Lec10-new

Saturday, January 24, 2009 5:25 PM

Remind the project proposal Oct. 9.

https://www.youtube.com/watch?v=9XL2oRoAii8

Necessary and sufficient conditions for optimality:

$$f_0'(\bar{x}; x-\bar{x}) \geq 0 \quad \forall x \in C$$

$$\Updownarrow$$

$$[\nabla f_0(\bar{x})]^T (x-\bar{x}) \geq 0 \quad \forall x \in C$$

$$\Leftrightarrow \begin{cases} \nabla f_0(\bar{x}) = 0 & \text{if } \bar{x} \text{ is in the interior of } C \\ -\nabla f_0(\bar{x}) \in N_C(\bar{x}). & \text{if } \bar{x} \text{ is at the boundary} \end{cases}$$

- If the constraint is defined by $f_i(x) \leq 0$,
  - At a point $\bar{x}$ such that $f_i(\bar{x}) = 0$ we have

$$(\nabla f_i(\bar{x}))^T (x-\bar{x}) \leq f(x) - f(\bar{x}) \leq 0$$

for all $x \in C$

$\Rightarrow$ $\nabla f_i(\bar{x})$ is the normal vector!

- Then $-\nabla f_0(\bar{x})$ must be a conic combination of these $\nabla f_i(\bar{x})$.

$$-\nabla f_0(\bar{x}) = \underset{0}{\underset{\vee\!\shortmid}{\lambda_1}} \cdot \nabla f_1(\bar{x}) + \underset{0}{\underset{\vee\!\shortmid}{\lambda_2}} \cdot \nabla f_2(\bar{x}) + \cdots$$

Only binding constraints matter!

Linear regression - unconstrained

Monday, January 19, 2009 5:27 PM

(1) Linear Regression / Signal estimation

$$\min_{w} \sum_{i=1}^{m} (y_i - x_i^T w)^2 = \min_{w} \|y - Xw\|^2$$

- In linear regression

w: parameter to be estimated
y: dep. variable

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

X: indep. variable,

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix}$$

- We want $y = X^T \cdot w$

The value $y_i - x_i^T w$ is called the residual (error), e.g., due to noise in the observations.

- Also useful for estimating a signal w
  - y: measurement
  - X: design matrix

Solution: (L_2-norm)

$$\min_{w} f(w) = \| Y - Xw \|_2^2$$

- $f(w) = (Y - Xw)^T (Y - Xw) = w^T X^T X w - 2 y^T X w + y^T y$

$$\nabla f(w) = 2X^T X w - 2X^T y = 0$$

$$\Rightarrow \quad \bar{w} = (X^T X)^{-1} X^T y.$$

- Boyd p293
- Other distributions, see Boyd p352

55

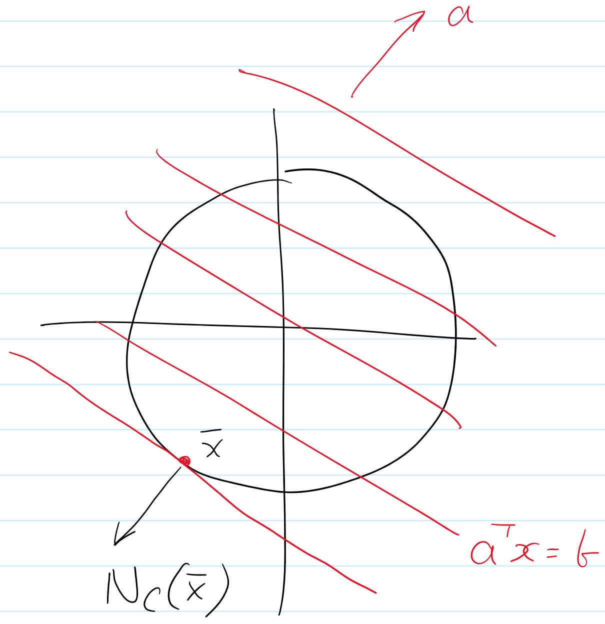
Constrained problems

Monday, January 26, 2009 10:47 PM

Single convex constraint

(1) $\min \; a^T x$

sub to $x^T x \le 1$



- Since the objective is linear, the optimum point must lie at the boundary!

