

# Chapter 3: Convex Optimization Terminology

---

## convex optimization problem (or program)

---

$$\begin{array}{ll} \min_{x \in D} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, \dots, m \\ & Ax = b \end{array}$$

where  $f$  and  $g_i, i = 1, \dots, m$  are all convex, and the optimization domain is  $D = \text{dom}(f) \cap \bigcap_{i=1}^m \text{dom}(g_i)$  (often we do not write  $D$ )

- $f$  is called criterion or objective function
  - $g_i$  is called inequality constraint function
  - If  $x \in D, g_i(x) \leq 0, i = 1, \dots, m$ , and  $Ax = b$  then  $x$  is called a feasible point
  - The minimum of  $f(x)$  over all feasible points  $x$  is called the optimal value, written  $f^*$
  - If  $x$  is feasible and  $f(x) = f^*$ , then  $x$  is called optimal; also called a solution, or a minimizer.
  - If  $x$  is feasible and  $f(x) \leq f^* + \epsilon$ , then  $x$  is called  $\epsilon$ -suboptimal
  - If  $x$  is feasible and  $g_i(x) = 0$ , then we say  $g_i$  is active at  $x$
- 

## Difference between $\inf_D f(x)$ and $\min_D f(x)$ .

- Give an example of a strictly convex function that does not attain its infimum.  $f(x) = \exp(-x)$  is strictly convex and  $\inf_{x \in \mathbb{R}} f(x) = 0$ , but no  $x^*$  such that  $f(x^*) = 0$ .
- Strongly convex function always attains its infimum.
- A continuous function  $f$  on a closed set  $D$  always attains its infimum and supremum.

## Solution set

---

Let  $X_{\text{opt}}$  be the set of all solutions of convex problem, written

$$\begin{aligned} X_{\text{opt}} = \operatorname{argmin} f(x) \\ \text{subject to } g_i(x) \leq 0, i = 1, \dots, m \\ Ax = b \end{aligned}$$

### Key property 1 : $X_{\text{opt}}$ is a convex set

Proof: use definitions. If  $x, y$  are solutions, then for  $0 \leq t \leq 1$ .

- $g_i(tx + (1 - t)y) \leq tg_i(x) + (1 - t)g_i(y) \leq 0$
- $A(tx + (1 - t)y) = tAx + (1 - t)Ay = b$
- $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) = f^*$  Therefore  $tx + (1 - t)y$  is also a solution

### Key property 2: if $f$ is strictly convex, then the solution is unique, i.e., $X_{\text{opt}}$ contains one element $x^*$

If  $x$  and  $y$  are both optimal, then  $x_t = tx + (1 - t)y$  has a lower function value than the minimum  $f^*$ :  $f(x_t) < tf(x) + (1 - t)f(y) = f^*$  when  $t \in (0, 1)$ .

## Properties and first-order optimality

---

### First order optimality condition for convex problem in general

---

For a convex problem

$$\min_x f(x) \quad \text{subject to } x \in C$$

where the feasible set  $C$  is convex and the convex  $f$  is *differentiable*, a feasible point  $x^*$  is optimal **if and only if**

$$\nabla f(x^*)^T(y - x^*) \geq 0, \forall y \in C$$

Proof: If the above first order condition holds, then by the convexity of  $f$  at  $x^*$ , we have

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

So  $x^*$  is optimal.

If  $x^*$  is optimal, i.e.,

$$f(y) \geq f(x^*), \quad \forall y \in C,$$

to show the first order optimality is true. We assume that there exists one point  $z \in C$  such that  $\nabla f(x^*)^T(z - x^*) < 0$ . Define  $y_t = (1 - t)x^* + tz$  where  $t \in [0, 1]$ , we have

$$\frac{d}{dt}f(y_t) = \nabla f(y_t)(x^* - z)|_{t=0} < 0.$$

So by the [mean-value theorem from calculus](#) for very small  $t$ , we have  $f(y_t) < f(y_0) = f(x^*)$ , which is a contradiction.

