

# Proofs in SS'17 Convex Optimization Lectures\*

Last updated: June 21, 2017

## 1 Convex Analysis

**Theorem 1.1** (separation of convex sets). *Let  $C_1, C_2$  be nonempty convex subsets in  $\mathbb{E}$  such that  $C_1 \cap C_2 = \emptyset$  and  $C_1$  is open. Then there exists a hyperplane separating  $C_1$  and  $C_2$ , i.e.  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  such that*

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

*Proof.* (i) Claim: Let  $C \subset \mathbb{E}$  be closed, convex set, and  $w \in \mathbb{E} \setminus C$ . Then  $\exists v \in \mathbb{E}, v \neq 0, \alpha \in \mathbb{R}$  s.t.  $\langle v, w \rangle > \alpha \geq \langle v, u \rangle \quad \forall u \in C$ .

Consider the projection of  $w$  onto  $C$ , i.e. set  $u^* := \arg \min_{u \in C} \frac{1}{2} \|u - w\|^2$  or, equivalently, let  $\langle u - u^*, u^* - w \rangle \geq 0 \quad \forall u \in C$ .

Now set  $v := w - u^* \neq 0$ . Then  $\forall u \in C$ , 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  $\alpha := \sup\{\langle v, u \rangle : u \in C\}$ . Note  $\alpha < \infty$  since  $\langle v, u \rangle \leq \langle v, u^* \rangle \quad \forall u \in C$ . Thus  $\langle v, w \rangle > \alpha \geq \langle v, u \rangle \quad \forall u \in C$ , which proves the claim.

(ii) Assume  $C_2 = \{\bar{w}\}$  with  $\bar{w} \in C_1$ . Since  $\mathbb{E} \setminus C_1$  is closed,  $\exists w^k \in \mathbb{E} \setminus C_1$  s.t.  $w^k \rightarrow \bar{w}$ . For each  $w^k$ , by (i),  $\exists v^k \in \mathbb{E}$  with  $\|v^k\| = 1$  s.t.  $\langle v^k, w^k \rangle \leq \langle v^k, u^1 \rangle \quad \forall u^1 \in C_1$ . Hence  $v^k \rightarrow \bar{v} \in \mathbb{E}$  along a subsequence s.t.  $\|\bar{v}\| = 1$  and  $\langle \bar{v}, \bar{w} \rangle \leq \langle \bar{v}, u^1 \rangle \quad \forall u^1 \in C_1$ .

(iii) Consider  $C_2$  as a general convex subset of  $\mathbb{E}$ . Set  $C := C_2 - C_1 = \{u^2 - u^1 : u^1 \in C_1, u^2 \in C_2\}$ . Note that  $C$  is a convex, open set, and  $0 \in C$ . By (ii),  $\exists \bar{v}$  with  $\|\bar{v}\| = 1$  s.t.  $\langle -\bar{v}, u^2 - u^1 \rangle \geq \langle -\bar{v}, 0 \rangle = 0$  or, equivalently,  $\langle \bar{v}, u^1 \rangle \geq \langle \bar{v}, u^2 \rangle \quad \forall u^1 \in C_1, u^2 \in C_2$ . Set  $\alpha := \sup\{\langle \bar{v}, u^2 \rangle : u^2 \in C_2\}$ , then we conclude that  $\langle \bar{v}, u^1 \rangle \geq \alpha \geq \langle \bar{v}, u^2 \rangle \quad \forall u^1 \in C_1, u^2 \in C_2$ .  $\square$

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

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

Let  $M := \sup\{J(v) : v \in B_\epsilon(u)\} < \infty$ . By convexity of  $J$ ,  $\forall v \in B_\epsilon(u) : J(v) \geq 2J(u) - J(2u - v) \geq 2J(u) - M$ . Thus,  $\|J\|_{B_\epsilon(u)} := \sup\{|J(v)| : v \in B_\epsilon(u)\} \leq M + 2|J(u)|$ .

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  $t \leq \|w - v\| / \|z - v\| \leq 2\|w - v\| / \epsilon$  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  $u \in \text{ri dom } J$  and  $n$  be the dimension of the affine hull of  $\text{dom } J$ , then  $\exists \{\alpha^i\}_{i=1}^{n+1} \subset (0, 1), \{u^i\}_{i=1}^{n+1} \subset \text{dom } J$  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 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  $u$ .  $\square$

---

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

**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  $J(\alpha \bar{u} + (1 - \alpha)u^*) \leq J(u^*) - \alpha(J(u^*) - J(\bar{u})) \forall \alpha \in [0, 1]$ . 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.  $\lim_{k \rightarrow \infty} J(u^k) = \inf_{u \in \mathbb{E}} J(u) > -\infty$ . Since  $\{J(u^k)\}$  is uniformly bounded from above, by coercivity of  $J$ ,  $\{u^k\}$  is 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  $\partial J$  is a monotone operator, i.e.  $\forall u^1, u^2 \in \text{dom } J$ ,  $\xi^1 \in \partial J(u^1)$ ,  $\xi^2 \in \partial J(u^2)$ :

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

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

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

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

**Theorem 1.7.** 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,  $\exists (\xi, -\alpha) \in \mathbb{E} \times \mathbb{R}$  s.t.  $(\xi, -\alpha) \neq (0, 0)$ ,  $\alpha \geq 0$  by our choice, and  $\langle (\xi, -\alpha), (u - v, J(u) - J(v)) \rangle \geq 0 \forall v \in \text{dom } J$ . In fact, we must have  $\alpha > 0$  since otherwise  $\xi = 0$ . Thus, we conclude that  $\xi/\alpha \in \partial J(u)$ .

