801 Recitation 3 Duality, KKT Condition, dual certificate, and examples

Primal: $\min f(x)$

s.t. $h_i(x) \le 0$ for $i = 1, \dots, m$

$\ell_j(x) = 0$ for $j = 1 \dots p$

Dual: $\max g(u,v) = \inf_x \underbrace{f(x) + u^T h(x) + v^T \ell(x)}_{\text{Lagragian}}$

s.t. $u \ge 0$

($g(u,v)$ may contain indicator functions, so there will often be "implicit" constraint here when we derive $g(u,v)$.)

Notes. ① $u \ge 0$, $h(x) \le 0$, $\ell(x) = 0$, for ∀ feasible pair of (x, u)

so $f(x) \ge L(x,u,v)$. in particular $f^*(x) \ge \inf_{x \in X} L(x,u,v) \ge \inf_{x \in \mathbb{R}^n} L(x,u,v) = g(u,v)$

* so any feasible (u,v) will give us a lower bound of $f^*(x)$. (Weak Duality!)

② By optimality, $g(u,v) \le g^*$, $f^* \le f(x)$

③ concatenate the inequality we get:

$$g(u,v) \le g^*(u,v) = g(u^*,v^*) = \inf_{x \in \mathbb{R}^n} L(x,u^*,v^*) \le \inf_{x \in X} L(x,u^*,v^*) \le f^*(x) \le f(x)$$

Ⓐ Stationarity / saddle point condition:

$$0 \in \partial L(x^*, u^*, v^*)$$

Ⓑ complementary slackness

$u^T h(x) = 0$

so either $u_i = 0$ or $h_i(x) = 0$

Ⓒ primal and dual feasibility.

KKT Conditions

For convex optimization, KKT condition is sufficient for optimality, and if strong duality holds, it is also necessary.

Exp 1 Entropy Minimization

$$\boxed{\begin{array}{l} \min \sum_{i=1}^n x_i \log x_i \quad (\text{min negative Entropy}) \\ \text{s.t. } Ax \le b,\ 1^T x = 1 \end{array}}$$

$L = \sum_{i=1}^n x_i \log x_i + u^T(Ax - b) + v(1^T x - 1)$

$g(u,v) = \inf_{x_i} L(x,u,v)$

$\frac{\partial L}{\partial x_i} = \log x_i + x_i \frac{1}{x_i} + u^T A_i + v_i$

$\therefore x_i = e^{-u^T A_i - v_i - 1}$

plug it back in

$$g(u,v) = \sum \left[x_i(-u^T A_i - v_i - 1) + u^T(A_i x_i - b_i) + v_i x_i \right] - v$$

$$= -u^T b - v - \sum_{i=1}^n x_i$$

$$= -u^T b - v \sum_{i=1}^n e^{-v_i - 1} e^{-u^T A_i}$$

$$\boxed{\text{Dual: } \max\ g(u,v) \quad \text{s.t. } u \ge 0}$$

Projected gradient easily solvable!

Using Fenchel Conjugate. $f^*(x) = \sup_y \{\langle x, y\rangle - f(y)\}$

(3) Lasso duals (what if the primal is unconstrained?)

$$\min \tfrac{1}{2}\|y - Ax\|^2 + \lambda\|x\|_1$$

Version 1: $\min \tfrac{1}{2}\|y - z\|^2 + \lambda\|x\|_1$

s.t. $z = Ax$

$$g(u) = \min_{x,z} \tfrac{1}{2}\|y - z\|^2 + \lambda\|x\|_1 + u^T(z - Ax)$$

$$= \min_z u^T z + \tfrac{1}{2}\|y - z\|^2 + \min_x \lambda\|x\|_1 - (A^T u)^T x$$

Fenchel conjugate $(\|x\|_1)^* = I_{\|u\|_\infty \le 1}(\cdot)$

so $\min_x \lambda\|x\|_1 - (A^T u)^T x = -I_{\|v\|_\infty \le 1}(A^T u/\lambda)$

$\min_z u^T z + \tfrac{1}{2}\|y - z\|^2$ can be easily solved by taking derivative.

$z = y - u$

so $g(u,v) = u^T y - \tfrac{1}{2}\|u\|^2 - I(\|A^T u\|_\infty \le \lambda)$

So Dual program:

$$\max_u \langle u, y\rangle - \tfrac{1}{2} u^T u$$

s.t. $\|A^T u\|_\infty \le \lambda$

Version 2:

$\min \tfrac{1}{2}\|e\|^2 + \lambda\|x\|_1$

s.t. $y = Ax + e$

$$L = \tfrac{1}{2}\|e\|^2 + \lambda\|x\|_1 + u^T(y - Ax - e)$$

$$g(u) = \min_{x,e} \tfrac{1}{2}\|e\|^2 - u^T e + u^T y + \lambda\|x\|_1 - (A^T u)^T x$$

$$= u^T y + \min_e \tfrac{1}{2}\|e\|^2 - u^T e + \min_x \lambda\|x\|_1 - (A^T u)^T x$$

$$= u^T y - \tfrac{1}{2}\|u\|^2 - I(\|A^T u\|_\infty \le \lambda)$$

Dual $\max \langle u, y\rangle - \tfrac{1}{2} u^T u$

s.t. $\|A^T u\|_\infty \le \lambda$

(3) matrix compressive sensing, error correction

$\min_{X,e} \|X\|_* + \lambda\|e\|_1$

s.t. $\mathcal{A}(X) = b + e$

$$g(x,e) = \min_{x,e} \|X\|_* + \lambda\|e\|_1 + u^T(b + e - \mathcal{A}(X))$$

$$= u^T b - \max_X(\langle \mathcal{A}^*(u), X\rangle - \|X\|_*) - \max_e(\langle -u, e\rangle - \lambda\|e\|_1)$$

