I: Multivariate functions – norms, continuity, differentials.

II: Convex sets and unconstrained minimization.

Part I.

$\boxed{\text{Norms:}}$ $f(x)$ is a norm if

1) $f(x) \geq 0$, $f(x) = 0 \Rightarrow x = 0$
2) $f(x+y) \leq f(x) + f(y) \quad \forall x, y$
3) $f(\alpha x) = |\alpha| f(x)$.

Vector norms ($\mathbb{R}^n$)

def. p-norms: $\|x\|_p = \left[\sum_{i=1}^n |x_i|^p\right]^{1/p}$

important examples: $\|x\|_1 = \sum_{i=1}^n |x_i|$

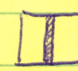
$\|x\|_2 = (x^T x)^{\frac{1}{2}}$

$\|x\|_\infty = \max_i \{|x_i|\}$.

Matrix norms ($\mathbb{R}^{m \times n}$)

$\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$.

$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p = 1} \|Ax\|_p$

$\|A\|_1$; $x = e_i$ $\|Ae_i\|_1$ maximum $\rightarrow$ 

$\|A\|_\infty$; $x_i = \pm 1$ $\|Ax\|_\infty$ maximum $\rightarrow$ 

$\|A\|_2$: square $\rightarrow$ $x^T A^T A x$, max when x is eigenvector with max eigen value, so $\|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$.

ineq. Holder ineq. $|x^T y| \leq \|x\|_p \cdot \|y\|_q \quad (\frac{1}{p} + \frac{1}{q} = 1)$

$p = q = 2$: Cauchy-Schwartz: $|x^T y| \leq \|x\|_2 \cdot \|y\|_2$.

$\|AB\|_p \leq \|A\|_p \cdot \|B\|_2$

$\|Ax\|_p \leq \|A\|_p \cdot \|x\|_p$ (by definition)

$\|Ax\|_\beta \leq \|A\|_{\alpha,\beta} \cdot \|x\|_\alpha$, $\|A\|_{\alpha,\beta} := \sup_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}$.

Norm equivalence: two norms $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$.

$c_1 \|x\|_\alpha \leq \|x\|_\beta \leq c_2 \|x\|_\alpha$

e.g. $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\, \|x\|_2$

$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\, \|x\|_\infty$

$\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty$.

(optional)

$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\, \|A\|_2$

$\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq m \cdot n \cdot \max_{i,j} |a_{ij}|$

$\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\, \|A\|_\infty$

$\frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\, \|A\|_1$

$\|A\|_2 \leq \sqrt{\|A\|_1 \cdot \|A\|_\infty}$.

$\|A\|_2 = \|A^T\|_2$.

- Invariance under rotations: Q, P orthogonal : $\|QAP\|_F = \|A\|_F$, $\|QAP\|_2 = \|A\|_2$

Continuity

Let function f map from (closed) domain $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$. Then f is said to be continuous <u>at</u> $x_0 \in D$ if

$$\lim_{x \to x_0} f(x) = f(x_0)$$ ~~holds with~~

which holds if for all $\varepsilon > 0$, $\exists\, \delta > 0$:

$$\|x - x_0\| < \delta \text{ and } x \in D \Rightarrow \|f(x) - f(x_0)\| < \varepsilon.$$

Continuous on D if this holds for all $x_0 \in D$.

<u>Lipschitz - continuity</u> -

Function f is Lipschitz continuous, if $\exists\, L > 0$:

$$\|f(x_1) - f(x_0)\| \leq L \|x_1 - x_0\| \quad \forall\, x_0, x_1 \in D.$$

Locally Lipschitz at x_0 if there exists a neighborhood $\mathcal{N}$ around x_0:

$$\|f(x_1) - f(x_0)\| \leq L \|x_1 - x_0\| \quad \forall\, x \in \mathcal{N} \cap D.$$

Derivatives

<u>Univariate</u> ($\mathbb{R}^1$)

$$\frac{df}{dx} = f'(x) := \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

$$\frac{d^2 f}{dx^2} = f''(x) := \lim_{\varepsilon \to 0} \frac{f'(x+\varepsilon) - f'(x)}{\varepsilon}.$$

<u>Multivariate</u> ($\mathbb{R}^n$, $\mathbb{R}^{m \times n}$)

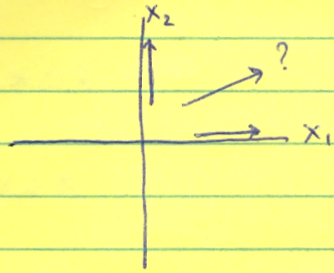
Gradient: $\nabla f(x) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}$, $\dfrac{\partial f}{\partial x_i} := \lim\limits_{\varepsilon \to 0} \dfrac{f(x+\varepsilon e_i) - f(x)}{\varepsilon}$.

e_i = i-th column of $n \times n$ identity mat.

