Convex Optimization 101: Theory

Contents:

More specifically,

Section 2 contains: line, affine set, convex set, convex combination, convex hull, hyperplane, halfspace, separating hyperplane theorem, and supporting hyperplane theorem.

Section 3 contains: convex function, first order condition, second order condition, sublevel set, epigraph, pointwise maximum, pointwise supremum, and conjugate function

Section 4 contains: Lagrangian, dual function, lower bound property, dual problem, weak and strong duality, Slater's constraint qualification, complementary slackness, Karush-Kuhn-Tucker (KKT) condition, and Lagrange dual and conjugate function.

Optimization problem

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \leq 0, \ i = 1, \cdots, m \\ &&& h_i(x) = 0, \ i = 1, \cdots, p \end{aligned}$$

$x = (x_1, \cdots, x_n)^T \in \mathbb{R}^n$; optimization variables.

$f_0 : \mathbb{R}^n \to \mathbb{R}$; objective function

$f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, m$: inequality constraint functions

$h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, p$: equality constraint functions.

* Domain of the optimization problem is

$$D = \left[\bigcap_{i=0}^{m} \operatorname{dom} f_i \right] \cap \left[\bigcap_{i=1}^{p} \operatorname{dom} h_i \right]$$

* A point $x \in D$ is feasible if it satisfies the constraints. The set of all feasible points is called the feasible set.

* If feasible set has at least one point, the optimization problem is feasible ; if not, the problem is infeasible.

* The optimal value p^* of the problem is

$$p^* = \inf\{f_0(x) \mid f_i(x) \leq 0, i=1,\cdots,m, h_i(x)=0, i=1,\cdots,p\}$$

* ~~The~~ An optimal point x^* of the problem is, if x^* is feasible and $f_0(x^*) = p^*$. The set of all optimal points is the optimal set, denoted

$$X_{opt} = \{x \mid f_i(x) \leq 0, i=1,\cdots,m, h_i(x)=0, i=1,\cdots,p, f_0(x)=p^*\}$$

Examples

① variable: model parameters.
② constraints: prior information, parameter limits
③ objective: - likelihood, loss function

Examples

① variable: amounts invested in different assets
② constraints: budget, minimum return
③ objective: overall risk, portfolio variance

Least-squares problem

$$\text{minimize} \quad f_0(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{k} (a_i^T x - b_i)^2$$

where $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $A = (a_1, \cdots, a_k)^T \in \mathbb{R}^{k \times n}$. $(k \geq n)$

* Analytical solution : $x^* = (A^T A)^{-1} A^T b$ (unique)

Linear Programming (LP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x - b_i \leq 0 \ , \quad i = 1, \cdots, m \end{array}$$

where $x \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$.

* No analytical solution
* Reliable and efficient algorithms, e.g. Simplex algorithm.
* Linear means

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y)$$

where $\alpha + \beta = 1$

Convex optimization problem

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \cdots, m \\
& h_i(x) = 0, \quad i = 1, \cdots, p
\end{aligned}
$$

$X_i \in \mathbb{R}^n$.

* Both objective and constraint functions are convex

$$f_i(\alpha X + \beta y) \le \alpha f_i(x) + \beta f_i(y)$$
$$h_i(\alpha X + \beta y) \le \alpha h_i(x) + \beta \overset{h_i}{\cancel{f_i}}(y)$$

where $\alpha + \beta = 1$, $\alpha \ge 0$, $\beta \ge 0$

* Includes least-squares problems and linear programmings as special case.

* No analytical solution.

* hard to recognize

* many tricks for transforming problems into convex form.

Selected basic concepts of convex sets

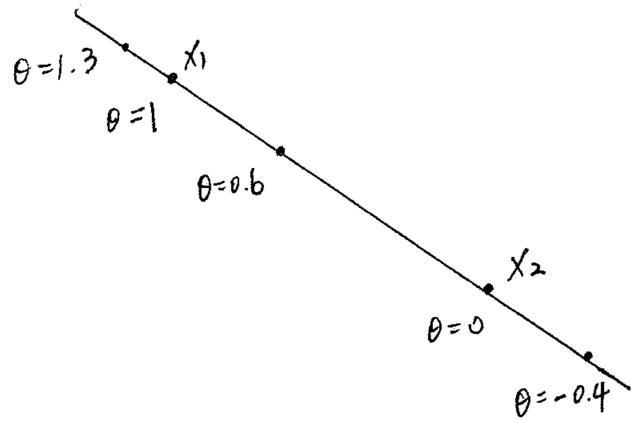
1. Line.

