Week 2

Nicolo Michelusi

I. Convex functions

- Definition:

- Geometric interpretation:

- Counter-example

- Concave functions

- Strictly convex (concave) functions

- Jensen's inequality: if f convex, then $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y), \ \forall \theta \in [0,1], \ \forall x, y, \in \mathrm{dom}(f)$

  can be extended to infinite sums, integrals, expectations:

II. RELATIONSHIP BETWEEN CONVEX FUNCTIONS AND CONVEX SETS

- Epigraph of function, $\mathrm{epi}(f)$

- f convex iff $\mathrm{epi}(f)$ convex

- sub-level set C_α

- f convex $\Rightarrow$ C_α convex for all α

Q: is the converse true?

III. Restriction to a line

- f convex iff convex when restricted to any line intersecting its domain

  Not surprising since, to check convexity, we only need to check straight lines.
- Useful because we can reduce the problem of checking convexity of any function to a one-dimensional problem
- Example: is $f(x) = x_1 x_2$ convex?

IV. Conditions for convexity

- Often, not easy to check convexity. The following will be useful conditions:
  - First order conditions, ∇f
  - Second order conditions, $\nabla^2 f$ (Hessian)
  - Compositions that preserve convexity
- First order condition (f differentiable):

  f convex iff $\mathrm{dom}(f)$ is convex and $f(y) \geq f(x) + \nabla f(x)^T (y - x)$, $\forall x, y$

  In other words, the rest of the function must be above the *supporting hyperplane*

  Proof of $\Rightarrow$

.

Proof of $\Leftarrow$

- Second order condition (f twice differentiable):
  f convex iff $\mathrm{dom}(f)$ is convex and $\nabla^2 f(x) \succeq 0, \forall x$ (strictly if $\succ 0$)

  Special case, $f : \mathbb{R} \mapsto \mathbb{R}$

  More in general, $f : \mathbb{R}^n \mapsto \mathbb{R}$

  Proof of $\Rightarrow$

.

Proof of $\Leftarrow$

V. Operations that preserve convexity

- non-negative weighted sums (also, infinite sums and integrals):

Q: what if some weights are negative?

- affine mapping of the argument: f convex, $g(x) = f(Ax + b)$

- Example: $\log(\sum_i e^{x_i})$

- Example: x^2

- **Q:** if $f(x)$ convex, is $f(-x)$ convex? Is $-f(x)$ convex?
- Using the above two properties, whenever we check for convexity, we can ignore the affine change of variables and non-negative weights, and focus on the simplest function form.

VI. Examples of convex functions

- To use convex optimization, it is very important to be able to quickly identify/convert to convex functions. For some problems, the key to success is to find/identify convex functions. Here is a list of commonly used convex functions with applications.
- Exponent $e^{\alpha x}$

- negative log $-\ln(x)$

- Example: Shannon capacity: $C = W \log_2(1 + P/N)$

  Capacity increment due to increased power diminishes as P increases; *Principle of diminishing returns*, often assumed for utility functions

- Example: Power-rate: $P = N(2^{C/W} - 1)$

  To get the same increment in capacity, the required increase in power grows exponentially; in power systems, the cost of generation is often assumed to be convex.

- Powers, x^α

- Power of absolute values, $|x|^p,\ p \geq 1$

- Norms $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p},\ p \geq 1$

- Ellipsoid $(x - x_0)^T P (x - x_0),\ P \succeq 0$

- Example: mean squared errors, regression; given $(x_i, y_i), i = 1, \ldots, N$, find a, b such that $y = ax + b$ has the smallest error, defined as

$$\sum_i |y_i - (ax_i + b)|^p,\ p \geq 1$$

- Example: detection and likelihood functions: estimate x from $y = x + w,\ w \sim \mathcal{CN}(0, \sigma^2)$ via maximum likelihood

- Example: entropy $x \log_2 x,\ x > 0$; if a source generates symbols S according to distribution $p_i = \mathbb{P}(S = s_i),\ i = 1, \ldots, N$, then entropy is

$$H(p) = -\sum_i p_i \log_2(p_i)$$

  Entropy measures the uncertainty of the source, i.e. the amount of information generated by the source.

- Max: $\max_i x_i$ (non-differentiable!) (proof with epigraph)

- log-sum-exp: $\ln(\sum_i e^{x_i})$ (restrict to a line and use 2nd order condition)

- Example: log-moment generating function: $f(s) = \ln \mathbb{E}[e^{sX}]$ (restrict to a line and use 2nd order condition)

- Example: high-SNR approximation of Shannon capacity

$$C_i = W \log_2 \left(1 + \frac{P_i}{\sum_{j \neq i} P_j + N} \right) \geq W \log_2 \left(\frac{P_i}{\sum_{j \neq i} P_j + N} \right)$$

(not convex nor concave; use change of variables $P_i = e^{x_i}$)

- Low-SNR approximation,

$$C_i = W\log_2\left(1+\frac{P_i}{\sum_{j\neq i}P_j+N}\right)\leq \frac{W}{\ln(2)}\frac{P_i}{\sum_{j\neq i}P_j+N}$$

(not convex nor concave; take $y_i=\ln(C_i)$ and change of variables $P_i=e^{x_i}$)

- Geometric mean: $\left(\prod_i x_i\right)^{1/N}$ is concave (see Boyd page 74)
- Quadratic over linear: $x^2/y,\ y>0$ (check 2nd order condition)

VII. More operations that preserve convexity

- In addition to non-negative weighted sums and affine mappings of arguments
- Pointwise maximum and supremum of convex functions: $\max_i f_i(x)$ or $\sup_y f(x,y)$ (use epigraph)

- Pointwise minimum and infimum of concave functions: $\min_i f_i(x)$ or $\inf_y f(x, y)$
- Examples:

  $\max_i x_i$

  $\max_i a_i^T x_i + b_i$

  $\min_i x_i$

  $\min_i a_i^T x_i + b_i$
- Sum of the largest r components of $x \in \mathbb{R}^n$

- Distance to the farthest point: let $x \in \mathbb{R}^n$, $\mathcal{C} \subseteq \mathbb{R}^n$, $f(x) = \sup_{y \in \mathcal{C}} \|x - y\|_2$

- What about the infimum of convex functions? $f(x) = \inf_{y \in \mathcal{C}} g(x, y)$?

- Example: distance to a set, $\text{dist}(x, \mathcal{C}) = \inf_{y \in \mathcal{C}} \|x - y\|_2$

- Perspective of a function $g(x, t) = tf(x/t)$

- Example: find channel capacity one a user is served a portion $t < 1$ of the time; is it convex?

- Example: $g(x) = (c^T x + d) f \left(\frac{Ax+b}{c^T x+d} \right)$

- Compositions: $f(x) = h(g(x))$

- Examples:

  g convex, what about $e^{g(x)}$?

  g convex, what about $g(x)^2$?

  g convex and non-negative, what about $g(x)^2$?

  g concave, what about $\ln g(x)$?

  g concave, what about $1/g(x)$?

- Vector composition: $f(x) = h(g(x))$ with $h : \mathbb{R}^k \mapsto \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}^k$, what are the conditions? (assume differentiable)