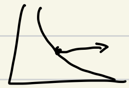
General optimization Problem

$$\min_{x \in K \subseteq \mathbb{R}^n} f(x)$$

- K admissible set
- f is called the cost functional

<u>Def</u>: A vector v is said to be an admissible direction at a point $x \in K$ if $\exists\, \varepsilon > 0$, such that $x + tv \in K$, $\forall t \in [0, \varepsilon]$



<u>Note</u>: if $x \in \text{Interior}(K)$ the all directions are admissible

First Order

Necessary condition for local optimality, but does not guarantee

<u>Lemma</u>: A point x is a local min for a function f, then $\forall$ admissable direction v at x
$\nabla f(x)^T v \geq 0$

<u>Proof</u>: $g(t) = f(x+tv)$ $\qquad \dfrac{f(x+tv) - f(x)}{t} \geq \dfrac{0}{t}$

$$\lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t} = \nabla f(x)^T v \geq 0$$

$$g(t) = g(0) + t\underbrace{g'(0)}_{0} + \underbrace{\frac{t^2}{2} g''(0)} + R(t) \qquad \text{Taylor therem}$$

$$g'(0) = \nabla f(x)^T v \qquad g''(0) = v^T \cdot \underbrace{H_f(x)}_{\text{Hessian}} \cdot v$$

<u>Lemma</u> (Second order NC), If x is a local min of f, then for any admissable direction if $\nabla f(x)^T v = 0$, then $v^T H_f(x) v \geq 0$

Let $f(x) = c^T x + d$, $x \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^1$

If $x_0 \in \text{Interior}(K)$, the x_0 cannot be a local min of f.

<u>Observation</u>: If $x \in \text{int}(K)$, then FONC implies if x is a local min for f then $\nabla f(x) = 0$

Second Order Sufficient Condition:

Theorem. If x is an interior point of K and $\nabla f(x) = 0$, then $H_f(x) > 0$ implies x is a local min of f

$H_f(x) > 0 \iff H_f(x)$ is positive definite

Proof: $f(x+v) = f(x) + \frac{1}{2} v^T H_f(x) v + o(\|v\|)^2$

↳ small o-notation.

$f(x+v) > f(x)$

Note: If $H_f(x) > 0$, $\exists\, c > 0$, st $v^T H_f(x) v \geq c v^T v = c\|v\|_2^2$

Algorithm for unconstrained minimization $\iff$ searching for critical point $\nabla f(x) = 0$

Gradient Descent Methods. Let x_0 be given $x_{k+1} = x_k - \lambda_k \nabla f(x_k)$

λ_k : step size

1) steepest descent: find λ_k st $f(x_k - \lambda_k \nabla f(x_k))$ is minimized.
2) Fixed step size, $\lambda_k = \lambda^*$ (what we'll use)

Quadratic Functional

let A be a positive definite matrix. A quadratic function has the form.

$f(x) = \frac{1}{2} x^T A x - b^T x$, $b \in \mathbb{R}^n$

$\nabla f(x) = Ax - b$ $\quad x^* = A^{-1} b$ $\quad H_f(x) = A$

Another one: $f(x) = \frac{1}{2}(x - x^*)^T A (x - x^*) - \frac{1}{2} x^* A x^*$

$\nabla f(x_{k+1})^T \nabla f(x_k) = 0$

$(A x_{k+1} - b)^T (A x_k - b) = 0$

Algorithm for Unconstrained Optimization.

Gradient Descent Method

x_0 given

$x_{k+1} = x_k - \lambda_k \nabla f(x_k)$

Special case for $f(x) = \frac{1}{2} x^T A x - b^T x = \underbrace{\frac{1}{2}(x - x^*)^T A (x - x^*)}_{V(x)} + c$

$x^* = A^{-1} b$

1) Steepest descent. Choose λ_k such that $f(x_{k+1}) = f(x_k - \lambda_k \nabla f(x_k))$ is minimized