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

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

(iv) convexity. Let  $v \in \mathbb{E}$  be arbitrarily given,  $\xi, \eta \in \partial J(u)$ . Then we have

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

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

**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.  $\xi^* \in \partial J(u^*)$  whenever

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

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

**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 all  $u \in \text{dom } J$ ,  $p \in \text{dom } J^*$ , we have  $J(u) + J^*(p) \geq \langle u, p \rangle$ . The equality holds iff  $p \in \partial J(u)$ .

*Proof.*  $J(u) + J^*(p) \geq \langle u, p \rangle$  follows directly from the definition of convex conjugate;  $p \in \partial J(u)$  is the sufficient and necessary condition for:  $u = \arg \min_{v \in \mathbb{E}} J(v) - \langle v, p \rangle$ .  $\square$

**Theorem 1.11.** Assume  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  and  $J^{**} = (J^*)^*$  is 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  $\langle p, u \rangle - J^*(p) \leq J(u)$  by Theorem 1.10, we assert  $J^{**}(\cdot) \leq J(\cdot)$ .

(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  $d := \frac{1}{2}(J(\bar{u}) - J^{**}(\bar{u})) > 0$ . Since  $(\bar{u}, J(\bar{u}) - d) \notin \text{epi } J$  and  $\text{epi } J$  is closed, by Theorem 1.1,  $\exists(\bar{p}, -1) \in \mathbb{E} \times \mathbb{R}$  s.t.  $\langle (\bar{p}, -1), (\bar{u}, J(\bar{u}) - d) \rangle \geq \langle (\xi, -1), (u, \alpha) \rangle \forall (u, \alpha) \in \text{epi } J$ . In particular,  $\langle \bar{p}, \bar{u} \rangle - J(\bar{u}) + d \geq \langle \bar{p}, \bar{u} \rangle - J(\bar{u}) \forall u$ . 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  $J^{**}(\bar{u}) \geq J(\bar{u})$  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 1.12.** Assume that  $J : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  is proper, convex, and lsc. Then  $J$  is  $\mu$ -strongly convex iff  $J^*$  has  $\frac{1}{\mu}$ -Lipschitz gradient.

*Proof.* 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 (Frechét-)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  $\|u - v\| \leq \frac{1}{\mu} \|p - q\|$ .  $\square$

**Theorem 1.13** (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 solution pair iff*

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

*Proof.* Define  $\Phi(\cdot) := \inf_u \{F(Ku + \cdot) + G(u)\}$ . Since  $\forall v^1, v^2 \in \mathbb{R}^m$ ,  $\forall \alpha \in [0, 1] : \alpha \Phi(v^1) + (1 - \alpha) \Phi(v^2) = \inf_{u^1} \{\alpha F(Ku^1 + v^1) + \alpha G(u^1)\} + \inf_{u^2} \{(1 - \alpha) F(Ku^2 + v^2) + (1 - \alpha) G(u^2)\} = \inf_{u^1, u^2} \{\alpha F(Ku^1 + v^1) + (1 - \alpha) F(Ku^2 + v^2) + \alpha G(u^1) + (1 - \alpha) G(u^2)\} \geq \inf_u \{\alpha F(Ku + v^1) + (1 - \alpha) F(Ku + v^2) + G(u)\} \geq \inf_u \{F(Ku + \alpha v^1 + (1 - \alpha)v^2) + G(u)\} = \Phi(\alpha v^1 + (1 - \alpha)v^2)$ , 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  $M < \infty$ . Hence, 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 1.14** (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 1.15.** *Let  $F, G : \mathbb{E} \rightarrow \overline{\mathbb{R}}$  be proper, convex, and lsc. Then*

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

*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 directly from the 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 the curvature conditions are both satisfied with  $0 < c_1 < c_2 < 1$  for each  $k$ . In addition,  $\nabla J(\cdot)$  is  $\mu$ -Lipschitz on  $\{u \in \mathbb{E} : J(u) \leq J(u^0)\}$ . Then we have  $\sum_{k=0}^{\infty} \frac{|\langle \nabla J(u^k), d^k \rangle|^2}{\|d^k\|^2} < \infty$ .*

*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  $J : \mathbb{E} \rightarrow \mathbb{R}$  is convex with  $\mu$ -Lipschitz gradient. 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  $J : \mathbb{E} \rightarrow \mathbb{R}$  is convex with  $\mu$ -Lipschitz gradient. 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 a (finite) minimum 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,  $\langle \nabla J(v) - \nabla J(u), v - u \rangle = D_J(u, v) + D_J(v, u) \geq \|\nabla J(v) - \nabla J(u)\|^2$  and  $\nabla J$  is firmly nonexpansive by Proposition 2.6.  $\square$

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