$x_1 \neq x_2$ are two points in R^n.

$$y = \theta x_1 + (1-\theta) x_2, \qquad \theta \in \mathbb{R}$$

form the line passing through x_1 and x_2.

$0 \le \theta \le 1 \Rightarrow$ Line segment



2. Affine sets.

A set $C \subseteq \mathbb{R}^n$ is affine if the line through any two distinct points in C lies in C, i.e. if for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have

$$\theta x_1 + (1-\theta) x_2 \in C$$

If we have more than two points, denote

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

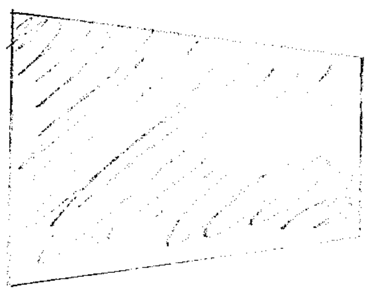
where $\theta_1 + \cdots + \theta_k = 1$, to be an affine combination.

If C is an affine set, $x_1, \cdots, x_k \in C$, and $\theta_1 + \cdots + \theta_k = 1$, then the point $\theta_1 x_1 + \cdots + \theta_k x_k \in C$.

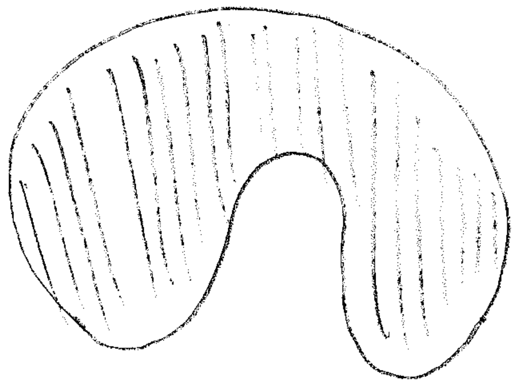
3. Convex Set.

A set C is convex if the line segment between any two points in C lines in C, i.e., if for any $x_1, x_2 \in C$ and any θ with $0 \le \theta \le 1$, we have

$$\theta x_1 + (1-\theta) x_2 \in C.$$



(convex)



(nonconvex)

4. Convex combination

For any $x_1, \cdots, x_k$, $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0, i = 1, \cdots, k$

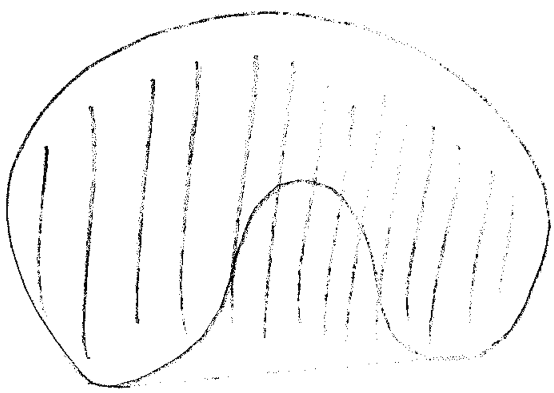
$$y = \theta_1 x_1 + \cdots + \theta_k x_k$$

is called convex combination.

5. Convex hull

The convex hull of a set C, denoted $\operatorname{conv} C$, is the set of all convex combinations of points in C:

$$\operatorname{conv} C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C, \theta_i \geq 0, i = 1, \cdots, k, \theta_1 + \cdots + \theta_k = 1\}$$



6. Hyperplanes

A hyperplane is a set of the form

$$\{x \mid a^T x = b\}$$

where $x \in \mathbb{R}^n$, $a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$. Geometrically speaking, for any x_0 with $a^T x_0 = b$.