Sufficient Condition. $\nabla f(x_{k+1})^T \underbrace{\nabla f(x)}_{g(x)} = 0$

$$\nabla f(x) = Ax - b$$

$$(A(x_k - \lambda_k (Ax_k - b)) - b)^T$$

$$Ax_k - b = 0$$

$$(A(x_k - \lambda_k g_k) - b)^T g_k = 0$$

$$(Ax_k - b - \lambda_k A g_k)^T g_k = 0$$

λ^*_k : step size for steepest decent.

$$(g_k - \lambda^*_k A g_k)^T g_k = 0$$

$$\boxed{\lambda^*_k = \frac{g_k^T g_k}{g_k^T A g_k}}$$

$V(x_{k+1}) = V(x_k)(1 - r_k)$

Note $V(x) = (x - x^*)^T A (x - x^*)$ with $A > 0$ positive definite. $V(x) = 0 \iff x = x^*$

For the algorithm to produce a converging sequence of x_k, we need $\lim_{k \to \infty} V(x_k) = 0$

$\boxed{V(x_k) = V(x_0) \prod_{j=1}^{k-1} (1 - r_j)}$ convergence requires $\lim_{k \to \infty} \prod_{j=1}^{k-1} (1 - r_j) = 0$

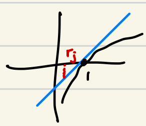
Theorem for $0 \le r_k < 1$, the $\lim_{k \to \infty} \prod_{j=1}^{k-1} (1 - r_j) = 0$ iff $\sum_{k=1}^{\infty} r_k = +\infty$

Proof:

$\lim_{k \to \infty} \prod_{j=1}^{k} (1 - r_j) = 0 \Rightarrow \lim_{k \to \infty} \log\left(\prod_{j=1}^{k} (1 - r_j)\right) = -\infty$

$\therefore \sum_{j=1}^{k} \log(1 - r_j) \ge -\sum_{j=1}^{k} r_j$

$\Rightarrow \lim_{k \to \infty} \sum_{j=1}^{k} r_j = +\infty$



$\log(1 - r_j) \ge -r_j$

$$V(x) = (x-x^*) A (x - x^*) , \quad g_k = \nabla f(x_k) = A(x_k - x^*)$$
$$= (A\underbrace{(x-x^*)}_{g})^T A^{-1} \underbrace{A(x-x^*)}_{g_k}$$

$$V(x_{k+1}) = V(x_k)(1-r_k)$$

$$(x_k - \lambda_k g_k - x^*) A (x_k - \lambda_k g_k - x^*) = (x-x^*)^T A (x-x^*) - 2\lambda_k g_k^T A(x-x^*) + \lambda_k^2 g_k^T A g_k$$
$$= V(x_k)\Big(1 - \underbrace{\frac{2\lambda_k g_k^T A(x-x^*) - \lambda_k^2 g_k^T A g_k}{V(x_k)}}_{r_k}\Big)$$

$$r_k = \frac{2\lambda_k g_k^T A(x-x^*) - \lambda_k^2 g_k^T A g_k}{g_k^T A^{-1} g_k} = \lambda_k \frac{g_k^T A g_k}{g_k^T A^{-1} g_k}\left(2 \underbrace{\boxed{\frac{g_k^T g_k}{g_k^T A g_k}}}_{\text{steepest decent}} - \lambda_k\right)$$

For steepest decent approach

$$\lambda_k = \frac{g_k^T g_k}{g_k^T A g_k} \qquad r_k = \left(\frac{g_k^T g_k}{g_k^T A g_k}\right)^2 \cdot \frac{g_k^T A g_k}{g_k^T A^{-1} g_k} = \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)} \geq \frac{1}{\lambda_{max}} \lambda_{min}$$

Let v be any vector in $\mathbb{R}^n$

$$v = \sum_{k=1}^n \alpha_k \cdot v_k , \; v_k \text{ are eigenvectors of } A \qquad \begin{pmatrix} v_k^T v_k = 1 \\ v_k^T v_j = 0 \end{pmatrix}$$
$$v^T A v = \sum_{k=1}^n \alpha_k^2 \lambda_k$$
$$Av = \sum_{k=1}^n \alpha_k A v_k = \sum_{k=1}^n \alpha_k \lambda_k v_k$$

