1 Convex sets

1.1 Definitions

Convex hull of S: $\text{conv}\, S = \left\{\sum_{i=1}^k \theta_i x_i : k \in [\![1, |S|]\!], \forall i, x_i \in S, \theta \succeq 0, \mathbf{1}_k^T \theta = 1\right\}$.
Hyperplane: $\left\{x : a^T x = b\right\}$ (affine). ***Halfspace***: $\left\{x : a^T x \leq b\right\}$ (convex).
Euclidean ball (or ball) in $\mathbb{R}^n$: $B(x_c, r) = \{x : ||x - x_c||_2 \leq r\}, r > 0$ (convex).
Ellipsoid: $\mathcal{E} = \left\{x : (x - x_c)^T P^{-1}(x - x_c) \leq 1\right\}$ where $P \in \mathcal{S}_n^{++}$. Another representation is $\{x_c + Au : ||u||_2 \leq 1\}$ with A square and nonsingular (convex).
Norm cone: $C = \{(x, t) : ||x|| \leq t\} \subset \mathbb{R}^{n+1}$. $\|\cdot\|_2$: second-order cone (convex).
Polyhedron: $\mathcal{P} = \{x : Ax \preceq b, Cx = d\}$ (convex).

1.2 Operations that preserve convexity

Show convexity of C: (i) definition, (ii) operations that preserve convexity.
Intersection: the intersection of (any number of) convex sets is convex.
Affine functions: the image or inverse image by an affine function preserves convexity (e.g. scaling, translation, projection).
Perspective function: $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\text{dom}\, P = \mathbb{R}^n \times \mathbb{R}_+^*$, $P(z, t) = z/t$. If $C \subset \text{dom}\, P$ convex, then $P(C)$ convex. If $C \subset \mathbb{R}^n$ convex, then $P^{-1}(C)$ convex.
Linear-fractional function: compose the perspective function with an affine function, i.e. $f(x) = \frac{Ax+b}{c^T x + d}$, $\text{dom}\, f = \left\{x : c^T x + d > 0\right\}$ (same results about convexity).

1.3 Separating and supporting hyperplanes

Separating hyperplane theorem. C, D nonempty disjoint convex sets. Then there exist $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.
Definition. The ***supporting hyperplane*** to set C at a boundary point x_0 is defined by $\left\{x : a^T x = a^T x_0\right\}$ where $a \neq 0$ and for all $x \in C$, $a^T x \leq a^T x_0$.
Supporting hyperplane theorem. C convex. Then there exists a supporting hyperplane at every boundary point of C.

1.4 Dual cones and generalized inequalities

Proper cone: $K \subset \mathbb{R}^n$ convex, closed, solid ($\textbf{int}\, K \neq \emptyset$), pointed (contains no line i.e., if $x, -x \in K$, then $x = 0$)). ***Generalized inequality***, $x \preceq_K y$ iff $y - x \in K$.
Dual cone. K a cone. $K^* = \left\{y : \forall x \in K, x^T y \geq 0\right\}$ is the ***dual cone*** of K. K^* is a cone, and is always convex. Geometrically, $y \in K^*$ if and only if $-y$ is the normal of a hyperplane that supports K at the origin.
Properties. K, K_1, K_2 cones. (i) K^* closed and convex, (ii) $K_1 \subset K_2$ implies $K_2^* \subset K_1^*$, (iii) if $\textbf{int} K \neq \emptyset$, then K^* pointed, (iv) if $\overline{K}$ is pointed then $\textbf{int} K^* \neq \emptyset$, (v) $K^{**} = \overline{\text{conv}\, K}$ (hence if K is convex and closed, $K^{**} = K$). These properties show that if K is a proper cone, then so is its dual K^*, and moreover, that $K^{**} = K$.

