Convex Optimization

(EE227BT: UC Berkeley)

Lecture 1
(Convex sets and functions)

August 28, 2014

Laurent El Ghaoui (slides by Suvrit Sra)

Course organization

- Course material and interaction: use **bCourse**
- Relevant texts / references:
  - ♡ *Convex optimization* – Boyd & Vandenberghe (BV)
  - ♡ *Introductory lectures on convex optimisation* – Nesterov
  - ♡ *Nonlinear programming* – Bertsekas
  - ♡ *Convex Analysis* – Rockafellar
  - ♡ *Numerical optimization* – Nocedal & Wright
  - ♡ *Lectures on modern convex optimization* – Nemirovski
  - ♡ *Optimization for Machine Learning* – Sra, Nowozin, Wright
  - ♡ *Optimization Models* – Calafiore, El Ghaoui (to appear in November)
- Instructor: Laurent El Ghaoui (elghaoui@berkeley.edu)
- HW + Quizzes (40%); Midterm (30%); Project (30%)
- TA: Vu Pham (ptvu@berkeley.edu)
- Office hours: Thu 9-10am (El Ghaoui), TBA (Vu Pham)
- If you email me, please put **EE227BT** in **Subject:**

Linear algebra recap

Eigenvalues and Eigenvectors

Def. If $A \in \mathbb{C}^{n\times n}$ and $x \in \mathbb{C}^n$. Consider the equation

$$Ax = \lambda x, \qquad x \neq 0, \quad \lambda \in \mathbb{C}.$$

If scalar λ and vector x satisfy this equation, then λ is called an **eigenvalue** and x and **eigenvector** of A.

Above equation may be rewritten equivalently as

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

Thus, λ is an eigenvalue, if and only if

$$\det(\lambda I - A) = 0.$$

Def. $p_A(t) := \det(tI - A)$ is called **characteristic polynomial**.

Eigenvalues are roots of characteristic polynomial.

Eigenvalues and Eigenvectors

Theorem Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of $A \in \mathbb{C}^{n\times n}$. Then,

$$\mathrm{Tr}(A) = \sum_i a_{ii} = \sum_i \lambda_i, \qquad \det(A) = \prod_i \lambda_i.$$

Def. Matrix $U \in \mathbb{C}^{n\times n}$ **unitary** if $U^*U = I$ ($[U^*]_{ij} = [\bar{u}_{ji}]$)

Theorem (Schur factorization). If $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, then there is a unitary matrix $U \in \mathbb{C}^{n\times n}$ (i.e., $U^*U = I$), such that

$$U^*AU = T = [t_{ij}]$$

is upper triangular with diagonal entries $t_{ii} = \lambda_i$.

Corollary. If $A^*A = AA^*$, then there exists a unitary U such that $A = U\Lambda U^*$. We will call this the **Eigenvector Decomposition**.

Proof. $A = VTV^*$, $A^* = VT^*V^*$, so $AA^* = TT^* = T^*T = A^*A$. But T is upper triangular, so only way for $TT^* = T^*T$, some easy but tedious induction shows that T must be diagonal. Hence, $T = \Lambda$.

Singular value decomposition

Theorem (SVD) Let $A \in \mathbb{C}^{m \times n}$. There are unitaries s.t. U and V

$$U^* A V = \mathrm{Diag}(\sigma_1, \ldots, \sigma_p), \quad p = \min(m, n),$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_p \geq 0$. Usually written as

$$A = U \Sigma V^*.$$

left singular vectors U are eigenvectors of AA^*

right singular vectors V are eigenvectors of A^*A

nonzero **singular values** $\sigma_i = \sqrt{\lambda_i(AA^*)} = \sqrt{\lambda_i(A^*A)}$

Positive definite matrices

Def. Let $A \in \mathbb{R}^{n \times n}$ be **symmetric**, i.e., $a_{ij} = a_{ji}$. Then, A is called **positive definite** if

$$x^T A x = \sum_{ij} x_i a_{ij} x_j > 0, \quad \forall\, x \neq 0.$$

If $>$ replaced by $\geq$, we call A **positive semidefinite**.

Theorem A symmetric real matrix is positive semidefinite (positive definite) iff all its eigenvalues are nonnegative (positive).

Theorem Every semidefinite matrix can be written as $B^T B$

Exercise: Prove this claim. Also prove converse.

Notation: $A \succ 0$ (posdef) or $A \succeq 0$ (semidef)

Amongst most important objects in convex optimization!