$$A^{-1} = \frac{1}{\lambda} \text{ eigen}$$

$$\lambda_{min} \leq \frac{v^T A v}{v^T v} = \frac{\sum_{k=1}^n \alpha_k^2 \lambda_k}{\sum_{k=1}^n \alpha_k^2} \leq \lambda_{max}$$

smallest eigen — biggest eigen value

$$\frac{1}{\lambda_{max}} \leq \frac{v^T v}{v^T A v} \leq \frac{1}{\lambda_{min}}$$

$$\lambda_{min} \leq \frac{v^T v}{v^T A^{-1} v} \leq \lambda_{max}$$

Gradient Descent Method

$$x_{k+1} = x_k - \lambda_k \underbrace{\nabla f(x_k)}_{g_k}$$

Quadratic case

$$f(x) = \frac{1}{2} x^T A x - b^T x , \quad A > 0$$

① Steepest Descent

$$\lambda_k = \frac{g_k^T g_k}{g_k^T A g_k}$$

Always converge $r_k \geq \frac{\lambda_{min}}{\lambda_{max}}$

$$V(x) = (x-x^*)^T A (x-x^*)$$
$$V(x_{k+1}) = V(x_k)(1-r_k)$$
$$r_k = \lambda_k \frac{g_k^T A g_k}{g_k^T A^{-1} g_k}\left(2 \frac{g_k^T g_k}{g_k^T A g_k} - \lambda_k\right)$$

$$\frac{g_k^T g_k}{g_k^T A g_k} \geq \frac{1}{\lambda_{max}} \quad \text{if } \lambda < \frac{2}{\lambda_{max}}, \ r_k > 0$$

Proof.

$$\|x_{k+1} - x^*\| = \|x_k - \lambda g_k - x^*\|$$
$$= \|x_k - x^* - \lambda A(x_k - x^*)\|$$
$$= \|(I - \lambda A)(x_k - x^*)\|$$

Eigenvalue of $I - \lambda A$ has the form $1 - \lambda\lambda_k$

Then the eigenvalues of $I - \lambda A$ satisfy $|1 - \lambda\lambda_k| < 1$

Let $\alpha = \min_{1 \le k \le n} |1 - \lambda\lambda_k|$. $\|x_{k+1} - x^*\| \le \alpha \|x_k - x^*\|$

So if $\lambda = \frac{2}{\lambda_{max}}$, the solution converges.

Conjugate Gradient Method

Definition: Let A be a positive definite matrix. A set of vectors $\{q_1, \dots, q_n\}$ is said to be A-conjugate if $q_k \neq 0$, $q_k^T A q_j = 0$ when $k \neq j$

"Conjugate Gradient Method": Let $q_1 \dots q_n$ be A-conjugate.

Let us consider minimizing $f(x) = \frac{1}{2} x^T A x - b^T x$

Let x_1 be given, we define $x_{k+1} = x_k - \lambda_k q_k$. λ_k is selected to minimize $f(x_{k+1})$

Then $x_{n+1} = x^*$

How to find λ_k : λ_k is optimal if $\nabla f(x_{k+1})^T \cdot q_k = 0$ (perpendicular).

$$\nabla f(x_{k+1}) = A x_{k+1} - b = A(x_{k+1} - x^*)$$

$$A(x_k - \lambda_k q_k - x^*) = \underbrace{A(x_k - x^*)}_{g_k} - \lambda_k A q_k = 0$$

$$(g_k - \lambda_k A q_k)^T q_k = 0$$

$$\lambda_k = \frac{g_k^T q_k}{q_k^T A q_k}$$

$$\lambda_1 = \frac{g^T \cdot q_1}{q_1^T A q_1} = \frac{(A(x_1 - x^*))^T q_1}{q_1^T A q_1} = \frac{(\sum_{k=1}^{n} \alpha_k A_k q_k)^T q_1}{q_1^T A q_1} = \frac{\sum_{k=1}^{n} (\alpha_k q_k^T A q_1)}{q_1^T A q_T} = \frac{\alpha_1 q_1^T A q_1}{q_1^T A q_1} = \alpha_1$$

Let x_1 be given, then $x_1 - x^* = \sum_{k=1}^{n} \alpha_k q_k$

$$x_2 - x^* = x_1 - x^* - \lambda_1 q_1$$
$$= x_1 - x^* - \alpha_1 q_1$$
$$= \sum_{k=2}^{n} \alpha_k q_k$$
$$x_3 - x^* = \sum_{k=3}^{n} \alpha_k q_k$$

q_k linearly independent. $x_{n+1} - x^* = 0$

Nonlinear Programming Algorithm

- Gradient Descent Method,
- Conjugate Gradient Descent Method.