$$= u^T b - I(\|\mathcal{A}^*(u)\| \le 1) - I(\|u\|_\infty \le \lambda)$$

Nuclear norm / ~~matrix compressive sensing~~

$\min \|X\|_*$

s.t. $\mathrm{tr}(A^T X) = b$

Dual

$$g(u) = \min_X \|X\|_* + u(b - \mathrm{tr}(A^T X))$$

$$= ub - \max_X \langle uA, X\rangle - \|X\|_*$$

$$= ub - I(\|uA\|_{op} \le 1)$$

$\max ub$

s.t. $\|uA\|_{op} \le 1$ $= \max\{u \mid \|uA\|_{op} \le 1\}$

$= \max\{u \mid u\|A\|_{op} \le 1\}$

$= \dfrac{1}{\sigma_{\max}(A)}$!

(4) Norm approximation (Different way to derive duals lead to different duals)

Primal: minimize $\|Ax-b\|$

Reformulation 1: $\min \|y\|$ s.t. $Ax-b=y$

$$g(u) = \min_{x,y} \|y\| + u^T(Ax-b-y)$$

$$= -u^T b + \min_x u^T A x + \min_y \|y\| - \langle u, y\rangle$$

$f(x)=0$

By Fenchel Conjugate

$$\min_x u^T A x = -\max_x 0 + \langle A^T u, x\rangle$$
$$= 0^*(A^T u) = I(A^T u = 0)$$

$$\min_y \|y\| - \langle u,y\rangle = -\max \langle u,y\rangle - \|y\|$$
$$= I(\|u\|^* \le 1)$$

Dual: $\max_u -u^T b$ s.t. $\|u\|^* \le 1$, $A^T u = 0$ $=$ (sign doesn't matter) $\max u^T b$ s.t. $\|u\|^* \le 1$, $A^T u = 0$

Reformulation 2:

$$\min_{x,y} \frac{1}{2}\|y\|^2 \quad \text{s.t. } Ax-b=y$$

$$g(u) = \min_{x,y} \frac{1}{2}\|y\|^2 + u^T(Ax-b-y)$$

Immediate difference: differentiable now!

We use Fenchel Conjugate anyway

$$= \min_x u^T A x + \min_y \frac{1}{2}\|y\|^2 - u^T y - u^T b$$

$$= I(A^T u = 0) - u^T b - \max (u^T y - \frac{1}{2}\|y\|^2)$$

$$= I(A^T u = 0) - u^T b - \frac{1}{2}(\|u\|^*)^2$$

Dual: $\max -u^T b - \frac{1}{2}(\|u\|^*)^2$ s.t. $A^T u = 0$

We get rid of one constraint!

Example for KKT condition: (why do we need it?)

(1) Analytic solutions for some problem

$$\boxed{\begin{aligned} &\min \frac{1}{2} x^T P x + q^T x + r \qquad P \succeq 0 \\ &\text{s.t. } Ax = b \end{aligned}}$$

$$L = f(x) + v^T(Ax-b)$$

$$g(v) = \inf_x L(x,v)$$

Stationarity

is given by $\nabla f(x) + A^T v = 0$

$$\begin{cases} Px + q + A^T v = 0 \\ Ax = b \leftarrow \text{Primal Feasibility} \end{cases}$$

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$ a linear system of equations!

* For convex, twice differentiable $f(x)$, we can iteratively solve its taylor expansion with this formula.

This gives us a Newton's algorithm for constrained optimization!

~~Note~~

Dual Certificate

Suppose we have x^* ($\in \mathcal{X}$ (feasible)). How can we tell if it is the optimal solution for the primal?, How about uniqueness?

One way is to compare $f(x^*)$ against all other $f(x)$ ...

A better way is to test the KKT condition. But we do not have a dual variable!

Theorem. If we can construct v^* such that v^* is dual feasible and the duality gap vanishes, then x^* is the optimal solution to primal.
Moreover, if we have "strict complementary" then x^* is the unique optimal solution.

We say v^* certifies the optimality of x^*, or v^* is a dual certificate.

Example

Compressive sensing / $\ell 1$ support recovery

$$\boxed{\begin{array}{l} \min \|x\|_0 \ \text{s.t.}\ Ax=b \\ \quad \text{convex} \Downarrow \text{relaxation} \\ P: \min \|x\|_1 \ \text{s.t.}\ Ax=b \\ D: \max b^T u \ \text{s.t.}\ \|A^T u\|_\infty \le 1 \end{array}}$$

Given an x^*, $\text{supp}(x^*) = I$ (support set), $x^*_{I^c} = 0$

How to tell if this is the unique optimal?

Conditions

① Vanishing duality gap: $\|x\|_1 = b^T u = x^T A^T u$

$\|x_I\|_1 = \langle x_I, [A^T u]_I \rangle \Leftrightarrow [A^T u]_I = \text{Sgn}(x_I)$

② Feasibility: $\|A^T u\|_\infty \le 1$

Now it boils down to constructing a u s.t.

$$\text{So} \begin{cases} [A^T u]_I = \text{Sgn}(x_I) \\ |[A^T u]_{I^c}| \le 1 \end{cases}$$

are satisfied.

if this is strictly $<$ we say we have "strict complementary" then x^* is the unique solution.

Often non-trivial, but it reveals the key properties of the problem.

Here, we can construct such a u if A is RIP.

$$(1-\delta_s)\|c\| \le \|Ac\| \le (1+\delta_s)\|c\|$$