- Optimality condition is

  - $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$

  - $\nabla f_0(\bar{x}) = a$

  - To get the normal vector, we use

$$\nabla f_1(\bar{x}) = \bar{x}$$

Hence, $-a$ must be in the direction of $\bar{x}$. Equivalently, there exists $\lambda \ge 0$ such that

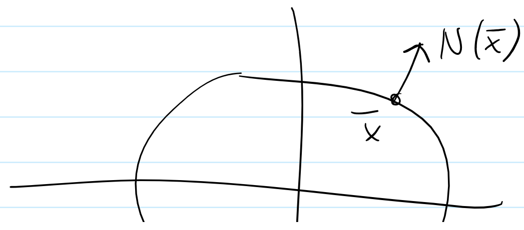
$$-a = \lambda \bar{x}$$

$$\& \quad \bar{x}^T \bar{x} = 1$$

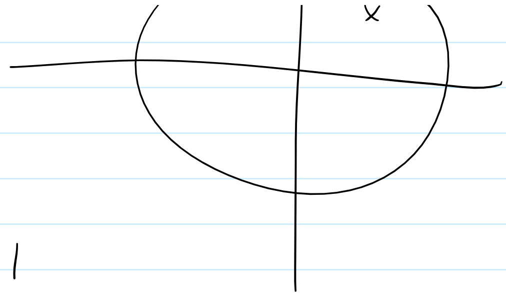
(2) What if we do not know whether $\bar{x}$ is at the boundary.

$\min \; f_0(x)$

sub to $x^T x \le 1$



subj to $\quad x^T x \leq 1$



- either at the interior

$$\nabla f_0(\bar{x}) = 0 \quad \& \quad \bar{x}^T \bar{x} < 1$$

- or at the boundary $\quad \bar{x}^T \bar{x} = 1$

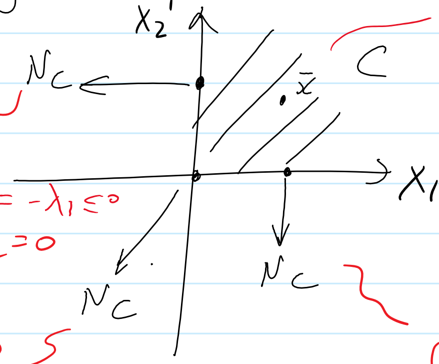
$$-\nabla f_0(\bar{x}) = \lambda \bar{x}, \quad \lambda \geq 0$$

$$\& \quad \bar{x}^T \bar{x} = 1.$$

③ Minimization over the non-negative quadrants

$$\min f_0(x)$$
$$\text{sub to} \quad x \geq 0$$

$\nabla f_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$



$X_1 = 0$
$X_2 > 0$
$-[\nabla f(\bar{x})]_1 = -\lambda_1 \leq 0$
$-[\nabla f(\bar{x})]_2 = 0$

$[\nabla f(\bar{x})]_1 = 0$
$[\nabla f(\bar{x})]_2 = 0$
$x_1 > 0, \; x_2 > 0$

$X_1 = 0, X_2 = 0$
$-[\nabla f(\bar{x})]_1 = -\lambda_1 \leq 0$
$-[\nabla f(\bar{x})]_2 = -\lambda_2 \leq 0$

$X_1 > 0, \; X_2 = 0$
$-[\nabla f(\bar{x})]_1 = 0$
$-[\nabla f(\bar{x})]_2 = -\lambda_2 \leq 0$

$\nabla f_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

- Four optimality conditions!
- They can be collected into 1 equation

$$\nabla f(\bar{x}) + \lambda_1 \nabla f_1(\bar{x}) + \lambda_2 \nabla f_2(\bar{x}) = 0$$

Lagrange multiplier $\rightarrow$

$$\left.\begin{array}{ll} \lambda_i = 0 & \text{if } f_i(\bar{x}) < 0 \\ \lambda_i \geq 0 & \text{if } f_i(\bar{x}) = 0 \end{array}\right\} \Rightarrow \lambda_i f_i(\bar{x}) = 0$$

known as "complementary slackness"

- Later in duality, we will see this as the KKT condition.

