Dual of Linear Program (LP)

LP in "standard" form:

$$\text{minimize } \; c^T x \qquad \text{——— } (P)$$
$$\text{s.t. } \; Ax \leq b, \; x \geq 0$$

$$D(\mu_1, \mu_2) = \min_{x \in \mathbb{R}^n} \; c^T x + \mu_1^T (Ax - b) - \mu_2^T x$$

$\mu_1 \geq 0$
$\mu_2 \geq 0$

$$= \min_{x \in \mathbb{R}^n} \; (c^T + \mu_1^T A - \mu_2^T) x - \mu_1^T b$$

$$= \begin{cases} -\infty & \text{if } c^T + \mu_1^T A - \mu_2^T \neq 0 \\ -\mu_1^T b & \text{if } c^T + \mu_1^T A - \mu_2^T = 0 \end{cases}$$

Note $\quad c^T + \mu_1^T A - \mu_2^T = 0 \quad \equiv \quad A^T \mu_1 + c = \mu_2$

Therefore the dual problem is:

$$\text{maximize } \; -\mu_1^T b$$
$$\text{s.t. } \; \mu_1 \geq 0, \; \mu_2 \geq 0, \; A^T \mu_1 + c = \mu_2$$

$$\equiv \quad \text{minimize } \; \mu_1^T b$$
$$\text{s.t. } \; \mu_1 \geq 0, \; A^T \mu_1 + c \geq 0$$

$$\equiv \quad \text{minimize } \; \tilde{x}^T b \qquad \text{— same form as } (P)$$
$$\text{s.t. } \; -A^T \tilde{x} \leq c, \; \tilde{x} \geq 0$$

Dual of Dual: $\quad \text{minimize } \; x^T c \qquad$ Primal!
$$\text{s.t. } \; Ax \leq b, \; x \geq 0$$

Augmented Lagrangian and Method of Multipliers

(Recall) Penalty Method:

$$\min f(x) \quad \text{s.t. } h(x) = 0 \qquad \text{—} \qquad (P)$$

solve sequence of unconstrained problems:

$$\text{minimize } f(x) + c_k \|h(x)\|^2 \qquad \text{—} \qquad (P_k)$$

$c_k > 0$ and $c_k \uparrow \infty$ as $k \to \infty$.

If $x^{(k)}$ is solution to (P_k), we showed in lec 18 that every limit point $\bar{x}$ of $\{x^{(k)}\}$ is solution to (P)

Apply penalty method to:

$$\text{minimize } x^T Q x \qquad Q \succ 0$$
$$\text{s.t. } Ax = b \qquad A_{m \times n}, \; m < n$$

Penalty problem sequence:

$$\text{minimize } x^T Q x + c_k \| Ax - b \|^2$$

$$\| Ax - b \|^2 = (Ax - b)^T (Ax - b)$$
$$= x^T A^T A x - 2 x^T A^T b + b^T b$$

Note: $A^T A \succeq 0$

Since $m < n$, $\text{rank}(A^T A) \leq m < n$

$$\Rightarrow \lambda_{\min}(A^T A) = 0.$$

$$(P_k): \text{minimize } \overbrace{x^TQx + c_k(x^TA^TAx - 2x^TA^Tb + b^Tb)}^{g^{(k)}(x)}$$

i.e. minimize $x^T(Q + c_kA^TA)x - 2c_kx^TA^Tb + c_kb^Tb$

Since $Q \succ 0$, $c_k > 0$, $A^TA \succeq 0$, $(Q + c_kA^TA) \succ 0$.

(P_k) is a (strictly) convex optimization problem

$\Rightarrow$ solution $x^{(k)}$ satisfies $\nabla g^{(k)}(x^{(k)}) = 0$

i.e.

$$(Q + c_kA^TA)x^{(k)} - c_kA^Tb = 0$$

$$\Rightarrow \quad x^{(k)} = (Q + c_kA^TA)^{-1}c_kA^Tb$$

$$= \left(\frac{Q}{c_k} + A^TA\right)^{-1}A^Tb$$

If we use gradient descent to solve (P_k) then rate of convergence depends on

$$\text{condition number} = \frac{\lambda_{max}\left(\frac{Q}{c_k} + A^TA\right)}{\lambda_{min}\left(\frac{Q}{c_k} + A^TA\right)}$$

As $c_k \to \infty$,

$$\lambda_{min}\left(\frac{Q}{c_k} + A^TA\right) \approx \lambda_{min}(A^TA) = 0$$

i.e. optimization problem (P_k) becomes ill-conditioned as $k \to \infty$.