Hessian: $\nabla^2 f(x) = \begin{bmatrix} \partial f/\partial x_1 \partial x_1 & \partial f/\partial x_1 \partial x_2 & \cdots & \partial f/\partial x_1 \partial x_n \\ \partial f/\partial x_2 \partial x_1 & & \cdots & \partial f/\partial x_2 \partial x_n \\ \vdots & & & \vdots \\ \partial f/\partial x_n \partial x_1 & & \cdots & \partial f/\partial x_n \partial x_n \end{bmatrix}$.

Chain rule: $\nabla f(g(x)) = \sum_{i=1}^{n} \dfrac{\partial f}{\partial x_i} \nabla g_i(x)$.

Directional derivative



$$D(f(x);p) := \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon p) - f(x)}{\varepsilon} \quad \boxed{= \nabla f(x)^T p.}$$

Evaluating the gradient/Hessian

Ex. 1 $\quad f(x) = v^T x \quad = \sum v_i \cdot x_i \quad \rightarrow \quad \frac{\partial f}{\partial x_i} = v_i \quad$ or $\quad \boxed{\nabla f(x) = v}$.

Ex. 2 $\quad f(x) = x^T M x, \quad M = [\vec{m}_1, \vec{m}_2, \dots, \vec{m}_n], \quad f(x) = x^T \sum_i \vec{m}_i \cdot x_i = \sum_i x^T \vec{m}_i \, x_i$

$$= \sum_i x_i \Big(\sum_j x_j M_{ij} \Big)$$

For $\frac{\partial f}{\partial x_k}$ consider: $\sum_i x_i (x_k M_{ik}) + \sum x_k \sum_j x_j M_{kj} - x_k^2 M_{kk}$

$$\Rightarrow \sum_i x_i M_{ki} + \sum_j x_j M_{jk} \quad \Rightarrow \boxed{\nabla f(x) = Mx + M^T x}.$$

Special case: $M = M^T$: $\quad 2Mx$

$M = I$: $\quad f(x) = x^T x, \quad \nabla f(x) = 2x.$

Alternative view: $\quad \underset{\text{fix}}{\underline{x^T}} M x \qquad v = (M^T x)$

$x^T M \underset{\text{fix}}{\underline{x}} \qquad v = (Mx)$

$\Bigg\} \; M^T x + Mx.$

Ex. 3 $\quad f(x) = \frac{1}{2} \|Ax - b\|_2^2 \quad = \frac{1}{2} x^T A^T A x - \frac{1}{2} b^T A x - \frac{1}{2} x^T A^T b + \frac{1}{2} b^T b$

$$= \underbrace{\tfrac{1}{2} x^T A^T A x}_{} - \underbrace{b^T A}_{v} x + \tfrac{1}{2} b^T b.$$

$$\tfrac{1}{2} \cdot (2 A^T A x) - A^T b = A^T A x - A^T b = A^T (Ax - b).$$

See further: wikipedia & links: matrix calculus.

Mean value theorem

$x \in \mathbb{R}^1$: $\quad f(x_1) = f(x_0) + f'(y)(x_1 - x_0) \qquad x_0 < y < x_1$.

$x \in \mathbb{R}^n$: $\quad f(x+p) = f(x) + \nabla f(x+\alpha p)^T p \qquad f$ along line $x + \gamma p$
for $\alpha \in (0,1)$

twice differentiable functions:

$$f(x+p) = f(x) + \nabla f(x)^T p + \tfrac{1}{2} p^T \left(\nabla^2 f(x+\alpha p)\right)^T p, \qquad \alpha \in (0,1)$$

(one form of Taylor's theorem).

$$f(x+\alpha p) = f(x) + \alpha \nabla f(x)^T p + \tfrac{1}{2}\alpha^2 p^T \left(\nabla^2 f(x)\right)^T p + O(\alpha^3)$$

where $\|p\|_2 = 1$.

II. Convex Sets.

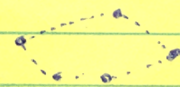
~~Examples: $\mathbb{R}^n$, $\{x \mid \|x\|_2 \le 1\}$, ...~~

Characterization: convex combination of any two points in ~~D~~ C lies in ~~D~~ C:

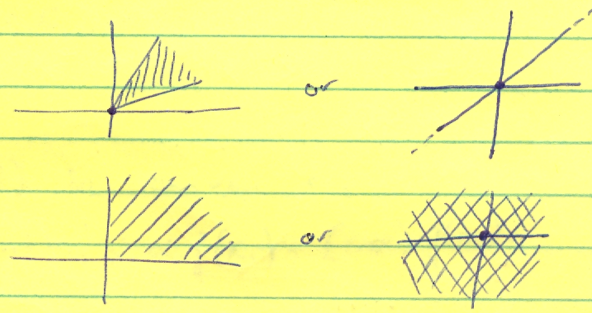
$$\alpha x_1 + (1-\alpha) x_2 \in C \qquad 0 \le \alpha \le 1.$$

Examples

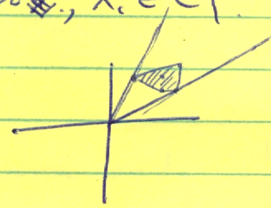
- Convex hull: $\{\sum \alpha_i x_i \mid \alpha_i \geq 0, \ \sum \alpha_i = 1\}$.



