Solutions to *Convex Analysis and Optimization*

Yunwei Ren

Contents

1 Basic Convexity Concepts

1.2 Convex Sets and Functions

1.

Proof. For every $y \in (\lambda_1 + \lambda_2)C$, there is an $x \in C$ such that

$$y = (\lambda_1 + \lambda_2)x = \lambda_1 x + \lambda_2 x.$$

Since $\lambda_i x \in \lambda_i C$, $(i = 1, 2,)$ $y \in \lambda_1 C + \lambda_2 C$. Thus, $(\lambda_1 + \lambda_2)C \subset \lambda_1 C + \lambda_2 C$. For the converse, suppose that $y_i = \lambda_i x_i \in \lambda_i C$. Then

$$\lambda_1 x_1 + \lambda_2 x_2 = (\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} x_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_2\right) = (\lambda_1 + \lambda_2) z.$$

By the convexity of C, $z \in C$. Hence, $\lambda_1 x_1 + \lambda_2 x_2 \in (\lambda_1 + \lambda_2)C$. Namely, $(\lambda_1 + \lambda_2)C \supset \lambda_1 C + \lambda_2 C$.

If C is not convex, the statement may be false. For example, put $n = 1$, $C = \{0, 1\}$ and $\lambda_1 = \lambda_2 = 1$. Then $(\lambda_1 + \lambda_2)C = \{0, 2\}$ but $\lambda_1 C + \lambda_2 C = \{0, 1, 2\}$. □

2.(d, e)

Proof.

(d) Let C be a cone and $x \in \bar{C}$. Then there is a sequence $\{x_k\} \subset C$ with $x_k \to x$. For every positive λ, Clear that $\lambda x = \lim_{k \to \infty} \lambda x_k$ and $\lambda x_k \in C$. Namely, $\{\lambda x_k\} \subset C$ converges to λx. Hence, $\lambda x \in \bar{C}$. Thus, $\bar{C}$ is a cone.

(e) Let T a linear transformation on $\mathbb{R}^n$. Suppose $y = Tx$ for some $x \in C$. Then $\lambda y = \lambda T x = T(\lambda x) \in T(C)$. Hence, $T(C)$ is a cone. Suppose that there is an $v \in C$ such that $Tu = v$. Then $T(\lambda u) = \lambda v \in C$. Hence, the inverse image is also a cone. □

3. Lower Semicontinuity under Composition

Proof.

(a) For every $x \in \mathbb{R}^n$ and $\{x_k\}$ converging to x, put $y_k = f(x_k)$. Since f is continuous, $y_k \to y = f(x)$. Hence,

$$\liminf_{k \to \infty} h(x) = \liminf_{k \to \infty} g(y_k) \geq g(y) = h(x).$$

Namely, h is lower semicontinuous.

(b) First we show that for every $\{y_k\} \subset \mathbb{R}$,

$$\liminf_{k \to \infty} g(y_k) \geq g\left(\liminf_{k \to \infty} y_k\right). \tag{1}$$

Put $y = \liminf y_k$. Since $y_k \geq y$ for every k and g is nondecreasing, $g(y_k) \geq g(y)$. Hence, $\liminf g(y_k) \geq g(y)$.

For every $x \in \mathbb{R}^n$ and $\{x_k\}$ converging to x, put $y_k = f(x_k)$. Since f is lower semicontinuous, $\liminf y_k \geq f(x)$. Hence,

$$\liminf_{k \to \infty} h(x) = \liminf_{k \to \infty} g(y_k) \geq g\left(\liminf_{k \to \infty} y_k\right) \geq g(f(x)),$$

where the second and third inequalities come from (1) and the monotonicity of g respectively. Thus, h is lower semicontinuous.

To show that the monotonic nondecrease assumption is essential, put $n = 1$ and define both f and g by

$$f(x) = g(x) = \begin{cases} 1, & x < 0 \\ -1, & x \geq 0 \end{cases}.$$

Clear that both f and g are lower semicontinuous but $h = g \circ f$ takes value -1 for $x < 0$ and 1 for $x \geq 0$ and therefore is not lower semicontinuous. $\square$

4. Convexity under Composition

Proof.

