Optimization.

Chapter 1. Introduction.

The basic problem: $\begin{cases} \text{minimize } f(x) \\ \text{subject to } x \in \mathcal{F} \end{cases}$ where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

$f \to$ objective function, $\mathcal{F} \to$ feasible set.

Linear programming: $\begin{cases} \text{minimize } c^T x \\ \text{subject to } Ax \geq b \end{cases}$

Quadratic optimization: $\begin{cases} \text{minimize } \frac{1}{2} x^T H x + c^T x \\ \text{subject to } Ax \geq b. \end{cases}$

Nonlinear optimization: $\begin{cases} \text{minimize } f(x) \\ \text{subject to } g_i(x) \leq 0,\ i = 1, \dots, m. \end{cases}$

Def: $u \in \mathbb{R}$ is upper bound of S if $\forall x \in S$, $x \leq u$. If $\exists u$, S is bounded above.

$l \in \mathbb{R}$ is lower bound of S if $\forall x \in S$, $l \leq x$. If $\exists l$, S is bounded below.

Least upper bound $\to$ supremum. Greatest lower bound $\to$ Infimum.

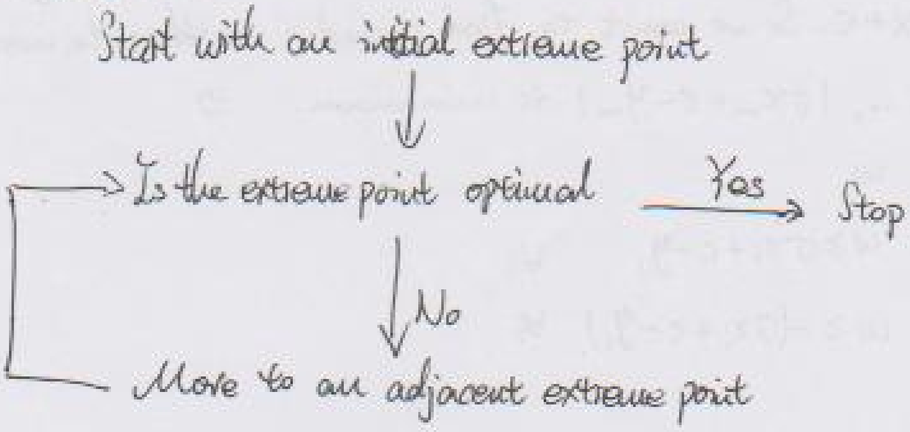
If $\sup S \in S \Rightarrow \max S$, if $\inf S \in S \Rightarrow \min S$.

If S not bounded above, $\sup S = +\infty$. If S not bounded below, $\inf S = -\infty$.

Chapter 2. Introduction to linear programming.

$f(x) = c_1 x_1 + \dots + c_n x_n = c^T x$ where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. $\mathcal{F}: a_{i1} x_1 + \dots + a_{in} x_n \geq b_i$

Simplex Method:



Chapter 3. The standard form.

Minimize $f: \mathcal{F} \to \mathbb{R}$ where $f(x) = c^T x$ $(x \in \mathcal{F})$ and

$$\mathcal{F} = \{x \in \mathbb{R}^n : Ax = b \text{ and } x \geq 0\}. \Rightarrow$$

$$\begin{cases} \text{minimize} & c_1 x_1 + \dots + c_n x_n \\ \text{subject to} & a_{11} x_1 + \dots + a_{1n} x_n = b_1 \\ & \vdots \\ & a_{m1} x_1 + \dots + a_{mn} x_n = b_m \\ & x_1 \geq 0, \dots, x_n \geq 0 \end{cases}$$

Slack variables

Consider:

$$\begin{cases} \text{minimize} & c_1 x_1 + \dots + c_n x_n \\ \text{subject to} & a_{11} x_1 + \dots + a_{1n} x_n \leq b_1 \\ & \vdots \\ & a_{m1} x_1 + \dots + a_{mn} x_n \leq b_m \\ & x_1 \geq 0, \dots, x_n \geq 0 \end{cases} \Rightarrow \begin{cases} c_1 x_1 + \dots + c_n x_n \\ a_{11} x_1 + \dots + a_{1n} x_n + y_1 = b_1 \\ \vdots \\ a_{m1} x_1 + \dots + a_{mn} x_n + y_m = b_m \\ x_1 \geq 0, \dots, x_n \geq 0 \\ y_1 \geq 0, \dots, y_m \geq 0 \end{cases}$$

Slack variables.

Free variables

Consider:

$$\begin{aligned} &\text{minimize } x_1 + 3x_2 + 4x_3 \\ &\text{subject to } x_1 + 2x_2 + x_3 = 5 \\ &\qquad 2x_1 + 3x_2 + x_3 = 6 \\ &\text{and } x_2 \geq 0, x_3 \geq 0 \end{aligned} \Rightarrow \begin{aligned} &\text{minimize } u_1 - v_1 + 3x_2 + 4x_3 \\ &\text{subject to } u_1 - v_1 + 2x_2 + x_3 = 5 \\ &\qquad 2u_1 - 2v_1 + 3x_2 + x_3 = 6 \\ &\text{and } u_1 \geq 0, v_1 \geq 0, x_2 \geq 0, x_3 \geq 0. \end{aligned}$$

Example: (Line fitting). m observation points. $(x_1, y_1), \dots, (x_m, y_m)$.
Line fit: $y = \sigma x + c$. So we want to find σ and c so that the max of $|\sigma x_1 + c - y_1|, \dots, |\sigma x_m + c - y_m|$ is minimum. $\Rightarrow$

$$\begin{cases} \text{minimize} & \omega \\ \text{subject to} & \omega \geq \sigma x_i + c - y_i \quad \forall i \\ & \omega \geq -(\sigma x_i + c - y_i) \quad \forall i. \end{cases}$$

Chapter 4. Basic feasible solutions and extreme points

Standard assumptions: A has rank m, that is, A has independent rows.

