nabla^2 G(u^k)} \\
& \leq \|(u^k - [\nabla^2 G(u^k)]^{-1} \nabla G(u^k)) - (u^* - [\nabla^2 G(u^k)]^{-1} \nabla G(u^*))\|_{\nabla^2 G(u^k)} \\
& \leq \frac{1}{\sqrt{\mu}} \|\nabla^2 G(u^k)(u^k - u^*) - (\nabla G(u^k) - \nabla G(u^*))\| \\
& = \frac{1}{\sqrt{\mu}} \left\| \int_0^1 (\nabla^2 G(u^k) - \nabla^2 G(u^* + t(u^k - u^*))) dt \cdot (u^k - u^*) \right\| \\
& \leq \frac{L}{2\sqrt{\mu}} \|u^k - u^*\|^2.
\end{aligned}$$

□

**Theorem 2.14.** *Assume  $\forall k : \mu I \preceq \nabla^2 J(\tilde{u}^k) \preceq LI$  for some constants  $\mu, L > 0$ . If  $\theta \geq \max\{|1 - \sqrt{\tau\mu}|, |1 - \sqrt{\tau L}|\}^2$ , then  $\text{sr}(A^k) = \sqrt{\theta} \forall k$ .*

*Proof.* For each  $k$ , let  $\nabla^2 J(\tilde{u}^k) = U^k \Lambda^k (U^k)^\top$  be the eigendecomposition of the spd matrix  $\nabla^2 J(\tilde{u}^k)$ , where  $U^k$  is orthogonal and  $\Lambda^k = \text{diag}\{\lambda_i^k\}$ . Then we have

$$\begin{bmatrix} U^k & 0 \\ 0 & U^k \end{bmatrix} \begin{bmatrix} (1+\theta)I - \tau \nabla^2 J(\tilde{u}^k) & -\theta I \\ I & 0 \end{bmatrix} \begin{bmatrix} U^k & 0 \\ 0 & U^k \end{bmatrix}^\top = \begin{bmatrix} (1+\theta)I - \tau \Lambda^k & -\theta I \\ I & 0 \end{bmatrix}, \quad (2)$$

whose eigenvalues consists of those eigenvalues of 2-by-2 blocks:

$$\begin{bmatrix} 1 + \theta - \tau \lambda_i^k & -\theta \\ 1 & 0 \end{bmatrix},$$

i.e., the roots of  $t^2 - (1 + \theta - \tau \lambda_i^k)t + \theta = 0$ . By the assumption, we have  $1 - \sqrt{\theta} \leq \sqrt{\tau \lambda_i^k} \leq 1 + \sqrt{\theta}$ , which implies  $|1 + \theta - \tau \lambda_i^k|^2 - 4\theta \leq 0$ . Therefore, both roots have the same magnitude  $\sqrt{\theta}$ . □

**Proposition 2.15.** *Assume  $S, T$  are spd matrices, and  $\widehat{K} = T^{-1/2} K S^{-1/2}$ . Then*

$$M_{S,T} = \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \succ 0 \Leftrightarrow \widehat{M}_{S,T} = \begin{bmatrix} I & -\widehat{K}^\top \\ -\widehat{K} & I \end{bmatrix} \succ 0 \Leftrightarrow \|T^{-1/2} K S^{-1/2}\| < 1.$$

*Proof.* Note that

$$\begin{aligned}
\begin{bmatrix} S^{1/2} & 0 \\ 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} I & -\widehat{K}^\top \\ -\widehat{K} & I \end{bmatrix} \begin{bmatrix} S^{1/2} & 0 \\ 0 & T^{1/2} \end{bmatrix} &= \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \\
&= \begin{bmatrix} I & -K^\top T^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} S - K^\top T^{-1} K & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} I & 0 \\ -T^{-1} K & I \end{bmatrix}.
\end{aligned}$$

Hence

$$\begin{aligned} \widehat{M}_{S,T} \succ 0 &\Leftrightarrow M_{S,T} \succ 0 \Leftrightarrow S - K^\top T^{-1} K \succ 0 \Leftrightarrow I - S^{-1/2} K^\top T^{-1} K S^{-1/2} \succ 0 \\ &\Leftrightarrow (T^{-1/2} K S^{-1/2})^\top T^{-1/2} K S^{-1/2} \prec I \Leftrightarrow \|T^{-1/2} K S^{-1/2}\| < 1. \end{aligned}$$

□

**Proposition 2.16.** *Given matrix  $K$ , define*

$$S = \text{diag}(\{s_j\}), \quad s_j = \sum_i |K_{ij}|^{2-\theta}, \quad T = \text{diag}(\{t_i\}), \quad t_i = \sum_j |K_{ij}|^\theta,$$

where  $\theta \in [0, 2]$ . Then  $S$  and  $T$  satisfy  $M_{S,T} = \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \succeq 0$ .

*Proof.* For an arbitrary vector  $(u, p)$ , we have

$$\begin{aligned} \begin{bmatrix} u \\ p \end{bmatrix}^\top \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} &= \sum_{i,j} |K_{ij}|^{2-\theta} |u_j|^2 + \sum_{i,j} |K_{ij}|^\theta |p_i|^2 - 2 \sum_{i,j} K_{ij} u_j p_i \\ &\geq \sum_{i,j} \left| |K_{ij}|^{(2-\theta)/2} u_j - |K_{ij}|^{\theta/2} p_i \right|^2 \geq 0. \end{aligned}$$

Hence the conclusion follows.

□