15.093 Optimization Methods

Lecture 20: The Conjugate Gradient Algorithm
Optimality conditions for constrained optimization

1 Outline

Slide 1

1. The Conjugate Gradient Algorithm
2. Necessary Optimality Conditions
3. Sufficient Optimality Conditions
4. Convex Optimization
5. Applications

2 The Conjugate Gradient Algorithm

2.1 Quadratic functions

Slide 2

$$\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x'Qx} + \boldsymbol{c'x}$$

Definition: $\boldsymbol{d_1}, \ldots, \boldsymbol{d_n}$ are $\boldsymbol{Q}$-conjugate if

$$\boldsymbol{d_i} \neq \boldsymbol{0}, \qquad \boldsymbol{d_i'Qd_j} = 0, \quad i \neq j$$

Proposition: If $\boldsymbol{d_1}, \ldots, \boldsymbol{d_n}$ are $\boldsymbol{Q}$-conjugate, then $\boldsymbol{d_1}, \ldots, \boldsymbol{d_n}$ are linearly independent.

2.2 Motivation

Slide 3

Given $\boldsymbol{Q}$-conjugate $\boldsymbol{d_1}, \ldots, \boldsymbol{d_n}$, and $\boldsymbol{x}^k$, compute

$$\min_{\alpha} f(\boldsymbol{x}^k + \alpha \boldsymbol{d}_k) = \boldsymbol{c'x}^k + \alpha \boldsymbol{c'd_k} +$$

$$\frac{1}{2}(\boldsymbol{x}^k + \alpha \boldsymbol{d}_k)'\boldsymbol{Q}(\boldsymbol{x}^k + \alpha \boldsymbol{d}_k) =$$

$$f(\boldsymbol{x}^k) + \alpha \nabla f(\boldsymbol{x}^k)'\boldsymbol{d}_k + \frac{1}{2}\alpha^2 \boldsymbol{d_k'Qd_k}$$

Solution:

$$\hat{\alpha}_k = \frac{-\nabla f(\boldsymbol{x}^k)'\boldsymbol{d}_k}{\boldsymbol{d_k'Qd_k}}, \qquad \boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \hat{\alpha}_k \boldsymbol{d}_k$$

2.3 Expanding Subspace Theorem

Slide 4

$\boldsymbol{d_1}, \ldots, \boldsymbol{d_n}$ are $\boldsymbol{Q}$-conjugate. Then, $\boldsymbol{x}^{k+1}$ solves

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} = \boldsymbol{x}_1 + \sum_{j=1}^{k} \alpha_j \boldsymbol{d}_j \end{aligned}$$

Moreover, $\boldsymbol{x}^{n+1} = \boldsymbol{x}^*$.

2.4 The Algorithm

SLIDE 5

Step 0 Given $\boldsymbol{x}^1$, set $k := 1$, $\boldsymbol{d}_1 = -\nabla f(\boldsymbol{x}^0)$

Step 1 For $k = 1, \ldots, n$ do:
If $||\nabla f(\boldsymbol{x}^k)|| \leq \epsilon$, stop; else:

$$\hat{\alpha}_k = \text{argmin}_\alpha f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k) = \frac{-\nabla f(\boldsymbol{x}^k)' \boldsymbol{d}_k}{\boldsymbol{d}_{\boldsymbol{k}}' \boldsymbol{Q} \boldsymbol{d}_{\boldsymbol{k}}}$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \hat{\alpha}_k \boldsymbol{d}_k$$

$$\boldsymbol{d}_{k+1} = -\nabla f(\boldsymbol{x}^{k+1}) + \lambda_k \boldsymbol{d}_k, \qquad \lambda_k = \frac{-\nabla f(\boldsymbol{x}^{k+1})' \boldsymbol{Q} \boldsymbol{d}_k}{\boldsymbol{d}_{\boldsymbol{k}}' \boldsymbol{Q} \boldsymbol{d}_{\boldsymbol{k}}}$$

2.5 Correctness

SLIDE 6

Theorem: The directions $\boldsymbol{d}_{\boldsymbol{1}}, \ldots, \boldsymbol{d}_{\boldsymbol{n}}$ are $\boldsymbol{Q}$-conjugate.