Consequences:

(1) $m \leq n$

(2) The column of A span $\mathbb{R}^m$. $\Rightarrow$ At least 1 $x \in \mathbb{R}^n$ st. $Ax = b$.

If non-trivial, then we assume in the continuous: $m < n$.

Let's denote: $A = [a_1 \dots a_n] \in \mathbb{R}^{m \times n}$. Assume that we select m independent columns $a_{\beta_1}, \dots, a_{\beta_2}$ from the n columns of A. These column forms a basis for $\mathbb{R}^m$. We have the following:

(1) The tuple $\beta = (\beta_1, \dots, \beta_m)$ is called the basic index tuple.

(2) Let A_β be the $m \times m$ matrix of the chosen columns, that is:

$$A_\beta = [a_{\beta_1} \dots a_{\beta_m}] \in \mathbb{R}^{m \times m} \Rightarrow \text{Basic matrix}.$$

(3) Let $x_\beta := [x_{\beta_1} \cdots x_{\beta_m}]^T$ be the basic variable vector.

We collect the $l = n - m$ columns of left over of A into A_V. $\Rightarrow$

$$A_V = [a_{v_1}\ a_{v_2} \cdots a_{v_l}] \in \mathbb{R}^{m \times l} \quad \text{and} \quad x_v := [x_{v_1}\ x_{v_2} \cdots x_{v_l}]^T.$$

Non-basic matrix

Non-basic variables vector. $\Rightarrow$

$$A_\beta x_\beta + A_V x_V = \sum_{i=1}^{m} x_{\beta_i} a_{\beta_i} + \sum_{i=1}^{l} x_{v_i} a_{v_i} = \sum_{i=1}^{n} x_i a_i = Ax = b.$$

Def: Suppose β is a basic index tuple:

(1) A basic solution corresponding to β is a solution x to $Ax = b$ such that $A_\beta x_\beta = b$ and $x_V = 0$

(2) A basic feasible solution corresponding to β is a basic solution x such that $x_\beta \geq 0$.

(3) A basic feasible solution x such that all x_β are positive is called a non-degenerate basic feasible solution.

Theorem (Fundamental theorem of linear programming)

Consider the linear programming problem (P).

(1) If there exists a feasible solution, then there exists a basic feasible solution.

(2) If there exists an optimal solution, then there exists an optimal basic feasible solution.

Geometric view of basic feasible solutions

Def: A set $C \subset \mathbb{R}^n$ is called convex if $\forall x, y \in C$ and all $t \in (0,1)$, we have that $(1-t)x + ty \in C$.

Def: Let C be a convex set. A point $x \in C$ is called an extreme point of C if there are no two distinct points $y, z \in C$ such that $x = (1-t)y + tz$ for some $t \in (0,1)$

Theorem: Let $\mathcal{F} = \{x \in \mathbb{R}^n : Ax = b \text{ and } x \geq 0\}$ and let $x \in \mathcal{F}$. Then x is an extreme point of $\mathcal{F}$ iff x is a basic feasible solution of (P).

Corollary: * $\mathcal{F}$ has only finitely many extreme points.

* If the convex set $\mathcal{F}$ is nonempty, it has at least one extreme point.
* If there is an optimal solution to (P), then there is an optimal solution to (P) which is an extreme point of $\mathcal{F}$.

Chapter 5. The simplex method

Suppose that we have chosen a basic index tuple β. We introduce the following:

Notation	element of	definition
$\bar{b}$	$\mathbb{R}^m$	$A_\beta \bar{b} = b$
$\bar{a}_j$ ($j=1 \ldots n$)	$\mathbb{R}^m$	$A_\beta \bar{a}_j = a_j$
y	$\mathbb{R}^m$	$A_\beta^T y = c_\beta$
r	$\mathbb{R}^n$	$r = c - A^T y$
$\bar{z}$	$\mathbb{R}$	$\bar{z} = c_\beta^T \bar{b} = y^T A_\beta \bar{b} = y^T b$

In particular:

$$r_\beta^T = c_\beta^T - y^T A_\beta$$

$$r_\nu^T = c_\nu^T - y^T A_\nu$$

Introduce: $z = c^T x$. $\Rightarrow$

$$z = c^T x = c_\beta^T x_\beta + c_\nu^T x_\nu = y^T A_\beta x_\beta + c_\nu^T x_\nu = y^T (b - A_\nu x_\nu) + c_\nu^T x_\nu$$

$$= y^T b + (c_\nu^T - y^T A_\nu) x_\nu = \bar{z} + r_\nu^T x_\nu \Rightarrow z = c^T x = \bar{z} + \sum_{i=1}^{l} r_{\nu_i} x_{\nu_i}$$

r_{ν_i} — reduced costs for the nonbasic variables. In basic solution, $x_\nu = 0$, $x_\beta = \bar{b}$.

Theorem: Suppose that $\bar{b} \geq 0$ and $r_\nu \geq 0$. Then the basic feasible solution x with $x_\beta = \bar{b}$ and $x_\nu = 0$ is an optimal solution to the linear programming problem (P).