Linear equality constraints

Monday, January 26, 2009 10:47 PM

(4) Equality - Constrained Convex Problems

- $C = \{x \mid Ax = b\}$ $\qquad a^T x = b$

- Easiest if there is only one equation

  $ax = b. \qquad a \in R^n, \quad b \in R$

  - C is a hyperplane

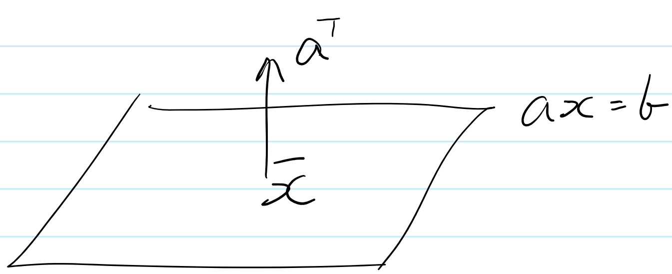
- Can think of this as two constraints

  $ax - b \le 0 \Rightarrow \nabla f_1(x) = a^T$

  $-ax - b \le 0 \Rightarrow \nabla f_2(x) = -a^T$

- The optimum $\bar{x}$ must be at the boundary!

  $\Rightarrow \quad -\nabla f_0(\bar{x}) = \underset{0}{\underset{\vee\!\!\mid}{\lambda_1}} a^T + \underset{0}{\underset{\vee\!\!\mid}{\lambda_2}}(-a^T) = \lambda \cdot a^T$



$\Rightarrow -\nabla f_0(\bar{x}) = \lambda a^T \quad \lambda \in R.$

- If C is defined by many equations

  $Ax = b \iff \begin{bmatrix} a_1 \\ \vdots \end{bmatrix} x = \begin{bmatrix} b_1 \\ \vdots \end{bmatrix}$

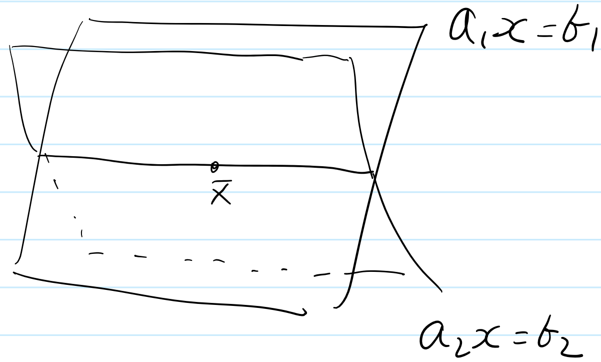
$$Ax = b \iff \begin{bmatrix} \vdots \\ a_m \end{bmatrix} x = \begin{bmatrix} \vdots \\ b_m \end{bmatrix}$$

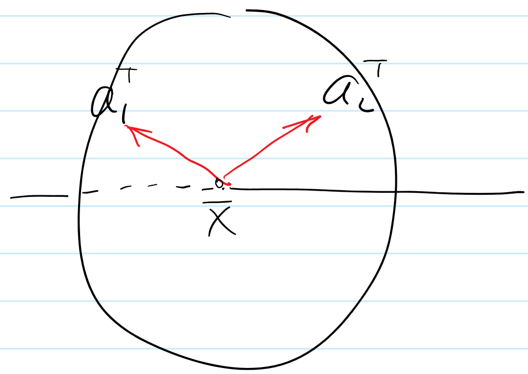
- The normal cone is spanned by

$$\pm a_1^T, \ \pm a_2^T, \ \cdots, \ \pm a_m^T$$

$$\Rightarrow \quad -\nabla f_0(\bar{x}) = \lambda_1 a_1^T + \lambda_2 a_2^T + \cdots + \lambda_m a_m^T$$
$$= A^T \vec{\lambda}$$

Ex)





- spans all linear combinations of a_1^T & a_2^T.

- We thus have

$$\begin{cases} \nabla f(\bar{x}) + A^T \lambda = 0 \\ A\bar{x} = b. \end{cases}$$

- More on this with Lagrange duality.

(30)

Sharing a single resource

Saturday, January 31, 2009 5:07 PM

Single Resource

- R: the amount of resource
  - bandwidth
  - stocks, etc

- N: number of users
- x_s: the amount of resource allocated to user s, $s=1,2,\dots,N$
  - Assume X_s is real number
  - The resource is infinitely divisible.