2.6 Convergence Properties

SLIDE 7

- This is a finite algorithm.
- If there are only k distinct eigenvalues of $\boldsymbol{Q}$, the CGA finds an optimal solution in k steps.
- Idea of pre-conditioning. Consider

$$\min f(\boldsymbol{S}\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{S}\boldsymbol{x})' \boldsymbol{Q} (\boldsymbol{S}\boldsymbol{x}) + \boldsymbol{c}' \boldsymbol{S}\boldsymbol{x}$$

  so that the number of distinct eigenvalues of $\boldsymbol{S}'\boldsymbol{Q}\boldsymbol{S}$ is small

2.7 Example

SLIDE 8

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}'\boldsymbol{x}$$

$$\boldsymbol{Q} = \begin{pmatrix} 35 & 19 & 22 & 28 & 16 & 3 & 16 & 6 & 4 & 4 \\ 19 & 43 & 33 & 19 & 5 & 2 & 5 & 4 & 0 & 0 \\ 22 & 33 & 40 & 29 & 12 & 7 & 6 & 2 & 2 & 4 \\ 28 & 19 & 29 & 39 & 16 & 7 & 14 & 6 & 2 & 4 \\ 16 & 5 & 12 & 16 & 12 & 4 & 8 & 2 & 4 & 8 \\ 3 & 2 & 7 & 7 & 4 & 5 & 1 & 0 & 1 & 4 \\ 16 & 5 & 6 & 14 & 8 & 1 & 12 & 2 & 2 & 4 \\ 6 & 4 & 2 & 6 & 2 & 0 & 2 & 4 & 0 & 0 \\ 4 & 0 & 2 & 2 & 4 & 1 & 2 & 0 & 2 & 4 \\ 4 & 0 & 4 & 4 & 8 & 4 & 4 & 0 & 4 & 16 \end{pmatrix} \quad \text{and} \quad \boldsymbol{c} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ -3 \\ 0 \\ -2 \\ 0 \\ -6 \\ -7 \\ -4 \end{pmatrix}$$

$\kappa(\boldsymbol{Q}) \approx 17,641$ $\qquad \delta = \left(\frac{\kappa(\boldsymbol{Q})-1}{\kappa(\boldsymbol{Q})+1}\right)^2 = 0.999774$ SLIDE 9

k	$f(\boldsymbol{x}^k)$	$f(\boldsymbol{x}^k) - f(\boldsymbol{x}^*)$	$\frac{f(\boldsymbol{x}^k) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}^{k-1}) - f(\boldsymbol{x}^*)}$
1	12.000000	2593.726852	1.000000
2	8.758578	2590.485430	0.998750
3	1.869218	2583.596069	0.997341
4	−12.777374	2568.949478	0.994331
5	−30.479483	2551.247369	0.993109
6	−187.804367	2393.922485	0.938334
7	−309.836907	2271.889945	0.949024
8	−408.590428	2173.136424	0.956532
9	−754.887518	1826.839334	0.840646
10	−2567.158421	14.568431	0.007975
11	−2581.711672	0.015180	0.001042
12	−2581.726852	−0.000000	−0.000000

2.8 General problems

SLIDE 10

Step 0 Given $\boldsymbol{x}^1$, set $k := 1$, $\boldsymbol{d}_1 = -\nabla f(\boldsymbol{x}^0)$

Step 1 If $||\nabla f(\boldsymbol{x}^k)|| \leq \epsilon$, stop; else:

$$\hat{\alpha}_k = \text{argmin}_\alpha f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k) = \frac{-\nabla f(\boldsymbol{x}^k)' \boldsymbol{d}_k}{\boldsymbol{d_k' Q d_k}}$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \hat{\alpha}_k \boldsymbol{d}_k$$

$$\boldsymbol{d}_{k+1} = -\nabla f(\boldsymbol{x}^{k+1}) + \lambda_k \boldsymbol{d}_k$$

$$\lambda_k = \frac{||\nabla f(\boldsymbol{x}^{k+1})||}{||\nabla f(\boldsymbol{x}^k)||}$$

Step 2 $k \leftarrow k + 1$, goto Step 1

3 Necessary Optimality Conditions

3.1 Nonlinear Optimization

SLIDE 11

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_j(\boldsymbol{x}) \leq 0, \qquad j = 1, \ldots, p \\ & h_i(\boldsymbol{x}) = 0, \qquad i = 1, \ldots, m \end{array}$$

$$P = \{\boldsymbol{x} |\ g_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, p,$$
$$h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m\}$$

3.2 The KKT conditions

SLIDE 12

Discovered by Karush-Kuhn-Tucker in 1950's.