(a) For every $\lambda \in [0, 1]$ and $x, y \in C$,

$$\begin{aligned} h(\lambda x + (1 - \lambda)y) &= g(f(\lambda x + (1 - \lambda)y)) \\ &\leq g(\lambda f(x) + (1 - \lambda)f(y)) \\ &\leq \lambda h(x) + (1 - \lambda)h(y), \end{aligned}$$

where the first inequality comes from the monotonicity of g and convexity of f, and the second one comes from the convexity of g. Thus, h is convex. If g is increasing and f is strictly convex, then the first inequality is strict, provided $\lambda \in (0, 1)$ and $x \neq y$. Therefore, h is strictly convex.

(b) It follows from a similar argument. $\square$

5. Examples of Convex Functions

Proof.

(a) For every $x \in \operatorname{dom} f$,

$$\nabla^2 f_1(x) = K \left\{ \left[\frac{1}{x_i x_j} \right]_{ij} - n \operatorname{diag} \left\{ \frac{1}{x_i^2} \right\}_i \right\},$$

where $K = -(x_1 \cdots x_n)^{1/n}/n^2 < 0$. For each $y \in \operatorname{dom} f_1$,

$$y^T \nabla^2 f_1(x) y = \frac{K}{n^2} \left\{ \left(\frac{\sum_{1=1}^n y_i/x_i}{n} \right)^2 - \frac{1}{n} \sum_{i=1}^n \frac{y_i^2}{x_i^2} \right\} \geq 0,$$

where the inequality comes from the RMS-AM inequality. Hence, f_1 is convex.

(b) For every $x \in \mathbb{R}^n$,

$$\nabla^2 f_2 = K \left\{ [e^{x_i} e^{x_j}]_{ij} - \left(\sum_{i=1}^n e^{x_i} \right) \operatorname{diag}\{e^{x_i}\} \right\},$$

where $K = -1/(e^{x_1} + \cdots + e^{x_n})^2 < 0$. For each $y \in \mathbb{R}^n$, put $a = (e^{x_1/2}, \ldots, e^{x_n/2})$ and $b = (y_1 e^{x_1/2}, \ldots, y_n e^{x_n/2})$. Then

$$\begin{aligned} y^T \nabla^2 f_2(x) y &= K \left\{ \left(\sum_{i=1}^n y_i e^{x_i} \right)^2 - \left(\sum_{i=1}^n e^{x_i} \right) \left(\sum_{i=1}^n y_i^2 e^{x_i} \right) \right\} \\ &= K\{(a^T b)^2 - (a^T a)(b^T b)\} \geq 0, \end{aligned}$$

where the inequality comes from the Cauchy-Schwarz inequality. Thus, f_2 is convex.

(c) Since $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ is convex over $\mathbb{R}^n$ and the function $x \mapsto x^p$ $(p \geq 1)$ is convex and nondecreasing on $[0, \infty)$, f_3 is convex by Prob. 1.4(a).

(d) $-f$ is convex and negative, and the function $x \mapsto -1/x$ is convex and nondecreasing on $(-\infty, 0)$, so, by Prob. 1.4(a), $f_4 = -1/(-f)$ is convex.

(e) The function $g : x \mapsto \alpha x + \beta$ is convex and nondecreasing on $\mathbb{R}$. Hence, $f_5 = g \circ f$ is convex by Prob. 1.4(a).

(f) The function $g : x \mapsto e^{\beta x}$ is convex and nondecreasing on $\mathbb{R}$ and the function $h : x \mapsto x^T A x$ is convex since A is positive semidefinite. Hence, by Prob. 1.4(a), $f_6 = g \circ h$ is convex.

(g) For every $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,

$$\begin{aligned} f_7(\lambda x + (1-\lambda)y) &= f(A(\lambda x + (1-\lambda)y) + b) \\ &= f(\lambda(Ax + b) + (1-\lambda)(Ay + b)) \\ &\leq \lambda f_7(x) + (1-\lambda) f_7(y), \end{aligned}$$

where the inequality comes from the convexity of f. Hence, f_7 is convex. $\square$

6. Ascent/Descent Behavior of a Convex Function

Proof.