Clearly, we must have

$$\sum_{s=1}^{N} x_s \le R.$$

Q How to allocate the resource among multiple users?

- Equally?
- Who pay highest amount of money?

---

a Utility Maximization (Welfare Maximization)

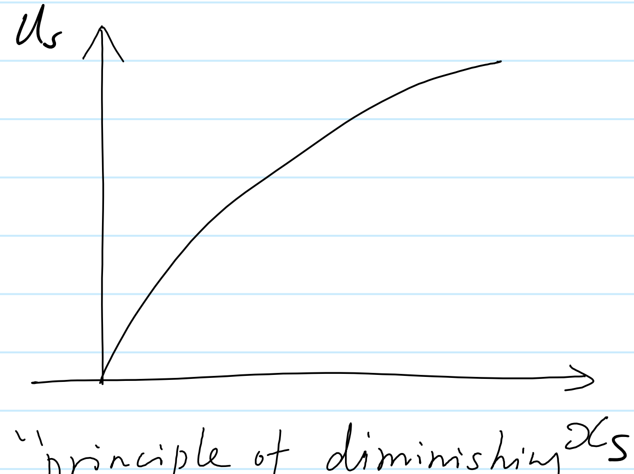
- $U_s(x_s)$: the utility to user s if he has x_s amount of resource

- Maximize total utility

$$\max \sum_{s=1}^{N} U(x_s)$$

$$\text{sub to} \quad \sum_{s=1}^{N} x_s \le R \quad , \quad x_s \ge 0$$

- A Convex problem if U_s's are concave



"principle of diminishing returns"

- A strictly-concave utility func, such as $\log x_s$, also is said to promote fairness
  - x_s will not be zero for any users.

Fairness - skip

Monday, January 30, 2023 4:02 PM

- The concavity of the utility func. can be used to model fairness

Examples

(a) $U_s(x_s) = w_s \log x_s$

$\uparrow$ constant weight

- Proportional Fair:
  - Let x_s^* be the optional solution.

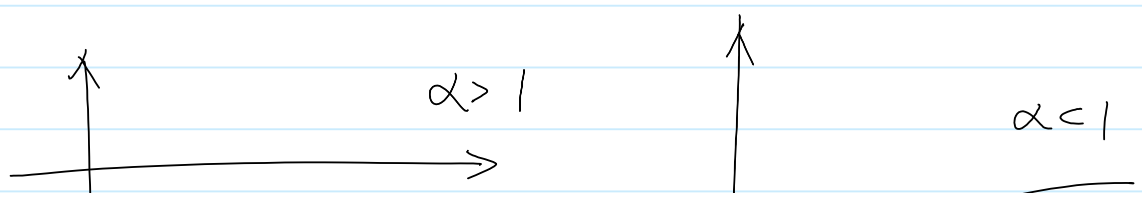
$$U_s'(x_s^*) = \frac{w_s}{x_s^*}$$

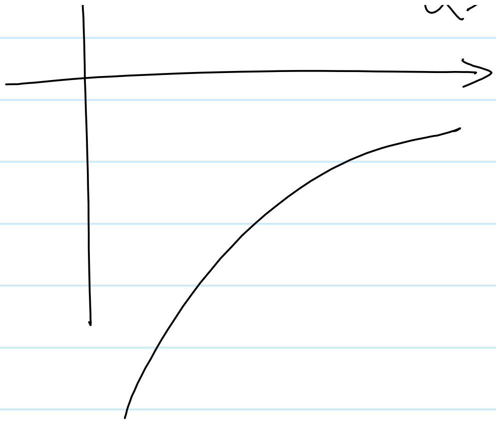
  - For any other feasible allocation x_s, we must have

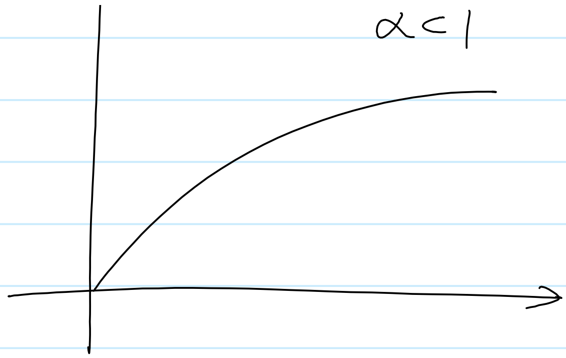
$$f'(x^*; x - x^*) = \sum_s \frac{w_s}{x_s^*}(x - x_s^*) \le 0$$

  - The sum of the relative rate-difference cannot be improved.