Matrix and vector calculus

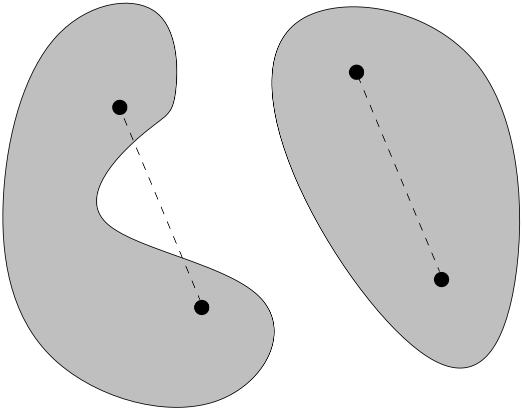
$f(x)$	$\nabla f(x)$
$x^T a = \sum_i x_i a_i$	a
$x^T A x = \sum_{ij} x_i a_{ij} x_j$	$(A + A^T)x$
$\log\det(X)$	X^{-1}
$\mathrm{Tr}(XA) = \sum_{ij} x_{ij} a_{ji}$	A^T
$\mathrm{Tr}(X^T A) = \sum_{ij} x_{ij} a_{ij}$	A
$\mathrm{Tr}(X^T A X)$	$(A + A^T)X$

Easily derived using "brute-force" rules

- ♣ Wikipedia
- ♣ Suvrit's ancient notes
- ♣ Matrix cookbook

Convex Sets

Convex sets



Convex sets

Def. A set $C \subset \mathbb{R}^n$ is called **convex**, if for any $x, y \in C$, the line-segment $\theta x + (1-\theta)y$ (here $\theta \geq 0$) also lies in C.

Combinations

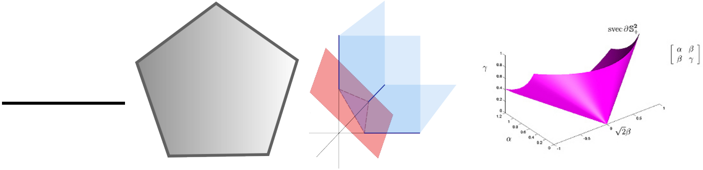
- Convex: $\theta_1 x + \theta_2 y \in C$, where $\theta_1, \theta_2 \geq 0$ and $\theta_1 + \theta_2 = 1$.
- Linear: if restrictions on θ_1, θ_2 are dropped
- Conic: if restriction $\theta_1 + \theta_2 = 1$ is dropped

Convex sets

Theorem (Intersection).
Let C_1, C_2 be convex sets. Then, $C_1 \cap C_2$ is also convex.

Proof. If $C_1 \cap C_2 = \emptyset$, then true vacuously.
Let $x, y \in C_1 \cap C_2$. Then, $x, y \in C_1$ and $x, y \in C_2$.
But C_1, C_2 are convex, hence $\theta x + (1-\theta)y \in C_1$, and also in C_2.
Thus, $\theta x + (1-\theta)y \in C_1 \cap C_2$.
Inductively follows that $\cap_{i=1}^m C_i$ is also convex.

Convex sets – more examples



(psdcone image from convexoptimization.com, Dattorro)

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$. Their **convex hull** is

$$\mathrm{co}(x_1, \ldots, x_m) := \left\{ \sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1 \right\}.$$

♡ Let $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. The set $\{x \mid Ax = b\}$ is convex (it is an *affine space* over subspace of solutions of $Ax = 0$).

♡ *halfspace* $\{x \mid a^T x \leq b\}$.

♡ *polyhedron* $\{x \mid Ax \leq b, Cx = d\}$.

♡ *ellipsoid* $\{x \mid (x - x_0)^T A(x - x_0) \leq 1\}$, ($A$: semidefinite)

♡ *probability simplex* $\{x \mid x \geq 0, \sum_i x_i = 1\}$

Quiz: Prove that these sets are convex.

Convex functions

Convex functions

Def. Function $f : I \to \mathbb{R}$ on interval I called **midpoint convex** if

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}, \qquad \text{whenever } x, y \in I.$$

Read: f of AM is less than or equal to AM of f.