Proof: The linear programming problem (P) can be rewritten as:

$$\begin{aligned} &\text{minimize} \quad \bar{z} + r_\nu^T x_\nu \\ &\text{subject to} \quad A_\beta x_\beta + A_\nu x_\nu = b \\ &\text{and} \quad x_\beta \geq 0 \text{ and } x_\nu \geq 0. \end{aligned}$$

Let $\tilde{x}$ be a feasible solution to the problem. Then we have in particular that $\tilde{x}_\nu \geq 0$. Together with the assumption that $r_\nu \geq 0$, this yields that the cost corresponding to $\tilde{x}$ is at least $\bar{z}$.

$$c^T \tilde{x} = \bar{z} + r_\nu^T \tilde{x}_\nu \geq \bar{z}$$

But $\bar{z}$ is precisely the cost corresponding to the basic feasible solution x with $x_\beta = \bar{b}$ and $x_\nu = 0$. Hence this basic feasible solution is optimal.

Here is the simplex method:

(1) Given is a partition of the variables, represented via the index-tuple β and ν, corresponding to a basic feasible solution x. Calculate the vector $\bar{b}, y, r_\nu$: $A_\beta \bar{b} = b$, $A_\beta^T y = c_\beta$, $r_\nu = c_\nu - A_\nu^T y$.

(Since x is a basic feasible solution, $\bar{b} \geq 0$)

(2) $\underline{1}^\circ$ If $r_\nu \geq 0$, then the algorithm terminates, and the basic feasible solution defined via $x_\beta = \bar{b}$ and $x_\nu = 0$ is an optimal solution to the linear programming problem (P).

$\underline{2}^\circ$ If $\neg[r_\nu \geq 0]$, then choose a q such that r_{ν_q} is the most negative of r_ν and calculate the vector $\bar{a}_{\nu_q}$: $A_\beta \bar{a}_{\nu_q} = a_{\nu_q}$

(3) $\underline{1}^\circ$ If $\bar{a}_{\nu_q} \leq 0$, then the algorithm terminates, and the problem has no optimal solution.

$\underline{2}^\circ$ If $\neg[\bar{a}_{\nu_q} \leq 0]$, then calculate t_{max} and determine a p:

$t_{max} = \min \left\{ \dfrac{\bar{b}_i}{\bar{a}_{i,\nu_q}} : \bar{a}_{i,\nu_q} > 0 \right\}$ and $p \in \{1, \dots, m\}$ is an index for which $\bar{a}_{p,\nu_q} > 0$ and $t_{max} = \dfrac{\bar{b}_q}{\bar{a}_{q,\nu_q}}$. Update β and ν by interchanging ν_q and β_p. Goto (1).

How do we find an initial basic feasible solution?

We assume that in our standard form of the linear programming problem, each component of b is nonnegative. This can be ensured by multiplying some of the equations in $Ax = b$ by -1 if necessary.

The key observation is the following. Consider the linear programming problem:

$$(P') : \begin{cases} \text{minimize} & y_1 + \dots + y_m \\ \text{subject to} & [A \ I_m]\begin{bmatrix} x \\ y \end{bmatrix} = b \\ \text{and} & x \geq 0,\ y \geq 0. \end{cases}$$

Then this problem has an obvious basic feasible solution, namely: $\begin{bmatrix} 0 \\ b \end{bmatrix}$

Theorem: The linear programming problem (P) has a basic feasible solution iff the associated artificial linear programming problem (P') has an optimal feasible solution with objective value 0.

The answer is then the following algorithm:

(1) We first set up the associated artificial linear programming problem (P')

(2) For (P'), we use the simplex method to find an optimal basic feasible solution, starting from the basic feasible solution $\begin{bmatrix} 0 \\ b \end{bmatrix}$.

We then have the following two possible cases:

$\underline{1}^\circ$ There is an optimal solution for (P') with a positive objective value. Then the problem (P) has no basic feasible solution.

$\underline{2}^\circ$ There is an optimal basic feasible solution for (P') with objective. Then it is in form $\begin{bmatrix} x \\ 0 \end{bmatrix}$, here x is the basic feasible solution for (P).

Theorem: If all of the basic feasible solutions are non-degenerate, then the simplex algorithm terminates after a finite number of iterations.

An example:

Let $n=4$, $m=2$, and solve it using the simplex method.

$$A=\begin{bmatrix}1&1&1&0\\2&1&0&1\end{bmatrix},\quad b=\begin{bmatrix}200\\300\end{bmatrix},\quad c=\begin{bmatrix}-400\\-300\\0\\0\end{bmatrix}.$$

We start with x_3, x_4 as the initial basic variables.

First iteration:

As x_3 and x_4 are basic variables, we have $\beta=(3,4)$ and $\nu=(1,2)$. Thus the basic matrix: $A_\beta=[a_3\ a_4]=\begin{bmatrix}1&0\\0&1\end{bmatrix}$ while $A_\nu=[a_1\ a_2]=\begin{bmatrix}1&1\\2&1\end{bmatrix}$

The basic variables takes the value $x_\beta=\bar{b}$ at the initial basic solution, where $\bar{b}$ is determined by $A_\beta\bar{b}=b$, that is:

$$\begin{bmatrix}1&0\\0&1\end{bmatrix}\bar{b}=\begin{bmatrix}200\\300\end{bmatrix}\Rightarrow\bar{b}=\begin{bmatrix}200\\300\end{bmatrix}$$ This gives a feasible basic solution $\bar{b}$.