2 Convex optimization problems

2.1 Basic properties and examples

Convex: $ax + b$, e^{ax}, x^α ($\alpha \in \mathbb{R} \setminus (0, 1)$), $|x|^p$ ($p \geq 1$), $x \log x$, norms (e.g. $\lambda_{\max}(X^T X)^{1/2}$).
Concave: $ax + b$, x^α on $\mathbb{R}_{++}$ ($\alpha \in [0, 1]$), $\log x$ on $\mathbb{R}_{++}$.
Property. $f : \mathbb{R}^n \to \mathbb{R}$ convex iff $t \mapsto f(x + tv)$ convex for any $x \in \text{dom}\, f, v \in \mathbb{R}^n$.
1st-order cond.: diff. f with convex domain is convex iff $f(y) \geq f(x) + \nabla f(x)^T (y - x)$.
2nd-order cond.: twice diff. f with convex domain is convex iff $\nabla^2 f(x) \succeq 0$.
α***-sublevel set***: $C_\alpha = \{x \in \text{dom}\, f, f(x) \leq \alpha\}$ (convex if f convex).

2.2 Operations that preserve convexity

Establishing convexity: (i) definition, (ii) Hessian, (iii) operations that preserve convexity.
Nonnegative weighted sum, composition with affine function (e.g. log barrier), **pointwise maximum/supremum, composition** ($h \circ g$ convex if g, h convex, $\tilde{h}$ nondecreasing or if g concave, h convex, $\tilde{h}$ nonincreasing), **minimization on a convex set** ($\inf_y x^T Ax + 2x^T By + y^T Cy = x^T (A - BC^{-1}B^T)x$), **perspective** ($g(x, t) = tf(x/t)$).

2.3 Definitions

Conjugate function: $f^*(y) = \sup_x \left(y^T x - f(x)\right)$ convex.
Quasiconvex: if sublevel sets are convex. **Jensen**: $f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}$.
Log-concave: $f(\theta x + (1 - \theta)y) \geq f(x)^\theta f(y)^{1-\theta}$ (e.g. $x^a, a \geq 0$, normal density, CDF Gaussian). Twice diff. f log-concave iff $f(x)\nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T$.

2.4 Convex optimization problem

Feasible set of a convex optimization problem is convex.

Linear program (LP):

$$\begin{array}{ll} \min_x & c^T x + d \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$$

Quadratic program (QP):

$$\begin{array}{ll} \min_x & \frac{1}{2} x^T P x + q^T x + r \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$$

Quadratically constrained quadratic program (QCQP):

$$\begin{array}{ll} \min_x & \frac{1}{2} x^T P x + q^T x + r \\ \text{s.t.} & \frac{1}{2} x^T P_i x + q_i^T x + r_i \preceq 0, Ax = b \end{array}$$

Second order cone program (SOCP):

$$\begin{array}{ll} \min_x & f^T x \\ \text{s.t.} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, Fx = g \end{array}$$

Chebyshev center: of $\mathcal{P} = \left\{x, a_i^T x \leq b_i, i = 1, \ldots, m\right\}$ is center of largest inscribed ball $\mathcal{B} = \{x_c + u, \|u\|_2 \leq r\}$. $a_i^T x \leq b_i$ for all $x \in \mathcal{B}$ iff $\sup\left\{a_i^T(x_c + u), \|u\|_2 \leq r\right\} = a_i^T x_c + r\|a_i\|_2 \leq b_i$. x_c, r determined by solving the LP: $\max r$ s.t. $a_i^T x_c + r\|a_i\|_2 \leq b_i$.
Perron-Frobenius eigenvalue: exists for (elementwise) positive square A. Real positive eigenvalue of A, equal to $\max_i |\lambda_i(A)|$. $A^k \sim \lambda_{pf}^k$.
Properties. $\lambda_{\max}(A) \leq t$ iff $A \preceq tI$. $\|A\|_2 \leq t$ iff $A^T A \preceq t^2 I$.