$\min f(x)$, $f(x) = \frac{1}{2} x^T A x - b^T x$

$A > 0$, positive definite

Def: A-conjugate $\{d_1, \dots, d_n\}$

$d_i^T A d_j = 0$

Why A-conjugates are L.I.

show $\sum_{k=1}^{n} \alpha_k d_k = 0 \Rightarrow \alpha_k = 0 \ \forall k$

$$d_j^T A \left(\sum_{k=1}^{n} \alpha_k d_k\right) = 0$$

$$\sum_{k=1}^{n} \alpha_k d_j^T A d_k = \alpha_j d_j^T A d_j \quad , \alpha_j \text{ must } = 0$$

Let $d_1 \dots d_n$ be A-conjugate

$x_{k+1} = x_k - \lambda_k d_k$

$g_{k+1} \perp d_k$, $\quad g_{k+1} = g_k - \lambda_k A d_k$

$g_{k+1}^T d_k = 0$, $\quad \lambda_k = \frac{g_k^T d_k}{d_k^T A d_k} \qquad x_{n+1} = x^*$

Conjugate Gradient Method that's practicable

x_1 given, $d_1 = \nabla f(x_1) = g_1$

At step k, $x_{k+1} = x_k - \lambda_k d_k$, $\lambda_k = \frac{g_k^T d_k}{d_k^T A d_k}$ => $g_{k+1} = g_k - \lambda_k A d_k$

=> $\frac{g_{k+1} - g_k}{\lambda_k} = A d_k$

$d_{k+1} = g_{k+1} - \left(\frac{g_{k+1}^T A d_k}{d_k^T A d_k}\right) d_k$ $\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}$

Motivated by $d_{k+1}^T A d_k = 0$, Q-conjugate d_{k+1} and d_k

Lemma: For all k, k=1,...,n, and $i \le k$, then

- $g_{k+1}^T g_i = 0$
- $g_{k+1}^T d_i = 0$
- $\boxed{d_{k+1}^T A d_i = 0}$

Proof: Induction hypothesis: for $k \le m$, all 3 equalities are true.

Take $k = m+1$, $i \le m+1$

$$g_{m+1}^T d_i = \begin{cases} 0, & i = m+1 \\ | & i < m+1 \end{cases}$$

$$d_i^T (g_{m+1} - \alpha_{m+1} A d_{m+1})$$

$$g_{m+2}^T g_i = 0, \quad i \le m+1$$

$$(g_{m+1} - \alpha_{m+1} A d_{m+1})^T (\beta_{i-1} d_{i-1} - d_i) = 0$$

$$g_{m+2}^T d_i = 0$$

$$d_{m+2}^T A d_i = \begin{cases} 0, & i = m+1 \\ (-g_{m+2} + \beta_{m+1} d_{m+1})^T Q d_i, & i < m+1 \end{cases}$$

$$\parallel$$
$$0$$

emm....

Conclusion: for conjugate gradient method, $x_{n+1} = x^*$

☆

For non-quadratic functions

$$X_2 = X_1 - \lambda_1 \cdot g_1 \qquad g_1 = d_1 = \nabla f(x)$$

Use line search technique to find λ_1

$$\lambda_1 = \frac{g_1^T g_1}{g_1^T A g_1}, \quad g_1^T A g_1 = \frac{g_1^T g_1}{\lambda_1}$$

for non-quadratic function

$$d_1 = g_1 = \nabla f(x_1)$$

$x_{k+1} = x_k - \alpha_k d_k$

$$\alpha_k = \alpha_{k-1} \frac{g_k^T d_k}{d_k^T (g_k - g_{k-1})}$$

$g_{k+1} = \nabla f(x_{k+1})$

$g_k = \nabla f(x_k)$

$$d_{k+1} = g_{k+1} - \beta_k d_k, \quad \beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k (g_{k+1} - g_k)}$$

turns out: $g_k - g_{k-1} = A d_k$

not the one we want to do. doesn't work well.