The solution is $\begin{bmatrix}0\\0\\200\\300\end{bmatrix}$. We now determine the simplex multipliers (components of y) by solving $A_\beta^T y=c_\beta$, that is $\begin{bmatrix}1&0\\0&1\end{bmatrix}y=\begin{bmatrix}0\\0\end{bmatrix}\Rightarrow y=\begin{bmatrix}0\\0\end{bmatrix}$.

The reduced costs for the non basic variables (components of r_ν) are determined by solving $r_\nu=c_\nu-A_\nu^T y$. $\Rightarrow r_\nu=\begin{bmatrix}-400\\-300\end{bmatrix}-\begin{bmatrix}1&2\\1&1\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix}=\begin{bmatrix}-400\\-300\end{bmatrix}$

(2) Since $\neg[r_\nu\geq 0]$, we must now choose q such that r_{ν_q} is the most negative component of r_ν. Since $r_{\nu_1}=r_1=-400<0$ and $r_{\nu_2}=r_2=-300<0$, we choose $q=1$. (Thus x_1 becomes a new basic variable).

We must also determine the vector $\bar{a}_{\nu_q}=\bar{a}_1$ by $A_\beta\bar{a}_1=a_1$, that is:

$$\begin{bmatrix}1&0\\0&1\end{bmatrix}\bar{a}_1=\begin{bmatrix}1\\2\end{bmatrix}\Rightarrow\bar{a}_1=\begin{bmatrix}1\\2\end{bmatrix}$$

(3) Since $\neg[\bar{a}_1\leq 0]$, we must now determine t_{max} and p. (Recall that t_{max} is the largest the new basic variable x_1 can grow.) We have:

$$t_{max}=\min\left\{\frac{\bar{b}_i}{\bar{a}_{i\nu_q}}:\bar{a}_{i,\nu_q}>0\right\}=\min\left\{\frac{200}{1},\frac{300}{2}\right\}=150.$$

while $p\in\{1,\dots,m\}=\{1,2\}$ is an index for which $\bar{a}_{p,\nu_q}>0$ and $t_{max}=\dfrac{\bar{b}_q}{\bar{a}_{p,\nu_q}}$ and so we see that $p=2$. So the basic variable $x_{\beta_p}=x_{\beta_2}=x_4$ leaves the set of basic variables. So $\beta=(3,1)$, $\nu=(2,4)$

Second iteration.

(1) Now $\beta = (3, 1)$ and $\nu = (2, 4)$. Thus the basic matrix:

$$A_\beta = [a_3 \ \ a_1] = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \quad ; \quad A_\nu = [a_2 \ \ a_4] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

$x_\beta = \bar{b}$ where $A_\beta \bar{b} = b \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \bar{b} = \begin{bmatrix} 200 \\ 300 \end{bmatrix} \Rightarrow \bar{b} = \begin{bmatrix} 50 \\ 150 \end{bmatrix} \Rightarrow$

Basic feasible solution $\begin{bmatrix} 150 \\ 0 \\ 50 \\ 0 \end{bmatrix}$.

Simplex multiplier: $A_\beta^T y = c_\beta \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} y = \begin{bmatrix} 0 \\ -400 \end{bmatrix} \Rightarrow y = \begin{bmatrix} 0 \\ -200 \end{bmatrix}$

The reduced costs for the non-basic variables: $r_\nu = c_\nu - A_\nu^T y. \Rightarrow$

$$r_\nu = \begin{bmatrix} -300 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -200 \end{bmatrix} = \begin{bmatrix} -100 \\ 200 \end{bmatrix}.$$

(2) Since $\neg[r_\nu > 0]$, we must now choose q such that r_{ν_q} is the most negative component of r_ν. Since $r_{\nu_2} = r_2 = -100 < 0$, we choose $q = 1$. (Thus x_2 becomes a new basic variables).

Vector $\bar{a}_{\nu_q} = \bar{a}_2$ by $A_\beta \bar{a}_2 = a_2 \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \bar{a}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \bar{a}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

Since $\neg[\bar{a}_2 \leq 0]$, determine t_{max} and p:

$$t_{max} = \min \left\{ \frac{\bar{b}_i}{\bar{a}_{i,\nu_q}} : \bar{a}_{i,\nu_q} > 0 \right\} = \min \left\{ \frac{50}{1/2}, \frac{150}{1/2} \right\} = 100.$$

$\Rightarrow p = 1$. So the basic variables $x_{\beta_1} = x_{\beta_2} = x_3$ leaves the set of basic variables.

Continuous until terminate.

Def: A vector $d \in \mathbb{R}^n$ is called a feasible direction at $x \in \mathcal{F}$ if there exist an $\varepsilon > 0$ such that $x + td \in \mathcal{F}$ $\forall t \in (0, \varepsilon)$.

A vector $d \in \mathbb{R}^n$ is called a descent direction for f at $x \in \mathcal{F}$ if there exists an $\varepsilon > 0$ s.t. $f(x+td) < f(x)$ $\forall t \in (0, \varepsilon)$.

A vector $d \in \mathbb{R}^n$ is called a feasible descent direction for f at $x \in \mathcal{F}$ if d is both a feasible direction and a descent direction.

Theorem: A point $\hat{x} \in \mathcal{F}$ is an optimal solution to the problem (CO) iff there does not exist a feasible descent direction for f at $\hat{x}$.

Chapter 9. Quadratic optimization: no constraints.