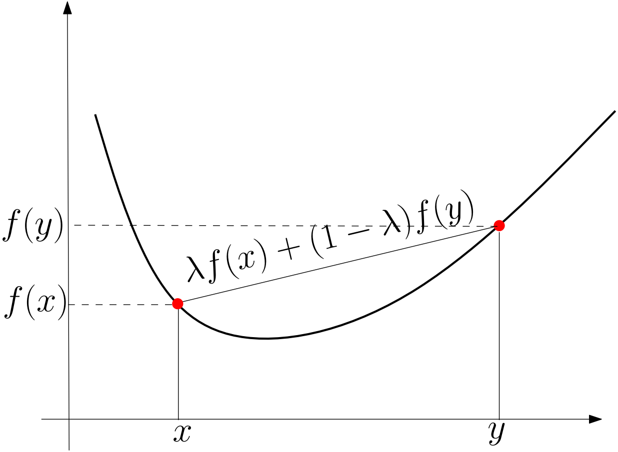
Def. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **convex** if its domain $\operatorname{dom}(f)$ is a convex set and for any $x, y \in \operatorname{dom}(f)$ and $\theta \geq 0$

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

Theorem (J.L.W.V. Jensen). Let $f : I \to \mathbb{R}$ be continuous. Then, f is convex *if and only if* it is midpoint convex.

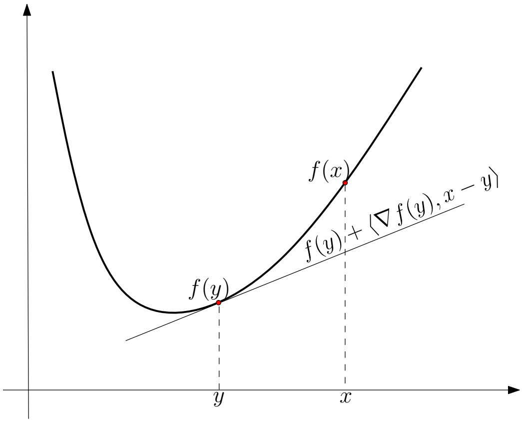
▶ Theorem extends to functions $f : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}$. Very useful to checking convexity of a given function.

Convex functions



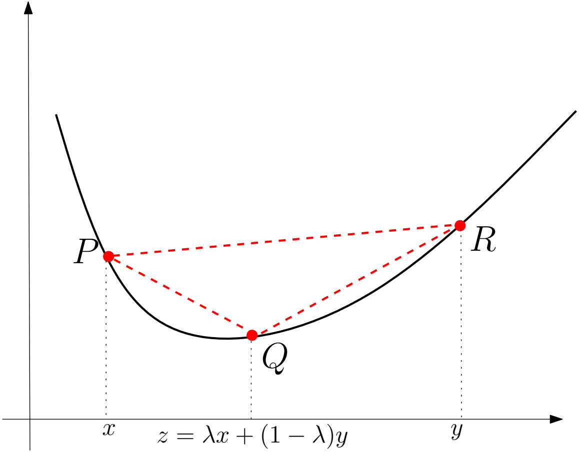
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Convex functions



$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

Convex functions



slope PQ $\leq$ slope PR $\leq$ slope QR

Recognizing convex functions

- ♠ If f is continuous and midpoint convex, then it is convex.
- ♠ If f is differentiable, then f is convex *if and only if* $\operatorname{dom} f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$ for all $x, y \in \operatorname{dom} f$.
- ♠ If f is twice differentiable, then f is convex *if and only if* $\operatorname{dom} f$ is convex and $\nabla^2 f(x) \succeq 0$ at every $x \in \operatorname{dom} f$.

Convex functions

- Linear: $f(\theta_1 x + \theta_2 y) = \theta_1 f(x) + \theta_2 f(y)$; θ_1, θ_2 unrestricted
- Concave: $f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta) f(y)$
- Strictly convex: If inequality is strict for $x \neq y$

Convex functions

Example The *pointwise maximum* of a family of convex functions is convex. That is, if $f(x; y)$ is a convex function of x for every y in some "index set" $\mathcal{Y}$, then

$$f(x) := \max_{y \in \mathcal{Y}} \; f(x; y)$$

is a convex function of x (set $\mathcal{Y}$ is arbitrary).

Example Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Let $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. Prove that $g(x) = f(Ax + b)$ is convex.

Exercise: Verify truth of above examples.

Convex functions

Theorem Let $\mathcal{Y}$ be a nonempty convex set. Suppose $L(x, y)$ is convex in (x, y), then,

$$f(x) := \inf_{y \in \mathcal{Y}} \quad L(x, y)$$

is a convex function of x, provided $f(x) > -\infty$.

