C&O 355
Mathematical Programming
Fall 2010
Lecture 15

N. Harvey

Topics

- Minimizing over a convex set: Necessary & Sufficient Conditions
- (Mini)-**KKT Theorem**
  Minimizing over a polyhedral set: Necessary & Sufficient Conditions
- Smallest Enclosing Ball Problem

Minimizing over a Convex Set: Necessary & Sufficient Conditions

- **Thm 3.12:** Let $C \subseteq \mathbb{R}^n$ be a convex set.
  Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex and differentiable.
  Then x minimizes f over C iff $\nabla f(x)^T(z-x) \geq 0 \ \forall z \in C$.

- **Proof:** $\Leftarrow$ direction

  Direct from subgradient inequality. (Theorem 3.5)

$$f(z) \geq f(x) + \nabla f(x)^T(z-x) \geq f(x)$$

Subgradient inequality

Our hypothesis

Minimizing over a Convex Set: Necessary & Sufficient Conditions

- **Thm 3.12:** Let $C \subseteq \mathbb{R}^n$ be a convex set.
  Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable.
  Then x minimizes f over C iff $\nabla f(x)^T(z-x) \geq 0 \ \forall z \in C$.

- **Proof:** $\Rightarrow$ direction

  Let x be a minimizer, let $z \in C$ and let $y = z-x$.

  Recall that $\nabla f(x)^T y = f'(x;y) = \lim_{t \to 0} \frac{f(x+ty)-f(x)}{t}$.

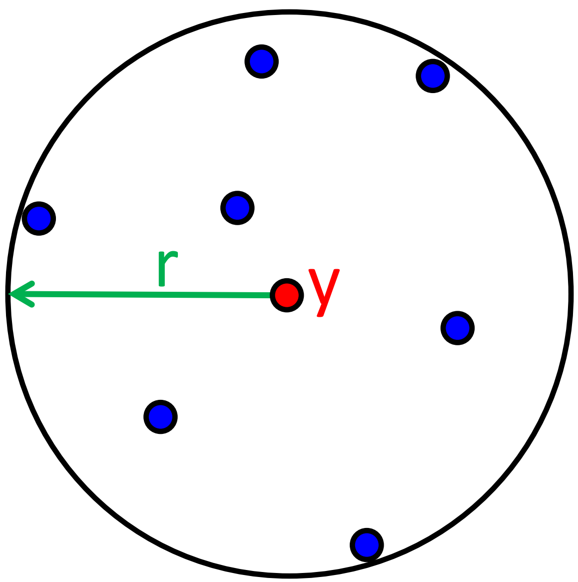
  If limit is negative then we have $f(x+ty)<f(x)$ for some $t \in [0,1]$, contradicting that x is a minimizer.

  So the limit is non-negative, and $\nabla f(x)^T y \geq 0$. ■

Smallest Ball Problem

- Let $\{p_1,\ldots,p_n\}$ be points in $\mathbb{R}^d$.
  Find (unique!) ball (not an ellipsoid!)
  of smallest volume that contains all the p_i's.
- In other words, we want to solve:

$$\min \{\, r : \exists y \in \mathbb{R}^d \text{ s.t. } p_i \in B(y,r)\ \forall i \,\}$$



Smallest Ball Problem

- Let $\{p_1,\ldots,p_n\}$ be points in $\mathbb{R}^d$.
  Find (unique!) ball (not an ellipsoid!)
  of smallest volume that contains all the p_i's.
- In other words, we want to solve:

$$\min \{ r : \exists y \in \mathbb{R}^d \text{ s.t. } p_i \in B(y,r)\ \forall i \}$$

- We will formulate this as a convex program.
- In fact, our convex program will be of the form

$$\min \{ f(x) : Ax=b,\ x \geq 0 \}, \quad \text{where } f \text{ is convex.}$$

Minimizing a convex function over a polyhedron

- To solve this, we will need **optimality conditions for convex programs**.

LP Optimality Conditions

Theorem:
Let $x \in \mathbb{R}^n$ be a feasible solution to the linear program

$$\max \{ c^T x : Ax=b, x \geq 0 \}$$

Then x is optimal iff $\exists$ dual solution $y \in \mathbb{R}^m$ s.t.