(b) $U_s(x_s) = w_s \frac{x_s^{1-\alpha}}{1-\alpha}, \alpha > 0$ α-fairness







- As $\alpha \to 1$, $\quad U_s(x_s) \to w_s \log x_s$

- For any optimal solution x_s^*

$$U_s'(x_s^*) = \frac{w_s}{(x_s^*)^\alpha}$$

$$f'(x^*; x - x^*) = \sum_s w_s \cdot \frac{x_s - x_s^*}{(x_s^*)^\alpha} \le 0$$

α- fairness

- What happens when $\alpha \to 0$?

  - $\max \sum_s w_s x_s$ weighted sum-rate

- What happens when $\alpha \to +\infty$?

  - Suppose $\alpha = 1000$

$$\max \sum_s - w_s \frac{1}{999\, x_s^{999}}$$

$$\Leftrightarrow \min \sum_s w_s \frac{1}{999\, x_s^{999}}$$

- Smallest x_s dominates!

$\Leftrightarrow$ $\max \min_s x_s$

- max-min fairness

- For a single resource, all of these allocations coincide when w_s is the same for all users.

  - That will change with multiple resources

Solution: Sharing a single resource

Monday, February 02, 2009 10:33 PM

- When $U_s(\cdot)$ is increasing

  $\Rightarrow \sum x_s = R$

  - optimum is at the boundary.

- Assume further that the optimal x^* satisfies $x_s^* > 0$ for all s.

  - True if $U_s(\cdot) = \log x_s$.

- Then the normal vector is simply $[1, \cdots, 1]$

We must then have

$$-\left[-U_s'(x_s)\right] = \lambda \cdot 1 \quad \text{for all } s$$

- This can be thought of as each user maximizing

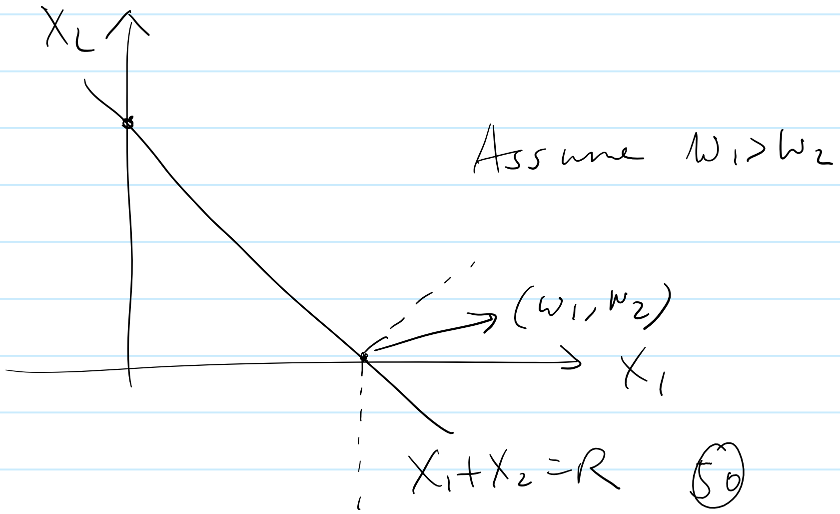
$$\max_{x_s} U_s(x_s) - \underset{\substack{\uparrow \\ \text{price}}}{\lambda} x_s \qquad \text{net-utility.}$$

- True even if $x_s^* = 0$ for some s (will be clear when we study duality).

- What if $U_s(x_s) = w_s x_s$?

- The optimal solution will always assign

R to the user with the largest w_s



(50)

- Following the gradient direction (of increasing $\sum_s U_s(x_s)$), the value of x will naturally go to $X_1 = R$, $X_2 = 0$

(More on this when we discuss gradient algorithms.)

- In general, the solution of LP will lie at the extreme points of the constraint.

- Not Fair!!!