Lec19-mwf

Sunday, March 06, 2011 9:40 PM

HW5 is assigned.

Benefits of Duality:

- Reveal addition structure of the optimal solution (KKT condition)

- The dual variables often have interesting engineering interpretations ("price")

- The original problem may be decomposed into sub-problems when we convert to the dual

  - distributed solutions.

- The constraint set of the dual problem is often simpler

  - easy to compute projection

Price interpretation

Sunday, February 15, 2009 9:22 AM

KKT conditions are important because the resulting information on the optimal solution often has very clear engineering interpretations.

We will discuss a few examples

- The Lagrange multiplier can be interpreted as the "price" of the constraints
- For separable problems,
  When the price are set right, each "user" can minimize the subproblem (from the Lagrangian) independently.

Utility max

Sunday, February 15, 2009 12:32 PM

<u>Utility maximization</u> of a Single resource

$$\max_{x_i \geq 0} \sum_{i=1}^{n} U_i(x_i)$$

$$\text{sub to} \quad \sum_{i=1}^{n} x_i \leq R$$

$$x_i \geq 0$$

- $U_i(\cdot)$: utility function of user i
- R : amount of resource

- Introduce a Lagrange multiplier λ for the constraint

$$L(\vec{x}, \lambda) = -\sum_{i=1}^{n} U_i(x_i) + \lambda\left(\sum_{i=1}^{n} x_i - R\right)$$

$$= -\sum_{i=1}^{n} \left[U_i(x_i) - \lambda x_i\right] - \lambda R$$

- To minimize the Lagrangian over $\vec{x}$, each x_i should be chosen to

$$\max_{x_i \geq 0} U_i(x_i) - \lambda x_i \qquad (*)$$

— Let $x_i(\lambda)$ denote the maximizer of $(*)$

λ can be viewed as the price of the resource.

Each user choose "demand" $x_i(\lambda)$ that maximizes the net-utility.

(Q) What is the optimal price?

(A) Either $\lambda = 0$ or $\sum x_i(\lambda) = R$ (from complementary slackness)

If $\lambda = 0$, & $\sum x_i(0) < R$, it means that the overall demand for the resource is low.

If $\lambda > 0$, then $\sum x_i(\lambda) = R$, demand = supply.

Projection in the dual

Wednesday, March 04, 2009 11:37 AM

- In general, the projection operation can be difficult to carry out if the constraint set is in a complex form.
- For the dual problem, the constraint set is always a quadrant

  $\Rightarrow$ projection becomes trivial.
- Also, the gradient (or subgradient) has a simple form.

Primal Problem:

$$\begin{array}{lll} \min & f_0(x) & \\ \text{sub to} & f_i(x) \le 0 & \lambda \\ & h_i(x) = 0 & \nu \end{array}$$

Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \Sigma \lambda_i f_i(x) + \Sigma \nu_i h_i(x)$$

$$g(\lambda, \nu) = \min_x L(x, \lambda, \nu)$$

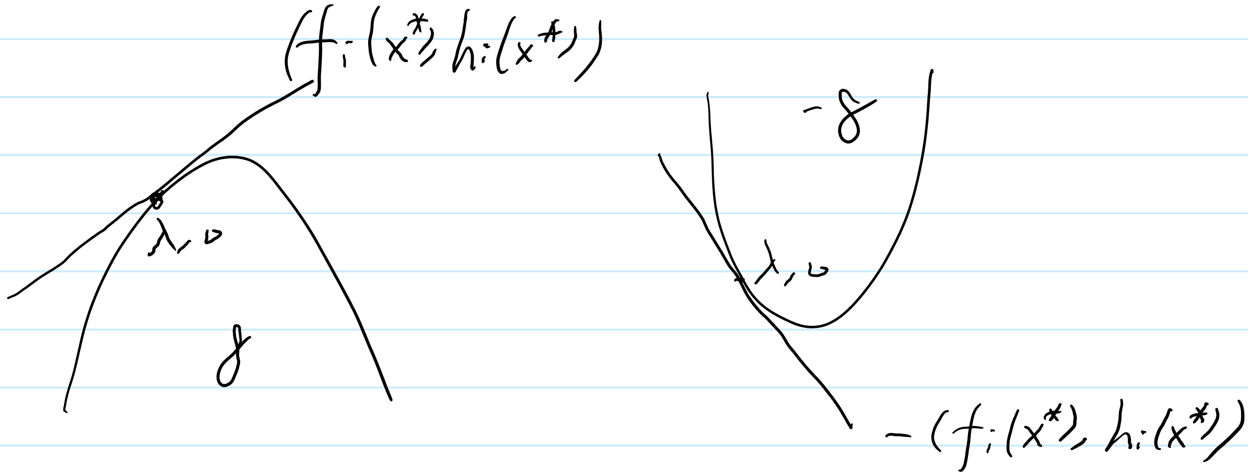
- $g(\lambda, \nu)$ is concave, so let us look at the subgradient of $-g(\lambda, \nu)$

(Q) What is the gradient / subgradient of $-g(\lambda, \nu)$?