Def: A function $f: \mathbb{R}^n \to \mathbb{R}$ is called a quadratic function if

$f(x) = \frac{1}{2} x^T H x + c^T x + c_0$ $(x \in \mathbb{R}^n)$ where $H \in \mathbb{R}^{n \times n}$ is symmetric.

$\Rightarrow$ $f(x+td) = f(x) + t(Hx+c)^T d + \frac{1}{2} t^2 d^T H d$

$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}(x) \ \dots \ \frac{\partial f}{\partial x_n}(x)\right] = (Hx+c)^T$

f is convex iff H is positive semi-definite.

If $(Hx+c)^T d < 0$ then d is a descent direction at x.

If H is not positive semi-definite then there is no lower bound.

If there is no constrain, i.e. $\mathcal{F} \in \mathbb{R}^n$ then optimal solution is $H\hat{x} = -c$.

If H is positive definite then $H\hat{x} = -c$ has a unique solution.

Chapter 10. Quadratic optimization: equality constraints.

Minimize $\frac{1}{2} x^T H x + c^T x + c_0$ subject to $Ax = b$.

To be non-trivial we assume: $b \in \operatorname{ran} A$, and $\ker A \neq \{0\}$

Let k be the dimension of $\ker A$ and let $z_1, \dots, z_k$ form a basis for $\ker A$.

Define the $n \times k$ matrix: $Z = [z_1 \ \dots \ z_k]$. Then $x - \bar{x} \in \ker A$ iff $x - \bar{x} = Zv$ for some $v \in \mathbb{R}^k$: $x \in \mathcal{F}$ iff $x = \bar{x} + Zv$ for some $v \in \mathbb{R}^k$. We can show that $\mathcal{F}$ is always convex.

f is convex iff $Z^T H Z$ is positive semi-definite. (We assume it here).

Optimal solution by the nullspace method: replace $\bar{x} + Z v$

$$f(\bar{x}+Zv) = f(\bar{x}) + (H\bar{x}+c)^T Z v + \frac{1}{2}(Zv)^T H (Zv)$$
$$= f(\bar{x}) + (Z^T(H\bar{x}+c))^T v + \frac{1}{2} v^T (Z^T H Z) v.$$

$\Rightarrow$ Equivalent to the following unconstrained problem:

$$\begin{cases} \text{minimize} & f(\bar{x}) + (Z^T(H\bar{x}+c))^T v + \frac{1}{2} v^T (Z^T H Z) v \\ \text{subject to} & v \in \mathbb{R}^k \end{cases} \Rightarrow$$

$(Z^T H Z)\hat{v} = -Z^T(H\bar{x}+c)$. So $\hat{x}$ is optimal solution iff

$(Z^T H Z)\hat{v} = -Z^T(H\bar{x}+c)$ and $\hat{x} = \bar{x} + Z\hat{v}$

Optimal solution by the Lagrange method.

If f is convex, then $x \in \mathcal{F}$ is an optimal solution iff

$(H\hat{x}+c)^T d = 0 \quad \forall d \in \ker A$. $\Rightarrow$

Theorem: f is convex, $\hat{x} \in \mathbb{R}^n$ is an optimal solution iff $A\hat{x} = b$ and $\exists u \in \mathbb{R}^m$. s.t. $H\hat{x} + c = A^T u$.

Chapter 11: Least-squares problems

$$\begin{cases} \text{minimize} & \frac{1}{2}(Ax-b)^T(Ax-b) \\ \text{subject to} & x \in \mathbb{R}^n \end{cases}$$

To minimize the square of the length of the "error vector" $Ax-b$.

If we want to fit the function $g(t) = s$ on measure points $(t_1, s_1) \dots (t_m, s_m)$:

$$g(t) \approx \alpha_1 \varphi_1(t) + \dots + \alpha_n \varphi_n(t) = \sum_{j=1}^{n} \alpha_j \varphi_j(t) \quad \text{s.t.} \sum_{j=1}^{n} \alpha_j \varphi_j(t_i) = s_i$$

But not always possible. Instead we want to minimize $\sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j \varphi_j(t_i) - s_i\right)$

$$\Rightarrow A = \begin{bmatrix} \varphi_1(t_1) & \cdots & \varphi_n(t_1) \\ & \vdots & \\ \varphi_1(t_m) & \cdots & \varphi_n(t_m) \end{bmatrix}, \quad b = \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}, \quad x = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Let $f(x) = \frac{1}{2}(Ax-b)^T(Ax-b) = \frac{1}{2} x^T H x + c^T x + c_0$ where $H = A^T A$, $c = -A^T b$.

Thus the least-squares problem always has an optimal solution.

From previous theorem, x minimize $f(x)$ when $Hx=-c \Rightarrow A^TAx=A^Tb$.

This system is called the normal equations for the least-squares problem.

If A^TA is not invertible, then we can try to find the shortest solution. $\Rightarrow$

$$\begin{cases} \text{minimize} & \frac{1}{2}\|x\|^2 \\ \text{subject to} & Ax=A\bar{x} \end{cases} \Rightarrow \hat{x}=A^T\hat{u} \text{ is the unique solution.}$$

Pseudo inverse: Suppose A has the singular value decomposition $A=USV^T$ where U has ortogonal columns and V has ortogonal rows. S.t. $U^TU=V^TV=I$. S is diagonal with strictly positive diagonal elements.

$\Rightarrow A^T=VSU^T$. $\Rightarrow A^TA=US^2U^T$, $AA^T=US^2U^T \Rightarrow VS^2V^Tx=VSU^Tb \Rightarrow$ $V^Tx=S^{-1}U^Tb$. If $\bar{x}$ is a solution to $V^Tx=S^{-1}U^Tb$, then the system $AA^Tu=A\bar{x}$ takes the form: $US^2U^Tu=USV^T\bar{x} \Rightarrow SU^Tu=V^T\bar{x}$.