Augmented Lagrangian Method

$$L_c(x,\lambda) = f(x) + \lambda^T h(x) + c\,\|h(x)\|^2$$

with $\lambda \in \mathbb{R}^m$

If $x^{(k)} \in \arg\min_x L_{c_k}(x,\lambda)$, $c_k \uparrow \infty$ as $k \uparrow \infty$

Then every limit point $\bar{x}$ of $\{x^{(k)}\}$ is a global min for (P) (HW # 6)

What is the advantage of adding $\lambda^T h(x)$?
If x^*, λ^* satisfy the second-order sufficiency condition for x^* being a strict local min for (P), $\exists\ \bar{c} > 0$, s.t. if $c \geq \bar{c}$, then for some $\gamma > 0$ and $\varepsilon > 0$,

$$L_c(x,\lambda^*) \geq L_c(x^*,\lambda^*) + \frac{\gamma}{2}\|x - x^*\|^2$$

for all x s.t. $\|x - x^*\| < \varepsilon$

i.e. x^* is a local min of $L_c(x,\lambda^*)$.
(See sec 4.2.1 of text)

Therefore the search for minima of (P) can be restricted to minima of $L_c(x,\lambda^*)$ if $c > \bar{c}$

Therefore if λ can be chosen close to λ^*, augmented Lagrangian method can work without $c_k \uparrow \infty$ as $k \uparrow \infty$
$c_k > \bar{c}$ enough.

Example minimize $f(x) = \frac{1}{2}(x_1^2 + x_2^2)$

s.t. $x_1 = 1$

optimal solution: $x^* = (1, 0)$, $\lambda^* = -1$

$$L_c(x, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + \frac{c}{2}(x_1 - 1)^2$$

– strictly convex in x

$\Rightarrow$ $\nabla_x L_c(x, \lambda) = 0$ gives minimizer:

$$x_1(\lambda, c) = \frac{c - \lambda}{c+1}, \quad x_2(\lambda, c) = 0$$

$\forall c > 0$, $\lim_{\lambda \to \lambda^*} x_1(\lambda, c) = \frac{c+1}{c+1} = 1$

$\lim_{\lambda \to \lambda^*} x_2(\lambda, c) = 0$

$\Rightarrow$ $\lim_{\lambda \to \lambda^*} x(\lambda, c) = x^*$.

For fixed λ, $\lim_{c \to \infty} x_1(\lambda, c) = 1$

$\Rightarrow$ $\lim_{c \to \infty} x(\lambda, c) = x^*$ as well.

If λ can be chosen close to λ^*, then we can avoid the ill-conditioning that may occur as $c_k \to \infty$ in penalty method.

How to make λ close to λ^* without knowing λ^*?

Duality!

(P) : minimize $f(x)$ s.t. $h(x) = 0$.

Lagrangian : $L(x, \lambda) = f(x) + \lambda^T h(x)$

Dual : $D(\lambda) = \min_x L(x, \lambda)$

Let $x(\lambda)$ be a minimizer of $L(x, \lambda)$. Then

$$\nabla_x f(x(\lambda)) + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x(\lambda)) = 0 \quad \text{--- (1)}$$

and

assume cont. diff.

$$D(\lambda) = f(x(\lambda)) + \sum_{i=1}^{m} \lambda_i h(x(\lambda))$$

$$\frac{\partial}{\partial \lambda_i} f(x(\lambda)) = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} f(x(\lambda)) \cdot \frac{\partial x_j(\lambda)}{\partial \lambda_i}$$

$$= \left[\frac{\partial x_1(\lambda)}{\partial \lambda_i} \cdots \frac{\partial x_n(\lambda)}{\partial \lambda_i} \right] \cdot \nabla_x f(x(\lambda))$$