1) $A^T y \geq c$,
2) For all j, if $x_j > 0$ then $A_j^T y = c_j$.

j^{th} column of A

- **Proof:** Dual is $\min \{ b^T y : A^T y \geq c \}$.
- x optimal $\Rightarrow$ dual has optimal solution y.
- So (1) holds by feasibility of y.
- By optimality of x & y, $c^T x = b^T y$. Weak duality says:

$$c^T x = \sum_{j=1}^{n} c_j x_j \leq \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} A_{i,j} y_i \Big) x_j = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{i,j} x_j \Big) y_i = \sum_{i=1}^{m} b_i y_i = b^T y$$

- Equality holds here $\Rightarrow$ (2) holds. (This is "complementary slackness")

LP Optimality Conditions

Theorem:
Let $x \in \mathbb{R}^n$ be a feasible solution to the linear program

$$\min \{ -c^T x : Ax=b,\ x \geq 0 \}$$

Then x is optimal iff $\exists$ dual solution $y \in \mathbb{R}^m$ s.t.

1) $-c^T + y^T A \geq 0$,
2) For all j, if $x_j > 0$ then $-c_j + y^T A_j = 0$.

j^{th} column of A

- **Proof:** Dual is $\min \{ b^T y : A^T y \geq c \}$.
- x optimal $\Rightarrow$ dual has optimal solution y.
- So (1) holds by feasibility of y.
- By optimality of x & y, $c^T x = b^T y$. Weak duality says:

$$c^T x = \sum_{j=1}^{n} c_j x_j \leq \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} A_{i,j} y_i \Big) x_j = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{i,j} x_j \Big) y_i = \sum_{i=1}^{m} b_i y_i = b^T y$$

- Equality holds here $\Rightarrow$ (2) holds. (This is "complementary slackness")

(Mini)-KKT Theorem

Theorem: Let $f:\mathbb{R}^n \rightarrow \mathbb{R}$ be a convex, C^2 function.
Let $x \in \mathbb{R}^n$ be a feasible solution to the convex program

$$\min \{ f(x) : Ax=b,\ x \geq 0 \}$$

Then x is optimal iff $\exists y \in \mathbb{R}^m$ s.t.

1) $\nabla f(x)^T + y^T A \geq 0$,
2) For all j, if $x_j > 0$ then $\nabla f(x)_j + y^T A_j = 0$.

j^{th} column of A

- Natural generalization of LP optimality conditions: approximation at $\bar{x}$ is $f(x) \approx f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$
- Proven by Karush in 1939 (his Master's thesis!), and by Kuhn and Tucker in 1951.

Full KKT Theorem

Even stating it requires a lot of details!

See Section 3.7 and 3.8 of the course notes

$$
(3.25) \quad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0 \quad (i = 1, 2, \ldots, p), \\ & h_j(x) = 0 \quad (j = 1, 2, \ldots, q), \\ & x \in S. \end{cases}
$$

Theorem 3.22 (Karush-Kuhn-Tucker Theorem) *Consider the nonlinear program (3.25). Suppose the Mangasarian-Fromowitz Constraint Qualification holds at the point* $\bar{x} \in \mathbf{R}^n$*, and assume furthermore that the objective function* $f : S \to \mathbf{R}$ *is differentiable there. Then a necessary condition for* $\bar{x}$ *to be a local minimizer is the existence of Lagrange multipliers* $\lambda_i \geq 0$ *in* $\mathbf{R}$ *(for the indices* $i \in I(\bar{x})$*) and* $\mu_j \in \mathbf{R}$ *(for the indices* $j = 1, 2, \ldots, q$*) with*

$$
(3.31) \qquad \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\bar{x}) = 0.
$$

Mangasarian-Fromowitz Constraint Qualification

(i) *The point* $\bar{x}$ *lies in the open set* $S \subseteq \mathbf{R}^n$.

(ii) *The continuous functions* $g_1, g_2, \ldots, g_p, h_1, h_2, \ldots, h_q : S \to \mathbf{R}$ *satisfy*

$$
\begin{aligned} g_i(\bar{x}) &= 0 \quad (i \in I(\bar{x})) \\ g_i(\bar{x}) &< 0 \quad (i \notin I(\bar{x})) \\ h_j(\bar{x}) &= 0 \quad (j = 1, 2, \ldots, q). \end{aligned}
$$

(iii) *The functions* $g_i, h_j : S \to \mathbf{R}$ *are continuously differentiable, for* $i \in I(\bar{x})$ *and* $j = 1, 2, \ldots, q$.