<u>Theorem</u>

If

- $\overline{\boldsymbol{x}}$ is local minimum of P
- $I = \{j|\ g_j(\overline{\boldsymbol{x}}) = 0\}$, set of tight constraints
- Constraint qualification condition (CQC): The vectors $\nabla g_j(\overline{\boldsymbol{x}})$, $j \in I$ and $\nabla h_i(\overline{\boldsymbol{x}})$, $i = 1, \ldots, m$, are linearly independent

SLIDE 13

Then, there exist vectors $(\boldsymbol{u}, \boldsymbol{v})$:

1. $\nabla f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i \nabla h_i(\overline{\boldsymbol{x}}) = \boldsymbol{0}$
2. $\boldsymbol{u} \geq \boldsymbol{0}$
3. $u_j g_j(\overline{\boldsymbol{x}}) = 0, \qquad j = 1, \ldots, p$

3.3 Some Intuition from LO

SLIDE 14

Linearize the functions in the neighborhood of the solution $\overline{\boldsymbol{x}}$. Problem becomes:

$$\begin{aligned} \min \quad & f(\overline{\boldsymbol{x}}) + \nabla f(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) \\ \text{s.t.} \quad & g_j(\overline{\boldsymbol{x}}) + \nabla g_j(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) \leq \boldsymbol{0}, \quad j \in I \\ & h_i(\overline{\boldsymbol{x}}) + \nabla h_i(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) = \boldsymbol{0}, \quad i = 1, \ldots, m \end{aligned}$$

SLIDE 15

This is a LO problem. Dual feasibility:

$$\sum_{j \in I} \hat{u}_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} \hat{v}_i \nabla h_i(\overline{\boldsymbol{x}}) = \nabla f(\overline{\boldsymbol{x}}), \quad \hat{u}_j \leq 0$$

Change to $u_j = -\hat{u}_j, \quad v_i = -\hat{v}_i$ to obtain:

$$\nabla f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i \nabla h_i(\overline{\boldsymbol{x}}) = \boldsymbol{0}, \quad u_j \geq 0$$

3.4 Example 1

SLIDE 16

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = (x_1 - 12)^2 + (x_2 + 6)^2 \\ \text{s.t.} \quad & h_1(\boldsymbol{x}) = 8x_1 + 4x_2 - 20 = 0 \\ & g_1(\boldsymbol{x}) = x_1^2 + x_2^2 + 3x_1 - 4.5x_2 - 6.5 \leq 0 \\ & g_2(\boldsymbol{x}) = (x_1 - 9)^2 + x_2^2 - 64 \leq 0 \end{aligned}$$

$$\overline{\boldsymbol{x}} = (2,1)'; \quad g_1(\overline{\boldsymbol{x}}) = 0, \quad g_2(\overline{\boldsymbol{x}}) = -14, \quad h_1(\overline{\boldsymbol{x}}) = 0.$$

SLIDE 17

- $I = \{1\}$
- $\nabla f(\overline{\boldsymbol{x}}) = (-20, 14)'$; $\nabla g_1(\overline{\boldsymbol{x}}) = (7, -2.5)'$
- $\nabla g_2(\overline{\boldsymbol{x}}) = (-14, 2)'$; $\nabla h_1(\overline{\boldsymbol{x}}) = (8, 4)'$
- $u_1 = 4$, $u_2 = 0$, $v_1 = -1$
- $\nabla g_1(\overline{\boldsymbol{x}})$, $\nabla h_1(\overline{\boldsymbol{x}})$ linearly independent
- $\nabla f(\overline{\boldsymbol{x}}) + u_1 \nabla g_1(\overline{\boldsymbol{x}}) + u_2 \nabla g_2(\overline{\boldsymbol{x}}) + v_1 \nabla h_1(\overline{\boldsymbol{x}}) = \boldsymbol{0}$

$$\begin{pmatrix} -20 \\ 14 \end{pmatrix} + 4 \begin{pmatrix} 7 \\ -2.5 \end{pmatrix} + 0 \begin{pmatrix} -14 \\ 2 \end{pmatrix} + (-1) \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

3.5 Example 2

SLIDE 18

$$\begin{array}{rl} \max & \boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{x}'\boldsymbol{x} \le 1 \end{array}$$

$\boldsymbol{Q}$ arbitrary; Not a convex optimization problem.