Define
$$\nabla_\lambda x(\lambda) = \begin{bmatrix} \frac{\partial x_1(\lambda)}{\partial \lambda_1} & \cdots & \frac{\partial x_n(\lambda)}{\partial \lambda_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1(\lambda)}{\partial \lambda_m} & \cdots & \frac{\partial x_n(\lambda)}{\partial \lambda_m} \end{bmatrix}$$

Then $\nabla_\lambda f(x(\lambda)) = \nabla_\lambda x(\lambda) \nabla_x f(x(\lambda))$

Similarly $\nabla_\lambda h_i(x(\lambda)) = \nabla_\lambda x(\lambda) \nabla_x h_i(x(\lambda))$

$$\nabla_\lambda D(\lambda) = \nabla_\lambda f(x(\lambda)) + \nabla_\lambda \left(\sum_{i=1}^{m} \lambda_i h_i(x(\lambda)) \right)$$

$$\nabla_\lambda D(\lambda) = \nabla_\lambda f(x(\lambda)) + \nabla_\lambda \left(\sum_{i=1}^{m} \lambda_i h_i(x(\lambda)) \right)$$

$$= \nabla_\lambda x(\lambda) \cdot \nabla_x f(x(\lambda)) + \sum_{i=1}^{m} \lambda_i \nabla_\lambda x(\lambda) \nabla_x h_i(x(\lambda))$$

$$+ h(x(\lambda))$$

$$= \nabla_\lambda x(\lambda) \underbrace{\left(\nabla_x f(x(\lambda)) + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x(\lambda)) \right)}_{= 0 \text{ by } (1)}$$

$$+ h(x(\lambda))$$

$$\Rightarrow \nabla_\lambda D(\lambda) = h(x(\lambda))$$

$D(\lambda)$ is concave in λ $\Rightarrow$ we can use gradient ascent to find λ^*

$$\lambda^{(k+1)} = \lambda^{(k)} + \alpha_k \nabla_\lambda D(\lambda^{(k)})$$

$$= \lambda^{(k)} + \alpha_k h(x(\lambda^{(k)}))$$

This leads to Method of Multipliers:

$$x^{(k)} \in \arg\min_x L_{c_k}(x, \lambda^{(k)})$$

$$\lambda^{(k+1)} = \lambda^{(k)} + c_k h(x^{(k)})$$

Can show that Method of Multipliers converges to min of (P) if $\{c_k\}$ is chosen carefully (don't need to take $c_k \to \infty$). See sec 5.2 of text.

Example minimize $f(x) = \frac{1}{2}(x_1^2 + x_2^2)$

s.t. $x_1 = 1$

$x^* = (1, 0)$ and $\lambda^* = -1$.

$$L_c(x, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + \frac{c}{2}(x_1 - 1)^2$$

$$x^{(k)} = \left(\frac{c_k - \lambda^{(k)}}{c_k + 1},\ 0 \right)$$

$$\lambda^{(k+1)} = \lambda^{(k)} + c_k \left(\frac{c_k - \lambda^{(k)}}{c_k + 1} - 1 \right)$$

$$= \frac{\lambda^{(k)}}{c_k + 1} - \frac{c_k}{c_k + 1}$$

$$(\lambda^{(k+1)} - \lambda^*) = \lambda^{(k+1)} + 1$$

$$= \frac{\lambda^{(k)}}{c_k + 1} - \frac{c_k}{c_k + 1} + 1$$

$$= \frac{\lambda^{(k)}}{c_k + 1} + \frac{1}{c_k + 1}$$

$$= \frac{\lambda^{(k)} - \lambda^*}{c_k + 1}$$

As long as $c_k \geq \bar{c} > 0 \quad \forall k$, for any $\bar{c} > 0$.

$\lambda^{(k)} \longrightarrow \lambda^*$ linearly since $\frac{1}{\bar{c} + 1} < 1$

Thus $\lambda^{(k)} \to \lambda^* \implies x^{(k)} \to (1, 0) = x^*$