INDENG 262A Notes, Fall 2020

Mathimatical Programming

Professor: Alper Atamtürk
Fengzhe Shi

Contents

1 Week 1 Lecture

1.1 Optimization Problem

An optimization problem (P) has the following format,

$$\begin{aligned} \max / \min \ & f(x) \\ \text{s.t. } & g_i(x) \le b_i, \ i = 1, \cdots, m \\ & x \in S \subseteq \mathbb{R}^n \end{aligned}$$

$F = \{x \in S : \ g_i(x) \le b_i, \ i = 1, \cdots, m\}$

1.2 Feasible Region/Sets

$x \in F$ is called feasible point.

Definition 1.1. *If $F \neq \phi$, then P is infeasible.*

Definition 1.2. *If $\exists x \in F$ s.t. $f(x) \ge \lambda$, $\forall \lambda \in \mathbb{R}$, then P is unbounded.*

Definition 1.3. *$\bar{x} \in F$ is a global maximizer of P if $f(\bar{x}) \ge f(x), \ \forall x \in F$.*

Example 1.1

Show that

$$\max\{f(x) : x \in F\} = -\min\{f(x) : x \in F\}$$

1.3 Definitions

Definition 1.4. *Let $x \in \mathbb{R}^n, \ 0 < \epsilon \in R$, the epsilon neighborhood of x in $\mathbb{R}^n$ is the set $N_\epsilon(x) = \{y \in \mathbb{R}^n : ||x - y|| \le \epsilon\}$.*

Definition 1.5. *$S \subseteq \mathbb{R}^n$ is bounded if $S \subseteq N_\epsilon(0)$ for some $\epsilon > 0$.*

Definition 1.6. *$x \in S \subseteq \mathbb{R}^n$ is an interior point of S if $N_\epsilon(x) \subset S$ for some $\epsilon > 0$.*

Definition 1.7. *$x \in S \subseteq \mathbb{R}^n$ is a boundary point if $N_\epsilon(x) \subset S$ contains at least one point in S and at least one point not in S for any $\epsilon > 0$.*

Definition 1.8. *Closure of $S \subseteq \mathbb{R}^n$ is the set $cl(S) \subset S \cup bd(S)$.*

Definition 1.9. *S is closed if $S = cl(S)$.*

Definition 1.10. *S is open if $S = int(S)$.*

Example 1.2

Let $S = \{(x, y) \in \mathbb{R}^2 : \ x^2 + y^2 \leq 1\}$,

$$int(S) = \{(x, y) \in \mathbb{R}^2 : \ x^2 + y^2 < 1\}$$
$$bd(S) = \{(x, y) \in \mathbb{R}^2 : \ x^2 + y^2 = 1\}$$

$\mathbb{R}^2$, ϕ are both open and closed.

Definition 1.11. *$\bar{x} \in F$ is called a local minimizer if there exists small $\epsilon > 0$ s.t. $f(\bar{x}) \leq f(x) \quad \forall x \in F : \ x \in N_\epsilon(\bar{x})$*

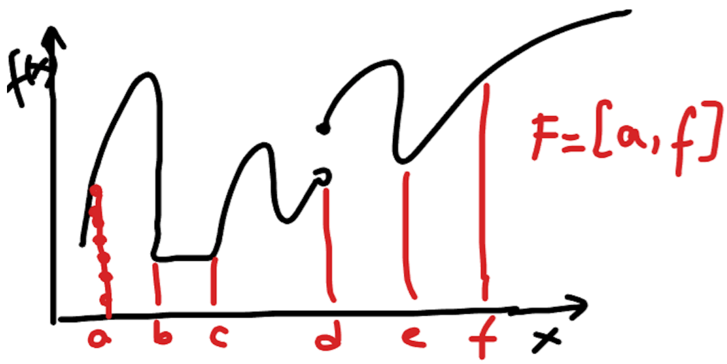


Figure 1.1: Some local minimizers a, $[b, c]$, e and global minimizers $[b, c]$

Non-existence of optima:

1) $F = \phi$

2) $F = R_+$, F unbounded

3) $F = (a, b)$, F not closed

4) $F = [a, b]$, f not continuous

Theorem 1.1 (Weierstrass Theorem)

Let F be a nonempty compact (bounded, closed) set, and $f : F \to \mathbb{R}$ be continuous on F. Then $\min\{f(x) : x \in F\}$ attains its minimum (there exists minimizer in F).

Proof. f continuous, F bounded, closed, $F \neq \phi$, $\exists \alpha \equiv inf\{f(x) : x \in F\}$ □

Definition 1.12. *α is the greatest lower bound in f on F: $\alpha \leq f(x)$, $\forall x \in F$ and $\nexists \bar{\alpha} > \alpha$ s.t. $\bar{\alpha} \leq f(x)$, $\forall x \in F$.*

1.4 Combinations

Let x^i be vectors in $\mathbb{R}^n$, $i = 1, \cdots, k$.

Definition 1.13. *$\bar{x} \in \mathbb{R}^n$ is a <u>convex combination</u> of $\{x^i\}$, if $\bar{x} = \sum_{i=1}^k \lambda_i x^i$ subject to $\sum_{i=1}^k \lambda_i = 1,\ \lambda \geq 0$.*

Definition 1.14. *$\bar{x} \in \mathbb{R}^n$ is a <u>conic combination</u> of $\{x^i\}$, if $\bar{x} = \sum_{i=1}^k \lambda_i x^i,\ \lambda \geq 0$.*

Definition 1.15. *The <u>convex hull</u> of $\{x^i\}$ is the set of all convex combinations of the vectors.*

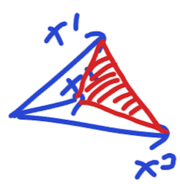


Figure 1.2: Convex combinations

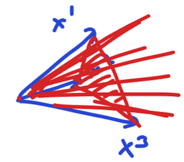


Figure 1.3: Conic combination

2 Week 2 Lecture

2.1 Convexity

Definition 2.1. *$S \subseteq \mathbb{R}^n$ is convex if $\lambda x^1 + (1-\lambda)x^2 \in S,\ \forall x^1, x^2 \in S,\ \lambda \in [0,1]$.*

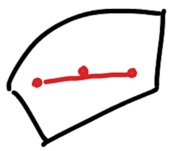


Figure 2.1: Convex

Figure 2.2: Not convex

Definition 2.2. *Let $S \subseteq \mathbb{R}^n$ be a convex set. Let $f : \mathbb{R}^n \to \mathbb{R}$.*

(i) f is a convex function on S if

$$f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2),\ \forall x^1, x^2 \in S,\ \lambda \in [0,1]$$

(ii) f is strictly convex on S if

$$f(\lambda x^1 + (1-\lambda)x^2) < \lambda f(x^1) + (1-\lambda) f(x^2),\ \forall x^1, x^2 \in S,\ \lambda \in (0,1)$$

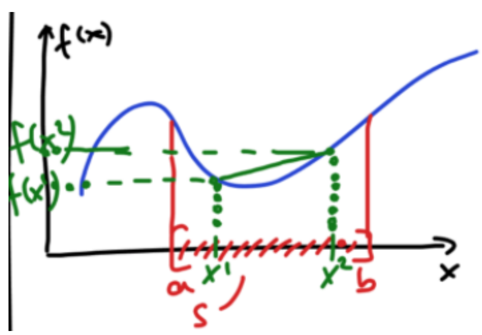


Figure 2.3: A convex function

Definition 2.3. *Let $S \subseteq \mathbb{R}^n$ be a convex set. $f : \mathbb{R}^n \to \mathbb{R}$ is concave on S if $f(\lambda x^1 + (1-\lambda)x^2) \geq \lambda f(x^1) + (1-\lambda) f(x^2),\ \forall x^1, x^2 \in S,\ \lambda \in [0,1]$.*

Remark: f is convex if and only if $-f$ is concave.

Definition 2.4. *$f : \mathbb{R}^n \to \mathbb{R}$ is an affine function if $f(x) = \sum_{j=1}^n a_j x_j + a_0,\ a_j \in \mathbb{R}$. If $a_0 = 0$, then f is linear.*

Definition 2.5. *Epigraph of function $f : \mathbb{R}^n \to \mathbb{R}$ is*

$$\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1} : f(x) \leq y\}$$

Definition 2.6. *Hypograph of function f is*

$$\text{hyp}(f) = \{(x, y) \in \mathbb{R}^{n+1} : f(x) \geq y\}$$

Definition 2.7. *The lower level set of f for $a \in \mathbb{R}$ is*

$$S_a = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$$

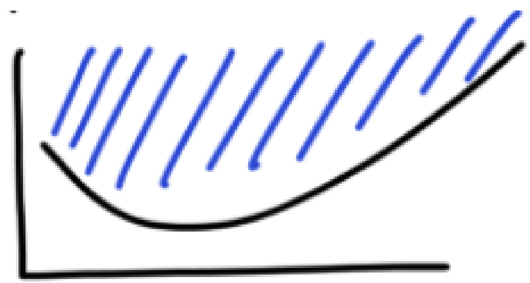


Figure 2.4: Epigraph

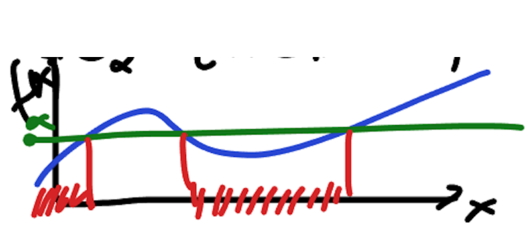


Figure 2.5: Lower level set

Definition 2.8. *Let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. x is an* <u>*extreme point*</u> *of S if $\nexists y, z \in S$ subject to for any $0 < \lambda < 1$,*

$$x = \lambda y + (1 - \lambda) z$$

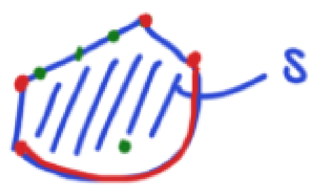


Figure 2.6: Extreme point

2.2 Why is convexity important?

Proposition 2.1

Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function.

(i) A local minimizer of f on S is also a global minimizer of f on S.

Proof. Let x be a local min of f on S. For contradiction, suppose x is not a global min of f on S. Then $\exists y \in S : f(y) < f(x)$. Let z be a strict convex combination of x and y: $z = \lambda x + (1 - \lambda) y,\ 0 < \lambda < 1$.

$$\begin{aligned} f(z) &\leq \lambda f(x) + (1 - \lambda) f(y) \\ &< \lambda f(x) + (1 - \lambda) f(x) \\ &= f(x) \end{aligned}$$

Observe as $\lambda \to 1$, $z \to x$, $\nexists$ no ϵ-neighborhood where x is a local min. Contradiction! $\square$

(ii) Moreover, if f is strictly convex, then there exists at most one global minimizer of f on S.

Proof. If $x \neq y$ are two global minimizers,

$$
\begin{aligned}
f(\frac{1}{2}x + \frac{1}{2}y &< \frac{1}{2}f(x) + \frac{1}{2}f(y)) \\
&= f(x) = f(y)
\end{aligned}
$$

$f(z) < f(x) = f(y)$. Contradiction! $\square$

3 Week 3 Lecture

3.1 Projection onto Closed Convex Sets

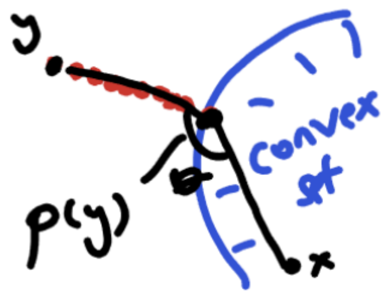


Figure 3.1: Projection

Theorem 3.1 (Projection Theorem)
Let C be a closed convex subset of $\mathbb{R}^n$.

(i) For any $y \in \mathbb{R}^n$, there exits a unique point $p(y) \in C$, that is closest to y, i.e.

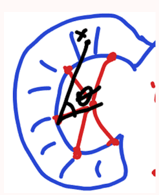
$$p(y) = \text{argmin}\{||y - x||^2\}$$

$p(y)$ is called the projection of y onto C.