$$\begin{array}{rl} \min & -\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{x}'\boldsymbol{x} \le 1 \end{array}$$

3.5.1 KKT

SLIDE 19

$$\begin{array}{rcl} -2\boldsymbol{Q}\boldsymbol{x} + 2u\boldsymbol{x} & = & \boldsymbol{0} \\ \boldsymbol{x}'\boldsymbol{x} & \le & 1 \\ u & \ge & 0 \\ u(1 - \boldsymbol{x}'\boldsymbol{x}) & = & 0 \end{array}$$

SLIDE 20

3.5.2 Solutions of KKT

- $\overline{\boldsymbol{x}} = \boldsymbol{0}$, $\overline{u} = 0$, Obj $= 0$.
- $\overline{\boldsymbol{x}} \neq \boldsymbol{0} \Rightarrow \boldsymbol{Q}\overline{\boldsymbol{x}} = \overline{u}\ \overline{\boldsymbol{x}} \Rightarrow \overline{\boldsymbol{x}}$ eigenvector of $\boldsymbol{Q}$ with non-negative eigenvalue $\overline{u}$.
- $\overline{\boldsymbol{x}}'\boldsymbol{Q}\overline{\boldsymbol{x}} = \overline{u}\ \overline{\boldsymbol{x}}'\overline{\boldsymbol{x}} = \overline{u}$.
- Thus, pick the largest nonnegative eigenvalue $\hat{u}$ of $\boldsymbol{Q}$. The solution is the corresponding eigenvector $\hat{\boldsymbol{x}}$ normalized such that $\hat{\boldsymbol{x}}'\hat{\boldsymbol{x}} = 1$. If all eigenvalues are negative, $\hat{\boldsymbol{x}} = \boldsymbol{0}$.

3.6 Are CQC Necessary?

SLIDE 21

$$\begin{aligned} \min \quad & x_1 \\ \text{s.t.} \quad & x_1^2 - x_2 \le 0 \\ & x_2 = 0 \end{aligned}$$

Feasible space is $(0,0)$.
KKT:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + u \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix} + v \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

KKT multipliers do not exist, while still $(0,0)'$ is local minimum. Check $\nabla g_1(0,0)$ and $\nabla h_1(0,0)$.

3.7 Constrained Qualification

SLIDE 22

Slater condition: There exists an $\boldsymbol{x}^0$ such that $g_j(\boldsymbol{x}^0) < 0$, $j = 1, \ldots, p$, and $h_i(\boldsymbol{x}^0) = 0$ for all $i = 1, \ldots, m$.

Theorem Under the Slater condition the KKT conditions are necessary.

4 Sufficient Optimality Conditions

SLIDE 23

Theorem If

- $\overline{\boldsymbol{x}}$ feasible for P
- Feasible set is P is convex and $f(\boldsymbol{x})$ convex
- There exist vectors $(\boldsymbol{u}, \boldsymbol{v})$, $\boldsymbol{u} \ge \boldsymbol{0}$:

$$\nabla f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i \nabla h_i(\overline{\boldsymbol{x}}) = \boldsymbol{0}$$

$$u_j g_j(\overline{\boldsymbol{x}}) = 0, \qquad j = 1, \ldots, p$$

Then, $\overline{\boldsymbol{x}}$ is a global minimum of P.

4.1 Proof

SLIDE 24

- Let $\boldsymbol{x} \in P$. Then $(1-\lambda)\overline{\boldsymbol{x}} + \lambda \boldsymbol{x} \in P$ for $\lambda \in [0,1]$.
- $g_j(\overline{\boldsymbol{x}} + \lambda(\boldsymbol{x} - \overline{\boldsymbol{x}})) \le 0 \;\Rightarrow$

$$\nabla g_j(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) \le 0$$

- Similarly, $h_i(\overline{\boldsymbol{x}} + \lambda(\boldsymbol{x} - \overline{\boldsymbol{x}})) \le 0 \;\Rightarrow$

$$\nabla h_i(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) = 0$$

SLIDE 25

- Thus,

$$\nabla f(\overline{\boldsymbol{x}})'(\boldsymbol{x} - \overline{\boldsymbol{x}}) =$$

$$-\left(\sum_{j=1}^{p} u_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i \nabla h_i(\overline{\boldsymbol{x}})\right)' (\boldsymbol{x} - \overline{\boldsymbol{x}}) \ge 0$$

$$\Rightarrow f(\boldsymbol{x}) \ge f(\overline{\boldsymbol{x}}).$$

5 Convex Optimization

SLIDE 26

- The KKT conditions are always necessary under CQC.
- The KKT conditions are sufficient for convex optimization problems.
- The KKT conditions are necessary and sufficient for convex optimization problems under CQC.
- $\min f(\boldsymbol{x})$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \ge \boldsymbol{0},\ f(\boldsymbol{x})$ convex, KKT are necessary and sufficient even without CQC.

