Constrained optimization in standard form

$$\min_{x \in \mathbb{R}^n} f_0(x). \quad \text{s.t} \quad \left.\begin{array}{ll} f_i(x) \le 0, & i = 1, \dots, m \\ h_i(x) = 0, & i = 1, \dots, p. \end{array}\right\} \text{(original) primal formulation} \quad -①$$

Here, $D = \bigcap_{i=0}^{m} \underbrace{\mathrm{dom}(f_0)}_{\mathbb{R}^n} \cap \bigcap_{i=1}^{p} \underbrace{\mathrm{dom}(h_i)}_{\mathbb{R}^n}.$

Take the constraints and add to the objective with weighted sum of constraints. to get the Lagrangian.

$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ s.t

$$\underbrace{\mathcal{L}(x, \lambda, \nu)}_{\text{Lagrangian.}} = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h(x).$$

Lagrange multipliers.

Let $g : \mathbb{R}^m \times \mathbb{R}^p$, s.t.

$$g(\lambda, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu). \quad \rightarrow g \text{ is called dual function.}$$

motivation : ~~$g(\lambda,\nu)$ provides~~ Let x^* be the ~~opt~~ minimizer of ① and let $p^* = f_0(x^*)$ be the primal optimal objective.

$g(\lambda, \nu)$ provides a lower bound for certain λ, ν.

- This can be used to determine a condition λ.

$$\mathcal{L}(x^*,\lambda,\nu) = f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_0(x^*) + \sum_{i=1}^{p} \nu_i \underbrace{h_i(x^*)}_{=0}$$

because of feasibility.

want $p^* = f_0(x^*) \geq \mathcal{L}(x^*,\lambda,\nu)$

$$= f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_0(x^*).$$

$\Rightarrow \lambda_i f_0(x^*) \leq 0.$

$\Rightarrow \lambda_i \geq 0$ since $f_0(x^*) \leq 0$.

<u>properties of g:</u>

- g is concave wrt to (λ,ν) because $\mathcal{L}(x,\nu,\lambda)$ is ~~linear~~ affine wrt λ & ν. (and inf of concave function is concave).
- for some (λ,ν) g can be $-\infty$.
- for $\lambda \geq 0$, $p^* \geq g(\lambda,\nu)$.

<u>eg:</u> $\min_{x\in\mathbb{R}^n} \|x\|_2^2$ s.t. $Ax=b$, $A\in\mathbb{R}^{m\times n}$.

$$\mathcal{L}(x,\nu) = \|x\|_2^2 + \cancel{\nu^T(Ax-b)} \ \langle \nu, Ax-b\rangle.$$

$$g(\nu) = \inf_x \mathcal{L}(x,\nu).$$

$$\nabla_x \mathcal{L}(x,\nu) = \cancel{\tfrac{1}{2}x} \ 2x + A^T\nu = 0.$$

$$\Rightarrow x = -\frac{1}{2}A^T\nu.$$

$$\Rightarrow g(\nu) = \frac{1}{4}\|A^T\nu\|_2^2 + \langle \nu, -\tfrac{1}{2}AA^T\nu - b\rangle.$$

$$= \frac{1}{4}\|A^T\nu\|_2^2 - \frac{1}{2}\|A^T\nu\|_2^2 - \langle \nu, b\rangle.$$

$$= -\frac{1}{4}\|A^T\nu\|_2^2 - \langle \nu, b\rangle.$$

· By lower bound property:

$$g(\nu) = -\frac{1}{4}\|A^T\nu\|_2^2 - \langle \nu, b \rangle \leq p^* \quad \text{for all } \nu \in \mathbb{R}^m.$$

eg: <u>Linear program</u>

$$\min_{x \in \mathbb{R}^n} \; c^T x \quad \text{s.t.} \quad Ax = b, \; \underset{x_i \geq 0 \; \forall i}{\overset{\uparrow}{x \geq 0}} \quad \left(\underset{c \in \mathbb{R}^n}{c\text{-fixed}} \right)$$

$$\mathcal{L}(x, \nu, \lambda) = c^T x + \nu^T(Ax - b) - \lambda^T x \qquad (\lambda \geq 0).$$

$$= (c + A^T\nu - \lambda)^T x - \nu^T b$$

$$g(\nu, \lambda) = \inf_{x \in \mathbb{R}^n} (c + A^T\nu - \lambda)^T x - \nu^T b.$$

$$= \begin{cases} -\infty & \text{if } c + A^T\nu - \lambda \neq 0. \\ -\nu^T b & \text{if } c + A^T\nu - \lambda = 0. \end{cases}$$

and $p^* \geq -\nu^T b$ if $c + A^T\nu - \lambda = 0$

Finding the best lower bound for the primal optimal.

$$\left.\max_{(\lambda, \nu)} \; g(\lambda, \nu) \quad \text{s.t.} \quad \underbrace{\lambda \geq 0}_{\text{dual feasibility.}} \right\} \text{dual } \cancel{\text{problem}}. \text{ program}$$

· if (λ^*, ν^*) ~~maximize~~ is the maximizer of dual problem

Then (λ^*, ν^*) are ~~know~~ called dual optimal.