(a) Let $\lambda \in (0, 1)$ be such that $x_2 = \lambda_1 x_1 + (1-\lambda)x_3$. Then

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{\lambda f(x_1) + (1-\lambda) f(x_3) - f(x_1)}{\lambda x_1 + (1-\lambda)x_3 - x_1} = \frac{f(x_3) - f(x_1)}{x_3 - x_1}.$$

Similarly, we can show that

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} \geq \frac{f(x_3) - f(x_1)}{x_3 - x_1}.$$

Thus,

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}.$$

$\square$

7. Characterization of Differentiable Convex Functions

Proof. If f is convex over C, then by Proposition 1.2.5,

$$f(y) - f(x) \geq \nabla f(x)^T (y - x), \quad f(x) - f(y) \geq \nabla f(y)^T (x - y)$$

for every $x, y \in C$. Sum up these two inequalities and we get

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0. \tag{2}$$

For the converse, we first prove a lemma: If $h : (a, b) \to \mathbb{R}$ is differentiable and its derivative is nondecreasing, then it is convex. By the mean value theorem, for every $x, y \in (a, b)$, $h(y) - h(x) = h'(\xi)(y - x)$ where ξ is between x and y. Since h' is nondecreasing, this implies that $h(y) - h(x) \geq h'(x)(y - x)$. Thus, h is convex.

Now we suppose (2) holds for every $x, y \in C$. Define $h : [0,1] \to \mathbb{R}^n$ by $h(t) = x + t(y-x)$ and put $g = f \circ h$. Then

$$Dg(t) = \nabla f(h(t))^T (y-x).$$

Hence, for $1 \geq t_2 > t_1 \geq 0$,

$$Dg(t_2) - Dg(t_1) = (\nabla f(h(t_2)) - \nabla f(h(t_1))^T \frac{h(t_2) - h(t_1)}{t_2 - t_1} \geq 0.$$

Namely, Dg is nondecreasing. By our lemma, g is convex. Since the choice of $x, y \in C$ are arbitrary, we conclude that f is convex over C. □

8. Characterization of Twice Continuously Differentiable Convex Functions

Proof. We may assume without loss of generality that $0 \in C$ and, in consequence, $S = \operatorname{aff}(C)$. If $\dim S = 0$, then there is nothing to prove. Suppose $m = \dim S > 0$, let $Z \in \operatorname{Hom}(\mathbb{R}^m, S)$ be isometric[1] and define $g : \mathbb{R}^m \to \mathbb{R}$ by $u \mapsto f(Zu)$. Clear that g is also twice continuously differentiable and $\nabla^2 g = Z^T \nabla^2 f Z$.

First we suppose that $y^T \nabla^2 f(x) y \geq 0$ for all $x \in C$ and $y \in S$. Since Z is an isometry, this implies that $u^T Z^T \nabla^2 f(x) Z u \geq 0$ for all $u \in \mathbb{R}^m$. Namely, $\nabla^2 g(x)$ is positive semidefinite on $\mathbb{R}^m$. Therefore, by Prop. 1.2.6, g is convex. Thus, $f = g \circ Z^{-1}$ is also convex.

Now we suppose that f is convex over C and assume, to obtain a contradiction, that there is some $x \in C$ and $y \in S$ such that $y^T \nabla^2 f(x) y < 0$. Suppose $y = Zu$. Then this implies that $u^T \nabla^2 g(x) u < 0$. However, since g is convex (as f is) and $\mathbb{R}^m$ is open, by Prop. 1.2.6(c), $\nabla^2 g(x)$ should be positive semidefinite on $\mathbb{R}^m$. Contradiction. Thus, $y^T \nabla^2 f(x) y \geq 0$ for all $x \in C$ and $y \in S$. □

9. Strong Convexity

Proof.

(a) Note that (1.16) implies that when restricted to the line segment connecting x and y, the function f has strictly increasing gradient. Hence, the argument in Prob. 1.7, *mutatis mutandis*, gives a proof of (a).

