Lecture 8. Inequalities. Optimization problems.

Example of composition: log-sum-exp. norm (a,b) of matrices.

$\|A\|_{a,b} = \sup_{w\neq 0} \frac{\|Aw\|_a}{\|w\|_b}$ $\quad f_w(A) = \|Aw\|_a$ convex. $\|\cdot\|_{a,b} = \sup_{\|w\|_b=1} f_w(A)$.

Jensen's inequality: $f(\theta x + \bar{\theta} y) \leq \theta f(x) + \bar{\theta} f(y)$ convex.

generalization: $f(\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \theta_2 f(x_2) + \cdots + \theta_k f(x_k)$

where $\theta_1, \theta_2, \cdots \theta_k \geq 0$ and $\theta_1 + \theta_2 + \cdots + \theta_k = 1$.

integrals: if $p(x) \geq 0$ defined on $\Omega \subseteq \operatorname{dom} f$. $\int_\Omega p(x)\,dx = 1$ then.

$f\left(\int_\Omega p(x)\, x\, dx\right) \leq \int_\Omega f(x) p(x)\, dx$. provided the integrals exist.

probability: if x is a random variable. s.t. $\Pr[x \in \operatorname{dom} f] = 1$. then

$f(\mathbb{E}[x]) \leq \mathbb{E}[f(x)]$. provided the expectations exist.

Recall Cauchy-Schwarz. $\langle x, y\rangle \leq \|x\|_2 \cdot \|y\|_2$.

$\langle x, y\rangle \leq \|x\|_3 \|y\|_?$ $\qquad \|x\|_3 \leq \|x\|_2$?

proposition. given $1 \leq p_1 < p_2 \leq \infty$. $x_i \geq 0$. $\|x\|_{p_1} \geq \|x\|_{p_2}$.

Proof. if $p_2 = \infty$. $\|x\|_\infty = \max_{1\leq i\leq n} |x_i| \leq \left(\sum |x_i|^{p_1}\right)^{1/p_1}$.

let $q = \left(\sum |x_i|^{p_1}\right)^{1/p_1}$ $\quad \tilde{x}_i = x_i / q$. $\quad \|\tilde{x}\|_{p_1} = 1$.

$\|\tilde{x}\|_{p_2} = \Big(\sum \big(\underbrace{|\tilde{x}_i|^{p_1}}_{\in[0,1]}\big)^{p_2/p_1}\Big)^{1/p_2} \leq \left(\sum |\tilde{x}_i|^{p_1}\right)^{1/p_2} = 1$. $\square$

Hölder's inequality: let $p, q \in (1, \infty)$ be conjugate exponents.

i.e. $1/p + 1/q = 1$. then $\langle x, y \rangle \le \|x\|_p \|y\|_q \quad \forall x, y \in \mathbb{R}^n$.

or $\sum_{i=1}^{n} |x_i y_i| \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$.

Remark. let $p = q = 2$. it is Cauchy–Schwarz.

Proof. w.l.o.g assume $x_i \ge 0$, $y_i \ge 0$, and $\langle x, y \rangle > 0$.

let $\tilde{x} = x/\|x\|_p$, $\tilde{y} = y/\|y\|_q$. goal: $\sum |\tilde{x}_i \tilde{y}_i| \le 1$.

We first show that $u^{1/p} v^{1/q} \le u/p + v/q \quad \forall u, v \ge 0$.

if $uv = 0$. trivial. o.w. $\log(u/p + v/q) \ge \frac{1}{p}\log u + \frac{1}{q}\log v$.

$\Rightarrow |\tilde{x}_i| \cdot |\tilde{y}_i| \le \frac{1}{p}|\tilde{x}_i|^p + \frac{1}{q}|\tilde{y}_i|^q \Rightarrow \sum |\tilde{x}_i \tilde{y}_i| \le \frac{1}{p} + \frac{1}{q} = 1$. □

Minkowski's inequality: verify triangle inequality for L_p-norms.

for $1 \le p \le \infty$. $\|x+y\|_p \le \|x\|_p + \|y\|_p$

Proof. only consider $\|x+y\|_p > 0$. (the right hand side is nonnegative).

$\|x+y\|_p^p = \sum |x_i + y_i|^p \le \sum |x_i| \cdot |x_i + y_i|^{p-1} + |y_i| \cdot |x_i + y_i|^{p-1}$

By Hölder's inequality $\langle u, v \rangle \le \|u\|_p \|v\|_{p/(p-1)}$

$\Rightarrow \sum |x_i| \cdot |x_i + y_i|^{p-1} \le \|x\|_p \left(\sum |x_i + y_i|^p\right)^{(p-1)/p} = \|x\|_p \|x+y\|_p^{p-1}$

$\Rightarrow \|x+y\|_p^p \le (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1} \Rightarrow$ Minkowski's ineq. □

Convex Optimization Problems

standard forms: $\min f(x)$ — objective function

(P). s.t. $g_i(x) \le 0$, $i = 1, 2, \cdots, m$.