(ii) $z \in C$ is $p(y)$ if and only if

$$(y - z)^\top (x - z) \leq 0, \ \forall x \in C$$

Example 3.1



Observe if C is not convex:
there may be multiple projection points.

ii) part two is not satisfied either. θ may be acute as seen in the picture.

Proof. i) We may assume S is bounded WLOG. By Weierstrass Theorem (1.1), projection point exist. Since the objective is strictly convex, $p(y)$ is unique.

ii) ($\Longleftarrow$) Suppose $(y - z)^\top (x - z) \leq 0, \ \forall x \in C$. Let $x \in C$.

$$\begin{aligned} ||y - x||^2 &= ||y - z + z - x||^2 \\ &= ||y - z||^2 + ||z - x||^2 + 2(y - z)^\top (x - z) \end{aligned}$$

[1] So, $||y-x||^2 \geq ||y-z||^2, \ \forall x \in C$.

$\therefore z = p(y)$

$(\Longrightarrow)$Suppose $z = p(y)$, thus

$$||y-x||^2 \geq ||y-z||^2, \ \forall x \in C$$

Fix some $\forall x \in C$. Then $z + \lambda(x-z) \in C$ for $0 \geq \lambda \geq 1$.

$$||y-z-\lambda(x-z)||^2 = ||y-z||^2 + ||x-z||^2 - 2\lambda(y-z)^\top(x-z)$$

So,

$$\lambda^2||x-z||^2 - 2\lambda(y-z)^\top(x-z) \geq 0$$

For $\lambda > 0$, divide by λ, let $\lambda \to 0^+, \ (y-z)^\top(x-z) \leq 0$.

$\square$

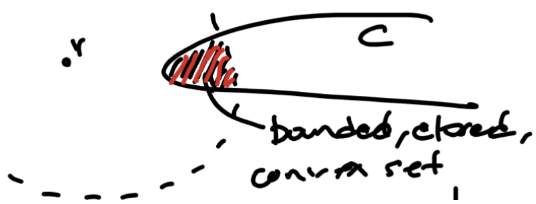


Figure 3.2: Projection theorem

3.2 Separating Hyperplanes

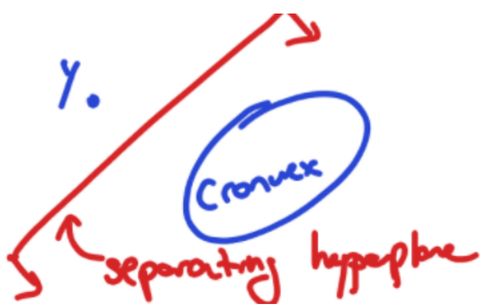


Figure 3.3: Separating hyperplane

Definition 3.1. *A hyperplane in $\mathbb{R}^n$ is the set $H = \{x \in \mathbb{R}^n : \ a^\top x = \alpha\}$ defined by $0 \neq a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.*

Definition 3.2. *A hyperplane divides $\mathbb{R}^n$ into two halfspaces:*

$$H^- = \{x \in \mathbb{R}^n : \ a^\top x \leq \alpha\} \text{ and}$$
$$H^+ = \{x \in \mathbb{R}^n : \ a^\top x \geq \alpha\}$$

[1]For $a, b \in \mathbb{R}^n$, we have $\|a \pm b\|^2 = \|a\|^2 + \|b\|^2 \pm 2a^\top b$

Theorem 3.2 (Separating Hyperplane Theorem)
Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed convex set and $y \in \mathbb{R}^n \setminus C$. Then there exists a hyperplane that separates them, i.e. there exists $0 \neq a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ subject to

$$a^\top x \leq \alpha < a^\top y, \ \forall x \in C$$

Proof. Let $z = p(y)$. By the projection theorem (3.1),

$$(y - z)^\top (x - z) \leq 0, \ \forall x \in C$$

Let $a = y - z$ and $\alpha = a^\top z$. Then $a^\top x \leq \alpha$ (Note: $a \neq 0, \ y \neq z$). To see $\alpha < a^\top y$, observe that

$$\alpha \equiv a^\top z < a^\top y,$$
$$a^\top (y - z) = a^\top a > 0 \quad (a \neq 0)$$

□

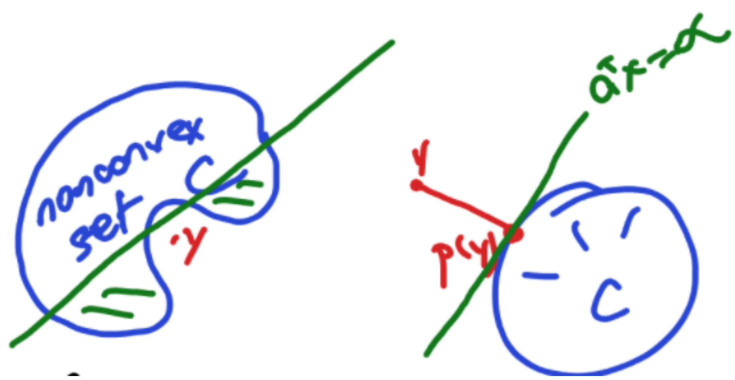


Figure 3.4: Separating hyperplane theorem

4 Week 4 Lecture

4.1 Supporting Hyperplanes

Theorem 4.1 (Supporting Hyperplane Theorem)
If $S \subsetneq \mathbb{R}^n$ is a nonempty, closed, convex set and z is a boundary point of S, then

$$\exists 0 \neq a \in \mathbb{R}^n \text{ s.t. } a^\top (x - z) \leq 0, \ \forall x \in S$$

Proposition 4.1
Let $f : \mathbb{R}^n \to R$ be a concave function and $S \subseteq \mathbb{R}^n$ be a closed, convex set with an extreme point. If $\min\{f(x) : x \in S\}$ has an optimal solution, then it has an optimal solution that is an extreme point of S.

Proof. Let O be the set of optimal sets. $O \subseteq S$. S has extreme point implies that O has extreme point. Let $\bar{x}$ be an extreme point of O. If $\bar{x}$ is an extreme point of S, we are done. Otherwise,

$$\begin{aligned} \bar{x} &= \lambda z + (1 - \lambda) y, \ 0 < \lambda < 1, \ z, y \in S \\ f(\bar{x}) &\geq \lambda z + (1 - \lambda) y \end{aligned}$$

Also

$$f(\bar{x}) \leq z, f(\bar{x}) \leq y$$

$f(\bar{x}) = f(z) = f(y)$ implies that $\bar{x}, y, z \in O$. $\bar{x}$ is not an extreme point of O. Contradiction! Therefore $\bar{x}$ must be an extreme point of S! □

Figure 4.1: Existence of extreme points

Theorem 4.2
Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed, convex set. Then C has an extreme point if and only if it contains no line.

Proof. [$\Longrightarrow$] Suppose x is an extreme point of C. For contradiction, suppose C contains a line

$$L \equiv \{\bar{x} + \alpha d : \alpha \in \mathbb{R}\}, \ d \neq 0$$

For positive integer n, consider

$$\begin{aligned} x^n &= (1-\frac{1}{n})x + \frac{1}{n}(\bar{x}+nd) \\ &= x + d + \frac{1}{n}(\bar{x}-x) \in C \\ \lim_{n\to\infty} x^n &= x+d \in C \end{aligned}$$

Similarly, $x-d \in C$.

$$x = \frac{1}{2}(x+d) + \frac{1}{2}(x-d)$$

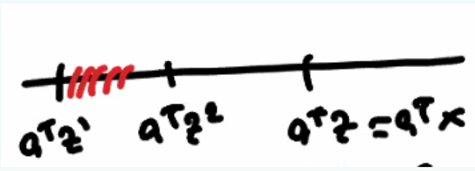
Contradiction with x is an extreme point of C.

[$\Longleftarrow$] Suppose C has no line. Induction on dimension.
$C \subseteq \mathbb{R}^1$. Trivial.
Assume true for sets of dimension up to $n-1$. C is nonempty, closed, convex set with no line. Then C has a boundary point $\bar{x}$. Consider a supporting hyperplane at $\bar{x}$.

$$\begin{aligned} H &= \{x \in \mathbb{R}^n : \ a^\top x = a^\top \bar{x}\} \\ C &\subseteq \{x \in \mathbb{R}^n : \ a^\top x \le a^\top \bar{x}\} \end{aligned}$$

$C \cap H$ has dimension $n-1$. By assumption, $C \cap H$ has an extreme point z. Z must be an extreme point of C as well. Otherwise,

$$\begin{aligned} & z = \lambda z^1 + (1-\lambda)z^2, \ 0<\lambda<1, z^1, z^2 \in C \\ & a^\top z^1 \le a^\top z = a^\top \bar{x} \\ & a^\top z^2 \le a^\top \bar{x} = a^\top z \end{aligned}$$



$\therefore a^\top z^1 = a^\top z^2 = a^\top z = a^\top \bar{x}$.
$\therefore z^1, z^2 \in C \cap H$. Contradiction with z is an extreme point. $\square$

4.2 Intro to Linear Optimization

$$
\begin{aligned}
\min / \max\ & c^\top x \\
s.t.\ & a_i^\top x \le b_i, \\
& x \in \mathbb{R}^n \\
& Ax \le b \ \text{(matrix form)}
\end{aligned}
$$

Definition 4.1. *A* <u>*polyhedron*</u> *is a set that can be described in the form*

$$\{x \in \mathbb{R}^n \mid Ax \ge b\}$$

It is the intersection of a finite number of halfspaces.

Observe: Any polyhedron is convex and closed.

Definition 4.2. *A bounded polyhedron is called a* <u>*polytope*</u>*.*

Theorem 4.3

(a) The intersection of convex sets is convex.

(b) Every polyhedron is a convex set.

(c) A convex combination of a finite number of elements of a convex set also belongs to that set.

(d) The convex hull of a finite number of vectors is a convex set.

Theorem 4.4

A nonempty and bounded polyhedron is the convex hull of its extreme points.

Example 4.1

$$
\begin{aligned}
\min\ & c^\top x \\
s.t.\ & Ax = b \\
& x \in \mathbb{R}^n
\end{aligned}
$$

$F = \{x \in \mathbb{R}^n : \ Ax = b\}$

Cases:

1) $F = \Phi$

2) $F = \{\bar{x}\}, \ \bar{x} = A^{-1}b$ (A nonsingular)

3) F is an affine subspace

    i) $c^\top d \neq 0$ for some $d \in \text{Null}(A)$, is unbounded

    ii) $c^\top d = 0\ \forall d \in \text{Null}(A)$. All points in F are optimal.

4.3 Linear Programming

$$\begin{aligned} &\min\ c^\top x \\ &\text{s.t. } Ax \leq b \end{aligned}$$

Equivalent Forms of LP:

1. $\max c^\top x = -\min -c^\top x$
2. $a^T x \leq b \iff -a^T x \geq -b$
3. $a^T x = b \iff \begin{cases} a^\top x & \leq b \\ a^\top x & \geq b \end{cases}$
4. $a^T x \leq b \iff \begin{cases} a^\top x + s & = b \\ s & \geq 0 \end{cases}$
5. $x \in \mathbb{R}$ (unrestricted in sign/free) $\iff \begin{cases} x = x^+ - x^- \\ x^+ \geqslant 0, x^- \geq 0 \end{cases}$

Canonical Form

$$\begin{aligned} &\min\ c^\top x \\ &s.t.\ Ax \leq b \end{aligned}$$

Standard Form

$$\begin{aligned} \min\ c^\top x^+ - c^\top x^- & \\ Ax^+ - Ax^- Js = & b \\ x^+, x^-, s \geqslant & 0 \end{aligned}$$

4.4 LP in $\mathbb{R}^2$

Example 4.2

$$\begin{aligned} \max\ -x_2 & \\ \text{s.t. } x_1 &\geq 2 \\ 3x_1 - x_2 &\geq 0 \\ x_1 + x_2 &\geq 6 \\ -x_1 + 2x_2 &\geq 0 \end{aligned}$$

Matrix Form

$$
\begin{aligned}
&\max\ [0,-1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&\text{s.t. } \begin{bmatrix} 1 & 0 \\ 3 & -1 \\ 1 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geqslant \begin{bmatrix} 2 \\ 0 \\ 6 \\ 0 \end{bmatrix}
\end{aligned}
$$

5 Week 5 Lecture

5.1 Important Questions on LP

1) When is an LP feasible?
2) When is an LP unbounded?
3) If there is an optimal solution:
   a) How do we characterize it?
   b) How do we find it?