(b) First we suppose that $\nabla^2 f(x) - \alpha I$ is positive semidefinite. Then for every $y, x \in \mathbb{R}^n$, there exists some $\theta \in (0,1)$ and $z = x + \theta(y-x)$ such that

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^T (y-x) + (y-x)^T \nabla^2 f(z)(y-x) \\ &= f(x) + \nabla f(x)^T (y-x) + (y-x)^T (\nabla^2 f(z) - \alpha I)(y-x) + \alpha \|y-x\|^2 \\ &\geq f(x) + \nabla f(x)^T (y-x) + \alpha \|y-x\|^2. \end{aligned} \tag{3}$$

Meanwhile, since $\nabla^2 f(x)$ is positive semidefinite, f is convex and therefore

$$f(y) - f(x) \leq \nabla f(y)^T (y-x). \tag{4}$$

The previous two inequalities imply (1.16), i.e., f is strongly convex with coefficient α.

[1] Consider the linear transformation X which maps an orthonormal basis of $\mathbb{R}^m$ to an orthonormal basis of S. It can be verified that X is an isometry and is bijective.

Now suppose that (1.16) holds. For fixed x, let $u \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Then there exists some $\theta_1, \theta_2 \in (0,1)$ such that

$$f(x+tu) = f(x) + \nabla f(x)^T tu + \frac{t^2}{2} u^T \nabla^2 f(x+\theta_1 tu) u,$$
$$f(x) = f(x+tu) - \nabla f(x+tu)^T tu + \frac{t^2}{2} u^T \nabla^2 f(x+\theta_2 tu) u.$$

Add these two equations and we get

$$\frac{t^2}{2} u^T (\nabla^2 f(x+\theta_1 tu) + \nabla^2 f(x+\theta_2 tu)) u = (\nabla f(x+tu) - \nabla f(x))^T tu \geq \alpha \|tu\|^2.$$

Namely,

$$\frac{1}{2} u^T (\nabla^2 f(x+\theta_1 tu) + \nabla^2 f(x+\theta_2 tu)) u \geq \alpha \|u\|^2.$$

Let $t \to 0$ and we obtain

$$u^T \nabla^2 f(x) u \geq \alpha \|u\|^2.$$

Hence, all eigenvalues of $\nabla^2 f(x)$ are no less than α and, in consequence, $\nabla^2 f(x) - \alpha I$ is positive semidefinite. □

11. Arithmetic-Geometric Mean Inequality

Proof. Since the function $x \mapsto -\log x$ is strictly convex on $(0, \infty)$.

$$\begin{aligned} -\log(\alpha_1 x_1 + \cdots + \alpha_n x_n) &\leq -\alpha_1 \log x_1 - \cdots - \alpha_n \log x_n \\ &= -\log(x_1^{\alpha_1} \cdots x_n^{\alpha_n}), \end{aligned}$$

where the equality is obtained when $x_1 = \cdots = x_n$. Thus, $x_1^{\alpha_1} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + \cdots + \alpha_n x_n$ with equality iff $x_1 = \cdots = x_n$. □

12.

Proof. If $x = 0$ or $y = 0$, then the inequality is trivial. If both x and y are nonzero, then, by Prob. 1.11, $x^{1/p} y^{1/q} \leq x/p + y/q$. Replace x and y with x^p and y^q respectively and we get $xy \leq x^p/p + y^q/q$.

If all y_i are zero or all x_i are zero, then the inequality is trivial. If there exists some nonzero y_i and some nonzero x_i, then, by the homogeneity, we may assume without loss of generality that

$$\sum_{i=1}^n |x_i|^p = \sum_{i=1}^n |y_i|^q = 1.$$

Then, by Young's inequality,

$$\sum_{i=1}^n |x_i y_i| \leq \frac{1}{p} \sum_{i=1}^n |x_i|^p + \frac{1}{q} \sum_{i=1}^n |y_i|^q = \frac{1}{p} + \frac{1}{q} = 1.$$

Namely, Holder's inequality holds. □

13.