(iv) *The set of gradients* $\{\nabla h_j(\bar{x}) : j = 1, 2, \ldots, q\}$ *is linearly independent.*

(v) *The set* H *of vectors* $d \in \mathbf{R}^n$ *satisfying*

$$
(3.29) \qquad \begin{cases} \nabla g_i(\bar{x})^T d < 0 \quad (i \in I(\bar{x})) \\ \nabla h_j(\bar{x})^T d = 0 \quad (j = 1, 2, \ldots, q), \end{cases}
$$

is nonempty.

Theorem: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex, C^2 function.
Let $x \in \mathbb{R}^n$ be a feasible solution to the convex program

$$\min \{ f(x) : Ax=b,\ x \geq 0 \}$$

Then x is optimal iff $\exists y \in \mathbb{R}^m$ s.t.

1) $\nabla f(x)^T + y^T A \geq 0$,
2) For all j, if $x_j > 0$ then $\nabla f(x)_j + y^T A_j = 0$.

Proof: $\Leftarrow$ direction. Suppose such a y exists. Then

$$(\nabla f(x)^T + y^T A)\, x = 0.$$ (Just like complementary slackness)

For any feasible $z \in \mathbb{R}^n$, we have

$$(\nabla f(x)^T + y^T A)\, z \geq 0.$$

Subtracting these, and using $Ax=Az=b$, we get

$$\nabla f(x)^T (z-x) \geq 0 \quad \forall \text{ feasible } z.$$

So x is optimal. (By Thm 3.12: "Minimizing over a Convex Set")

Theorem: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex, C^2 function.
Let $x \in \mathbb{R}^n$ be a feasible solution to the convex program

$$\min \{ f(x) : Ax=b,\ x \geq 0 \}$$

Then x is optimal iff $\exists y \in \mathbb{R}^m$ s.t.

1) $\nabla f(x)^T + y^T A \geq 0$,
2) For all j, if $x_j > 0$ then $\nabla f(x)_j + y^T A_j = 0$.

Proof: $\Rightarrow$ direction. Suppose x is optimal. Let $c = -\nabla f(x)$.
Then $\nabla f(x)^T (z-x) \geq 0 \Rightarrow c^T z \leq c^T x$ for all feasible points z.

By Thm 3.12: "Minimizing over a Convex Set"

Theorem: Let $f:\mathbb{R}^n \rightarrow \mathbb{R}$ be a convex, C^2 function.
Let $x \in \mathbb{R}^n$ be a feasible solution to the convex program

$$\min \{ f(x) : Ax=b,\ x \geq 0 \}$$

Then x is optimal iff $\exists y \in \mathbb{R}^m$ s.t.

1) $\nabla f(x)^T + y^T A \geq 0$,
2) For all j, if $x_j > 0$ then $\nabla f(x)_j + y^T A_j = 0$.

Proof: $\Rightarrow$ direction. Suppose x is optimal. Let $c = -\nabla f(x)$.
Then $\nabla f(x)^T (z-x) \geq 0 \quad \Rightarrow \quad c^T z \leq c^T x$ for all feasible points z.
So x is optimal for the LP $\max \{ c^T x : Ax=b,\ x \geq 0 \}$.
So there is an optimal solution y to dual LP $\min \{ b^T y : A^T y \geq c \}$.
So $\nabla f(x)^T + y^T A = -c^T + y^T A \geq 0 \quad \Rightarrow \quad$ (1) holds.

Furthermore, x and y are both optimal so C.S. holds.

$\Rightarrow$ whenever $x_j > 0$, the j^{th} dual constraint is tight
$\Rightarrow$ $y^T A_j = c_j \quad \Rightarrow \quad$ (2) holds. ∎