When is an LP feasible?

$$\min\left\{c'x : \ Ax = b, \ x \geq 0\right\}$$

For the moment, ignore $x \geq 0$. When is $Ax = b$ feasible?

Example 5.1

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ 2x + 3x_2 + x_3 &= 8 \\ 2x_1 + x_2 + 3x_0 &= 0 \end{aligned}$$

In general, we have:

Theorem 5.1

Exactly one of the following is true:

I) $\exists x \in \mathbb{R}^n : \ Ax = b$

II) $\exists p \in \mathbb{R}^m : \ p^\top A = 0 \ \& \ p^\top b \neq 0$

Lemma 5.1 (Farkas Lemma)

Exactly one of the following statements is true:

I) $\exists x \in \mathbb{R}^n : \ Ax = b, \ x \geq 0$

II) $\exists p \in \mathbb{R}^m : \ p^\top A \leq 0$ and $\ p^\top b > 0$

Proof. (I) true $\Longrightarrow$ (II) false. Equivalently, (II) true $\Longrightarrow$ (I) false.
Suppose (II) is true. For such p,

$$p^\top Ax \leq 0, \ \forall x \geq 0$$

Also, $p^\top b > 0$. Then $p^\top \underbrace{Ax}_{\neq b} \leq 0$.
(I) false $\implies$ (II) true. Equivalently, (II) false $\implies$ (I) true.
Suppose (I) is false. Let

$$S = \{y \in \mathbb{R}^m : y = Ax, x \geqslant 0\}$$
$$S \neq \Phi \ (0 \in S)$$

S is a polyhedral cone. Then, it is closed, convex.

By separating hyperplane theorem(3.2),

$$\exists 0 \neq p \in \mathbb{R}^m: \ \forall y \in S, \ p^\top y < p^\top b = \beta(b \notin S)$$

Since $0 \in S$, we have $\beta = p^\top b > 0$. All we need to show is $p^\top A \leq 0$.

$$p^\top y = p^\top Ax < \beta, \ \forall x \geq 0$$

If $p^\top A^j > 0$, then $p^\top A^j \to +\infty$ as $x^j \to +\infty$. Combination with $p^\top Ax < \beta$, contradiction! Therefore

$$p^\top A^j \leq 0, \ \forall j = 1, \cdots, n$$

$\therefore p^\top A \leq 0$. OED. $\square$

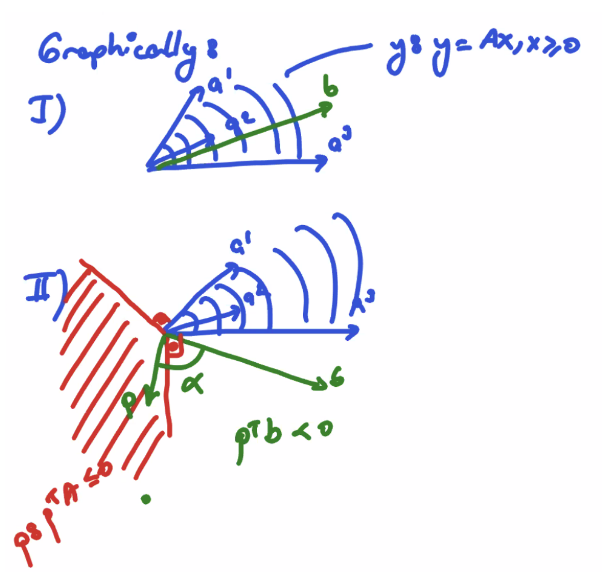


Figure 5.1: Graphically Farkas Lemma 5.1

Lemma 5.2 (Farkas in Canonical Form)

Exactly one of the following statements is true:

I) $\exists x \in \mathbb{R}^n : \ Ax \leq b$

II) $\exists p \in \mathbb{R}^m : \ p^\top A = 0, \ p \geq 0, \ p^\top b < 0$

Proof. (I) $Ax \leq b \iff \begin{array}{l} Ax^+ - Ax^- + J := b \ (\mathbf{p}) \\ x^+, x^-, s \geq 0 \end{array}$

$$\begin{array}{r} p^\top A = 0 \\ p \geq 0 \\ p^\top b < 0 \end{array} \iff \begin{array}{l} p^\top [A| - A|I] \geqslant 0 \\ p^\top b < 0 \end{array}$$

□

5.2 Blending Model

Example 5.2

Determine on optimal mix for animal feed.

Ingredients	Calcium(%)	Protein(%)	Fiber(%)	Cost($/lb)
Limestone	38	0	0	10
Corn	0.1	9	2	30
Soybean	0.2	50	8	90

Constraints:

1) Calcium content between 0.8% and 1.2%
2) Protein content at least 22%
3) Fiber content at most 5%

Goal: Find the least cost mix satisfying the constraints.

Solution. Variables:

- x_1: Proportion of limestone in the mix.
- x_2: Proportion of corn in the mix.
- x_3: Proportion of soybean in the mix.

Constraints:

$$\left.\begin{array}{l} x_1 + x_2 + x_2 = 1 \\ x_{41} x_2, x_2 \geqslant 0 \end{array}\right\} \text{Proportion Definition}$$

$0.38x_1 + 0.001x_2 + 0.002x_0 \geqslant 0.008$
$0.38x_1 + 0.001x_2 + 0.002x_0 \leqslant 0.012$ ■

5.3 Linear Programming Duality

Example 5.3

$$\begin{aligned} z = \min\ & x_1 + 2x_2 + 4x_3 \\ s.t.\ & \left.\begin{aligned} x_1 + x_2 + 2x_3 &= 5 \\ 2x_1 + x_2 + 3x_3 &= 8 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}\right\} F \end{aligned}$$

Any feasible solution provides an upper bound on z

$\bar{x} = (2, 1, 1)$ with objective value to be 8

2) $\bar{x} = (3, 2, 0)$ with objective value to be 7

Get a lower bound from constraints, $z = 7$

In general, if we find row multipliers $y \in \mathbb{R}^m$ subject to

$$y^\top a^j \leq cj,\ \forall j = 1, \cdots, n$$

then,

$$c^\top x \geqslant y^\top Ax = y^\top b$$

$$\begin{aligned} z = \min\ & c^\top x \\ s.t.\ & Ax = b \\ & x \geq 0 \text{ (PRIMAL LP), (P)} \end{aligned}$$

To find the best lower bound, solve an optimization problem as the following,

$$\begin{aligned} w = \max\ & y^\top b \\ s.t.\ & y^\top A \leq c^\top \text{ (DUAL LP), (D)} \end{aligned}$$

Theorem 5.2 (Weak Duality)

$$w \leq z$$

Proof. Let x be a feasible solution for (P). Let y be a feasible solution for (D). Then

$$y^\top b = y^\top Ax \leqslant c^\top x$$

□

In particular, the above equation holds for optimal x and optimal y as well! QED

Proposition 5.1
The dual of the dual problem (D) is the primal problem (P).

Proof.

$$
\begin{aligned}
& \max\ y^\top b \\
& \quad s.t.\ y^\top A \leq c^\top \\
\iff & -\min\ -b^\top y \\
& \quad s.t.\ y^\top A + s^\top = c^\top \\
& \qquad s \geq 0 \\
\iff & -\min\ -b^\top y^+ + b^\top y^- \\
& \quad s.t.\ A^\top y^+ - A^\top y^- + Is = c \text{ (u)} \\
& \qquad y^+, y^-, s \geq 0\ (\bar{D})
\end{aligned}
$$

Taking the dual of $(\bar{D})$:

$$
\begin{aligned}
& -\max\ u^\top c \\
& \quad s.t.\ u^\top \left[A^\top | -A^\top | I\right] \leq \left[-b^\top | b^\top | 0\right] \\
\iff & -\max\ u^\top c \\
& \quad s.t.\ Au \leq -b \\
& \qquad -Au \leq b \\
& \qquad u = 0
\end{aligned}
$$

Let $x = -u$,

$$
\begin{aligned}
z = \min\ & c^\top x \\
& s.t.\ Ax = b \\
& \quad x \geq 0 \text{ (P)}
\end{aligned}
$$

□

Corollary 5.1
1. If (P) is unbounded, i.e. $z = -\infty$, then (D) is infeasible.
2. If (D) is unbounded, i.e. $w = +\infty$, then (P) is infeasible.

Dual of (P) in canonical form:

$$\begin{aligned} z = \min\ & c^\top Ax \\ s.t.\ & Ax \geq b \end{aligned}$$

To get a lower bound we require:

$$\begin{aligned} y^\top A &= c^\top \\ y &\geq 0 \end{aligned}$$

To find the best lower bound value:

$$\begin{aligned} \max\ & y^\top b \\ s.t.\ & y^\top A = c^\top \\ & y \geq 0 \end{aligned}$$

In general, the following applies for stating the dual problem:

Primal	Dual
$\min c^\top x$	$\max y^\top b$
$s.t.\ a_i x \leq b_i$	$s.t.\ y_i \geqslant 0$
$a_i x \geq b_i$	$y_i \leq 0$
$a_i x = b_i$	y_i free
$x_j \geq 0$	$y^\top a^j \leq c_j$
$x_j \leq 0$	$y^\top a^j \geq c_j$
x_j free	$y^\top a^j = c_j$

6 Week 6 Lecture

6.1 LP Duality

Example 6.1

Primal

$$\begin{aligned} \min\ & 2x_1 - 3x_2 + 5x_3 \\ s.t.\ & x_1 + x_2 \geq 0 \\ & 3x_1 - x_2 - 2x_3 \leq 5 \\ & 5x_2 + x_3 = 3 \\ & x_1 \geq 0 \\ & x_2 \leq 0 \\ & x_3 \text{ free} \end{aligned}$$

Dual

$$\begin{aligned} \max\ & 5y_2 + 3y_3 \\ s.t.\ & y_1 + 3y_2 \geq 2 \\ & y_1 - y_2 + 5y_3 \geq -3 \\ & -2y_2 + y_3 = 5 \\ & y_1 \geq 0 \\ & y_2 \leq 0 \\ & y_3 \text{ free} \end{aligned}$$

Primal

$$\begin{aligned} z = \min\ & c^\top x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual

$$\begin{aligned} \max\ & y^\top b \\ \text{s.t. } & y \geq 0 \end{aligned}$$

Theorem 6.1 (Strong Duality Theorem)
If either (P) or (D) is feasible, then $w = z$.

Proof. Only need to show $w \geq z$. WLOG, suppose (P) is feasible. If (P) is unbounded, then (D) is infeasible.