Non quadratic case

Either λ_k is obtained through line search then,

this is out of no where

$$\beta_k = \lambda_k \frac{g_{k+1}^T d_k}{d_k^T (g_{k+1} - g_k)}$$

Derived from $g_k^T d_k = g_k^T (g_k - \beta_{k-1} d_{k-1})$

$g_k^T d_{k-1} = 0$

$$\beta_k = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)} = \frac{g_{k+1}^T g_{k+1}}{d_k^T g_k} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} = \beta_k$$

- Similar to quadratic case, the convergence rate of gradient descent & conjugate gradient method is linear.

Newton's Method:

x_0 is given. $x_{k+1} = x_k - H_f(x_k)^{-1} \circ \nabla f(x_k)$

<u>Theorem</u>: Let $f: \mathbb{R}^n \to \mathbb{R}^1$ to be 3 times continuously differentiable.
Let x^* be a local minimum for f with the properties
$mI \leq H_f(x^*) \leq MI$, $m, M > 0$
Then $\exists$ $\varepsilon > 0$ and $k > 0$, s.t. $\forall$ $\|x_0 - x^*\| \leq \varepsilon$,
the sequence x_k generated by the Newton's method satisfies.
$\|x_{k+1} - x^*\| \leq k\|x_k - x^*\|^2$

Advantage: converge faster

disadvantages: expensive to compute H_f.
hard to estimate $\underline{\varepsilon} > 0$
$H_f(x_k)^{-1} \circ \nabla f(x_k)$ may not be negative.

Quasi Newton's Method

Main Ideas:

1) Replace H_f with $H_k = H_f(x_k) + \alpha I$
Choose α large enough so that H_k is positive definite
2) Replace N by $x_{k+1} = x_k - \lambda_k H_k^{-1} \cdot \nabla f(x_k)$
3) Approximate H_f using g_k, x_k
$g_k = \nabla f(x_k)$, $g_{k-1} = \nabla f(x_{k-1})$
$\Delta g = g_k - g_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1}) = H_f(x_{k-1})(x_k - x_{k-1}) + o(\|x_k - x_{k-1}\|)$
usually drop this

Look for H_k such that $\Delta g_k = H_k \Delta x_k$
H_k has infinite possibility
We can generalize this requirement to
$\Delta g_i = H_k \Delta x_i$, $i \leq k$ constrain a bit to get an unique H_k

find a positive definite matrix B_k, s.t.
$B_k \Delta g_i = \Delta x_i$ $i \leq k$

inverse of Hessian

Requirement for updating H_{k+1}, B_{k+1}

- $H_{k+1} \Delta X_k = \Delta g_k$

  $B_{k+1} \Delta g_k = \Delta X_k$

- $H_{k+1} V = H_k V, \quad V \perp \Delta X_k$

  $B_{k+1} V = B_k V, \quad V \perp \Delta g_k$

- $H_{k+1} = H_k + uu^T, \quad u, v \in \mathbb{R}^n$

  $B_{k+1} = B_k + vu^T$

$$V = \frac{\Delta X_k - B_k \Delta g_k}{\Delta g_k^T \Delta g_k}$$

$u = \Delta g_k$

$H_{k+1} = H_k + vu^T$

Let w be given such that $w^T \Delta g_k = 0$

$B_{k+1} w = B_k w + v \Delta g_k^T w = B_k w$

not symmetric

$$B_{k+1} \Delta g_k = B_k \Delta g_k + \frac{1}{\Delta g_k^T \Delta g_k}\left(\Delta X_k \Delta g_k^T \Delta g_k - B_k \Delta g_k \Delta g_k^T \Delta g_k\right) = \Delta X_k$$

For the first step, need to find B_1 such that

$B_1 \Delta g_1 = \Delta X_1 \quad , \quad B_1 V = B_0 V, \quad V \perp \Delta g_1$