(Q) What is the gradient / subgradient of $-g(\lambda, \nu)$?

(A) Let x^* be the minimizer of the Lagrangian at λ, ν. Then $-(f_i(x^*), h_i(x^*))$ is a subgradient of $-g$ at λ, ν.

- Or, equivalently, $(f_i(x^*), h_i(x^*))$ is a super-gradient of g at λ, ν.



Proof: We need to show that for any λ', ν',

$$-g(\lambda', \nu') \geq -g(\lambda, \nu) + \sum (-f_i(x^*))(\lambda_i' - \lambda_i) + \sum (-h_i(x^*))(\nu_i' - \nu_i)$$

$$\Leftrightarrow \quad g(\lambda', \nu') \leq g(\lambda, \nu) + \sum f_i(x^*)(\lambda_i' - \lambda_i) + \sum h_i(x^*)(\nu_i' - \nu_i)$$

Note that by the definition of $g(\lambda, \nu)$ & x^*,

$$g(\lambda, \nu) = f_0(x^*) + \sum \lambda_i f_i(x^*) + \sum \nu_i h_i(x^*)$$

$$g(\lambda, \nu) = f_0(x^*) + \Sigma \lambda_i f_i(x^*) + \Sigma \nu_i h_i(x^*)$$

$$RHS = f_0(x^*) + \Sigma \lambda_i' f_i(x^*) + \Sigma \nu_i' h_i(x^*).$$

Hence the result follows. #

As a result, the gradient projection algorithm for the dual is of the following simple form

$$\lambda_i(t+1) = [\lambda_i(t) + \gamma f_i(x^*(t))]^+$$

$$\nu_i(t+1) = \nu_i(t) + \gamma h_i(x^*(t))$$

— gradient "accent" rather than "descent".

Utility maximization of a single resource

Sunday, March 08, 2009 3:56 PM

$$\max_{x_i \geq 0} \sum_{i=1}^{n} U_i(x_i)$$

$$\text{sub to} \quad \sum x_i \leq R$$

- U_i: utility function, concave

  R: amount of resource

- Lagrangian

$$L(\vec{x}, \lambda) = -\sum_{i=1}^{n} U_i(x_i) + \lambda\left(\sum x_i - R\right)$$

$$= -\sum_{i=1}^{n} \left[U_i(x_i) - \lambda x_i\right] - \lambda R$$

- To minimize the Lagrangian

$$\frac{\partial L}{\partial x_i} = U_i'(x_i) - \lambda = 0$$

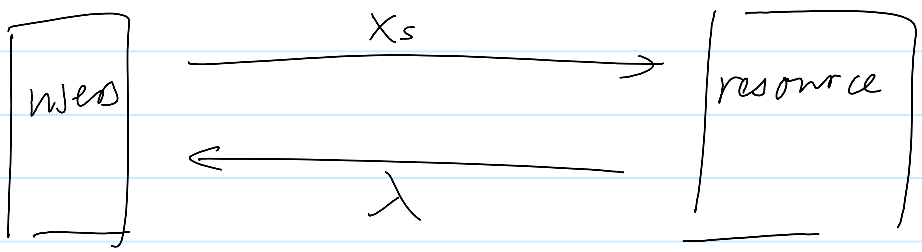
In other words, each user chooses x_i^* that maximize the net utility.

- Gradient descent on the dual

$$\lambda(t+1) = \left[\lambda(t) + \gamma \left(\sum x_i^*(t) - R \right) \right]^+$$

Interpretation: If the demand exceeds the supply, increase price λ. Otherwise, decrease price.

- Note the distributed structure of the solution



Water filling again

Wednesday, March 04, 2009 11:47 AM

① Water-filling for fading channels

$$\max_{p_1,\cdots p_m \geq 0} \sum_{k=1}^{m} q_k \log\left(1+\frac{g_k p_k}{N}\right)$$

$$\text{Sub to} \quad \sum_{k=1}^{M} q_k p_k \leq P_0$$

- g_k : channel gain at state k
  q_k : $\Pr\{g = g_k\}$
  p_k : power allocation at state k. (to be determined).