Suppose (P) has an optimal solution x^*, i.e., $A^x* = b$, $x^* \geq 0$, $c^\top x^* = z$

Claim: $\exists y \in \mathbb{R}^n : \ y^\top A = c^\top$ and $y^\top b \geq z$.

$$y^\top [A \mid -b] \leqslant \left[c^\top \mid -z\right] \text{ feasible}$$

$$\Updownarrow$$

$$[A \mid -b] \begin{bmatrix} x \\ \lambda \end{bmatrix} = 0, x, \lambda \geq 0, c^\top x = z\lambda < 0 \text{ infeasible}$$

$$\Updownarrow$$

$$\begin{array}{rcl} Ax & = & \lambda b \\ c^\top x & < & z\lambda \\ x & \geq & 0 \\ \lambda & \geq & 0 \end{array} \quad \text{infeasible}$$

□

		Primal		
		Optimal	Unbounded	Infeasible
Dual	Optimal	3	7	7
	Unbounded	7	7	3
	Infeasible	7	3	3

Table 1: The primal/dual pairs

Example 6.2 (Infeasible/Feasible)

Primal[a]

$$\min\ x_1 + 2x_2$$
$$s.t.\ x_1 + x_2 = 1$$
$$2x_1 + 2x_2 = 3$$

Dual

$$\min\ y_1 + 3y_2$$
$$s.t.\ y_1 + 2y_2 = 1$$
$$y_1 + 2y_2 = 2$$

[a] See more solutions at https://www.epfl.ch/labs/disopt/wp-content/uploads/2018/09/solution1.pdf

6.2 When is an LP unbounded?

Definition 6.1. *$C \subseteq \mathbb{R}^n$ is called a cone if for $\forall x \in C$ and $\forall \lambda \geq 0$, $\lambda x \in C$*

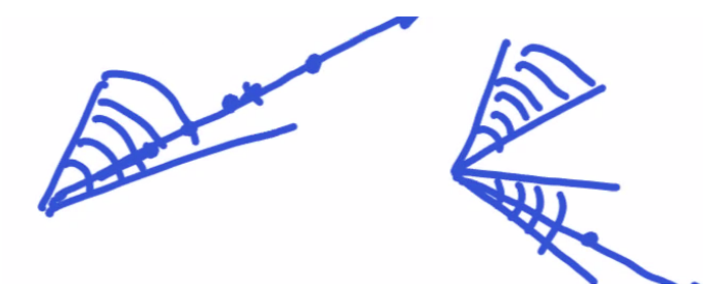


Figure 6.1: Cones

Definition 6.2. *Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a nonempty polyhedron.*

(a) The recession cone of P is defined as

$$P_0 = \{d \in \mathbb{R}^n : Ad \leq 0\}$$

(b) Any $d \in P_0 \backslash \{0\}$ is called a ray of P.

Definition 6.3.

(a) A nonzero element x of a polyhedral cone $C \subset \mathbb{R}^n$ is called an extreme ray if there are $n-1$ linearly independent constraints that are active at x.

(b) An extreme ray of the recession cone associated with a nonempty polyhedron P is also called an extreme ray of P.

Theorem 6.2

Let (P) $\min\{c^\top x : \ x \in S\}$, where S is a nonempty polyhedron. (P) is unbounded if and only if there exists $d \in S$, such that $c^\top d < 0$.

Proof. ($\Longleftarrow$) Let $x \in S$ and $d \in S$ *s.t.* $c^\top d < 0$. Then,

$$y = x + \mu d \in S, \ \forall \mu \geq 0$$

and $c^\top y = c^\top (x + \mu d) \to -\infty$ as $\mu \to \infty$

($\Longrightarrow$) (P) is unbounded.

$\therefore$

Primal	Dual
$\min\ c^\top x$ $s.t.\ Ax \geq b$ Unbounded	$\max\ w^\top b$ $s.t.\ w^\top A = c^\top$ $w \geq 0$ infeasible
$\Downarrow$	$\Downarrow$
$\min\ c^\top d$ $s.t.\ Ad \geq 0$ unbounded as $d = 0$ is feasible.	$\max\ w^\top \cdot 0$ $s.t.\ w^\top A = c^\top$ $w \geq 0$ infeasible

$\exists d : \ Ad \geq 0, \ c^\top d < 0$ $\square$

6.3 Complementary Slackness

For x feasible for (P) and (y, s) feasible for (D), the duality gap is

$$\begin{aligned} c'x - y'b &= c'x - y'Ax \\ &= (c' - y'A)x \\ &= s^\top x = x^\top s \end{aligned}$$

Primal	Dual
$\min\ c^\top x$ $s.t.\ Ax = b$ $x \geq 0$	$\max\ y^\top b$ $s.t.\ y^\top A + s^\top = b$ $s \geq 0$

Theorem 6.3 (Complementary Slackness)
For x, (y, s) feasible for (P) and (D) respectively, the following are equivalent:

1) x is optimal for (P), (y,s) is optimal for (D).
2) $x^\top s = 0$
3) $x_j \cdot s_j = 0,\ \forall j = 1, \cdots, n$ (complementary slackness condition)

Proof. 1) holds: $c^\top x = y^\top b \Leftrightarrow x^\top s = 0$
3) $\Rightarrow$ 2) immediate
2) $\Rightarrow$ 3) because $x \geq 0,\ s \geq 0$ □

In canonical form,

Primal	Dual
$\min\ c^\top x$ $s.t.\ Ax - p = b$ $p \geq 0$	$\max\ y^\top b$ $s.t.\ y^\top A = c^\top$ $y \geq 0$

Duality gap:

$$
\begin{aligned}
c^\top x - y^\top b &= y^\top Ax - y^\top b \\
&= y^\top (Ax - b) \\
&= y^\top p
\end{aligned}
$$

$$
y_i \cdot p_i = 0,\ \forall i = 1, \cdots, m \tag{6.1}
$$

In general, complementary slackness conditions can be stated as follows:

1. $x_j(c_j - y^\top a^j) = 0,\ \forall j = 1, \cdots, n$
2. $y_i(a_i x - b_i) = 0,\ \forall j = 1, \cdots, m$

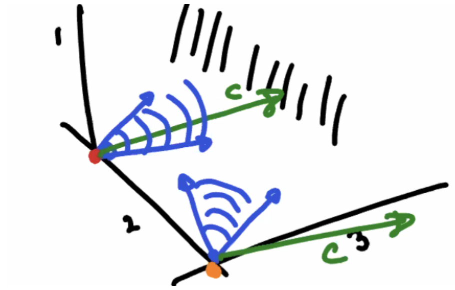


Figure 6.2: Geometric interpretation of 6.1: At an optimal solution x, the objective vector c is written as a conic combination of the constraints active at x

6.4 Extreme Points and Basic Feasible Solutions of Polyhedra

Definition 6.4. *Constraint $a^\top x \leq b$ is active (or binding) at $\bar{x}$ if $a^\top \bar{x} = b$.*

Definition 6.5. *Let $P = \{x \in \mathbb{R}^n : \ Ax \leq b\}$ be a nonempty polyhedron. $x \in \mathbb{R}^n$ is a basic solution for P if there are n linearly independent active constraints at $\bar{x}$.*

Definition 6.6. *If $\bar{x}$ is basic and feasible, then it is called a basic feasible solution (bfs).*

Definition 6.7. *$\bar{x}$ is degenerate if there are more than n constraints active at $\bar{x}$.*

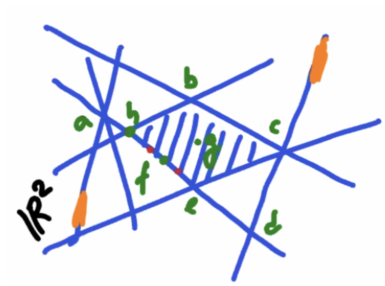


Figure 6.3: In the figure, there are basic solutions $a, \ b, \ c, \ d, \ e, \ h$, where $b, \ c, \ e, \ h$ are both basic feasible solutions and extreme solutions, $a, \ c$ are degenerate basic solutions

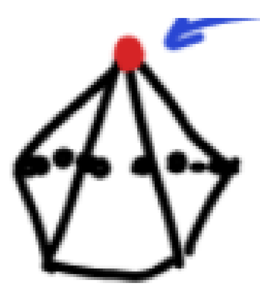


Figure 6.4: In the $\mathbb{R}^3$ space, the red point in the above figure is a degenerate point, but there are no redundant constraints

Theorem 6.4

Let $P = \{x \in \mathbb{R}^n : \ Ax \le b\}$. $x \in P$ is an extreme point of P if and only if x is a b.f.s. of P.

Proof. Suppose $x \in P$ is a b.f.s. of P. Let $(A^=, b^=)$ be the set of active constraints. Let $\tilde{A}$ be a nonsingular $n \times n$ submatrix of $A^=$. For contradiction, suppose x is not extreme point [a], i.e.,

$$x = \lambda x^1 + (1-\lambda)x^2, \text{ distinct } x^1, x^2 \in P \text{ and } 0 < \lambda < 1$$

$$\tilde{b} = \tilde{A}x = \lambda \underbrace{\tilde{A}x^1}_{\ge \tilde{b}} + (1-\lambda)\underbrace{\tilde{A}x^2}_{\ge \tilde{b}}$$

$\therefore \tilde{b} = \tilde{A}x^1 = \tilde{A}x^2$

$\therefore x = x^1 = x^2$!

Contradiction with our assumption about x^1, x^2!

$(\Longrightarrow)$ x extreme point implies that x b.f.s. or equivalent. x not b.f.s. implies that x not extreme point.

Suppose x is not a b.f.s implies that

$$\begin{aligned} &\text{rank}(A^=) < n \\ \Leftrightarrow\ &\dim(\text{Null}(A^=)) > 0 \\ \Leftrightarrow\ &\exists y \ne 0 : A^= y = 0 \end{aligned}$$

Then $A^=(x + \varepsilon y) = b$ for any $\varepsilon \ne 0$

For sufficiently small $\varepsilon > 0$, there is

$$A^<(x \mp \varepsilon y) = b$$

Then,

$$\begin{aligned} &(x \pm \varepsilon y) \in P \\ &x = \frac{1}{2}(x+\varepsilon) + \frac{1}{2}(x-\varepsilon) \end{aligned}$$

Thus x is not an extreme point! □

[a] A point p of a convex set S is an extreme point if each line segment that lies completely in S and contains p has p as an endpoint. An extreme point is also called a corner point.

BFS in standard form:

$$Ax = b, A \text{ is } m \times n, x \in \mathbb{R}^n, x \ge 0, m \le n$$

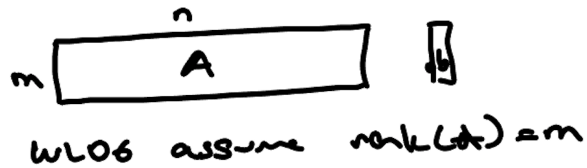


Figure 6.5: BFS

Proposition 6.1

For

$$P = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}$$

where rank $A = m$ (A is $m \times n$), $\bar{x}$ is a b.f.s. if and only if

1) $A\bar{x} = b$

2) There exists m linearly independent columns of A,

$$a^{B(1)}, a^{B(2)}, \cdots, a^{B(m)}$$

3) $\bar{x}_j = 0,\ \forall j \notin \{B(1), B(2), \cdots, B(m)\}$

4) $\bar{x}_B = B^{-1}b \geq 0$ (needed for feasibility)

Theorem 6.5

Let $P = \{x \mid Ax = b, x \geq 0\}$ be a nonempty polyhedron, where A is a matrix of dimensions $m \times n$. Suppose that rank(A) $= k < m$ and that the rows $a'_{i_1}, \ldots, a'_{i_k}$ are linearly independent. Consider the polyhedron

$$Q = \left\{x \mid a'_{i_1}x = b_{i_1}, \ldots, a'_{i_k}x = b_{i_k}, x \geq 0\right\}$$

Then $Q = P$

Theorem 6.6

Consider the linear programming problem of minimizing $\mathbf{c}^\top \mathbf{x}$ over a polyhedron P. Suppose that P has at least one extreme point. Then, either the optimal cost is equal to $-\infty$, or there exists an extreme point which is optimal.

6.5 Simplex Method

With an $m \times n$ matrix A,

$$\begin{aligned} \min\ & c^\top x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Let $A = [B \mid N]$, B is a nonsingular $m \times m$ matrix.

$$\begin{aligned} \min\ & c_B^\top x_B + c_N^\top x_N \\ \text{s.t. } & Bx_B + Nx_N = b \\ & x_B, x_N \geq 0 \end{aligned}$$

Equivalently,

$$\begin{aligned} \min\ & c_B^\top B^{-1} b + \left(c_N^\top - c_B^\top B^{-1} N \right) x_N \\ \text{s.t. } & x_B = B^{-1} b - B^{-1} N x_N \\ & x_B, x_N \geq 0 \end{aligned}$$

Recall (x_B, x_N) is b.f.s. if and only if

$$x_B = B^{-1} b \geq 0,\ \ x_N = 0$$

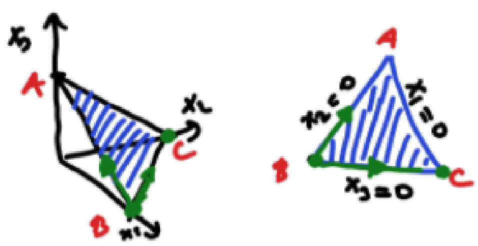


Figure 6.6: Geometry in standard form, where $x_1 + x_2 + x_3 = 1$ and $x_1, x_2, x_3 \geq 0$

At a b.f.s. simplex algorithm takes a basic direction d for a nonbasic variable x_j subject to

$$\begin{aligned} Ad &= 0, d_j = 1, d_i = 0 \quad \forall i \text{ nonbasic } \neq j \\ Ad &= Bd_B + Nd_N \\ &= Bd_B + a^j \ \Rightarrow\ d_B = -B^{-1} a^j \end{aligned}$$

Thus,

$$x_B = B^{-1} b - B^{-1} N x_n$$

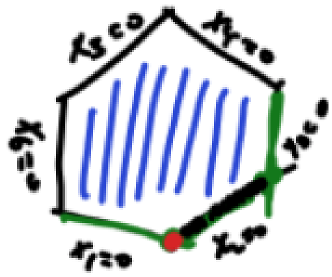


Figure 6.7: In the above figure, $x \in \mathbb{R}^6$. There are 2 nonbasic variables and 4 basic variables.