Proof. For $x \notin \operatorname{dom} f$, $f(x) = \inf \varnothing = \infty$. For every $x_1, x_2 \in \operatorname{dom}(f)$, since C is convex, $x_\theta = (1-\theta)x_1 + \theta x_2 \in \operatorname{dom}(f)$. By definition, for every $\varepsilon > 0$, there exists some $(x_1, w_1), (x_2, w_2) \in C$ such that $w_i < f(x_i) + \varepsilon$. Hence,

$$(1-\theta)w_1 + \theta w_2 < (1-\theta)f(x_1) + \theta f(x_2) + \varepsilon.$$

Since C is convex, $(1-\theta)(x_1, w_1) + \theta(x_2, w_2) \in C$ and therefore

$$f(x_\theta) \le (1-\theta)w_1 + \theta w_2.$$

These two inequalities, together with the fact that the choice of ε is arbitrary, imply that $f(x_\theta) \le (1-\theta)f(x_1) + \theta f(x_2)$. Thus, f is convex. □

1.3 Convex and Affine Hulls

14.

Proof. Given $\varnothing \neq X \subset \mathbb{R}^n$, let C be the collection of all convex combination of elements of X. Clear that $X \subset C$. Meanwhile, for every $x, y \in C$, they are the convex combination of points in X and therefore so is $(1-\theta)x + \theta y$ for every $\theta \in (0,1)$. Hence, C is a convex set containing X. Thus, $\operatorname{conv}(X) \subset C$. For every $x \in C$, x is a convex combination of points in X and therefore is contained in any convex set containing X; See Fig. 1.3.1. Hence, $x \in \operatorname{conv}(C)$. Thus, $C = \operatorname{conv}(C)$. □

15.

Proof. Let $D = \bigcup_{x \in C}\{\gamma x : \gamma \ge 0\}$. It follows immediately from the definition that $D \subset \operatorname{cone}(C)$. For every $x \in \operatorname{cone}(C)$. If $x = 0$, then clear that $x \in D$. If $x \neq 0$, then it can be written as $x = \alpha_1 x_1 + \cdots + \alpha_m x_m$ where $m > 0$, $\alpha_i > 0$ and $x_i \in C$. Hence

$$x = \frac{1}{\alpha}\sum \frac{\alpha_i}{\alpha} x_i \quad \text{where } \alpha = \sum \alpha_i.$$

Since C is convex, $\sum \alpha_i x_i / \alpha \in C$ and therefore $x \in D$. Thus, $D = \operatorname{cone}(C)$. □

16.

Proof.

(a) First we show that C is closed. Suppose that $\{x_k\} \subset C$ converges to some $x \in \mathbb{R}^n$. Then for every $i \in I$ and $k = 1, 2, \ldots$, $a_i^T x_k \le 0$. Let $k \to \infty$, by the continuity of the inner product, $a_i^T x \le 0$. Hence, C is closed.

For the convexity, let $x, y \in C$ and $\theta \in (0,1)$. Then for every $i \in I$,

$$a_i^T((1-\theta)x + \theta y) = (1-\theta)a_i^T x + \theta a_i^T y \le 0.$$

Namely, $(1-\theta)x + \theta y \in C$. Thus, C is convex.

Finally, since for all $\lambda > 0$, $a_i^T(\lambda x) \le 0$ as long as $a_i^T x \le 0$. Hence, C is cone. Thus, we conclude that C is a closed convex cone.

(b) Let C be a cone. Suppose that C is convex, then for every $x, y \in C$, $(x+y)/2 \in C$. Hence, $x + y = 2((x+y)/2) \in C$ as C is a cone. Namely, $C + C \subset C$. For the converse, suppose that $C + C \subset C$. For every $x, y \in C$ and $\theta \in (0, 1)$, since C is a cone, $(1-\theta)x, \theta y \in C$ and therefore $(1-\theta)x + \theta y \in C + C \subset C$. Hence, C is convex.

(c) For every $x \in C_1$ and $y \in C_2$,

$$x + y = \frac{1}{2}(2x) + \frac{1}{2}(2y) = \operatorname{conv}\{2x, 2y\} \subset \operatorname{conv}(C_1 \cup C_2).$$

Hence, $C_1 + C_2 \subset \operatorname{conv}(C_1 \cup C_2)$. For the converse, we show that $C_1 + C_2$ is a convex set containing $C_1 \cup C_2$. Since $0 \in C_1$, $C_2 \subset 0 + C_2 \subset C_1 + C_2$. Similarly, $C_1 \subset C_1 + C_2$. Meanwhile, by Prop. 1.2.1(b), $C_1 + C_2$ is convex. Hence, $\operatorname{conv}(C_1 \cup C_2) \subset C_1 + C_2$. Thus, $\operatorname{conv}(C_1 \cup C_2) = C_1 + C_2$.