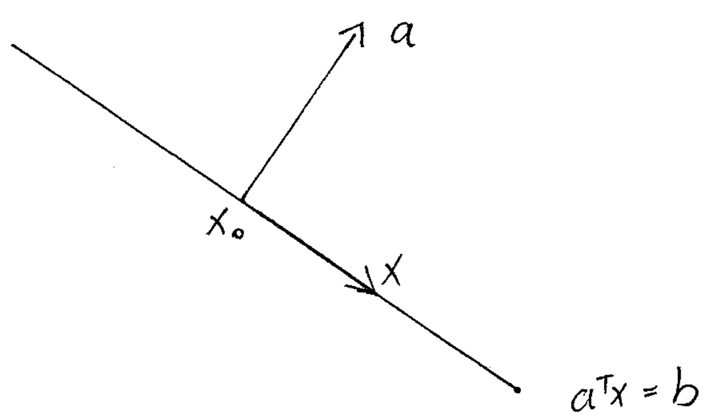
$$\{x \mid a^T x = b\} = \{x \mid a^T (x - x_0) = 0\}.$$

so that

$$\{x \mid a^T x = b\} = x_0 + a^{\perp}$$

where $a^{\perp}$ denotes the orthogonal complement of a.

$$a^{\perp} = \{v \mid a^T v = 0\}$$



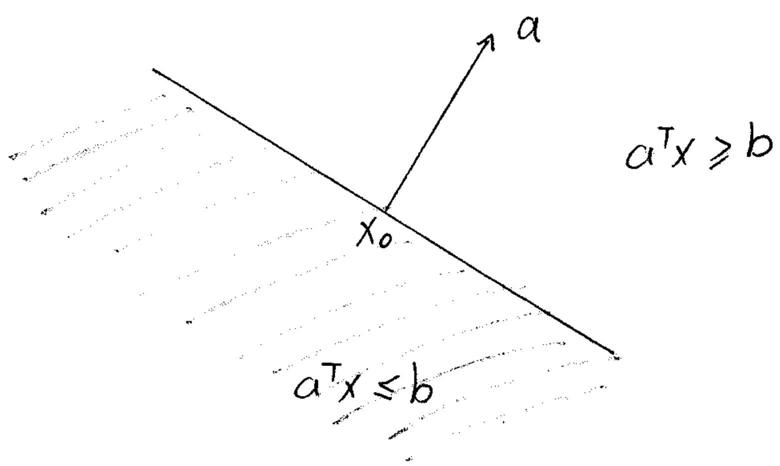
* Hyperplanes are affine and convex.

7. Halfspace.

A halfspace is a set of the form

$$\{ x \mid a^T x \le b \}$$

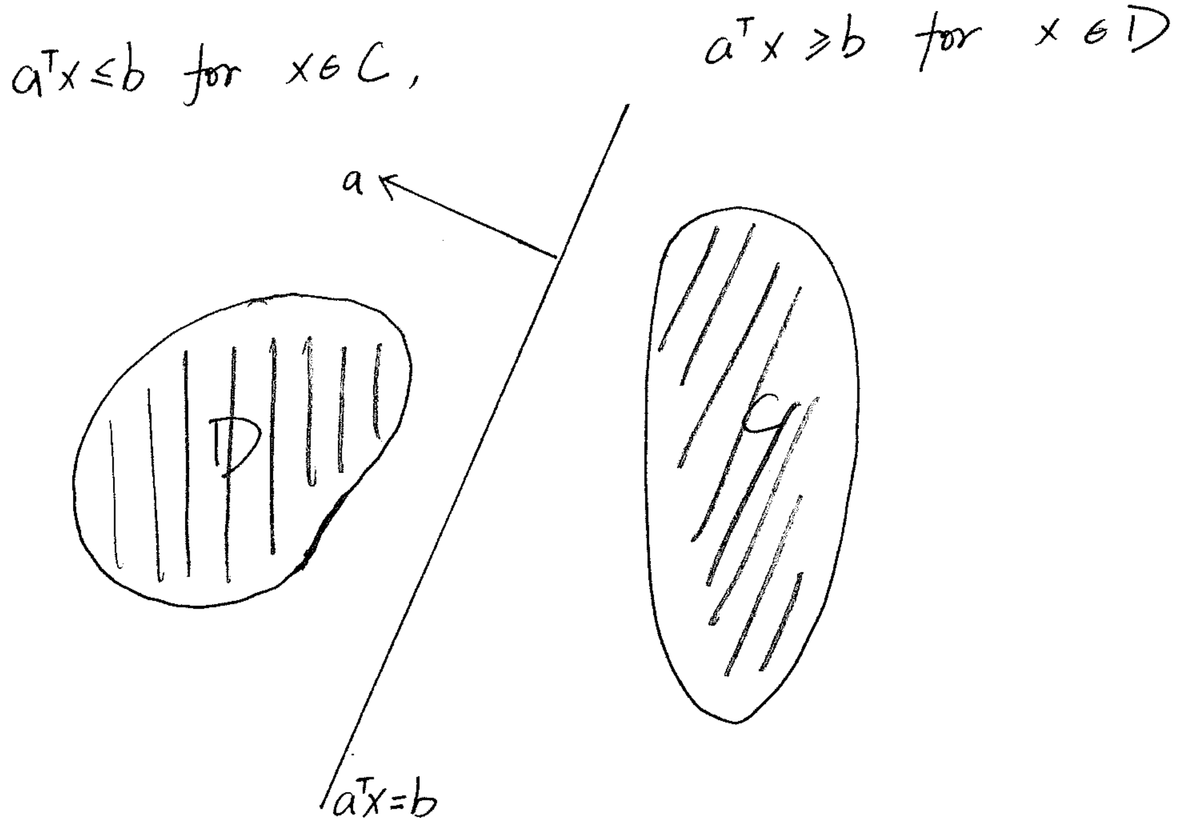
where $a \neq 0$.