7 Week 7 Lecture

7.1 Simplex Algorithm

0. Let x be a b.f.s.
1. Find a basic direction d at x such that $c^\top d < 0$
2. Move along $\hat{x} = x + \theta d$ until $\hat{x}$ is bfs. If $\hat{x}$ never becomes bfs, then LP is unbounded
3. Let $x \leftarrow \hat{x}$, got 1

Example 7.1

$$\begin{aligned} &\min 2x_1 \\ &s.t.\ x_1 + x_2 + x_3 + x_4 = 2 \\ &\qquad 2x_1 + 3x_3 + 4x_4 = 2 \\ &\qquad x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Let $B(1) = 1$, $B(2) = 2$. With $B = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$ and $N = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}$,

$$x_0 = B^{-1}b = \begin{bmatrix} 0 & 1/2 \\ 1 & -1/2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$x_3 = x_4 = 0$ (nonbasic)

$$\begin{aligned} \bar{c}_N^\top &= c_N^\top - c_B^\top B^{-1} N \\ (\bar{c}_3, \bar{c}_4) &= (0,0) - (2,0) \begin{bmatrix} 0 & 1/2 \\ 1 & -1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} \\ &= (-3, -4) < 0 \end{aligned}$$

Pick x_3 as "entering variable" since $\bar{c}_3 < 0$,

$$\begin{aligned} \bar{x}_B &= B^{-1}b - \theta B^{-1}a^3 \\ u &= \begin{bmatrix} 0 & v_2 \\ 1 & -y_2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \theta \begin{bmatrix} 3/2 \\ -1/2 \end{bmatrix} \end{aligned}$$

At $\hat{\theta} = 2/3, \ \hat{x_1} = 0$ leaves the basis.

$$x_0 = \begin{bmatrix} x_3 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 4/3 \end{bmatrix}$$

x_1, x_4 are nonbasic.

$$B = \begin{bmatrix} 1 & 1 \\ 3 & 0 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$$

7.2 Single Iteration of the Simplex Algorithm

0. Let $a^{B(1)}, a^{B(2)}, \cdots, a^{B(\cdots)}$ be the basic columns for B.

$$x_B = B^{-1}b \geq 0, \ x_N = 0$$

1. If $\bar{c}_j = c_j - c_B^\top B^{-1}a^j \geq 0$ for all nonbasic j, (x_b, x_N) is optimal.
2. Compute $u = B^{-1}a^j$, $\bar{b} = B^{-1}b$

$$\hat{\theta} = \min_{i=1,\cdots,m} \left\{ \frac{\bar{b}_i}{u_i} : \ u_i > 0 \right\}$$

   If all $u_i \leq 0$, then is unbounded.

3. Suppose $\hat{\theta} = \frac{\hat{b}_\ell}{u_\ell}$, then variable $B(\ell)$ leaves the basis. $B(\ell) \leftarrow j$.

$$\begin{aligned} x_{B(\ell)} &= \hat{\theta} \\ x_{B(i)} &= \bar{b}_i - \theta u_i, \ i = 1, \cdots, m.i \neq \ell \end{aligned}$$

Complexity: $O(m^3 + mn)$

7.3 Important Questions on the Simplex Algorithm

Q1 How to find an optimal b.f.s. ?

Q2 Why is (x_B, x_N) optimal when

$$\bar{c}_N^\top = c_N^\top - c_B^\top B^{-1}N \geq 0$$

Q3 Is it always true that $\hat{\theta} > 0$?

Q4 Why is the new matrix $\hat{B}$, obtained from B by swapping basis and nonbasic variables, is nonsingular?

- A1:

  (i) Two-phase method: Without loss of generality, $b \geq 0$

$$\begin{aligned} z_1 = \min \ & \sum_{i=1}^{m} y_i \\ \text{s.t. } & Ax + Iy = b \\ & x, y \geq 0 \end{aligned}$$

  It is feasible if and only if $z_1 = 0$. Continue to solve the following starting with the optimal solution of the above equation:

$$\begin{aligned} \min \ & c^\top x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

  (ii) Big-M Method:

$$\begin{aligned} \min \ & c^\top x + \sum_{i=1}^{m} M y_i \\ \text{s.t. } & Ax + Iy = b \\ & x, y \geq 0 \end{aligned}$$

- A2:

Proposition 7.1

If $\bar{c}_N^\top = c_N^\top - c_B^\top B^{-1} N \geq 0$, then $(\bar{x}_B, \bar{x}_N)$ is optimal for LP.

Proof. Let $\bar{y}^\top = c_B^\top B^{-1}$

$$\begin{aligned} \min \ & c^\top x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

(P):

$$\begin{aligned} \min \ & c_B^\top x_B + c_N^\top x_N \\ \text{s.t. } & B x_B + N x_N = b \\ & x_B, x_N \geq 0 \end{aligned}$$

(D):

$$\begin{aligned} &\max\ y^\top b \\ &\text{s.t. } y^\top B \le c_B^\top \\ &\quad\ \ y^\top N \in C_N^\top \end{aligned}$$

$\bar{y} = c_B^\top B^{-1}$ is feasible for (D). Check the primal, dual objectives:

$$\begin{aligned} c^\top \bar{x} &= c_B^\top \bar{x}_B + c_N^\top \bar{x}_N \\ &= c_B^\top B^{-1} b + 0 \\ y^\top b &= c_B^\top B^{-1} b \end{aligned}$$

Alternatively, observe that $\bar{x}^\top \bar{s} = 0$, where

$$\begin{aligned} &\bar{s}^\top = c^\top - \bar{y}^\top A \\ &\bar{x}_B = B^{-1} b,\ \bar{x}_N = 0 \\ &\bar{x}_B = 0,\ \bar{s}_N \ge 0 \end{aligned}$$

□

- A3: Degeneracy

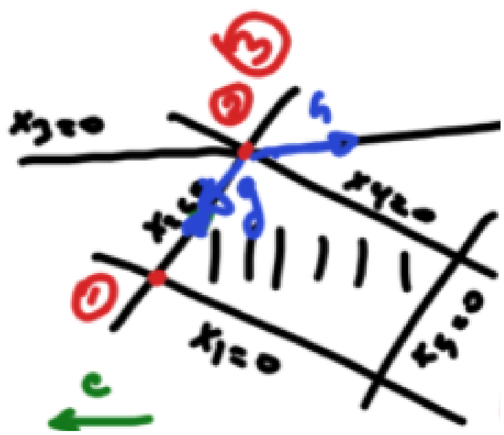


Figure 7.1: In the above plot, for ①, basics are x_3, x_4, x_5 and nonbasics are x_1 and x_2. For ②, basics are x_1, x_4, x_5 and nonbasics are x_2 and x_3. For ③, basics are x_1, x_2, x_5 and nonbasics are x_3 and x_4.

- A4:

Proposition 7.2

Let $\hat{B}$ be the matrix obtained from B after $x_{B(\ell)}$ leaves the basis and nonbasic x_j enters in its place. If B is nonsingular, so is $\hat{B}$.

Proof.

$$\hat{B} = B\,[e_1, e_2, \cdots, u, e_{l+1}, \cdots, e_m]$$

$\hat{B}$ is nonsingular $\Leftrightarrow$ $B^{-1}\hat{B}$ is nonsingular.

$$B^{-1}\hat{B} = [e_1, e_2, \cdots, u, e_{l+1}, \cdots, e_m]$$

which is nonsingular because $u_\ell > 0$ by ratio test! □

7.4 Stalling

Lexicographical Rule to Avoid Cycling:

1. Among nonbasic variables with $\varepsilon_j < 0$ choose the one with the smallest index to enter into the basis.

2. Among basic variables for leaving the basis (multiple minimizers in the ratio test), choose the variable with the smallest index to leave into the basis.

8 Week 8 Lecture

8.1 Sensitivity Analysis

Consider

$$\begin{aligned} &\min\ c^\top x \\ &s.t.\ Ax = b \\ &\quad x \geq 0 \end{aligned}$$

and its dual

$$\begin{aligned} &\max\ y^\top b \\ &s.t.\ y^\top A \leq c^\top \end{aligned}$$

B is an optimal basis [2] if and only if

$$\begin{aligned} &\bar{x}_B = B^{-1}b = \bar{b} \geq 0,\ \bar{x}_N = 0 \\ &\bar{c}_N^\top = c_N^\top - c_B^\top B^{-1} N \geq 0 \end{aligned}$$

8.1.1 Adding a new variable

$$\begin{aligned} &\min\ c^\top x + c_{n+1} x_{n+1} \\ &s.t.\ Ax + a^{n+1} x_{n+1} = b \\ &\quad x \geq 0,\ x_{n+1} \geq 0 \end{aligned}$$

$(x, x_{n+1}) = (\bar{x}, 0)$ is a b.f.s. of the above linear programming problem . It is optimal if

$$\bar{c}_{n+1} = c_{n+1} - c_B^\top B^{-1} a^{n+1} \geq 0$$

Otherwise $\bar{c}_{n+1} < 0$, enter x_{n+1} into the basis and continue with the simplex algorithm till optimality is verified.

8.1.2 Adding a new constraint

Suppose we add the constraint $a_{m+1}^\top x \geq b_{m+1}$. If $(\bar{x}_B, \bar{x}_N)$ satisfies the constraint, it is optimal. Otherwise, introduce a surplus variable.

$$\begin{aligned} &\min\ c^\top x + 0x_{n+1} \\ &s.t.\ Ax + 0x_{n+1} = b \\ &\quad a_{m+1}^\top x - x_{n+1} = b_{m+1} \\ &\quad x \geq 0,\ x_{n+1} \geq 0 \end{aligned}$$

[2] Notes: $\bar{c}_B^\top = c_B^\top - c_B^\top B^{-1} B = 0$

Let $\bar{B} = \begin{bmatrix} B & 0 \\ a^\top & -1 \end{bmatrix}$, $\det(\bar{B}) = -\det(B)$. $\bar{B}$ is nonsingular.

$$(\bar{x}, \bar{x}_{n+1}) = (\bar{x}, \underbrace{a_{m+1}^\top \bar{x}_{n+1} - b_{m+1}}_{<0})$$

Not primal feasible! Let's check dual feasibility:

$$\bar{B}^{-1} = \begin{bmatrix} B^{-1} & \mathbf{0} \\ a^\top B^{-1} & -1 \end{bmatrix}$$

$$(\bar{c}, \bar{c}_{n+1}) = (c^\top, 0) - (c_B^\top, 0) \begin{bmatrix} B^{-1} & \mathbf{0} \\ a^\top B^{-1} & -1 \end{bmatrix} \begin{bmatrix} A & \mathbf{0} \\ a_{m+1} & -1 \end{bmatrix}$$

$\bar{B}$ is a dual, feasible basis!

8.1.3 Changing the cost vector

Suppose c_j is changed to $c_j + \delta$. This has an effect on primal feasibility.

1. If x_j is a nonbasic variable: if

$$\begin{aligned} \bar{c}_j^\top &= (c_j + \delta) - c_B^\top B^{-1} a^j \geq 0 \\ &= \bar{c}_j + \delta \geq 0 \end{aligned}$$

   then $(\bar{x}_B, \bar{x}_N)$ remains optimal. Otherwise enter x_j into the basis and apply the simplex algorithm.