In other words, all feasible directions from  $x^*$  (i.e.  $y - x^*$ ) are aligned with gradient. When the optimization is unconstrained, this reduces to

$$\nabla f(x^*) = 0.$$

While this is a very general condition (and it is *sufficient* and *necessary*!), it is difficult to check for most problems and hence is not very useful.

## Examples

---

### Example: lasso

---

Given  $y \in \mathbb{R}^n$ ,  $X \in \mathbb{R}^{n \times p}$ ,  $\rho > 0$ , consider the lasso problem:

$$\begin{aligned} \min_{\beta} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq \rho \end{aligned}$$

- What is the criterion/objective function? The inequality and equality constraints? Feasible set?
- The first order optimality condition reads

$$[X^T(X\beta^* - y)]^T(\beta - \beta^*) \geq 0$$

for all  $\|\beta\|_1 \leq \rho$ .

- Is this a convex problem? – yes.
- Is the solution unique, when:
- $n \geq p$  and  $X$  has full column rank?

$X^T X \succ 0$ . the solution uniquely exists.

- $p > n$  (“high-dimensional” case)?

exists, may not unique.

## Example: exact solution of constraint Lasso in 1D

Given  $v \in \mathbb{R}$  and  $\rho > 0$ , consider

$$\begin{array}{ll} \min_{\beta} & |\beta - v|^2 \\ \text{subject to} & |\beta| \leq \rho \end{array}$$

The solution (draw a picture) is

$$\beta^* = \begin{cases} v, & |v| \leq \rho \\ \rho \operatorname{sign}(v), & |v| \geq \rho \end{cases}$$

Check the optimality  $(\beta^* - v)(\beta - \beta^*) \geq 0$  for all  $|\beta| \leq \rho$ .  
draw  $\beta^*$  as the function of  $v$ .

## Example: support vector machine

Given  $y \in \{-1, 1\}^n$ ,  $X \in \mathbb{R}^{n \times p}$  with rows  $x_1, \dots, x_n$ , and a positive constant  $C$ , consider the support vector machine or SVM problem (with slack variables  $\xi$ ):

$$\begin{array}{ll} \min_{\beta, \beta_0, \xi} & \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} & \xi_i \geq 0, i = 1, \dots, n \\ & y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i, i = 1, \dots, n \end{array}$$

- Is this convex? Yes.
- What is the criterion, constraints, feasible set?

- Is the solution  $(\beta, \beta_0, \xi)$  unique?

No. Just consider  $n = 1$  case.

Let  $x_1 = 1, y_1 = 1$ , the problem becomes

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\beta^2 + C\xi \\ \text{subject to} \quad & \xi \geq 0 \\ & \beta + \beta_0 \geq 1 - \xi \end{aligned}$$

where  $(\beta, \beta_0, \xi) = (0, 1, 0), (\beta, \beta_0, \xi) = (0, 2, 0)$  are both optimal solutions and the optimal value of objective function  $f^* = 0$ .

- What if changed the criterion to

$$\frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}\beta_0^2 + C \sum_{i=1}^n \xi_i^{1.01}?$$

yes

## hinge form of SVMs

Rewrite the constraint  $y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i$  and  $\xi_i > 0$  as

$$\xi_i \geq \max \{0, 1 - y_i (x_i^T \beta + \beta_0)\} = \left[ 1 - y_i (x_i^T \beta + \beta_0) \right]_+$$

So the hinge form of SVM is

$$\min_{\beta, \beta_0} \frac{1}{2}\|\beta\|_2^2 + C \sum_{i=1}^n \left[ 1 - y_i (x_i^T \beta + \beta_0) \right]_+$$

where  $a_+ = \max\{0, a\}$  is called the *hinge function* (RELU).

Is this a convex function of  $(\beta, \beta_0)$ ?

Yes. reason:  $t \rightarrow \max(0, 1 - t)$  is convex, affine maps, non-negative sum of convex funs.