Since C_1 and C_2 are cones, for $\alpha \in (0, 1)$, $C_1 = \alpha C_1$ and $C_2 = (1-\alpha)C_2$ and therefore $C_1 \cap C_2 = \alpha C_1 \cap (1-\alpha)C_2$. For $\alpha \in \{0, 1\}$, $\alpha C_1 \cap (1-\alpha)C_2 = \{0\} \in C_1 \cap C_2$. Thus, $C_1 \cap C_2 = \bigcup_{\alpha \in [0,1]} (\alpha C_1 \cap (1-\alpha)C_2)$. □

18. Convex Hulls, Affine Hulls, and Generated Cones

Proof.

(a) We may assume without loss of generality that $0 \in X$, so that the affine hulls are subspaces of $\mathbb{R}^n$. Since X is contained by $\operatorname{conv}(X)$ and $\operatorname{cl}(X)$, $\operatorname{aff}(X)$ is contained by $\operatorname{aff}(\operatorname{conv}(X))$ and $\operatorname{aff}(\operatorname{cl}(X))$. For the converse, note that a convex combination of points in X is also a linear combination, hence $\operatorname{conv}(X) \subset \operatorname{aff}(X)$ and therefore $\operatorname{aff}(\operatorname{conv}(X)) \subset \operatorname{aff}(X)$. Meanwhile, since finite dimensional vector spaces are all closed, $\operatorname{cl}(X) \subset \operatorname{aff}(X)$ and therefore $\operatorname{aff}(\operatorname{cl}(X)) \subset \operatorname{aff}(X)$. Thus, $\operatorname{aff}(X) = \operatorname{aff}(\operatorname{conv}(X)) = \operatorname{aff}(\operatorname{cl}(X))$.

(b) Clear that $\operatorname{cone}(X) \subset \operatorname{cone}(\operatorname{conv}(X))$. For the converse, suppose $x \in \operatorname{cone}(\operatorname{conv}(X))$. If $x = 0$, then $x \in \operatorname{cone}(X)$ in a trivial way. If $x \neq 0$, then $x = \alpha_1 x_1 + \cdots + \alpha_p x_p$ where $x_i \in \operatorname{conv}(X)$, $p > 0$ and $\alpha_i > 0$. Meanwhile, for each i, suppose that $x_i = \beta_{i,1} x_{i,1} + \cdots + \beta_{i,q} x_{i,q}$ where $q > 0$, $\beta_{i,j} > 0$ and $\sum_j \beta_{i,j} = 1$. Hence,

$$x = \sum_i \alpha_i \sum_j \beta_{i,j} x_{i,j} = \sum_{i,j} \alpha_i \beta_{i,j} x_{i,j}.$$

Namely, x is a positive combination of points in X and therefore $x \in \operatorname{cone}(X)$. Hence, $\operatorname{cone}(\operatorname{conv}(X)) \subset \operatorname{cone}(X)$. Thus, $\operatorname{cone}(\operatorname{conv}(X)) = \operatorname{cone}(X)$.

(c) Since $\operatorname{conv}(X) \subset \operatorname{cone}(X)$, $\operatorname{aff}(\operatorname{conv}(X)) \subset \operatorname{aff}(\operatorname{cone}(X))$. Let $X = [-1, 1] \times \{1\} \subset \mathbb{R}^2$. Then clear that $\operatorname{aff}(\operatorname{conv}(X))$ is the line crossing $(0, 1)$ and parallel to the x-axis while $\operatorname{aff}(\operatorname{cone}(X)) = \mathbb{R}^2$.