$h_i(x) = 0$. $i = 1, 2, \cdots, k$.

constraint function

domain of problem (P) is

$$D \triangleq \operatorname{dom} f \cap \left(\bigcap_{i=1}^{m} \operatorname{dom} g_i \right) \cap \left(\bigcap_{i=1}^{k} \operatorname{dom} h_i \right).$$

the feasible set is (P is feasible if $\Omega \neq \phi$).

$$\Omega \triangleq \{ x \in D : g_i(x) \le 0,\ 1 \le i \le m;\ h_i(x) = 0,\ 1 \le i \le k \}$$

the optimal value of P is $f^* \triangleq \inf_{x \in \Omega} f(x)$.

Remark. allow f^* to take the extended values $\pm\infty$.

- $f^* = \infty$ if P is infeasible. i.e. $\Omega = \phi$. $\sup \phi = -\infty$, $\inf \phi = \infty$.
- $f^* = -\infty$ if $f(x)$ is unbounded below.
- x^* is an optimal point solving (P). if $x^* \in \Omega$ s.t. $f(x^*) = f^*$
- x^* is a locally optimal if $f(x^*) \le f(x)$, $\forall \|x - x^*\| < \delta$ for some $\delta > 0$

optimization $\to$ convex optimization : f. g_i convex. h_i affine.

domain: $D = \operatorname{dom} f \cap (\cap \operatorname{dom} g_i)$ — convex sets.

feasible set: $\Omega = \{ x \in D : g_i(x) \le 0,\ h_j(x) = 0 \}$ — minimizing convex functions over convex sets.

α-sublevel sets.

Example. $\min f(x) = x_1^2 + x_2^2$. s.t. $\begin{cases} g(x) = x_1/(x_2^2+1) \le 0 \\ h(x) = (x_1+x_2)^2 = 0. \end{cases}$

- f is convex. domain D. feasible set $\Omega = \{x : x_1 + x_2 = 0.\ x_1 \le 0\}$ convex.
- but still not a convex opt. since g is not convex. h is not affine.

equivalent (but not identical) to a convex. opt.

$$\min f(x) = x_1^2 + x_2^2. \quad \text{s.t. } g(x) = x_1 \le 0,\ h(x) = x_1 + x_2 = 0.$$

Properties of convex optimization problems.

- any local minimum is a global minimum.
- the set of optimal points $\underbrace{\Omega_{opt} = \{x^* : f(x^*) \le f(x),\ \forall x\}}_{\text{is convex}}$
- in particular. if f strictly convex. at most one optimal point.

First-order optimality condition.

For a convex opt. whose objective f is differentiable in an open convex set. a feasible point x^* is optimal iff.

$$\nabla f(x^*)^T (x - x^*) \geq 0. \quad \forall x \in \Omega. \text{ feasible set.}$$

Corollary. in particular. if unconstrained. x^* optimal iff $\nabla f(x^*) = 0$.

Proof. "$\Leftarrow$" By the first-order condition for convexity.

$$f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*) \geq f(x^*).$$

"$\Leftarrow$". assume x^* is an optimal point. fix $x \in \Omega$. $d = x - x^*$

$\forall \alpha \in [0,1]$. $x^* + \alpha d = \alpha x + \bar{\alpha} x^* \in \Omega$. let $g(\alpha) = f(x^* + \alpha d)$.

x^* optimal $\Rightarrow g(\alpha) \geq g(0) \Rightarrow (g(\alpha) - g(0))/\alpha \geq 0$.

Taking the limit as $\alpha \downarrow 0 \Rightarrow g'(0) = \nabla f(x^*)^T d \geq 0$. $\square$

Canonical Convex Optimization. Problems.

Linear program: $\min_x c^T x$ s.t. $Bx \leq d$. $Ax = b$.

Standard form: $\min_x c^T x$ s.t. $Ax = b$. $x \geq 0$.

- adding slack variables. s. $\min_{x,s} c^T x$. s.t. $Bx + s = d$. $s \geq 0$.
- splitting variables into positive and negative parts $x = x^+ - x^-$.

$\min_{x^+, x^-, s} c^T x^+ - c^T x^-$. s.t. $Bx^+ - Bx^- + s = d$. $Ax^+ - Ax^- = b$. $x^+, x^-, s \geq 0$.

Quadratic program. and quadratically constrained quadratic program.

QP: $\min \frac{1}{2} x^T Q x + c^T x$. s.t. $Bx \leq d$. $Ax = b$.

QCQP: $\min \cdots$ s.t. $\frac{1}{2} x^T Q_i x + c_i^T x + d_i \leq 0$. $Ax = b$.

QCQP is convex if $Q \succeq 0$ and $Q_i \succeq 0$. $\forall i$.

Example: linear least squares regression: given $y \in \mathbb{R}^n$. $X \in \mathbb{R}^{n \times p}$.

goal: find $w \in \mathbb{R}^p$ s.t. $\min_w \|y - Xw\|_2^2$.