Rank 1 update

$$B_1 = B_0 + \alpha \underbrace{V u^T}_{[u_1 V,\ u_2 V,\ \ldots u_n V]}, \quad u, v \in \mathbb{R}^n$$

$[u_1 V, u_2 V, \ldots u_n V]$ rank one matrix. all linearly dependent

---

SGD (Stochastic Gradient Descent Method)

For project

nonlinear least square problem.

find parameter w^*

$$\min_{w} \underbrace{\sum_{k=1}^{N} |y_k - f(x_k; w)|^2}_{J(w)}$$

Gradient Descent

$$W_{k+1} = W_k - \lambda \nabla_w J(W_k)$$
$$= W_k + 2\lambda \sum_{i=1}^{N} (y_i - f(x_i; w_k)) \nabla_w f(x_i; w_k)$$

SGD

$$W_{k+1, i+1} = W_{k,i} + \lambda (y_i - f(x_i; w)) \nabla_w f(x_i, w_{k,i})$$

Non-linear programming

Quasi Newton's Method

Let x_0, H_0 be selected with $H_0 > 0$

For each step k, define

1) $g_k = \nabla f(x_k)$, $d_k = -H_k g_k$, stop if $g_k = 0$ or $\|g_k\| \le \varepsilon$
2) find α_k to minimize $f(x_k + \alpha d_k)$, $x_{k+1} = x_k + \alpha_k d_k$
3) $\Delta x_k = \alpha_k d_k$, $\Delta g_k = g^{k+1} - g^k$
4) Update H_k $\quad H_{k+1} = H_k + u_k u_k^T \quad u_k \in \mathbb{R}^n$

such that $H_{k+1} \Delta g_k = \Delta x_k$

$$H_{k+1} \Delta g_k = H_k \Delta g_k + u_k \underbrace{u_k^T \Delta g_k}_{\beta_k}$$

$$= \Delta x_k$$

$$H_{k+1} \Delta g_k - \Delta x_k = \beta_k u_k$$

$$u_k u_k^T \Delta g_k$$

$$= \beta_k^2 (H_k \Delta g_k - \Delta x_k)(H_k \Delta g_k - \Delta x_k)^T \Delta g_k$$

We need $\beta_k^2 (H_k \Delta g_k - \Delta x_k)^T \Delta g_k = -1$

$$\beta_k^2 = \frac{1}{\Delta x_k^T \Delta g_k - \Delta g_k^T H_k \Delta g_k}$$

<u>Theorem</u> If the construction of matrices H_k in the quasi-Newton's method has the property $H_{k+1} \Delta g_i = \Delta x_i$, for $i = 0, \ldots, k$ when applied to a quadratic function $f(x) = \frac{1}{2} x^T Q x - b^T x$ then the direction d_i are Q-conjugate for $i = 0 \ldots k$, if $\alpha_i \neq 0$, $i = 0, \ldots, k$

<u>Proof</u> by induction

$\nabla f(x) = Qx - b$

$\Delta g_0 = g_1 - g_0$

$= Q(x_1 - x_0)$

$= Q \Delta x_0$

<u>Base case</u> $k = 0$

$$H_1 \Delta g_0 = \Delta x_0$$

$$d_1^T \cdot Q d_0 = -g_1^T H_1 \cdot Q d_0 = -g_1^T H_1 Q \frac{\Delta x_0}{\alpha_0} = -g_1^T H_1 \frac{\Delta g_0}{\alpha_0} = -g_1^T \frac{\Delta x_0}{\alpha_0} = -g_1^T \cdot d_0 = 0$$

because α_0 is the minimal

Assuming the conclusion is valid up to $k-1$

$$d^{(k+1)} Q d^i = -g^{(k+1)} H_{k+1} Q d_i = -g^{(k+1)} H_{k+1} Q \frac{\Delta x_i}{\alpha_i} = -g^{(k+1)} H_{k+1} \frac{\Delta g_i}{\alpha_i} = -g^{(k+1)} \frac{\Delta x_i}{x_i} = -g^{(k+1)} d_i = 0$$