5.0.1 Separating hyperplanes

SLIDE 27

Theorem Let S be a nonempty closed convex subset of $\Re^n$ and let $\boldsymbol{x}^* \in \Re^n$ be a vector that does not belong to S. Then, there exists some vector $\boldsymbol{c} \in \Re^n$ such that $\boldsymbol{c}'\boldsymbol{x}^* < \boldsymbol{c}'\boldsymbol{x}$ for all $\boldsymbol{x} \in S$.
Proof in BT, p.170

5.1 Sketch of the Proof under convexity

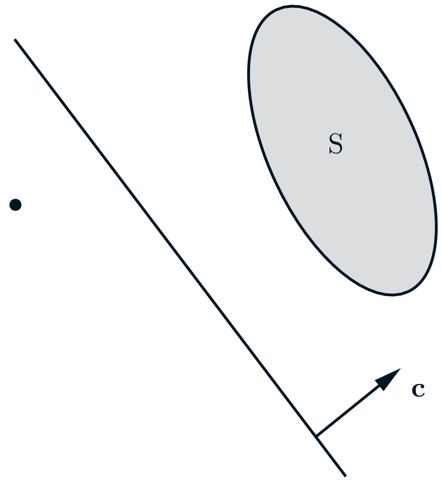
SLIDE 28

- Suppose $\overline{\boldsymbol{x}}$ is a local (and thus global) optimal solution.
- $f(\boldsymbol{x}) < f(\overline{\boldsymbol{x}}),\ g_j(\boldsymbol{x}) \le \boldsymbol{0},\ j = 1, \dots, p,\ h_i(\boldsymbol{x}) = \boldsymbol{0},\ i = 1, \dots, m$ is infeasible.
- Let $U = \{(u_0, \boldsymbol{u}, \boldsymbol{v}) |$ there exists $\boldsymbol{x}:\ f(\boldsymbol{x}) < u_0,\ g_j(\boldsymbol{x}) \le u_j,\ h_i(\boldsymbol{x}) = v_i\}$.
- $(f(\overline{\boldsymbol{x}}), \boldsymbol{0}, \boldsymbol{0}) \notin S$.
- U convex.

SLIDE 29

- By separating hyperplane theorem, there is a vector $(c_0, \boldsymbol{c}, \boldsymbol{d})$:

$$c_0 u_0 + \sum_{j=1}^{p} c_j u_j + \sum_{i=1}^{m} d_i v_i > c_0 f(\overline{\boldsymbol{x}})\ \forall\ (u_0, \boldsymbol{u}, \boldsymbol{v}) \in U.$$



- $c_0 \geq 0$ and $c_j \geq 0$ for $j \in I$ (constraint $g_j(\overline{\boldsymbol{x}}) \leq 0$ tight). Why?
  If $(u_0, \boldsymbol{u}, \boldsymbol{v}) \in U$, then $(u_0 + \lambda, \boldsymbol{u}, \boldsymbol{v}) \in U$ for $\lambda \geq 0$. Thus,

$$\forall\ \lambda \geq 0,\ \lambda c_0 + \sum_{j=1}^{p} c_j u_j + \sum_{i=1}^{m} d_i v_i > c_0 f(\overline{\boldsymbol{x}}) \ \Rightarrow\ c_0 \geq 0.$$

- Select $(u_0, \boldsymbol{u}, \boldsymbol{v}) = (f(\boldsymbol{x}) + \lambda, g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x}), h_1(\boldsymbol{x}), \ldots, h_m(\boldsymbol{x})) \in U$

- $$c_0(f(\boldsymbol{x}) + \lambda) + \sum_{j=1}^{p} c_j g_j(\boldsymbol{x}) + \sum_{i=1}^{m} d_i h_i(\boldsymbol{x}) > c_0 f(\overline{\boldsymbol{x}})$$

- Take $\lambda \to 0$:

$$c_0 f(\boldsymbol{x}) + \sum_{j=1}^{p} c_j g_j(\boldsymbol{x}) + \sum_{i=1}^{m} d_i h_i(\boldsymbol{x}) \geq c_0 f(\overline{\boldsymbol{x}})$$

- $c_0 > 0$ (constrained qualification needed here).