2. If x_j is a basic variable, i.e. $B(\ell) = j$, then we need to check:

$$\begin{aligned} \bar{c}_i &= c_i - \left(c_B^\top + \delta e_\ell^\top\right) B^{-1} a^i \\ &= \bar{c}_i - \delta \underbrace{e_\ell^\top B^{-1} a^i}_{\bar{a}_{\ell_i}} \geq 0 \\ \iff& \bar{c}_i \geq \delta \bar{a}_{\ell_i}, \ \forall i \text{ nonbasic} \end{aligned}$$

   $(\bar{x}_B, \bar{x}_N)$ remains optimal if δ satisfies

$$\max_{i:\ \bar{a}_{\ell_i} < 0} \left\{ \frac{\bar{c}_i}{\bar{a}_{\ell_i}} \right\} \leq \delta \leq \min_{i:\ \bar{a}_{\ell_i} > 0} \left\{ \frac{\bar{c}_i}{\bar{a}_{\ell_i}} \right\}$$

8.1.4 Changing the vector

Suppose b is changed to $b' = b + \delta e_i$. B remains optimal if

$$B^{-1} b' = B^{-1} b + \delta B^{-1} e_i \geq 0$$

If $\delta h_j \geq -\bar{X}_{B(j)},\ \forall j = 1, \cdots, m$, then B remains an optimal basis. This is true if

$$\max_{j:\ h_j>0} \left\{ \frac{-\bar{X}_{B(j)}}{h_j} \right\} \leq \delta \leq \min_{j:\ h_j<0} \left\{ \frac{-\bar{X}_{B(j)}}{h_j} \right\}$$

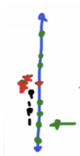
In this range, the optimal value changes as:

$$\begin{aligned} C^\top BB^{-1}(b + \delta e_i) &= \underbrace{C^\top BB^{-1}b}_{z} + \delta \underbrace{C^\top BB^{-1}}_{\bar{y}} e_i \\ &= z + \delta \bar{y}_i \end{aligned}$$

which is the "shadow price of constraint i"

8.2 The Dual Simplex Method

Consider the simplex tableau with a dual feasible, but primal infeasible basis:



	x_1	x_2	x_3	x_4	
0	2	1	0	0	0 $(-z)$
1	2	4	1	9	6 (x_3)
2	-1	-3	0	1	-4 (x_4)

Goals of the dual simplex:

1. Reduce $-z$
2. Maintain dual feasibility, i.e. $\bar{c} \geq 0$

For (1), $-\bar{c}_j \frac{\bar{b}_e}{v_j} \leq 0 \Rightarrow v_j < 0$
For (2), $\bar{c}_i - \bar{c}_j \frac{v_i}{v_j} \leq 0 \Rightarrow v_j \geq 0$
When $v_i < 0$, we need

$$\frac{\bar{c}_i}{|v_i|} \geq \frac{\bar{c}_j}{|v_j|},\ \forall i:\ v_i < 0$$

Entering variable:

$$j \leftarrow \arg\min_{i:\ v_i<0} \left\{ \frac{\bar{c}_i}{|v_i|} \right\}$$

Example 8.1

x_2 enters the basis,

$$\arg\min_{i:\ v_i<0} \left\{ \frac{\bar{c}_i}{|v_i|} \right\} = \arg\min \left\{ \frac{2}{|-1|}, \frac{1}{|-3|} \right\}$$

	x_1	x_2	x_3	x_4	
0	$\frac{5}{3}$	0	0	$\frac{1}{3}$	$\frac{4}{3}$ $(-z)$
1	$\frac{2}{3}$	0	1	$\frac{4}{3}$	$\frac{2}{3}$ (x_2)
2	$\frac{1}{3}$	1	0	$-\frac{1}{3}$	$-\frac{4}{3}$ (x_2)

Primal & dual feasible basic feasible solution $\bar{x} = (0, 4/3, 2/3, 0)$, optimal!

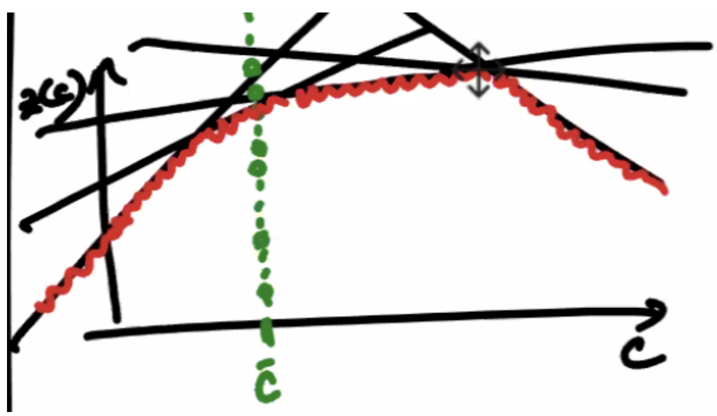
8.3 Global Dependence on the Objective Vectors

$$z(c) = \min \left\{ c^\top x : Ax = b, x \geq 0 \right\}$$

Suppose $\{x | \ Ax = b, x \geq 0\} \neq \phi$.

$$z(c) = \min_{i \in EXT} c^\top x^i$$

where $\left\{x^i\right\}_{i \in EXT}$ is the set of extreme points. z is a pointwise minimum of linear functions. Hence it is a piecewise linear concave function.



8.4 Global Dependence on the RHS Vectors

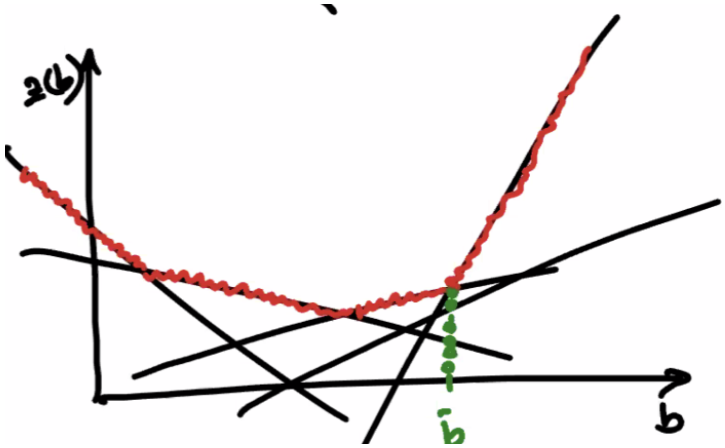
For

$$P(b) = \{x \in \mathbb{R}^n : Ax = b, \ x \geq 0\}$$

Let $S = \{b \in \mathbb{R}^m : \ P(b) \neq \phi\}$,

$$\begin{aligned} z(b) &= \min \left\{ c^\top x : \ x \in P(b) \right\} \\ &= \max \left\{ y^\top b : \ y^\top A \leq c^\top \right\} \\ &= \max_{i \in \ DUALEXT} \left\{ \left(y^i\right)^\top b \right\} \end{aligned}$$

where $\{y^i\}_{i \in DUAL\ EXT}$ is the set of dual extreme points.



z is pointwise maximum of linear functions. Hence it is a piecewise linear convex function.

9 Week 9 Lecture

9.1 Nonlinear Optimization

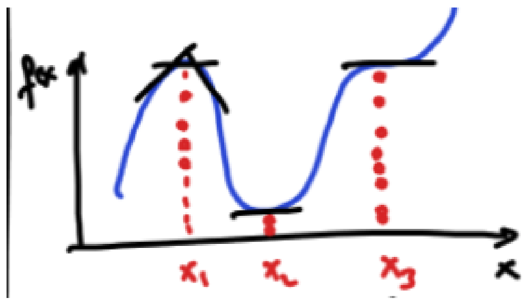


Figure 9.1: A nonlinear function

In the above function,

- x_1 is local maximizer: $\frac{\partial f(x_1)}{\partial x} = 0$, $\frac{\partial^2 f(x_1)}{\partial x} < 0$
- x_2 is local minimizer: $\frac{\partial f(x_2)}{\partial x} = 0$, $\frac{\partial^2 f(x_2)}{\partial x} > 0$
- x_3 is saddle point: $\frac{\partial f(x_3)}{\partial x} = 0$, $\frac{\partial^2 f(x_3)}{\partial x} = 0$

x_1, x_2, x_3 are all stationary points.

9.2 Taylor's Theorem (Calculus)

- First order Taylor expansion for differentiable $f: \mathbb{R} \to \mathbb{R}$ and $\bar{x} \in \mathbb{R}$:

  $$f(x) = f(\bar{x}) + f^1(\bar{x})(x - \bar{x}) + 0(\|x - \bar{x}\|)$$

  where

  $$\frac{0(\|x - \bar{x}\|)}{\|x - \bar{x}\|} \to 0$$

  as $x \to \bar{x}$. So,

  $$f(x) \approx f(\bar{x}) + f^1(\bar{x})(x - \bar{x})$$

  In higher dimension ($f: \mathbb{R}^n \to \mathbb{R}$):

  $$f(x) \approx f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) + 0(\|x - \bar{x}\|)$$

  where

  $$\nabla f(\bar{x}) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Then,

$$f(x) = f(\bar{x}) + \nabla f(z)^\top (x - \bar{x})$$

for some $z \in [x, \bar{x}]$.

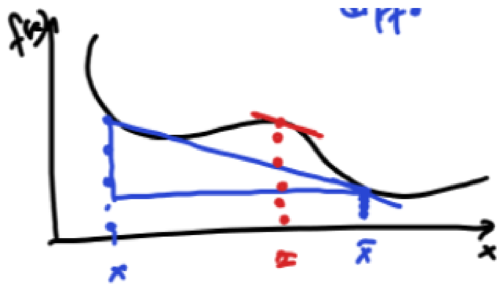


Figure 9.2: Graphical Taylor expansion

- Second order Taylor expansion for a twice differentiable function and $\bar{x} \in \mathbb{R}$:

$$f(x) = f(\bar{x}) + f^1(\bar{x})(x - \bar{x}) + f^2(\bar{x})(x - \bar{x})^2 + 0(\|x - \bar{x}\|)$$

where

$$\frac{0(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|} \to 0$$

as $x \to \bar{x}$. In higher dimensions ($f : \mathbb{R}^n \to \mathbb{R}$):

$$f(x) \approx f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) + (x - \bar{x})^\top \nabla f(\bar{x})(x - \bar{x}) + 0(\|x - \bar{x}\|^2)$$

where

$$\nabla^2 f(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

It is the Hessian matrix of f at $\bar{x}$.

Theorem 9.1 (Young's theorem)
If f is twice differentiable, then $\nabla^2 f(x)$ is symmetric.

Directional derivative:

$$f^1(x; d) = \lim_{\lambda \to 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

where $\|d\| = 1$.
Taylor's Expansion at x:

$$\begin{aligned}
f(x+\lambda d) &= f(x) + \lambda \nabla f(x)^\top d + o(\|\lambda d\|) \\
\frac{f(x+\lambda d) - (x)}{\lambda} &= \frac{\lambda \nabla f(x)^\top}{x} + \frac{o(\|\lambda d\|)}{x} \\
&\underset{\lambda \to 0}{\longrightarrow} \nabla f(x)^\top + 0 \\
f^1(x+\lambda d) &= \nabla f(x)^\top d
\end{aligned}$$

Example 9.1 (Gradients, Hessians, Directional Derivative)

$$\begin{aligned}
f(x) &= 2x_1^3 x_2 + x_2^2 x_3 - x_1 x_3^4 \\
\nabla f(x) &= \begin{bmatrix} 6x_1^2 x_2 - x_3^4 \\ 2x_1^3 + 2x_2 x_3 \\ x_2^2 - 4x_1 x_3^3 \end{bmatrix} \\
\nabla^2 f(x) &= \begin{bmatrix} 12x_1 x_2 & 6x_1^2 & -4x_3^3 \\ 6x_1^2 & 2x_3 & 2x_2 \\ -4x_3^3 & 2x_2 & -12x_1 x_3^2 \end{bmatrix}
\end{aligned}$$

9.3 Necessary Conditions for Local Optimality

Theorem 9.2
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. If $\bar{x}$ is a local minimizer (maximizer) of f, then $\nabla f(\bar{x}) = 0$.