3 Duality

3.1 Theory

Lagrangian: $(f_i \leq 0,\ h_i = 0)$. $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$.
Lagrange dual function: $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$ (concave). If $\lambda \succeq 0, g(\lambda, \nu) \leq p^*$.
Dual norm: $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$. $\|y\|^* = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise.
Lagrange dual problem: maximize $g(\lambda, \nu)$ s.t. $\lambda \succeq 0$.
Slater's condition: for a convex problem, if $\exists x \in \operatorname{relint} \mathcal{D},\ s.t.\ f_i(x) < 0,\ ,\ Ax = b$ (strict feasibility), then strong duality holds (can be relaxed to $f_i(x) \leq 0$ for affine constraints) and dual optimal value is attained.
Complementary slackness: suppose x^* primal optimal, (λ^*, ν^*) dual optimal, then x^* minimizes $L(x, \lambda^*, \nu^*)$ over x and $\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m$.
KKT conditions: suppose f_i and h_i are differentiable, let x^* primal optimal, (λ^*, ν^*) dual optimal, then

- $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^m \nu_i^* \nabla h_i(x^*) = 0$ (first-order condition).
- $f_i(x^*) \leq 0,\ i = 1, \ldots, m$ (primal feasibility).
- $h_i(x^*) = 0,\ i = 1, \ldots, m$ (primal feasibility).
- $\lambda_i^* \geq 0,\ i = 1, \ldots, m$ (dual feasibility).
- $\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m$ (complementary slackness).

KKT conditions for convex problem: KKT condition are also sufficient if the problem is convex, e.g, if $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ satisfy the KKT conditions, then $\tilde{x}$ is primal optimal, $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal and we have 0 duality gap.
KKT with Slater's condition : if Slater's condition holds, x is optimal i.i.f there are (λ, ν) s.t (x, λ, ν) satisfy the KKT conditions.

3.2 Examples

$$\begin{array}{ll} \min\limits_x & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \max\limits_x & -b^T \mu \\ \text{subject to} & A^T \mu + c \succeq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_x & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \qquad \begin{array}{ll} \max\limits_x & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0 \\ & \lambda \succeq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_x & \|x\| \\ \text{subject to} & Ax = b \end{array} \qquad \begin{array}{ll} \max\limits_x & -b^T \mu \\ \text{subject to} & \|A^T \lambda\| \leq 1 \end{array}$$

$$\begin{array}{ll} \min\limits_x & \log\det X^{-1} \\ \text{subject to} & a_i^T X a_i \leq 1,\ i = 1, \ldots, m \end{array} \qquad \begin{array}{ll} \max\limits_x & -\log\det(\sum_{i=1}^m \lambda_i a_i a_i^T) \\ & -1^T \lambda + n \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_x & \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0 \end{array} \qquad \begin{array}{ll} \max\limits_x & -\frac{1}{2} q(\lambda)^T P(\lambda) q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

4 Extras

Dual norm: $\|z\|_* = \sup\{z^T x \,|\, \|x\| \leq 1\}$. $\forall x, z \in \mathbb{R}^n,\ z^T x \leq \|x\| \|z\|_*$.
Singular values: $\sigma_{\max}(A) = \sup_{x \neq 0, y \neq 0} \frac{x^T A y}{\|x\|_2 \|y\|_2}$.
Schur complement: Let $X \in \mathbf{S}^n$, $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ where $A \in \mathcal{S}^k$, If $\det A \neq 0$, $S = C - B^T A^{-1} B$ is the Schur complement of A in X.

- $\det X = \det A \det S$, • $\inf_u \begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = v^T S v$
- $X \succ 0 \iff A \succ 0$ and $S \succ 0$, • if $A \succ 0$ then $X \succeq 0 \iff S \succeq 0$

Taylor's approximation: $\hat{f}(x + v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f v$.
Newton's method: $x_{n+1} \leftarrow x_n - \alpha_n (\nabla^2 f(x))^{-1} \nabla f(x)$.
Some Gradients:

- $\nabla_x(a^T x + b) = a$, • $\nabla_x(\frac{1}{2} x^T A x) = \frac{1}{2}(A^T + A)x$, • $\nabla_x(\operatorname{Tr}(A^T X + b) = A$.
- $\nabla_x(\det(X)) = \bar{X}$, $\bar{X}$ comatrix of X ($\bar{X} = \det(X) X^{-T}$).
- $\nabla_x(\log\det(X)) = X^{-1}$, • $f(X) = X^{-1} \Rightarrow \nabla_x f(H) = -X^{-1} H X^{-1}$.