* Halfspaces are convex.

8. Separating hyperplane theorem

Thm: Suppose C and D are two convex sets that do NOT intersect, i.e. $C \cap D = \phi$. 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 $x \in D$.



* The hyperplane $\{x \mid a^T x = b\}$ separates C and D.

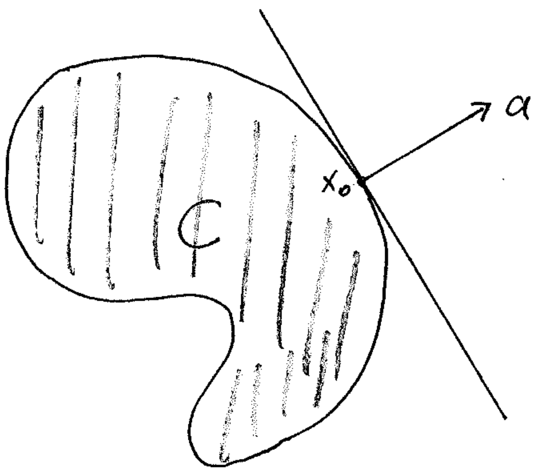
9. Supporting hyperplane theorem

Thm: For any nonempty convex set C, and any $x_0 \in \mathrm{bd}\, C$ (boundary of set C), there exists a supporting hyperplane to C at x_0.

Supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$.

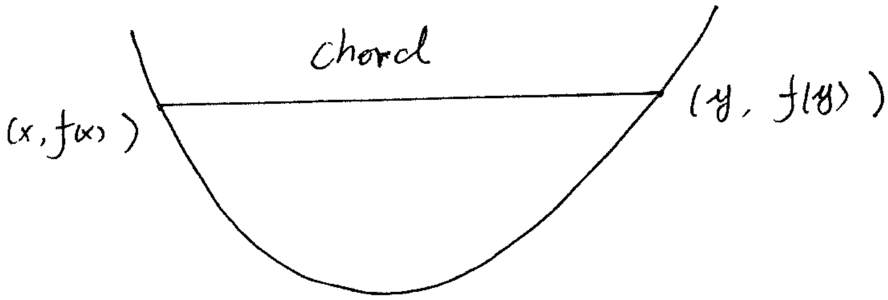


Selected basic concepts of convex functions

1. Convex function

A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathrm{dom} f$ is a convex set and if for all $x, y \in \mathrm{dom} f$, and θ with $0 \le \theta \le 1$, we have

$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$



* f is concave if $-f$ is convex.

* f is strictly convex if replace "$\le$" with "$<$" in both θ and f inequalities.

* A function is convex if and only if it is convex when restricted to any line that intersects its domain.

2. Convex function examples on $\mathbb{R}$.

* affine: $ax+b$ on $\mathbb{R}$, $\forall a,b \in \mathbb{R}$ (also concave)
* exponential: e^{ax}, $\forall a \in \mathbb{R}$.
* powers: x^{α} on $\mathbb{R}^{+}$, $\alpha \geq 1$ or $\alpha \leq 0$
* negative entropy: $x \log x$ on $\mathbb{R}^{+}$.