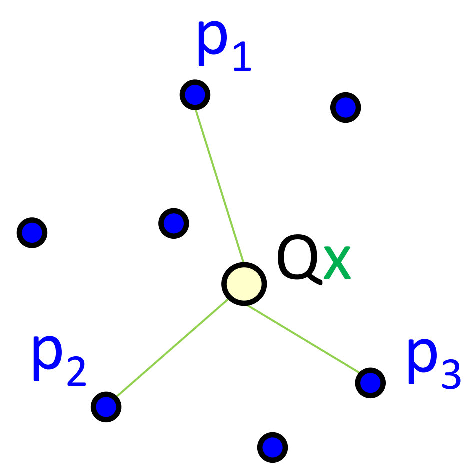
Smallest Ball Problem

- Let $P=\{p_1,\ldots,p_n\}$ be points in $\mathbb{R}^d$.
  Let Q be dxn matrix s.t. $Q_i=p_i$. (Q_i = i^{th} column of Q)
  Let $z \in \mathbb{R}^n$ satisfy $z_i=p_i^T p_i$.
  Define $f : \mathbb{R}^n \rightarrow \mathbb{R}$ by $f(x) = x^T Q^T Q x - x^T z$.
- **Claim 1:** f is convex. (Hessian is PSD)
- Consider the convex program

$$\min \{ f(x) : \textstyle\sum_j x_j = 1,\ x \geq 0 \}.$$

Interpretation

- Qx is an "average" (convex combination) of the p_i's
- $x^T Q^T Q x$ is norm² of this average point
- $x^T z$ is average norm² of the p_i's
- If $x^T Q^T Q x \ll x^T z$, the p_i's are "spread out"



x = [1/3, 1/3, 1/3, 0, 0, …]

Smallest Ball Problem

- Let $P=\{p_1,\ldots,p_n\}$ be points in $\mathbb{R}^d$.
  Let Q be $d\times n$ matrix s.t. $Q_i=p_i$. (Q_i = i^{th} column of Q)
  Let $z\in\mathbb{R}^n$ satisfy $z_i=p_i^T p_i$.
  Define $f : \mathbb{R}^n \rightarrow \mathbb{R}$ by $f(x) = x^T Q^T Q x - x^T z$.
- **Claim 1:** f is convex. (Hessian is $2Q^TQ$, which is PSD)
- Consider the convex program

$$\min \{ f(x) : \textstyle\sum_j x_j = 1, x \geq 0 \}.$$

- **Claim 2:** This program has an optimal solution x.
  (By Weierstrass' Theorem: the high-dimensional extreme value theorem)
- **Claim 3:** Assume x is optimal.
  Let $p^*=Qx$ and $r=\sqrt{-f(x)}$. Then $P\subset B(p^*,r)$.
- **Claim 4:** $B(p^*,r)$ is the smallest ball containing P.

- **Claim 3:** The ball $B(p^*, r)$ contains P.
- **Proof:** By KKT, $\exists y \in \mathbb{R}$ s.t. $\nabla f(x) + A^T y \geq 0$

For us $\nabla f(x) = 2Q^TQx - z = 2Q^Tp^* - z$ and $A = [1, \ldots, 1]$

So $y \geq p_j^T p_j - 2p_j^T p^* \;\; \forall j$. (Here $y \in \mathbb{R}$)

KKT also says: equality holds $\forall j$ s.t. $x_j > 0$.

So $y = \sum_j x_j y = \sum_j x_j p_j^T p_j - 2\sum_j x_j p_j^T p^* = \sum_j x_j p_j^T p_j - 2p^{*T}p^*$.

So $y + p^{*T}p^* = \sum_j x_j p_j^T p_j - p^{*T}p^* = -f(x) \Rightarrow r = \sqrt{y + p^{*T}p^*}$

It remains to show that $B(p^*, r)$ contains P.

This holds iff $\|p_j - p^*\| \leq r \;\; \forall j$.

Now $\|p_j - p^*\|^2 = (p_j - p^*)^T(p_j - p^*)$

$= p^{*T}p^* - 2p_j^T p^* + p_j^T p_j$

$\leq p^{*T}p^* + y = r^2 \;\; \forall j$. □

- **Claim 4:** $B(p^*, r)$ is the smallest ball containing P.
- **Proof:** See Matousek-Gartner Section 8.7.

Smallest Ball Problem: Summary

- Consider the convex program

$$\min \{ f(x) : \textstyle\sum_j x_j = 1, x \geq 0 \}.$$

- **Claim 2:** This program has an optimal solution x.
- **Claim 3:** Let $p^* = Qx$ and $r = \sqrt{-f(x)}$. Then $P \subset B(p^*, r)$.
- **Claim 4:** $B(p^*, r)$ is the smallest ball containing P.

- This example is a bit strange:
  - Not obvious how convex program relates to balls
  - Claim 3 is only valid when x is the **optimal** point
- Still, KKT tells us interesting things:
  - optimal value of convex program gives radius of smallest ball containing P