Proof. Let $\bar{x}$ be a local minimizer.

$$f(x) = f(\bar{x} + \lambda d), \quad \forall d \quad 0 \le \lambda < \varepsilon$$

Then, using first order Taylor expansion,

$$0 \le \lambda \nabla f(\bar{x})^\top d + 0(\|x - \bar{x}\|)$$

Divide by λ and let $\lambda \to 0^+$,

$$\begin{aligned}
&0 \le \nabla f(\bar{x})^\top d, \quad \forall d \\
&\Rightarrow 0 = f(\bar{x})^\top d, \quad \forall d \\
&\Rightarrow 0 = \nabla f(x)^\top = 0
\end{aligned}$$

□

Definition 9.1. *$\bar{x}$ is called a* <u>*stationary point*</u> *if* $\nabla f(\bar{x}) = 0$.

Definition 9.2. *A stationary point that is neither local maximizer nor local minimizer is called a* *saddle point*.

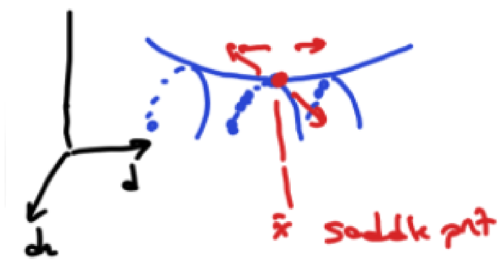


Figure 9.3: Saddle point

Definition 9.3. *Let A be an $n \times n$ symmetric matrix.*

1. *A is positive definite if*

$$x^\top A x > 0 \quad \forall x \in \mathbb{R}^n \backslash \{0\}$$

2. *A is negative definite if*

$$x^\top A x < 0 \quad \forall x \in \mathbb{R}^n \backslash \{0\}$$

3. *A is positive definite if*

$$x^\top A x \geq 0 \quad \forall x \in \mathbb{R}^n$$

4. *A is negative semidefinite if*

$$x^\top A x \neq 0 \quad \forall x \in \mathbb{R}^n$$

5. *Otherwise, A is indefinite.*

Example 9.2

$A = \begin{bmatrix} 4 & -1 \\ -1 & -2 \end{bmatrix}$

$$x^\top A x = 4x_1^2 - 2x_1x_2 - 2x_2^2$$

For $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x^\top A x = 4 > 0$.

For $x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $x^\top A x = -2 < 0$.

$\therefore A$ is indefinite.

9.4 Sufficient Conditions for Convex Optimality

Theorem 9.3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice, continuous differentiable function and let $\bar{x}$ be a stationary point of f.

1. If $\nabla^2 f(\bar{x})$ is positive definite, then $\bar{x}$ is a strict local minimizer.
2. If $\nabla^2 f(\bar{x})$ is negative definite, then $\bar{x}$ is a strict local maximizer.
3. If $\nabla^2 f(\bar{x})$ is indefinite, then $\bar{x}$ is a saddle maximizer.

Proof. Since $\frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}$ are continuous, there exists $\epsilon > 0$ subject to $\nabla^2 f(x)$ positive definite for all $\|x - \bar{x}\| < \epsilon$. Let x be a point in this neighborhood,

$$f(x) = f(x) + \nabla f(x)^\top (x - \bar{x}) + \underbrace{(x - \bar{x})^\top \nabla^2 f(z)(x - \bar{x})}_{>0}$$

for some $z \in [x, \bar{x}]$
$\therefore f(x) > f(\bar{x})$ □

Remark: If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is only positive semidefinite (rather than positive definite), then $\bar{x}$ is not necessarily a local minimizer

Example 9.3

$f(x_1, x_2) = x_1^3 + x_2^3$

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 \\ 3x_2^2 \end{bmatrix}, \; \nabla^2 f(x) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{bmatrix}$$

Then $\nabla^2 f \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is not a local minimizer.

Theorem 9.4

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function.

1. f is convex if and only if
$$f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y$$
2. f is strictly convex if and only if
$$f(y) > f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y, \; x \neq y$$

Proof. ($\Longrightarrow$) Suppose f is convex. Let $x, y \in \mathbb{R}^n$ and $x + \lambda(y - x), \; 0 \leq \lambda \leq 1$. Since f is convex,

$$f(x + \lambda(y - x)) \leq (1 - \lambda) f(x) + \lambda f(y)$$

remark as

$$\frac{f(x+\lambda(y-x))-f(x)}{\lambda} \leq \frac{\lambda\left[f(y)-f(x)\right]}{\lambda}$$

As $\lambda \to 0^+$, $\nabla f(x)^\top(y-x) \leq f(y)-f(x)$
($\Longleftarrow$) Let $x, y \in \mathbb{R}^n$ and $z = \lambda x + (1-\lambda)y$ for $0 \leq \lambda \leq 1$. Need to show:

$$f(z) \leq \lambda f(x) + (1-\lambda) f(y)$$

By assumption, we have

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

Regarding the sum of the above two equations,

$$\lambda f(x) + (1-\lambda) f(y) \geq f(z) + \nabla f(z)^\top \left[\underbrace{\lambda(x-z)+(1-\lambda)(y-z)}_{x+(1-\lambda)y-z=0}\right]$$

$\square$

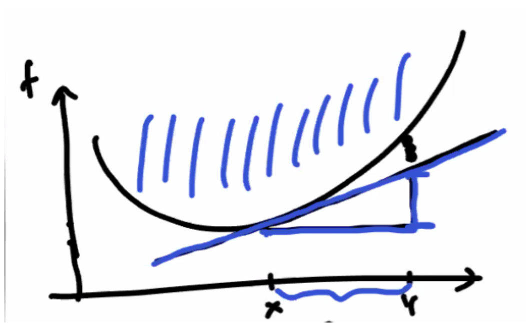


Figure 9.4: Convex function

Definition 9.4. *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuous function. $\gamma \in \mathbb{R}^n$ is called a subgradient of f at x if:*

$$f(y) \geq f(x) + \gamma^\top (y-x) \quad \forall y \in \mathbb{R}^n$$

The set of all subgradients of f at x is called the <u>*subdifferential*</u>.

Reversely, if 0 is a subgradient of f at x, then x is a global minimizer.

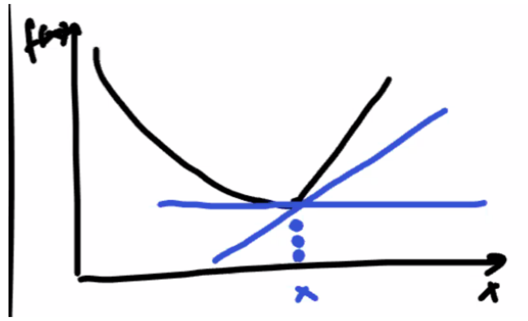


Figure 9.5: Subgradient

Corollary 9.1

If $f: \mathbb{R}^n \to \mathbb{R}$ is a (strictly) convex function and $\bar{x}$ is a stationary point of f, then $\bar{x}$ is a (strict) global minimizer of f.

Proof.

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) \quad \forall x$$

□

9.5 Checking Convexity of Functions

Proposition 9.1

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a twice, continuous differentiable function and $S \subseteq \mathbb{R}^n$ be a nonempty convex set.

1. If $\nabla^2 f(x)$ positive (negative) semidefinite forall $x \in S$, then f is convex (concave) on S.
2. If $\nabla^2 f(x)$ positive (negative) definite forall $x \in S$, then f is strictly convex (concave) on S.
3. If $\nabla^2 f(x)$ indefinite on some $x \in S$, then f is neither convex nor concave on S.

Proof. 1. By second order Taylor's expansion:

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} \underbrace{(y - x)^\top \nabla^2 f(z)(y - x)}_{\geq 0}$$

for some $z \in [x, y]$, $\forall x, y \in S$. Hence,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in S$$

Thus, f is convex on S!

$\square$

10 Week 10 Lecture

10.1 Determining Semidefiniteness

Proposition 10.1

Let A be an $n \times n$ symmetric matrix:

1. A is positive (negative) definite if and only if all eigenvalues of A are positive (negative).
2. A is positive (negative) semidefinite if and only if all eigenvalues of A are nonnegative (nonpositive).
3. A is indefinite if and only if it has at least one positive and at least one negative eigenvalue.

Proposition 10.2

Let A be an $n \times n$ symmetric matrix:

1. A is positive definite if and only if all leading principal minors of A are positive.
2. A is positive semidefinite if and only if all principal minors of A are nonnegative.
3. A is negative definite if and only if kth leading principal minor of A has the sign $(-1)^k$, $\forall k = 1, 2, \cdots, n$.
4. A is negative semidefinite if and only if kth nonzero minor of A has the sign $(-1)^k$, $\forall k = 1, 2, \cdots, n$.
5. A is indefinite otherwise.

Example 10.1

With

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 1 & 4 \end{bmatrix}$$

- 1st principal minors:

$$|1|, |1|, |4|$$

- 2nd principal minors:
  - indices $\{1, 2\}$: $\left|\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\right| = 1$
  - indices $\{1, 3\}$: $\left|\begin{bmatrix} 1 & 0 \\ 2 & 4 \end{bmatrix}\right| = 4$

- indices $\{2,3\}$: $\left|\begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}\right| = 2$

- Leading principal minors:

$$\triangle_1 = |1| \quad \triangle_2 = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = 1 \quad \triangle_3 = |A| = 10$$

Example 10.2

Check whether

$$f(x) = -x_1^2 - x_1x_2 - 2x_2^2$$

is convex/concave.

$$\nabla f(x) = \begin{bmatrix} -2x_1 - x_2 \\ -x_1 - 4x_2 \end{bmatrix}$$
$$\nabla^2 f(x) = \begin{bmatrix} -2 & -1 \\ -1 & -4 \end{bmatrix}$$
$$\triangle_1 = |[-2]| = -2 < 0$$
$$\triangle_2 = 7 > 0$$

$\therefore$ f is strictly concave over $\mathbb{R}^2$.

Example 10.3

Check whether

$$f(x) = x_1^3 + 2x_1x_2 + x_2^2$$

is convex/concave.

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 + 2x_2 \\ 2x_1 + 2x_2 \end{bmatrix}$$
$$\nabla^2 f(x) = \begin{bmatrix} 6x_1 & 2 \\ 2 & 2 \end{bmatrix}$$
$$\triangle_1 = 6x_1$$
$$\triangle_2 = 12x_1 - 4$$

$\triangle_1 > 0$ & $\triangle_2 > 0$ if $x_1 > 1/3$. Hence f is strictly convex over

$$\delta = \left\{x \in \mathbb{R}^2 : x_1 > 1/3\right\}$$

this neither convex nor concave over $\mathbb{R}^2$.

Example 10.4

Find all local minimizers / maximizers and saddle points if

$$f(x) = x_1^2 + x_2^2 + x_3^2 - 4x_1x_2$$

$$\nabla f(x) = \begin{bmatrix} 2x_1 - 4x_2 \\ 2x_2 - 4x_1 \\ 2x_3 \end{bmatrix} = 0 \Rightarrow \bar{x} = 0 \text{ is the only stationary point}$$

$$\nabla^2 f(x) = \begin{bmatrix} 2 & -4 & 0 \\ -4 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\det\left[\nabla^2 f(x) - \lambda I\right] = 0$$

$$\det \begin{bmatrix} 2-\lambda & -4 & 0 \\ -4 & 2-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{bmatrix} = 0$$

$$(2-\lambda)\left[(2-\lambda)^2 - 16\right] = 0$$

Solving for λ:

$$\lambda_1 = 2,\ (2-\lambda) = \pm 4 \Rightarrow \lambda_2 = 6,\ \lambda_3 = -2$$

$\nabla^2 f(0)$ is indefinite! $\bar{x} = 0$ is a saddle point.

10.2 Modeling Exercise

A TV manufacturer has two products: 19” set, 21” set.