3. Convex function examples on $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$

* affine : $a^T x + b$ on $\mathbb{R}^n$
* norms: $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$ for $p \geq 1$, on $\mathbb{R}^n$
* matrix affine: $tr(A^T X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij} X_{ij} + b$ on $\mathbb{R}^{m\times n}$

4. First-order condition

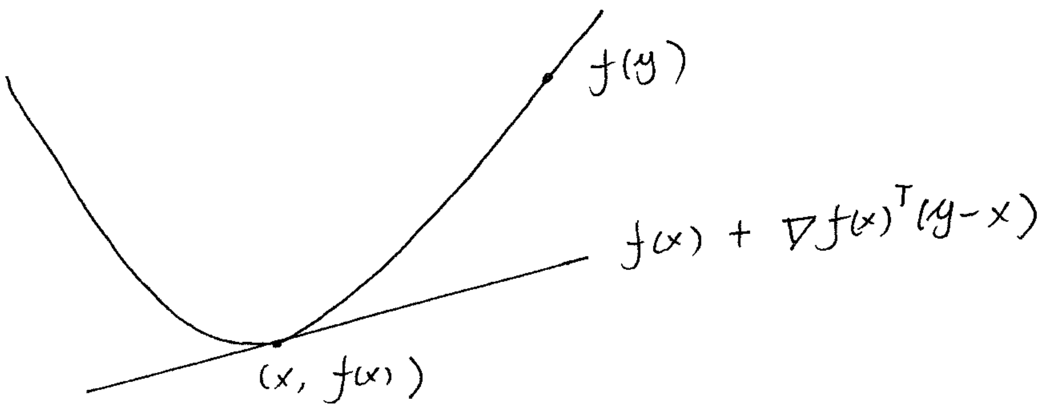
~~Suppose~~ f is differentiable ~~and~~ if $\mathrm{dom}\, f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

exists at each $x \in \mathrm{dom}\, f$. The 1st-order condition is: differentiable f with convex domain is convex

iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \mathrm{dom}\, f$$



* 1st-order approximation of f is global underestimator.

* Think about supporting hyperplane.

5. Second-order conditions

f is twice differentiable if $\mathrm{dom} f$ is open and the Hessian $\nabla^2 f(x)$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i,j = 1, \cdots, n$$

exists at each $x \in \mathrm{dom} f$.

The 2nd-order conditions: for twice differentiable f with convex domain, f is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathrm{dom} f.$$

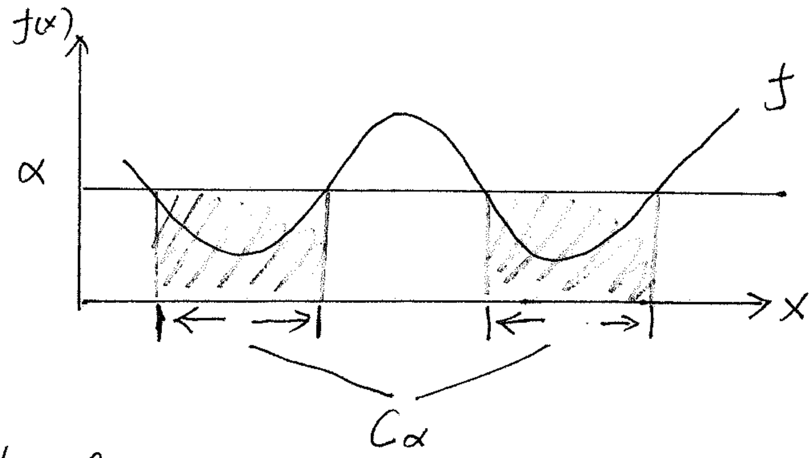
* If $\nabla^2 f(x) \succ 0$ for all $x \in \mathrm{dom} f$, then f is strictly convex.

* " $\succeq$ " and " $\succ$ " are component-wise inequality.