Proof. Let $u, v \in \operatorname{dom} f$. Since $f(u) = \inf_y L(u, y)$, for each $\epsilon > 0$, there is a $y_1 \in \mathcal{Y}$, s.t. $f(u) + \frac{\epsilon}{2}$ is not the infimum. Thus, $L(u, y_1) \leq f(u) + \frac{\epsilon}{2}$.
Similarly, there is $y_2 \in \mathcal{Y}$, such that $L(v, y_2) \leq f(v) + \frac{\epsilon}{2}$.
Now we prove that $f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda) f(v)$ directly.

$$\begin{aligned} f(\lambda u + (1-\lambda)v) &= \inf_{y \in \mathcal{Y}} L(\lambda u + (1-\lambda)v, y) \\ &\leq L(\lambda u + (1-\lambda)v, \lambda y_1 + (1-\lambda) y_2) \\ &\leq \lambda L(u, y_1) + (1-\lambda) L(v, y_2) \\ &\leq \lambda f(u) + (1-\lambda) f(v) + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, claim follows.

Example: Schur complement

Let A, B, C be matrices such that $C \succ 0$, and let

$$Z := \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0,$$

then the **Schur complement** $A - BC^{-1}B^T \succeq 0$.
Proof. $L(x, y) = [x, y]^T Z[x, y]$ is convex in (x, y) since $Z \succeq 0$

Observe that $f(x) = \inf_y L(x, y) = x^T(A - BC^{-1}B^T)x$ is convex.

(We skipped ahead and solved $\nabla_y L(x, y) = 0$ to minimize).

Recognizing convex functions

- ♠ If f is continuous and midpoint convex, then it is convex.
- ♠ If f is differentiable, then f is convex *if and only if* $\mathrm{dom}\, f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$ for all $x, y \in \mathrm{dom}\, f$.
- ♠ If f is twice differentiable, then f is convex *if and only if* $\mathrm{dom}\, f$ is convex and $\nabla^2 f(x) \succeq 0$ at every $x \in \mathrm{dom}\, f$.
- ♠ By showing f to be a pointwise max of convex functions
- ♠ By showing $f : \mathrm{dom}(f) \to \mathbb{R}$ is convex *if and only if* its restriction to **any** line that intersects $\mathrm{dom}(f)$ is convex. That is, for any $x \in \mathrm{dom}(f)$ and any v, the function $g(t) = f(x + tv)$ is convex (on its domain $\{t \mid x + tv \in \mathrm{dom}(f)\}$).
- ♠ See exercises (Ch. 3) in Boyd & Vandenberghe for more ways

Operations preserving convexity

Operations preserving convexity

Pointwise maximum: $f(x) = \sup_{y \in \mathcal{Y}} f(y; x)$

Conic combination: Let $a_1, \ldots, a_n \geq 0$; let $f_1, \ldots, f_n$ be convex functions. Then, $f(x) := \sum_i a_i f_i(x)$ is convex.

Remark: The set of all convex functions is a *convex cone*.

Affine composition: $f(x) := g(Ax + b)$, where g is convex.

Operations preserving convexity

Theorem Let $f : I_1 \to \mathbb{R}$ and $g : I_2 \to \mathbb{R}$, where range$(f) \subseteq I_2$. If f and g are convex, and g is increasing, then $g \circ f$ is convex on I_1

Proof. Let $x, y \in I_1$, and let $\lambda \in (0, 1)$.

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \lambda f(x) + (1-\lambda)f(y) \\ g(f(\lambda x + (1-\lambda)y)) &\leq g\big(\lambda f(x) + (1-\lambda)f(y)\big) \\ &\leq \lambda g\big(f(x)\big) + (1-\lambda)g\big(f(y)\big). \end{aligned}$$

Read Section 3.2.4 of BV for more

Examples

Quadratic

Let $f(x) = x^T Ax + b^T x + c$, where $A \succeq 0$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

What is: $\nabla^2 f(x)$?

$\nabla f(x) = 2Ax + b$, $\nabla^2 f(x) = 2A \succeq 0$, hence f is convex.

Indicator

Let $\mathbb{I}_{\mathcal{X}}$ be the *indicator function* for $\mathcal{X}$ defined as:

$$\mathbb{I}_{\mathcal{X}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{X}, \\ \infty & \text{otherwise.} \end{cases}$$

Note: $\mathbb{I}_{\mathcal{X}}(x)$ is convex **if and only if** $\mathcal{X}$ is convex.

Distance to a set

Example Let $\mathcal{Y}$ be a convex set. Let $x \in \mathbb{R}^n$ be some point. The distance of x to the set $\mathcal{Y}$ is defined as

$$\mathrm{dist}(x, \mathcal{Y}) := \inf_{y \in \mathcal{Y}} \quad \|x - y\|.$$

Because $\|x - y\|$ is jointly convex in (x, y), the function $\mathrm{dist}(x, \mathcal{Y})$ is a convex function of x.