	19” set	21” set
MSRP [3]	\$339	\$399
Cost	\$195	\$225

- For each set, any setting price drops by 1 cent for each unit sold.
- Average selling price of 19” set further drops by 0.3 cent for each 21” set sold.
- Average selling price of 21” set further drops by 0.4 cent for each 19” set sold.

How many of each product to produce and sell.
Declare variables:

- x_1: the number of 19” sets sold

[3] manufacturer’s suggested retail price

- x_2: the number of 21" sets sold

Objective: max profit of $p(x_1, x_2)$

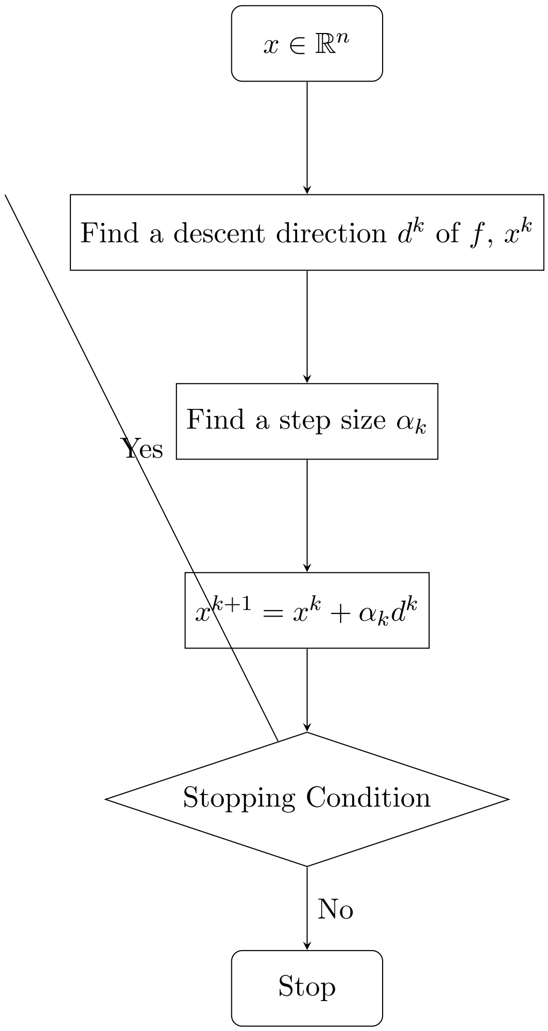
$$\begin{aligned} p(x_1, x_2) &= (339 - 0.01x_1 - 0.003x_2 - 195)\, x_1 + (399 - 0.01x_2 - 0.004x_2 - 225)\, x_2 \\ &= 144x_1 - 0.01x_1^2 - 0.007x_1x_2 + 174x_2 - 0.01x_2^2 \end{aligned}$$

$$\begin{aligned} &\max\ p(x_1, x_2) \\ &\ \text{s.t. } x_1, x_2 \geq 0 \end{aligned}$$

$$\begin{aligned} \nabla p(x) &= \begin{bmatrix} 144 - 0.02x_1 - 0.007x_2 \\ 174 - 0.02x_2 - 0.007x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ x_1 &\approx 4.735 \quad x_2 \approx 7.043 \\ \nabla^2 p(x) &= \begin{bmatrix} -0.02 & -0.007 \\ -0.007 & -0.02 \end{bmatrix} \\ \triangle_1 &= -0.02 < 0 \quad \triangle_2 > 0 \end{aligned}$$

$\therefore$ p is a strictly concave function.
$\therefore$ $\bar{x}$ is the unique maximizer!

10.3 Minimization Algorithm for Unconstrained Nonlinear Linear Programming



1. Descent Direction: For $f: \mathbb{R}^n \to \mathbb{R}$, d is a descent direction at x

$$f^1(x; d) = \nabla f(x)^\top d < 0$$

How about steepest descent direction? With $d \in \mathbb{R}^n$,

$$\min \nabla f(x)^\top d \text{s.t.} \qquad \|d\| = 1$$

Cauchy-Schwartz:

$$\nabla f^\top(x) d = \|\nabla f(x)\| \cdot \|d\| \cdot \cos\alpha$$

10.4 Steepest Descent Method for Minimize function

- Step 0: Start with $x^0 \in \mathbb{R}^n$, pick $\varepsilon > 0$ small.

- Step k:

$$x^{k+1} = x^k - \alpha_k \nabla f\left(x^k\right)$$

where

$$\begin{aligned} \alpha_k = \arg\min\ & f(x^k - \alpha \nabla f(x^k)) \\ \text{s.t. } & \alpha \geq 0 \end{aligned}$$

If $\|\nabla f(x^{k+1})\| < \varepsilon$, stop. Else $k \leftarrow k+1$

Example 10.5

$\min f(x) = 4x_1^2 - 4x_1x_2 + 2x_2^2$

$x^0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$\nabla f(x) = \begin{bmatrix} 8x_1 - 4x_2 \\ 4x_1 + 4x_2 \end{bmatrix}$ $x^1 = x^0 - \alpha_0 \nabla f(x^0)$, where α_0 solves

$$\min f(x^0 - \alpha \nabla f(x^0)) = \min \theta(x)$$

$$\begin{aligned} \theta^1(\alpha) &= \nabla f(x^0 - \alpha \nabla f(x^0))^\top \nabla f(x^0) \\ &= -\nabla f(2 - 4\alpha, 3 - 4\alpha)^\top \begin{pmatrix} 4 \\ 4 \end{pmatrix} \\ &= -16(2 - 4\alpha) = 0 \\ &\Rightarrow \bar{\alpha} = \frac{1}{2} \\ \theta^2(\alpha) &= 64 > 0 \end{aligned}$$

$\therefore \theta$ is convex, $\bar{\alpha} = \frac{1}{2}$ is minimizer. Then, $x^1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$f \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 10,\ f \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 2$

$\nabla f(x^1) = \begin{pmatrix} -4 \\ 4 \end{pmatrix}$

$\nabla f(x^2)^\top \nabla f(x^1) = 0$

$$\begin{aligned}
\theta(\alpha) &= f(0-4\alpha; 1-4\alpha) \\
\theta'(\alpha) &= -16(2-20\alpha) = 0 \\
&\Rightarrow \alpha_1 = \frac{1}{10} \\
x^2 &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \frac{1}{10}\begin{pmatrix} -4 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \end{pmatrix}
\end{aligned}$$

$$f\begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \end{pmatrix} = \frac{2}{5}$$

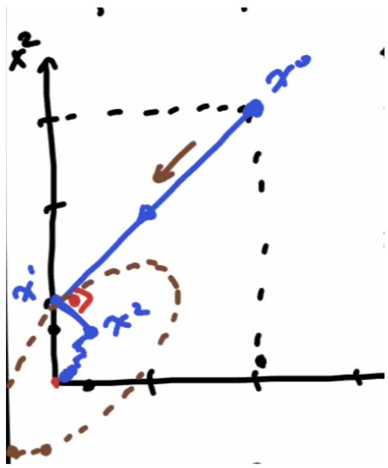


Figure 10.1: Zigzagging

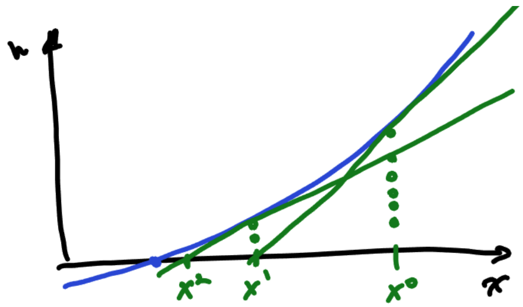


Figure 11.1

11 Week 12 Lecture

11.1 Newton's Method for Solving a System of Nonlinear Equations

Let $H : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Say we want to solve $h(x) = 0$, with Newton's Method:

$$x^{k+1} = x^k - \frac{h(x^k)}{h'(x^k)}$$

The equation of the target line at $(x^k, h(x^k))$ is

$$y = h(x^k) + h'(x^k)(x - x^k)$$

Next iterate is given by

$$(y, x) = (0, x^{k+1});$$
$$-h\left(x^k\right) = h'\left(x^k\right)\left(x^{k-1} - x^k\right)$$

More generally, for a differentiable function

$$g = (g_1, g_2, \cdots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$$

Newton's iterations satisfy

$$-g\left(x^k\right) = \nabla g\left(x^k\right)\left(x^{k-1} - x^k\right)$$

or

$$x^{k+1} = x^k - \left[\nabla g(x^k)\right]^{-1} g(x^k),$$

where ∇g is the Jacobian matrix of g at x. It is an $n \times n$ matrix where the (i, j)th entry is $\frac{\partial g_i(x)}{\partial x_j}$.

11.2 Using Newton's Method to Minimize a Convex function

$\bar{x}$ minimizes convex f if

$$\nabla f(\bar{x}) = 0 \qquad (\star)$$

Apply Newton's Method to solve $(\star)$. At iteration k,

$$-\nabla f\left(x^k\right) = \nabla^2 f\left(x^k\right)\left(x^{k+1} - x^k\right)$$

or

$$x^{k+1} = x^k - \left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right).$$

Example 11.1

Use Newton's Method to minimize

$$f\left(x_1, x_1\right) = x_1^4 + 2x_1^2 x_2^2 + x_2^4.$$

Clearly $\bar{x} = 0$ is the minimizer a: $f(x) \geq 0$. Start $x = (a, a)$:

$$\begin{aligned}
-\nabla f\left(x\right) &= \nabla^2 f\left(x\right)\left(x^{k+1} - x^k\right) \\
\nabla f(x) &= \begin{bmatrix} 4x_1^3 + 4x_1 x_2^2 \\ 4x_2^3 + 4x_1^2 x_2 \end{bmatrix} \\
\nabla^2 f(x) &= \begin{bmatrix} 12x_1^2 + 4x_n^2 & 8x_1 x_2 \\ 8x_1 x_2 & 4x_1^2 + 12x_2^2 \end{bmatrix} \\
-\nabla f\begin{pmatrix} a \\ a \end{pmatrix} &= \nabla^2 f\begin{pmatrix} a \\ a \end{pmatrix}\begin{pmatrix} x_1 - a \\ x_2 - a \end{pmatrix} \\
\begin{bmatrix} 8a^3 \\ 8a^3 \end{bmatrix} &= \begin{bmatrix} 16a^2 & 8a^2 \\ 8a^2 & 16a^2 \end{bmatrix}\begin{bmatrix} x_1 - a \\ x_2 - a \end{bmatrix}
\end{aligned}$$

$\therefore$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2a/3 \\ 2a/3 \end{bmatrix}$$

Thus $x^k = \left[(\frac{2}{3})^k a, (\frac{2}{3})^k a\right]$, $k \geq 0$.

Questions:

1. Is $-\left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right)$ descent direction?
2. Is $f(x^{k+1}) < f(x^k)$?

Answers:

1. *Proof.* If $\nabla^2 f(x^k)$ is positive definite and $\nabla f\left(x^k\right) \neq 0$, then

$$d = -\left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right)$$

is a descent direction:

$$-\nabla f\left(x^k\right)\left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right) < 0$$

□

2. Newton's Method may not converge to any point, let alone local minimizer!

Example 11.2

With $f(x) = \frac{2}{3}|x|^{3/2}$, start at $x^2 = 1$.

$$f'(x) = \begin{cases} x^{1/2} & x \geq 0 \\ (-x)^{1/2} & x < 0 \end{cases}$$
$$f''(x) = \begin{cases} \frac{1}{2}x^{-1/2} & x \geq 0 \\ -\frac{1}{2}x^{-1/2} & x < 0 \end{cases}$$

$$x^1 = x^0 - 2x^{1/2}x^{1/2} = -x^0$$
$$x^2 = x' - \left(-2\left(x^1\right)^{1/2}\left(x'\right)^{(x)}\right) = -x'$$

Then

$$x^k = (-1)^k, \ \forall k$$

Method does not converge!

Advantage:

- If it converges, it converges fast!

Disadvantage:

- It may not converge at all.
- $\nabla^2 f\left(x^k\right)$ may not be positive definite.
- Computational were complex than first-order methods.

11.3 Interior Point Method for Linear Programming