6. Sublevel set

The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is

$$C_\alpha = \{ x \in \operatorname{dom} f \mid f(x) \le \alpha \}$$



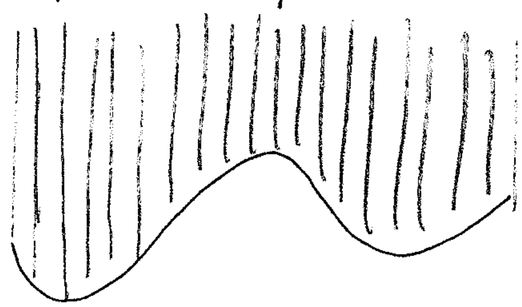
* Sublevel sets of a convex function are convex

7. Epigraph

The graph of f is $\{ (x, f(x)) \mid x \in \operatorname{dom} f \}$

The epigraph of f is

$$\operatorname{epi} f = \{ (x, t) \mid x \in \operatorname{dom} f,\ f(x) \le t \}$$



* "Epi" means "above".

* f is convex iff $\operatorname{epi} f$ is a convex set.

8. Pointwise Maximum preserves convexity

If $f_1, \cdots, f_m$ are convex functions, then

$$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$$

is convex.

9. Pointwise Supremum preserves convexity

If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

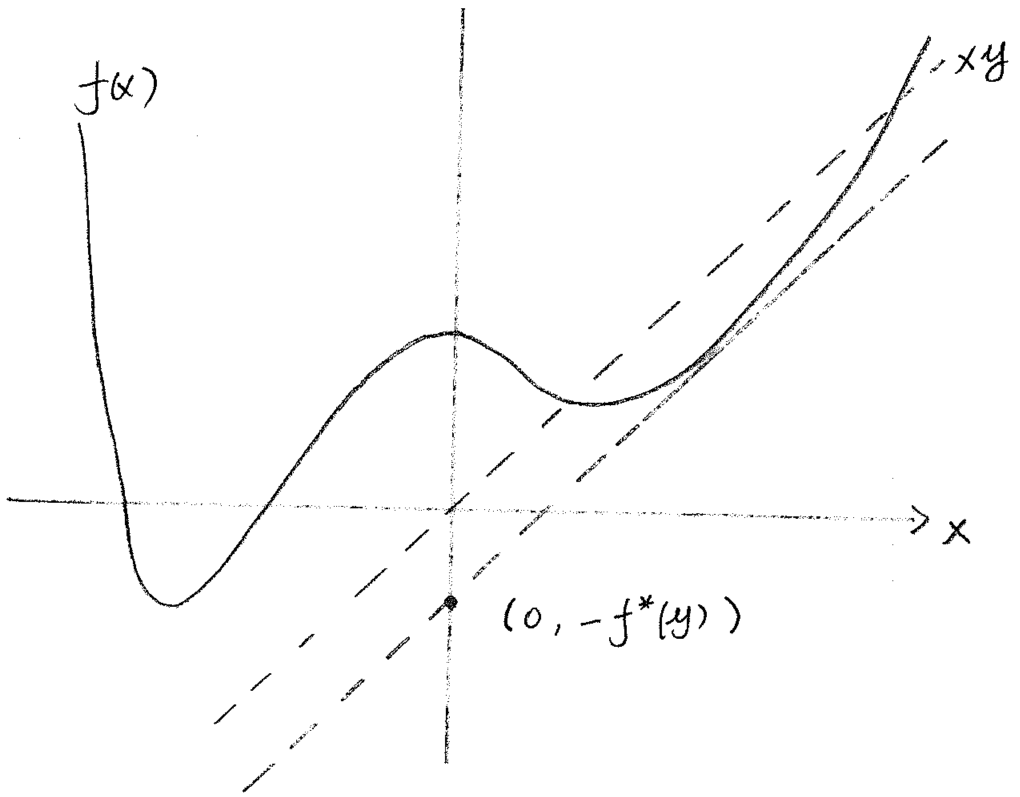
is convex.

10. The conjugate function

Let $f : \mathbb{R}^n \to \mathbb{R}$. The function $f^* : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) = \sup_{x \in \operatorname{dom} f} \left(y^T x - f(x) \right)$$

is called the conjugate of the function f.



* f^* is convex (even if f is not)
* If f is convex and closed, then $(f^*)^* = f$.
* For any f, $\operatorname{epi}((f^*)^*) = \operatorname{conv}(\operatorname{epi} f)$
* $f^{env} = (f^*)^*$ is called the convex envelope of f.

Duality Problem

Recall the standard form optimization problem:

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0, & i = 1, \cdots, m \\ & h_i(x) = 0, & i = 1, \cdots, p \end{array}$$

Variable $x \in \mathbb{R}^n$, domain D, optimal value p^*, optimal point x^*.