## Linear Programs

Fundamental problem in convex optimization, which can be solved using the simplex algorithm (noniteratively) and interior point methods (iteratively). Any LP described in the Basic Form below can be re-written in the Standard Form.

- Basic Form:  $\min_x c^T x$ , s.t.  $Dx \leq d$  and  $Ax = b$
- Standard Form:  $\min_x c^T x$ , s.t.  $Ax = b$  and  $x \geq 0$

## Quadratic minimization

Consider minimizing  $f(x) = \frac{1}{2}x^T Qx + b^T x + c$  where  $Q \succeq 0$ . Then  $f$  is convex. Optimality condition is

$$\nabla f(x) = Qx + b = 0$$

If  $Q \succ 0$ ,  $x = -Q^{-1}b$  is a unique solution

- If  $Q$  is singular and  $b \notin \text{col}(Q)$ , there are no solutions
- If  $Q$  is singular and  $b \in \text{col}(Q)$ , there are infinite solutions  $x = Q^+b + z$ , where  $Q^+$  is the pseudoinverse and  $z \in \text{null}(Q)$ .

## Projection on a convex set

Consider projection on convex set  $C$ : given a point  $a \notin C$ ,

$$\min_x \|x - a\|_2^2 \text{ subject to } x \in C.$$

The solution, denoted as  $P_C(a)$ , is unique since  $\|x - a\|_2^2$  is strongly convex.

According to first order optimality condition,

$$\nabla f(x^*)^T (y - x^*) = -(a - x^*)^T (y - x^*) \geq 0, \forall y \in C.$$

This says that  $a - x^*$  lies in the normal cone  $\mathcal{N}_C(x^*) = \{g : g^T x^* \geq g^T y, \text{ for all } y \in C\}$

## Example: linear-equality-constrained minimization

Consider the equality-constrained convex problem:

$$\min_x f(x) \text{ subject to } Ax = b$$

with  $f$  differentiable. Let's prove Lagrange multiplier optimality condition

$$\nabla f(x) + A^T u = 0 \quad \text{for some } u.$$

According to first-order optimality, solution  $x$  satisfies  $Ax = b$  and

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{for all } y \text{ such that } Ay = b$$

This is equivalent to

$$\nabla f(x)^T v = 0 \quad \text{for all } v \in \text{null}(A)$$

Result follows because  $\text{null}(A)^\perp = \text{row}(A)$

## Example: basis pursuit

---

Given  $y \in \mathbb{R}^n$  and  $X \in \mathbb{R}^{n \times p}$ , where  $p > n$ . Suppose that we seek the sparsest solution to underdetermined linear system  $X\beta = y$

Nonconvex formulation:

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_0 \\ \text{subject to} & X\beta = y \end{array}$$

where  $\|\beta\|_0 = \sum_{j=1}^p 1\{\beta_j \neq 0\}$ , the  $\ell_0$  "norm"

The  $\ell_1$  approximation, often called **basis pursuit**:

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_1 \\ \text{subject to} & X\beta = y \end{array}$$


---

Basis pursuit is a linear program. Reformulation:

$$\min_{\beta, z} 1^T z$$

subject to  $z \geq \beta, z \geq -\beta, X\beta = y.$

## Example: Dantzig selector (optional)

---

Modification of previous problem, where we allow for  $X\beta \approx y$  (we don't require exact equality), the [Dantzig selector](#):

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_1 \\ \text{subject to} & \|X^T(y - X\beta)\|_{\infty} \leq \lambda \end{array}$$

Here  $\lambda \geq 0$  is a tuning parameter

Again, this can be reformulated as a linear program (check this!)

### References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 4
- O. Guler (2010), "Foundations of optimization", Chapter 4
- D. Bertsimas and J. Tsitsiklis (1997), "Introduction to linear optimization," Chapters 1, 2
- A. Nemirovski and A. Ben-Tal (2001), "Lectures on modern convex optimization," Chapters 1–4