(d) Since $0 \in \operatorname{conv}(X) \subset \operatorname{cone}(X)$, both $\operatorname{aff}(\operatorname{conv}(X))$ and $\operatorname{aff}(\operatorname{cone}(X))$ are subspaces of $\mathbb{R}^n$. By part (c), we already have $\operatorname{aff}(\operatorname{conv}(X)) \subset \operatorname{aff}(\operatorname{cone}(X))$. Hence, we only need to show that $\dim \operatorname{aff}(\operatorname{conv}(X)) \geq \dim \operatorname{aff}(\operatorname{cone}(X))$ to complete the proof. Suppose that $\dim \operatorname{aff}(\operatorname{cone}(X)) = m$. By Prop. 1.3.1, there exists $b_1, \ldots, b_m \in X$ such that linearly independent and span $\operatorname{aff}(\operatorname{cone}(X))$. Note that $\{b_1, \ldots, b_m\}$ is also a set of linearly independent set in $\operatorname{aff}(\operatorname{conv}(X))$. Hence, $\dim \operatorname{aff}(\operatorname{conv}(X)) \geq m$. Thus, $\operatorname{aff}(\operatorname{conv}(X)) = \operatorname{aff}(\operatorname{cone}(X))$. □

19.

Proof. We denote these two representation by f and g respectively. For every $(x, w) \in \operatorname{conv}(\bigcup_{i \in I} \operatorname{epi}(f_i))$, there exists some positive $\alpha_1, \ldots, \alpha_m$ with $\sum \alpha_j = 1$ and $(x_1, w_1), \ldots,$

$(x_m, w_m) \in \bigcup \operatorname{epi}(f_i)$ such that $(x, w) = \sum_j \alpha_j (x_j, w_j)$. Namely, for fix x,

$$f(x) = \inf \left\{ \sum_j \alpha_j w_j : x = \sum_j \alpha_j x_j,\ (x_j, w_j) \in \bigcup_i \operatorname{epi}(f_i),\ \alpha_j \geq 0, \sum_j \alpha_j = 1, m > 0 \right\}.$$

By the definition of epi, $(x_j, w_j) \in \bigcup_i \operatorname{epi}(f_i)$ implies $f_{i_j}(x_j) \leq w_j$ for some i_j. Hence, $f(x) \geq g(x)$. Meanwhile, since the union of graphs of f_i is contained in $\bigcup \operatorname{epi}(f_i)$, $f(x) \leq g(x)$. Thus, $f(x) = g(x)$. □

20. Convexification of Nonconvex Functions

Proof.

(a) The convexity follows from Prob. 13 immediately. For each x, let f_x takes value $f(x)$ and ∞ for other points. Then $\{f_x\}$ is a collection of convex functions. Then, by Prob. 19, F has the representation given.

(b) Put $M = \inf_{x \in \operatorname{conv}(X)} F(x)$. By definition, for all $y \in X \subset \operatorname{conv}(X)$, $M \leq F(y)$ and $F(y) \leq f(y)$. Hence, $M \leq \inf_{y \in X} f(y)$. For the converse, again by definition, for every $\varepsilon > 0$, there exists some $x \in \operatorname{conv}(X)$ such that $M + \varepsilon \geq F(x)$. By part (a), this implies there exists nonnegative $\alpha_1, \ldots, \alpha_m$ with $\sum \alpha_i = 1$ and $x_1, \ldots, x_m \in X$ such that $\sum \alpha_i x_i = x$ and $M + \varepsilon \geq \sum \alpha_i f(x_i)$. Since $\sum \alpha_i f(x_i)$ is a weighted average of values of f, it is no less than $\inf_{x \in X} f(x)$. Since the choice of $\varepsilon > 0$ is arbitrary, we conclude that $M \geq \inf_{x \in X} f(x)$. Thus, $\inf_{x \in \operatorname{conv}(X)} F(x) = \inf_{x \in X} f(x)$.

(c) It follows immediately from part (b). □

21. Minimization of Linear Functions

Proof. Note that the convexification of $f : X \to \mathbb{R}$ is just $c^T x$ with domain $\operatorname{conv}(X)$. Hence, the equation follows from Prob. 20. Suppose that the infimum of the left-hand side is attained, that is, there is some $x^* \in \operatorname{conv}(X)$ such that $c^T x^* = \inf_{x \in \operatorname{conv}(X)} c^T x$. Then by the definition of the convex hull, x^* is the convex combination of some points $x_1, \ldots, x_m$ of X and, as $c^T x$ is linear, $c^T x^*$ is the weighted average of $c^T x_1, \ldots, c^T x_m$. As a consequence, $c^T x^* \geq \min\{c^T x_1, \ldots, c^T x_m\}$. Thus, the infimum in the right-hand side can also be attained. For the converse, it is obvious. □

22. Extension of Caratheodory's Theorem

Proof. TODO □

23.