If $\hat{u}$ is a solution to $SU^Tu=V^T\bar{x}$ then the least norm solution shall be:

$$\hat{x}=A^T\hat{u}=VSU^T\hat{u}=VV^T\bar{x}=VS^{-1}U^Tb=A^+b$$

where $A^+=VS^{-1}U^T$ is called the pseudo inverse of A.

Part 3. Nonlinear optimization.

Chapter 12. Introduction.

minimize $f(x)$ — continuely differentiable.

subject to $g_i(x)\le 0$, $h_i(x)$

Chapter 13. The one variable case

Theorem:

1) A nessary condition for $\hat{x}$ to be a local minimizer of f is that $f'(\hat{x})=0$ and $f''(\hat{x})\ge 0$

2) A suffiaient condition for $\hat{x}$ to be a local minimizer of f is that $f'(\hat{x})=0$ and $f''(\hat{x})>0$.

Chapter 14. The multivariable case

$$\begin{cases} \text{minimize } f(x) \\ \text{subject to } x \in \mathbb{R}^{x} \end{cases}$$

Theorem: A necessary condition for $\hat{x}$ to be a local minimizer of f is that $\nabla f(\hat{x}) = 0$ and that $F(\hat{x})$ is positive semi-definite.

Sufficient if $F(\hat{x})$ is positive definite.

Chapter 15. Convexity revisited.

If f is convex and continuously differentiable on $\mathbb{R}^n$, then global minimum is given by $\nabla f(\hat{x}) = 0$.

Chapter 16. Newton's method.

Assume $f: \mathbb{R}^n \to \mathbb{R}^n$, twice continuously differentiable, Hessian $F(x)$ is positive definite.

Given $x^{(k)}$, we calculate $\nabla f(x^{(k)})$ and Hessian $F(x^{(k)})$.

If $\nabla f(x^{(k)}) = 0$, then it is the minimizer. Search terminate.

If $\nabla f(x^{(k)}) \neq 0$, then we do Taylor-approximation, let $d = x - x^{(k)} \in \mathbb{R}^N$:

$$f(x^{(k)} + d) \approx f(x^{(k)}) + \nabla f(x^{(k)}) d + \frac{1}{2} d^T F(x^{(k)}) d$$ this is minimized

by the unique solution: $F(x^{(k)}) d = -(\nabla (f(x^{(k)}))^T \Rightarrow d^{(k)} \Rightarrow$

$\nabla f(x^{(k)}) d^{(k)} = -(d^{(k)})^T F(x^{(k)}) d^{(k)} < 0$. Then the next iteration point should be $x^{(k)} + d^{(k)}$.

In the special case $n = 1$: $x^{(k+1)} = x^{(k)} - \dfrac{f'(x^{(k)})}{f''(x^{(k)})}$

Chapter 17. Nonlinear least square problem: Gauss-Newton

Minimize $f(x) = \frac{1}{2} \Sigma (h_i(x))^2$. Gauss-Newton's method:

Linearlize $h_i(x) \approx h_i(x^{(k)}) + \nabla h_i(x^{(k)})(x - x^{(k)})$. Let $d := x - x^{(k)} \Rightarrow$

$h_i(x^{(k)} + d) \approx h_i(x^{(k)}) + \nabla h_i(x^{(k)}) d$. $h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}$; $\nabla h(x) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x) & \cdots & \frac{\partial h_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}(x) & \cdots & \frac{\partial h_m}{\partial x_n}(x) \end{bmatrix}$

$\Rightarrow$ Objective function: $f(x) = \frac{1}{2} \|h(x)\|^2 \approx \frac{1}{2} \|h(x^{(k)}) + \nabla h(x^{(k)}) d\|$

Let $A^{(k)} := \nabla h(x^{(k)})$, $b^{(k)} := -h(x^{(k)}) \Rightarrow$ minimize $\frac{1}{2} \|A^{(k)} d - b^{(k)}\|^2 \Leftrightarrow$ linear. $\Rightarrow$

$(\nabla h(x^{(k)}))^T \nabla h(x^{(k)}) d = -(\nabla h(x^{(k)}))^T h(x^{(k)}) \Rightarrow x^{(k+1)} = x^{(k)} + d^{(k)}$

Second iteration of the Gauss - Newton's method.

Notice: $\nabla f(x) = h(x)^T \nabla h(x)$; $F(x) = (\nabla h(x))^T \nabla h(x) + \sum_{i=1}^{m} h_i(x) H_i(x)$

where $H_i(x)$ is the Hessian to $h_i(x)$.

If $F(x^{(k)})$ is positive definite, then $F(x^{(k)}) = -\nabla f(x^{(k)})$

Chapter 18. Optimization with constraints: Introduction

$\begin{cases} \text{minimize } f(x) \\ \text{subject to } x \in \mathcal{F}. \end{cases}$ Local optimal solution $\Rightarrow$ not exist feasible descent direction

Chapter 19. Optimality conditions: equality constraints.

Assuming: $\mathcal{F} = \{x \in \mathbb{R}^n : h_i(x) = 0, \ i = 1, \dots, m\}$, f, h_i continuously differentiable.

If rows of $\nabla h(x)$ are linearly independent, then $x \in \mathcal{F}$ is a regular point.