Norms

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that satisfies

1. $f(x) \geq 0$, and $f(x) = 0$ if and only if $x = 0$ (definiteness)
2. $f(\lambda x) = |\lambda| f(x)$ for any $\lambda \in \mathbb{R}$ (positive homogeneity)
3. $f(x + y) \leq f(x) + f(y)$ (subadditivity)

Such a function is called a *norm*. We usually denote norms by $\|x\|$.

Theorem Norms are convex.

Proof. Immediate from subadditivity and positive homogeneity.

Vector norms

Example (ℓ_2-norm): Let $x \in \mathbb{R}^n$. The **Euclidean** or ℓ_2-norm is $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$

Example (ℓ_p-norm): Let $p \geq 1$. $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$

Exercise: Verify that $\|x\|_p$ is indeed a norm.

Example (ℓ_∞-norm): $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

Example (Frobenius-norm): Let $A \in \mathbb{R}^{m \times n}$. The **Frobenius** norm of A is $\|A\|_{\mathrm{F}} := \sqrt{\sum_{ij} |a_{ij}|^2}$; that is, $\|A\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(A^* A)}$.

Mixed norms

Def. Let $x \in \mathbb{R}^{n_1+n_2+\cdots+n_G}$ be a vector partitioned into subvectors $x_j \in \mathbb{R}^{n_j}$, $1 \leq j \leq G$. Let $\mathbf{p} := (p_0, p_1, p_2, \ldots, p_G)$, where $p_j \geq 1$. Consider the vector $\xi := (\|x_1\|_{p_1}, \cdots, \|x_G\|_{p_G})$. Then, we define the **mixed-norm** of x as

$$\|x\|_{\mathbf{p}} := \|\xi\|_{p_0}.$$

Example $\ell_{1,q}$-norm: Let x be as above.

$$\|x\|_{1,q} := \sum_{i=1}^{G} \|x_i\|_q.$$

This norm is popular in machine learning, statistics.

Matrix Norms

Induced norm

Let $A \in \mathbb{R}^{m \times n}$, and let $\|\cdot\|$ be any vector norm. We define an **induced matrix norm** as

$$\|A\| := \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Verify that above definition yields a norm.

- Clearly, $\|A\| = 0$ iff $A = 0$ (definiteness)
- $\|\alpha A\| = |\alpha| \; \|A\|$ (homogeneity)
- $\|A + B\| = \sup \frac{\|(A+B)x\|}{\|x\|} \leq \sup \frac{\|Ax\|+\|Bx\|}{\|x\|} \leq \|A\| + \|B\|.$

Operator norm

Example Let A be any matrix. Then, the **operator norm** of A is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

$\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue of a matrix is **not** a norm!
- $\|A\|_1$ and $\|A\|_\infty$—max-abs-column and max-abs-row sums.
- $\|A\|_p$ generally NP-Hard to compute for $p \notin \{1, 2, \infty\}$
- **Schatten p-norm:** ℓ_p-norm of vector of singular value.
- **Exercise:** Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ be singular values of a matrix $A \in \mathbb{R}^{m \times n}$. Prove that

  $$\|A\|_{(k)} := \sum\nolimits_{i=1}^{k} \sigma_i(A),$$

  is a norm; $1 \leq k \leq n$.

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Exercise: Verify that $\|u\|_*$ is a norm.

Exercise: Let $1/p + 1/q = 1$, where $p, q \geq 1$. Show that $\|\cdot\|_q$ is dual to $\|\cdot\|_p$. In particular, the ℓ_2-norm is self-dual.

Misc Convexity

Other forms of convexity

- **Log-convex:** $\log f$ is convex; log-cvx $\implies$ cvx;
- **Log-concavity:** $\log f$ concave; **not** closed under addition!
- **Exponentially convex:** $[f(x_i + x_j)] \succeq 0$, for $x_1, \ldots, x_n$
- **Operator convex:** $f(\lambda X + (1-\lambda)Y) \preceq \lambda f(X) + (1-\lambda)f(Y)$
- **Quasiconvex:** $f(\lambda x + (1 - \lambda y)) \leq \max\{f(x), f(y)\}$
- **Pseudoconvex:** $\langle \nabla f(y), x - y \rangle \geq 0 \implies f(x) \geq f(y)$
- **Discrete convexity:** $f : \mathbb{Z}^n \to \mathbb{Z}$; "convexity + matroid theory."