Proof. Since X is bounded, $\operatorname{cl}(X)$ is also bounded and therefore compact. Hence, by Prop. 1.3.2, $\operatorname{conv}(\operatorname{cl}(X))$ is compact. In consequence, $\operatorname{cl}(\operatorname{conv}(\operatorname{cl}(X))) = \operatorname{conv}(\operatorname{cl}(X))$. Thus, $\operatorname{cl}(\operatorname{conv}(X)) \subset \operatorname{cl}(\operatorname{conv}(\operatorname{cl}(X))) = \operatorname{conv}(\operatorname{cl}(X))$. For the converse, it follows from the fact that $\operatorname{conv}(\operatorname{cl}(\operatorname{conv}(X))) = \operatorname{cl}(\operatorname{conv}(X))$ and $\operatorname{conv}(\operatorname{cl}(X)) \subset \operatorname{conv}(\operatorname{cl}(\operatorname{conv}(X)))$. Thus, $\operatorname{cl}(\operatorname{conv}(X)) = \operatorname{conv}(\operatorname{cl}(X))$.

If X is compact, then it is bounded and closed. Hence, $\operatorname{conv}(X) = \operatorname{conv}(\operatorname{cl}(X)) = \operatorname{cl}(\operatorname{conv}(X))$. Namely, $\operatorname{conv}(X)$ is also closed. Meanwhile, $\operatorname{conv}(X)$ is bounded as X is. Thus, $\operatorname{conv}(X)$ is compact. □

24. Radon's Theorem

Proof. TODO $\square$

25. Helly's Theorem [Hel21]

Proof. We use induction on the size of the collection. If the size is no more than $n+1$, then the statement clearly holds. Assume that, for all collection of no more than M sets, the statement holds. We show that the statement holds for every collection of $M+1$ sets.

Let $C_1, \ldots, C_{m+1}$ be a collection of $M+1$ convex sets. For each $j = 1, \ldots, M+1$, put $B_j = \bigcap_{i \neq j} C_i$. By the induction hypothesis, all B_j are nonempty. Choose $x_j \in B_j$ ($j = 1, \ldots, M+1$). Note that $M+1 \geq n+2$. Hence, by Radon's Theorem, we can partition $\{1, \ldots, M+1\}$ into to two sets P and Q such that

$$D = \operatorname{conv}(\{x_p : p \in P\}) \cap \operatorname{conv}(\{x_q : q \in Q\}) \neq \varnothing.$$

Let $x \in D$ and we show that $x \in \bigcap C_j$ to complete the proof. By the construction of B_j, we know that for each $p \in P$, $x_p \in C_q$ for every $q \in Q$. Since all C_q are convex, x, a convex combination of x_p, belongs to all C_q. Similarly, we can show that x belongs to all C_p. Thus, $x \in \bigcap C_j$. Namely, the intersection of $C_1, \ldots, C_{M+1}$ is nonempty. $\square$

26.

Proof. First, clear that for any I, $\inf_x \max_i f_i(x) \leq f^*$. For the converse, we assume, to obtain a contradiction, that for all index set I with no more than $n+1$ indices, $\inf_x \max_i f_i(x) < f^*$. Then, putting $X_i = \{x : f_i(x) < f^*\}$, $i = 1, \ldots, M$, this implies that every subcollection of $X_1, \ldots, X_M$, provided it contains no more than $n+1$ sets, has nonempty intersection. Meanwhile, X_i are convex sets as f_i are convex functions. Hence, by Helly's theorem, $\bigcap_{i=1}^M X_i$ is nonempty, which contradicts the infimum assumption of f^*. Thus, there exists some I such that $\inf_x \max_i f_i(x) \geq f^*$ and therefore the two values coincide. $\square$