If $\hat{x}$ is regular and local optimal, then there exist a vector $\hat{u} \in \mathbb{R}^m$ s.t.

$$\nabla f(\hat{x}) + \hat{u} \nabla h(\hat{x}) = 0 \quad \Rightarrow \quad \begin{cases} \nabla f(\hat{x})^T + (\nabla h(\hat{x}))^T \hat{u} = 0 \\ h(\hat{x}) = 0. \end{cases}$$

Example quadratic optimization

$$\begin{array}{l} \text{minimize } \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{subject to } Ax = b. \end{array} \quad \Rightarrow \quad \begin{array}{l} f(x) = \frac{1}{2} x^T H x + c^T x + c_0 \\ h(x) = b - Ax \end{array}$$

Chapter 20. Optimality conditions: inequality constraints

$\mathcal{F} = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \ i = 1, \dots, m\}$

Def: $I_a(x) = \{i : g_i(x) = 0\}$ = active index set.

Theorem: suppose $\hat{x} \in \mathcal{F}$ with $I_a(\hat{x}) = \phi$ is a local optimal solution. Then

$\nabla f(\hat{x}) = 0$.

Lemma: if $\hat{x} \in \mathcal{F}$ Local optimal solution, $I_a(\hat{x}) \neq \phi$ Then NOT exist $d \in \mathbb{R}^n$ s.t. $\nabla f(\hat{x}) d < 0$ and $\nabla g_i(\hat{x}) d < 0 \ \forall i \in I_a(\hat{x})$

Def: $x \in \mathcal{F}$ with $I_a(x) \neq \phi$ is called regular point if Not exist $v_i \geq 0$, $i \in I_a(x)$ s.t.: $\sum_{i \in I_a(x)} v_i > 0$ and $\sum_{i \in I_a(x)} v_i \nabla g_i(x) = 0$. If $I_a(x) = \phi$ then it is regular.

Theorem: Suppose $\hat{x} \in \mathcal{F}$ is both a regular point and a local optimal solution, then there exists a vector $\hat{y} \in \mathbb{R}^m$ such that:

(1) $\nabla f(\hat{x}) + \sum_{i=1}^{m} \hat{y}_i \nabla g_i(\hat{x}) = 0$

(2) $g_i(\hat{x}) \leq 0$, $i = 1, \dots, m$

(3) $\hat{y}_i \geq 0$, $i = 1, \dots, m$

(4) $\hat{y}_i g_i(\hat{x}) = 0$, $i = 1, \dots, m$.

(4) $\sum_{i=1}^{m} \hat{y}_i g_i(\hat{x}) = 0$

$\Downarrow$

Suppose $\hat{x} \in \mathcal{F}$ is regular point and a local optimal solution, then there exist $\hat{y} \in \mathbb{R}^m$ s.t. (1) $\nabla f(\hat{x}) + \hat{y}^T \nabla g(\hat{x}) = 0$ (2) $g(\hat{x}) \leq 0$

(3) $\hat{y} \geq 0$ (4) $\hat{y} g(\hat{x}) = 0$.

Chapter 21. Optimality conditions for convex optimization

$$\begin{cases} \text{minimize } f(x) & \rightarrow \text{Convex function} \\ \text{subject to } x \in \mathcal{F} & \rightarrow \text{Convex set} \end{cases}$$

We instead consider: $\begin{cases} \text{minimize } f(x) \\ \text{subject to } g_i(x) \leq 0, \ i = 1, \dots, m. \end{cases} \Rightarrow$

Theorem: If $\hat{x} \in \mathbb{R}$ and $\hat{y} \in \mathbb{R}^m$ satisfy following KKT-condition:

(1) $\nabla f(\hat{x}) + \sum_{i=1}^{m} \hat{y}_i \nabla g_i(\hat{x}) = 0$ (2) $g_i(\hat{x}) \leq 0$, $i = 1, \dots, m$

(3) $\hat{y}_i \geq 0$, $i = 1, \dots, m$ (4) $\hat{y}_i g_i(\hat{x}) = 0$, $i = 1, \dots, m$

Then $\hat{x}$ is a (global) optimal solution to the problem.

Def: The convex problem is said to be regular if $\exists x_0 \in \mathbb{R}^n$ s.t. $g_i(x_0) < 0 \ \forall i$.

If the problem is regular, then all feasible solutions are regular points. $\Rightarrow$ The KKT-conditions are sufficient and necessary.

Chapter 22. Lagrange relaxation

$$(P): \begin{cases} \text{minimize } f(x) \\ \text{subject to } g_i(x) \leq 0, \ i = 1, \dots, m \\ \qquad\qquad x \in X \end{cases}$$

The following problem (PR_y) constitutes a relaxed Lagrange problem with respect to the explicit constraints $g(x) \le 0$ of the original problem (P) given by:

$$(PR_y): \begin{cases} \text{minimize} & f(x) + y^T g(x) \\ \text{subject to} & x \in X \end{cases} \qquad \text{where } y^T g(x) = \sum_{i=1}^{m} y_i g_i(x).$$

Here y_i is Lagrange multiplies.

Def: Function $L: X \times \mathbb{R}^m \to \mathbb{R}$ defined by $L(x,y) = f(x) + y^T g(x)$ is called Lagrangian associated with the problem (P).

Def: A pair $(\hat{x}, \hat{y}) \in X \times \mathbb{R}^m$ is said to ~~be~~ satisfy the global optimality conditions associated with (P) if:

(1) $L(\hat{x}, \hat{y}) = \min_{x \in X} L(x, \hat{y})$ (2) $g(\hat{x}) \le 0$

(3) $\hat{y} \ge 0$ (4) $\hat{y}^T g(\hat{x}) = 0$.