- Because $g(\lambda, \nu)$ is concave, dual program is a convex program.

eg: Linear program:

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax = b, \ x \succeq 0.$$

$$g(\lambda, \nu) = \begin{cases} -\infty & \text{if } c + A^T\nu - \lambda \neq 0 \rightarrow \text{trivial} \\ -\nu^T b & \text{if } c + A^T\nu - \lambda = 0 \end{cases}$$

dual program:

$$\max \ g(\lambda, \nu) \quad \text{s.t.} \ \lambda \succeq 0.$$

$$\Leftrightarrow \max \ -\nu^T b \quad \text{s.t.} \ c + A^T\nu - \lambda = 0, \ \lambda \succeq 0.$$

$$\Leftrightarrow \min \ \nu^T b \quad \text{s.t.} \ c + A^T\nu - \lambda = 0, \ \lambda \succeq 0.$$

Note that $p^* = f(x^*) \geq g(\nu^*, \lambda^*)$. → best lower bound

The gap $f(x^*) - g(\nu^*, \lambda^*)$ is called the duality gap.

Q: When is the duality gap zero?

~~Holds for convex problem.~~

Slater's condition:

For convex problem,

$$\min \ f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \quad i = 1, \dots, m. \qquad Ax = b$$

strong duality holds (i.e. duality gap is zero) if it is strictly feasible i.e.

$\exists \; x \in \text{interior}(\mathcal{D})$ s.t. $f_i(x) < 0 \quad \forall i,$

$$Ax = b.$$

eg: <u>Quadratic program</u>. $(P \succ 0)$ → positive definite matrix.

$$\min \; x^T P x \quad \text{s.t.} \quad Ax \preceq b.$$

exercise →

strong duality if $\exists \, x$ s.t. $Ax \prec b$.

$$\left[g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \right]$$

<u>Complementary slackness</u>

want: $$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right)$$

$$\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \underbrace{\sum_{i=1}^p \nu_i^* h_i(x^*)}_{=0}$$

$$\Rightarrow \sum_{i=1}^m \lambda_i^* f_i(x^*) = 0.$$

Complementary slackness: $\lambda_i^* f_i(x^*) = 0 \quad \forall \, i = 1, \ldots, m.$

or equivalently

$$\left.\begin{array}{l} \lambda_i^* > 0 \Rightarrow f_i(x^*) = 0 \\ f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0 \end{array}\right\} \text{shows which inequality constraint is active at the minimizer}$$

Also, note that x^* minimizes $\mathcal{L}(x, \lambda^*, \nu^*)$ over $x \in$

$$\Rightarrow \nabla \mathcal{L}(x^*, \lambda^*, \nu^*) = 0$$

KKT condition for convex program

~~let x^* be~~ primal optimal ~~& let (λ^*, ν^*) be dual optimal Then~~.

x^* is primal optimal and (λ^*, ν^*) are dual optimal iff.

KKT conditions:

$$\begin{cases} \nabla \mathcal{L}(x^*, \lambda^*, \nu^*) = 0 & \rightarrow \text{stationary condition.} \\ f_i(x^*) \le 0, \quad i = 1, \dots, m \\ h_i(x^*) = 0, \quad i = 1, \dots, p \end{cases}$$

$\left.\begin{matrix} f_i(x^*) \le 0, & i=1,\dots,m \\ h_i(x^*) = 0, & i=1,\dots,p \end{matrix}\right\} \rightarrow$ primal feasibility

$\lambda_i^* \geq 0, \quad i = 1, \dots, m \rightarrow$ dual feasibility.

$\lambda_i^* f_i^* = 0, \quad i = 1, \dots, m \rightarrow$ complementary slackness.

- Note that duality gap is 0.
- For non convex program ~~the~~ KKT condition are necessary for (x^*, λ^*, ν^*) to be optimal.

eg: $\min \frac{1}{2} x^T P x \qquad$ s.t. $\quad Ax = b$

where $P \in S_+^n$.

KKT conditions are: $\nabla_x \mathcal{L}(x, \nu) = 0$

$$\Rightarrow \cancel{\tfrac{1}{2}} P x + A^T \nu = 0.$$

primal feasibility $\Rightarrow Ax = b$.

so,

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

Solving this provides us with optimal x, ν.