<u>Lagrangian</u>: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\text{dom}\, L = D \times \mathbb{R}^m \times \mathbb{R}^p$, is

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

* weighted sum of objective and constraint functions.
* λ_i is Lagrange multiplier associated with $f_i(x) \le 0$
* ν_i is Lagrange multiplier associated with $h_i(x) = 0$

<u>Lagrange dual function</u>: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$.

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in D} L(x, \lambda, \nu) \\ &= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

* g is concave, can be $-\infty$ for some λ, ν.

Lower bound property

If $\lambda \succeq 0$, the $g(\lambda, \nu) \le p^*$

Proof: If $\tilde{x}$ is any feasible point and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu).$$

minimizing over all feasible $\tilde{x}$ gives $p^* \ge g(\lambda, \nu)$.

Example

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

$$L(x, \nu) = x^T x + \nu^T (Ax - b)$$

$$\nabla_x L(x, \nu) = 2x + A^T \nu \equiv 0 \quad \Rightarrow \quad x = -\tfrac{1}{2} A^T \nu$$

plug $x = -\frac{1}{2} A^T \nu$ into L to obtain g:

$$g(\nu) = L\left(-\tfrac{1}{2} A^T \nu, \nu\right) = \tfrac{1}{4} \nu^T A A^T \nu - \tfrac{1}{2} \nu^T A A^T \nu - \nu^T b$$

$$= -\tfrac{1}{4} \nu^T A A^T \nu - \nu^T b$$

a concave function of ν.

$$\Rightarrow p^* \ge -\tfrac{1}{4} \nu^T A A^T \nu - \nu^T b \quad \text{for all } \nu.$$

The dual Problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

* Find best lower bound for p^*
* It is always a convex optimization problem.
* Let its optimal value to be d^*.

Weak and Strong duality

weak duality : $d^* \leq p^*$

* always holds (for convex and nonconvex problems)

Strong duality : $d^* = p^*$

* ~~does~~ Not hold in general.
* (usually) holds for convex problems.
* conditions that guarantee strong duality in convex problems are called constraint qualifications. i.e., slater's constraint qualification, KKT conditions.

Slater's constraint qualification

Strong duality holds for a convex problem

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \leq 0, \quad i=1,\cdots,m \\ &&& Ax = b \end{aligned}$$

if it is strictly feasible. i.e.,

$\exists x \in \text{int} D \; ; \; f_i(x) < 0, \; i=1,\cdots,m, \quad Ax=b$

Complementary slackness

Assume strong duality holds. x^* is primal optimal, (λ^*, ν^*) is dual optimal

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

* two inequalities hold with equality.
* x^* minimize $L(x, \lambda^*, \nu^*)$
* $\lambda_i^* f_i(x^*) = 0$ for $i=1,\cdots,m$ (complementary slackness)

i.e.,

$$\begin{cases} \lambda_i^* > 0 \Rightarrow f_i(x^*) = 0 \\ f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0 \end{cases}$$

Karush - Kuhn - Tucker (KKT) conditions

If f_i, h_i differentiable, the following four conditions are called KKT conditions.

① Primal constraints: $f_i(x) \leq 0,\ i=1,\cdots,m$; $h_i(x)=0,\ i=1,\cdots,p$

② dual constraints: $\lambda \succeq 0$

③ complementary slackness: $\lambda_i f_i(x) = 0,\ i=1,\cdots,m$

④ gradient of Lagrangian with respect to x vanishes

$$\nabla_x L(x,\lambda,\nu) = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

* If strong duality ~~to~~ ($d^* = p^*$) holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

* If $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal.

* If Slater's constraint qualification is satisfied, x is optimal <u>iff</u> there exist λ, ν that satisfy KKT conditions

Lagrange dual and conjugate function

For problem

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad Ax \preceq b, \quad Cx = d \end{aligned}$$

The dual function is

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -\sup_x \left[(-A^T\lambda - C^T\nu)^T x - f_0(x) \right] - b^T\lambda - d^T\nu \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

* Recall the conjugate $f^*(y) = \sup_x (y^T x - f(x))$
* Simplifies derivation of dual function if the conjugate of f_0, say f_0^*, is known.