Theorem: If $(\hat{x}, \hat{y}) \in X \times \mathbb{R}^m$ satisfy the global optimality conditions associated with (P), then $\hat{x}$ is an optimal solution to (P).

Dual problem: Let $\mathbb{R}_+^m = \{ y \in \mathbb{R}^m : y \ge 0 \}$. The dual objective function $\varphi: \mathbb{R}_+^m \to \mathbb{R}$ is defined as follows: $\varphi(y) = \min_{x \in X} (f(x) + y^T g(x)) = \min_{x \in X} L(x,y)$.

For each $y \ge 0$, $\varphi(y)$ gives a lower bound for the optimal value of the problem (P). The dual problem to (P) is the problem of finding the best (greatest) lower bound, that is the following:

$$(D) \begin{cases} \text{maximize} & \varphi(y) \\ \text{subject to} & y \ge 0 \end{cases}$$

Theorem: $(\hat{x}, \hat{y}) \in X \times \mathbb{R}^m$ satisfy the global optimality conditions associated with (P) iff:

(1) $\hat{x}$ is an optimal solution to (P) (2) $\hat{y}$ is an optimal solution to (D)

(3) $\varphi(\hat{y}) = f(\hat{x})$.

Theorem: φ is a concave function ($-\varphi$ is a convex function) on $\mathbb{R}_+^m$

Part 4. Some Linear Algebra

Chapter 23. Subspaces.

A subspace S to vector space V if:

S1. $0 \in S$. S2. If $v_1, v_2 \in S$, then $v_1 + v_2 \in S$. S3. If $v \in S$ and $\alpha \in \mathbb{R}$, then $\alpha \cdot v \in S$.

Basis B to S is a subset of S that span $B = S$ and B is linear independent.

If $X \in \mathbb{R}^n$, then the orthogonal complement of $x \Rightarrow X^{\perp} = \{y \in \mathbb{R}^n : y^T x = 0 \ \ \forall x \in X\}$

Theorem: S be a subspace of $\mathbb{R}^n$. Then:

(1) For each $x \in \mathbb{R}^n$, there exist a unique $z \in S$ and a unique $y \in S^{\perp}$ such that $x = z + y$

(2) $(S^{\perp})^{\perp} = S$. (3) If $\dim S = k$, then $\dim(S^{\perp}) = n - k$.

Chapter 24. Four fundamental subspaces.

Let $A \in \mathbb{R}^{m \times n} \Rightarrow A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [C_1 \cdots C_n] = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \Rightarrow$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} C_1^T \\ \vdots \\ C_n^T \end{bmatrix} = [r_1^T \cdots r_m^T]$$

Column space of A = range of A = $\operatorname{ran} A = \{Ax : x \in \mathbb{R}^n\}$

$$= \left\{ \sum_{j=1}^{n} x_j C_j : x_j \in \mathbb{R}, \ j = 1, \cdots, n \right\}$$

Row space of A = Column space of $A^T = \operatorname{ran} A^T = \{A^T y : y \in \mathbb{R}^m\}$

$$= \left\{ \sum_{i=1}^{m} y_i r_i^T : y_i \in \mathbb{R}, \ i = 1, \ldots, m \right\}$$

Kernel of $A = \ker A = \{x \in \mathbb{R}^n : Ax = 0\} = \{x \in \mathbb{R}^n : r_i x = 0, \ i = 1, \ldots m\}$

Left kernel of $A = \ker A^T = \{y \in \mathbb{R}^m : A^T y = 0\} = \{y \in \mathbb{R}^m : y^T A = 0\}$

Theorem: $(\operatorname{ran} A)^{\perp} = \ker A^T$; $\operatorname{ran} A = (\ker A^T)^{\perp}$

$\dim(\operatorname{ran} A) = r$; $\dim(\operatorname{ran} A^T) = r$; $\dim(\ker A) = n - r$; $\dim(\ker A^T) = m - r$

Chapter 25. Bases for fundamental subspaces

Theorem: Let $A \in \mathbb{R}^{m\times n}$ and $A = BC$ where $B \in \mathbb{R}^{m\times r}$ has linearly independent column and $C \in \mathbb{R}^{r\times n}$ has linearly independent rows. Then A and A^T both have ranges of dimension r.

$\ker A = \ker C$; $\ker A^T = \ker B^T$; $\operatorname{ran} A = \operatorname{ran} B$; $\operatorname{ran} A^T = \operatorname{ran} C^T$.

Chapter 26: Positive definite and semidefinite matrices

Positive definite: $\forall x \quad x^T H x > 0$ Positive semidefinite: $\forall x \quad x^T H x \geq 0$

If H is symmetric and positive definite, **then H is invertible** and diagonal entries >0

A diagonal matrix D is positive definite iff all $d_i > 0$

If H is symmetric and positive definite and B has linearly independent column, then

$G = B^T H B$ is symmetric and positive definite.

A symmetric matrix is positive definite iff all eigenvalues are positive.

Both $A^T A$ and AA^T are symmetric and positive semidefinite. If A has linearly independent column then $A^T A$ and AA^T are positive definite and then:

$\ker(A^T A) = \ker A$; $\ker(AA^T) = \ker A^T$; $\operatorname{ran}(A^T A) = \operatorname{ran}(A^T A)$; $\operatorname{ran}(AA^T) = \operatorname{ran} A$