- Cones: set C is a convex cone if for any $x_1, x_2 \in C$, $\alpha_1, \alpha_2 \geq 0$, $\alpha_1 x_1 + \alpha_2 x_2 \in C$.

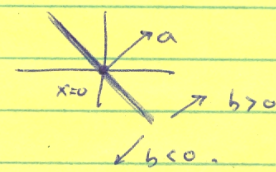


- Conic hull: $\{\sum_i \alpha_i x_i \mid \alpha_i \geq 0, \ x_i \in C\}$.



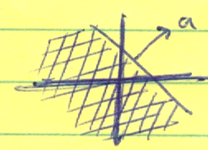
① convex hull
② cone of the hull

- Hyper planes: $\{x \mid a^T x = b\}$ $b=0$

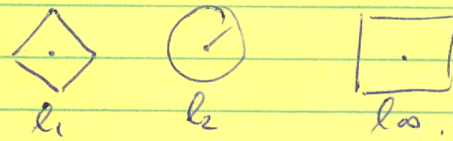


$x \perp a$.

- Half spaces: $\{x \mid a^T x \leq b\}$



- Balls: $\{x \mid \|x\| \leq r\}$



- Polyhedra: $\{x \mid a_j^T x \leq b_j$ or $A^T x \leq b\}$
(Geometry of linear inequalities)

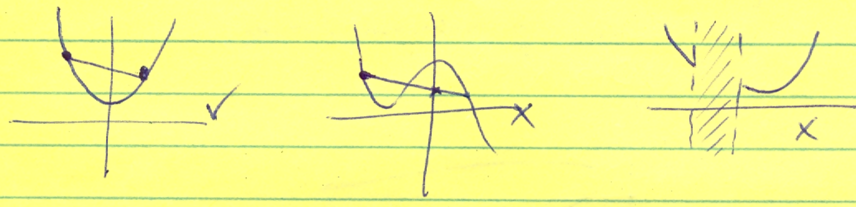
Operations that preserve convexity.

- intersection $S_1 \cap S_2$
- Affine functions: $f(x) = Ax + b$. $\{f(x) \mid x \in C\}$.
- Scaling and translation: $\alpha x + d$
- sum: $\{x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2\}$
- Cartesian product $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\forall x_1 \in C_1, x_2 \in C_2$.

Convex functions.

- $f(x)$ is convex if $dom(f)$ is convex and

$$f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2).$$



- Differentiable f: $dom(f)$ is convex and $f(y) \geq f(x) + \nabla f(x)^T (y-x)$ (NS)

- Second order: $dom(f)$ is convex, $\nabla^2 f(x) \succeq 0$: symmetric pos. def. $x^T H x \geq 0$.

Operations on convex functions

Unconstrained minimization $\min_x f(x)$.

- Minimizers:
  - global: $f(x^*) \leq f(x)$ $\forall x$
  - local: $f(x^*) \leq f(x)$ $\forall x \in N$, N is neighbourhood around x^*
  - strict local: " " , $x \neq x^*$.
- Stationary point: $\nabla f(x) = 0$.
- Conditions.

first-order necessary: If x^* is a local minimizer and f is cont. diff. in an open neighbourhood around x^*, then $\nabla f(x^*) = 0$.

$\nabla f(x^*) \neq 0 \Rightarrow x^*$ is not local min

- Second-order necessary

  If x^* is a local minimizer and $\nabla^2 f$ is cont. in open neighbourhood around x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$.

- Second-order sufficient

  Suppose $\nabla^2 f$ is cont. in an open neighbourhood of x^*, and $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$. Then x^* is a <u>strict</u> local minimizer of f.

* Convex functions:
- any local minimizer x^* is a global minimizer of f.
- f differentiable $\Rightarrow$ any stationary point x^* ($\nabla f(x^*) = 0$) is a global minimizer of f.

<u>Operations on convex functions.</u>

Examples: norms, max, ...
- weighted sums (≥ 0)
- $g(x) = f(Ax + b)$
- pointwise max.

<u>Convergence rates.</u>

Sequence $\{x_k\}$

Converge with order-r, largest r such that

$$0 \leq \lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} < \infty$$

(asymptotic convergence rate)

$r = 1$: linear,
$r = 2$: quadratic.

Limit is zero for $r = 1 \rightarrow$ superlinear convergence.