- Lagrangian

$$L(\vec{p}, \lambda) = -\sum_{k=1}^{m} q_k \log\left(1+\frac{g_k p_k}{N}\right) + \lambda\left(\sum_{k=1}^{M} q_k p_k - P_0\right)$$

$$= -\sum_{k=1}^{m} q_k \left[\log\left(1+\frac{g_k p_k}{N}\right) - \lambda p_k\right] - \lambda P_0$$

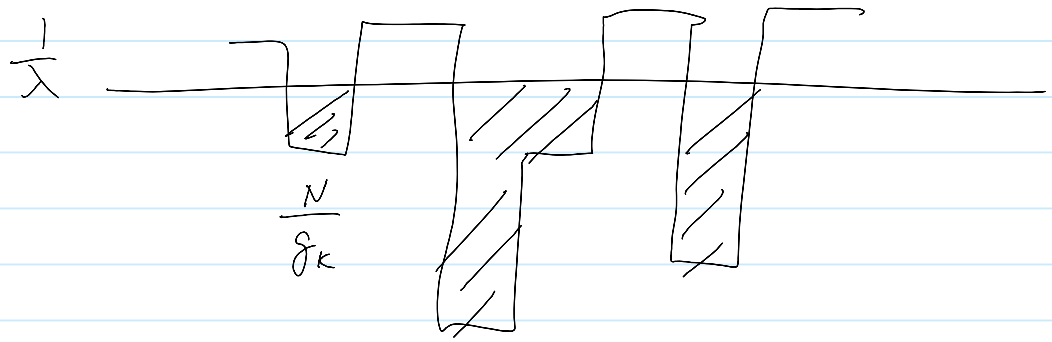
- To minimize $L(\vec{p}, \lambda)$ over $\vec{p}$, either $\frac{\partial L}{\partial p_k} = 0$ or $p_k = 0$ & $\frac{\partial L}{\partial p_k} > 0$

$$\frac{\partial L}{\partial p_k} = -q_k \frac{\frac{g_k}{N}}{1+\frac{g_k p_k}{N}} + \lambda q_k = 0$$

$$\Rightarrow \quad \frac{g_k}{N} = \lambda\left(1+\frac{g_k p_k}{N}\right)$$

$$\Rightarrow \quad p_k^* = \frac{1}{\lambda} - \frac{N}{g_k} \quad \text{if } \frac{1}{\lambda} - \frac{N}{g_k} \geq 0$$

or $p_k^* = 0$



(Q) What is the optimal λ?

(A) From complementary slackness.

either $\lambda = 0$ or $\sum_K g_K P_K = P_0$

It is trivial to exclude $\lambda = 0$ (since $P_K = +\infty$)

- The gradient-descent algorithm for the dual

$$\lambda(t+1) = \left[\lambda(t) + \gamma \left\{ \sum g_K P_K^*(t) - P_0 \right\} \right]^+$$

- Interpretation: If power is too high, drop the 'water-level'.

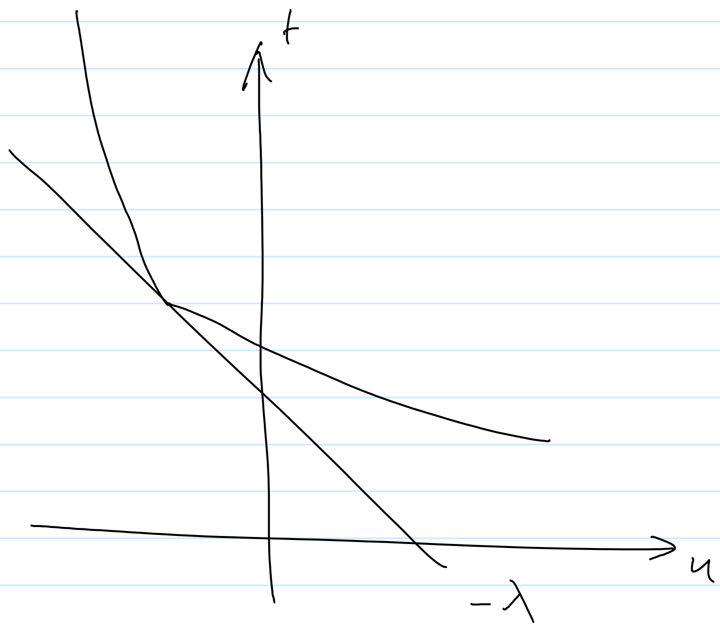
- Note the distributed structure of the solution.

