

## Dual of Linear Program (LP)

LP in "standard" form:

$$\text{minimize } C^T x$$

$$\text{s.t. } Ax \leq b, x \geq 0$$

—— (P)

$$\begin{aligned} D(\mu_1, \mu_2) &= \min_{x \in \mathbb{R}^n} C^T x + \mu_1^T (Ax - b) - \mu_2^T x \\ &\stackrel{\substack{\uparrow \\ \mu_1 \geq 0 \\ \mu_2 \geq 0}}{=} \min_{x \in \mathbb{R}^n} (C^T + \mu_1^T A - \mu_2^T) x - \mu_1^T b \\ &= \begin{cases} -\infty & \text{if } C^T + \mu_1^T A - \mu_2^T \neq 0 \\ -\mu_1^T b & \text{if } C^T + \mu_1^T A - \mu_2^T = 0 \end{cases} \end{aligned}$$

$$\text{Note } C^T + \mu_1^T A - \mu_2^T = 0 \equiv A^T \mu_1 + C = \mu_2$$

Therefore the dual problem is:

$$\text{maximize } -\mu_1^T b$$

$$\text{s.t. } \mu_1 \geq 0, \mu_2 \geq 0, A^T \mu_1 + C = \mu_2$$

$$\equiv \text{minimize } \mu_1^T b$$

$$\text{s.t. } \mu_1 \geq 0, A^T \mu_1 + C \geq 0$$

$$\equiv \text{minimize } \tilde{x}^T b \quad - \text{Same form as (P)}$$

$$\text{s.t. } -A^T \tilde{x} \leq C, \tilde{x} \geq 0$$

$$\text{Dual of Dual: } \text{minimize } x^T C$$

Primal!

$$\text{s.t. } Ax \leq b, x \geq 0$$

## Augmented Lagrangian and Method of Multipliers

(Recall) Penalty Method:

$$\begin{aligned} \min f(x) \\ \text{s.t. } h(x) = 0 \end{aligned} \quad \text{---} \quad (P)$$

Solve sequence of unconstrained problems:

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

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

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:

$$\begin{aligned} \text{minimize } x^T Q x \\ \text{s.t. } Ax = b \end{aligned} \quad \begin{aligned} Q > 0 \\ A_{m \times n}, m < n \end{aligned}$$

Penalty problem sequence:

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

$$\begin{aligned} \|Ax - b\|^2 &= (Ax - b)^T (Ax - b) \\ &= x^T A^T A x - 2 x^T A^T b + b^T b \end{aligned}$$

Note:  $A^T A \succcurlyeq 0$

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

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

$$\textcircled{P_k} : \text{minimize } \underbrace{g^{(k)}(x)}_{x^T Q x + c_k (x^T A^T A x - 2x^T A^T b + b^T b)}$$

i.e. minimize  $x^T (\Phi + c_k A^T A) x - 2c_k x^T A^T b + c_k b^T b$

Since  $\Phi \geq 0$ ,  $c_k > 0$ ,  $A^T A \succeq 0$ ,  $(\Phi + c_k A^T A) \succeq 0$ .

$\textcircled{P_k}$  is a (strictly) convex optimization problem

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

i.e.

$$(\Phi + c_k A^T A) x^{(k)} - c_k A^T b = 0$$

$$\Rightarrow x^{(k)} = (\Phi + c_k A^T A)^{-1} c_k A^T b$$

$$= \left( \frac{\Phi}{c_k} + A^T A \right)^{-1} A^T b$$

If we use gradient descent to solve  $\textcircled{P_k}$   
then rate of convergence depends on

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

As  $c_k \rightarrow \infty$ ,

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

i.e. optimization problem  $\textcircled{P_k}$  becomes  
ill-conditioned as  $k \rightarrow \infty$ .

## 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  $x^*, \lambda^*$  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  $\lambda$  can be chosen close to  $\lambda^*$ , 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:

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

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

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

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

$$\text{For fixed } \lambda, \quad \lim_{c \rightarrow \infty} x_1(\lambda, c) = 1$$

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

---

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

How to make  $\lambda$  close to  $\lambda^*$  without knowing  $\lambda^*$ ?

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

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

and  
 Assume cont. diff.  $\rightarrow D(\lambda) = f(x(\lambda)) + \sum_{i=1}^m \lambda_i h(x(\lambda))$

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} f(x(\lambda)) &= \sum_{j=1}^n \frac{\partial}{\partial x_j} f(x(\lambda)) \cdot \frac{\partial x_j(\lambda)}{\partial \lambda_i} \\ &= \left[ \frac{\partial x_1(\lambda)}{\partial \lambda_i} \dots \frac{\partial x_n(\lambda)}{\partial \lambda_i} \right] \cdot \nabla_x f(x(\lambda)) \end{aligned}$$

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

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

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

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

$$\begin{aligned}
 \nabla_{\lambda} D(\lambda) &= \nabla_{\lambda} f(x(\lambda)) + \nabla_{\lambda} \left( \sum_{i=1}^m \lambda_i h_i(x(\lambda)) \right) \\
 &= \nabla_{\lambda} x(\lambda) \cdot \nabla_x f(x(\lambda)) + \sum_{i=1}^m \lambda_i \nabla_{\lambda} x(\lambda) \nabla_x h_i(x(\lambda)) \\
 &\quad + h(x(\lambda)) \\
 &= \nabla_{\lambda} x(\lambda) \left( \nabla_x f(x(\lambda)) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x(\lambda)) \right) \\
 &\quad + h(x(\lambda)) \quad = 0 \text{ by (1)}
 \end{aligned}$$

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

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

$$\begin{aligned}
 \lambda^{(k+1)} &= \lambda^{(k)} + \alpha_k \nabla_{\lambda} D(\lambda^{(k)}) \\
 &= \lambda^{(k)} + \alpha_k h(x(\lambda^{(k)}))
 \end{aligned}$$

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) under certain conditions.  
(don't need to take  $c_k \rightarrow \infty$ ). See sec 5.2 of text.

Example

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

$$\text{s.t. } x_1 = 1$$

$$x^* = (1, 0) \quad \text{and} \quad \lambda^* = -1.$$

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

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

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

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

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

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

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

Thus

$$\lambda^{(k)} \rightarrow \lambda^* \Rightarrow x^{(k)} \rightarrow (1, 0) = x^*$$