- $$f(\boldsymbol{x}) + \sum_{j=1}^{p} u_j g_j(\boldsymbol{x}) + \sum_{i=1}^{m} v_i h_i(\boldsymbol{x}) \geq f(\overline{\boldsymbol{x}}), \quad u_j \geq 0$$

- $$f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i h_i(\overline{\boldsymbol{x}}) \leq f(\overline{\boldsymbol{x}}) \leq$$

$$f(\boldsymbol{x}) + \sum_{j=1}^{p} u_j g_j(\boldsymbol{x}) + \sum_{i=1}^{m} v_i h_i(\boldsymbol{x}).$$

- Thus,

$$f(\overline{\boldsymbol{x}}) = \min\left(f(\boldsymbol{x}) + \sum_{j=1}^{p} u_j g_j(\boldsymbol{x}) + \sum_{i=1}^{m} v_i h_i(\boldsymbol{x})\right)$$

$$\sum_{j=1}^{p} u_j g_j(\overline{\boldsymbol{x}}) = 0 \quad \Rightarrow \quad u_j g_j(\overline{\boldsymbol{x}}) = 0$$

- Unconstrained optimality conditions:

$$\nabla f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j \nabla g_j(\overline{\boldsymbol{x}}) + \sum_{i=1}^{m} v_i \nabla h_i(\overline{\boldsymbol{x}}) = \mathbf{0}$$

6 Applications

6.1 Linear Optimization

Slide 30

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \mathbf{0} \end{aligned}$$

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \mathbf{0} \\ & -\boldsymbol{x} \leq \mathbf{0} \end{aligned}$$

6.1.1 KKT

Slide 31

$$\begin{aligned} \boldsymbol{c} + \boldsymbol{A}'\hat{\boldsymbol{u}} - \boldsymbol{v} &= \mathbf{0} \\ \boldsymbol{v} &\geq \mathbf{0} \\ v_j x_j &= 0 \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} &= \mathbf{0} \\ \boldsymbol{x} &\geq \mathbf{0} \end{aligned}$$

$\boldsymbol{u} = -\hat{\boldsymbol{u}}$

$$\begin{array}{rcll} \boldsymbol{A}'\boldsymbol{u} & \leq & \boldsymbol{c} & \text{dual feasibility} \\ (c_j - \boldsymbol{A}_j'\boldsymbol{u})x_j & = & 0 & \text{complementarity} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} & = & \mathbf{0} & \text{primal feasibility} \\ \boldsymbol{x} & \geq & \mathbf{0} & \text{primal feasibility} \end{array}$$

6.2 Portfolio Optimization

SLIDE 32

$\boldsymbol{x}$ =weights of the portfolio

$$\begin{aligned} \max \quad & \boldsymbol{r'x} - \lambda \frac{1}{2} \boldsymbol{x'Qx} \\ \text{s.t.} \quad & \boldsymbol{e'x} = 1 \end{aligned}$$

$$\begin{aligned} \min \quad & -\boldsymbol{r'x} + \lambda \frac{1}{2} \boldsymbol{x'Qx} \\ \text{s.t.} \quad & \boldsymbol{e'x} = 1 \end{aligned}$$

6.2.1 KKT

SLIDE 33

$$-\boldsymbol{r} + \lambda \boldsymbol{Qx} + u\boldsymbol{e} = 0$$

$$\boldsymbol{x} = \frac{1}{\lambda} \boldsymbol{Q}^{-1}(\boldsymbol{r} - u\boldsymbol{e})$$

$$\boldsymbol{e'x} = 1 \quad \Rightarrow \quad \boldsymbol{e'Q}^{-1}(\boldsymbol{r} - u\boldsymbol{e}) = \lambda$$

$$u = \frac{\boldsymbol{e'Q}^{-1}\boldsymbol{r} - \lambda}{\boldsymbol{e'Q}^{-1}\boldsymbol{e}}$$

As λ changes, tradeoff of risk and return changes. The allocation changes as well. This is the essense of modern portfolio theory.