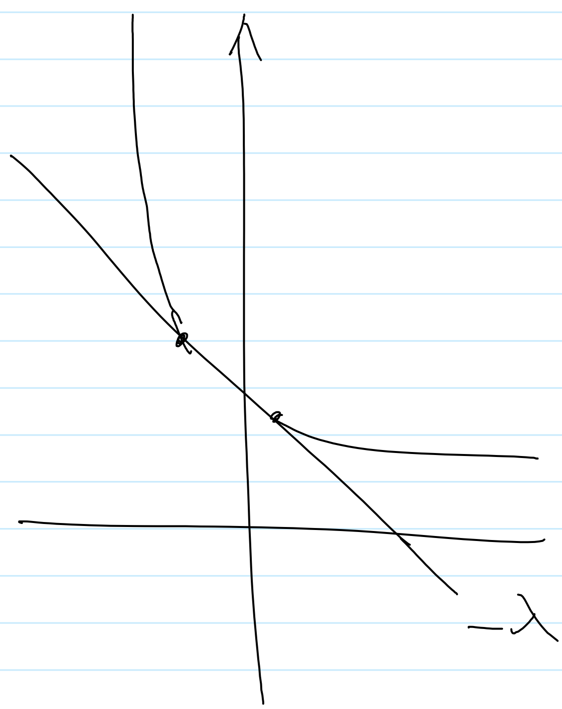
Smoothness condition of the dual

Sunday, March 08, 2009 4:10 PM

What is the condition on the primal problem that corresponds to the smoothness condition on the dual?

- If the primal problem is strongly convex, then the dual satisfies a smoothness condition.
- Vice versa.

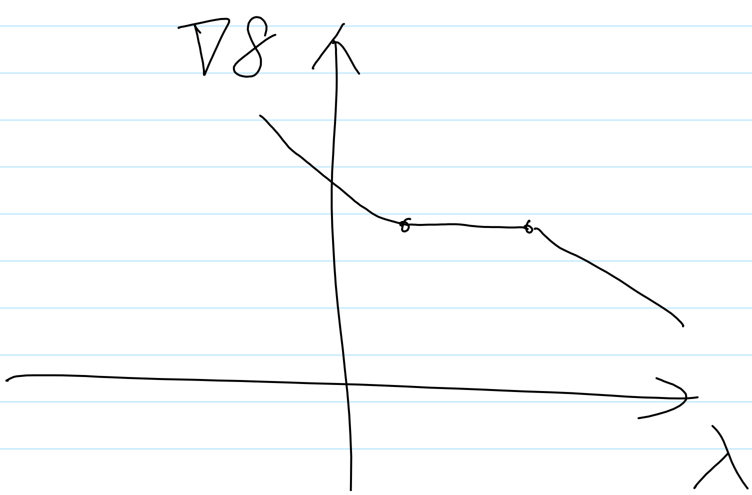


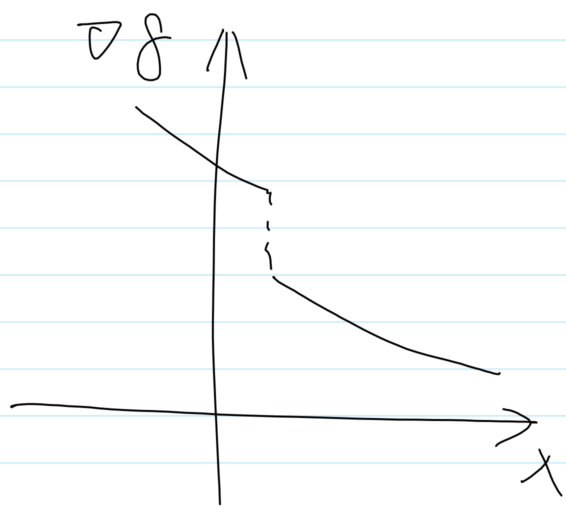


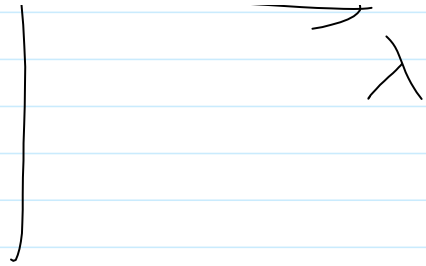
primal problem has discontinuous ∇f
(not-smooth)

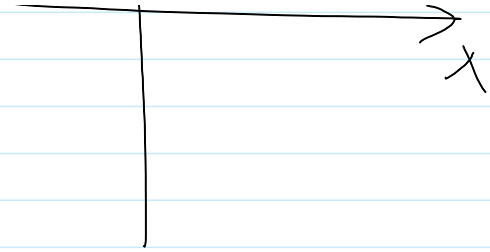
primal problem has linear components.
(not-strongly convex)

$\Downarrow$ $\Downarrow$









Dual problem has non-strongly convex ∇f

Dual problem has non-smooth ∇f