For quadratic functions

if $x_0 - x^* = \sum_{k=1}^{n} \alpha_k q_k$

$x_i - x^* = \sum_{k=i+1}^{n} \alpha_k q_k$

$g_i = A(x_i - x^*) = \sum_{k=i+1}^{n} \alpha_k A q_k$

$g_i^T q_j = 0, \; j \le i$

DFP Algorithm

1) x_0 and $H_0 > 0$ are selected
2) For each step $g_k = \nabla f(x_k)$, stop if $\|g_k\| \le \varepsilon$

$d_k = -H_k g_k$

find $\alpha_k = \text{argmin } f(x_k + \alpha d_k)$

$\Delta x_k = \alpha_k d_k, \quad \Delta g_k = g_{k+1} - g_k$

$$H_{k+1} = H_k + \frac{\Delta x_k \Delta x_k^T}{\Delta x_k^T \Delta g_k} - \frac{(H_k \Delta g_k)(H_k \Delta g_k)^T}{(H_k \Delta g_k)^T (H_k \Delta g_k)}$$

may not be positive definite

To avoid non-positive definite matrix H_{k+1}, we replace H_{k+1} by $H_{k+1} + \lambda I \Rightarrow$ BFGS

Theorem for DFP algorithms $H_{k+1} \Delta g_i = \Delta x_i$, when f is quadratic

Summary

- Gradient Descent
  - steepest descent
  - fixed step
- Newton's Algorithm
  - Quasi-Newton's Method
- Conjugate Gradient Method

Linear Programming

Linear Programming Problems.

minimize $c^T x$, $c, x \in \mathbb{R}^n$

subject to

$Gx + f \leq 0$, $G \cdot \mathbb{R}^{m\times n}$, $f \in \mathbb{R}^m$,

$Ax - b = 0$, $A \cdot \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$

Ex. A vendor is making fruit juice. He has 3 main ingredients, passion fruit juice, orange juice and honey. He makes two kinds of drinks passion-orange and sweet orange

Let x_1, x_2 be the amount of each kind of drink to make. Let C_1, and C_2 be the unit price. The total revenue is given by

$$C_1 x_1 + C_2 x_2$$

The use of ingredients for each kind of drink is given by

	P	O	H
passion-orange	a_{11}	a_{21}	a_{31}
sweet-orange	a_{12}	a_{22}	a_{32}

$a_{11}x_1 + a_{12}x_2 \leq b_1$, passion fruit

$a_{21}x_1 + a_{22}x_2 \leq b_2$ orange

$a_{31}x_1 + a_{32}x_2 \leq b_3$ honey

$x_1, x_2 \geq 0$

total amount available

Standard LP

min $C^T x$ subject to $Ax = b$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $x \geq 0$

and A is rank m. ($\Rightarrow n \geq m$)

Techniques of transforming LP to standard form

1) Add slack variables

min $c^T x$ subject to $Ax \geq b$, $x \geq 0$ $\iff$ min $\hat{c}^T \hat{x}$, $\hat{x} = \binom{x}{y}$ $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ $\hat{c} = \binom{c}{0}$

$Ax + y = b$, $x \geq 0, y \geq 0$ $\iff$ $\hat{A} = [A \ \ I_m]\hat{x} = b$

inequality to equality.

2) min $c^T x$ subject to $Ax = b$

$\iff x = x^+ - x^-$, $x^+, x^- \in \mathbb{R}^n$, $x^+, x^- \geq 0$, min $c^T x^+ - c^T x^-$ subject to $Ax^+ - Ax^- = b$

Basic Solution

Definition: A feasible solution x, i.e. x satisfies all constraints. is said to be a basic solution if it has no more than m non-zero components.

Theorem (Fundamental Theorem of Linear Programming)

1) If a LP has a feasible solution, it must have a basic feasible solution.

2) If a LP has an optimal solution, it must have a basic optimal solution.

Proof. Let x be a feasible solution. That is $x \geqslant 0$, $Ax = b$. suppose there are p components of x that are non-zero. If $p \leq m$, then x is basic. If $p > m$, WLOG we assume $x_1, \ldots, x_p \neq 0$ Let $A_1, \ldots, A_p$ be the first p columns of matrix A. Since $A_1, \ldots, A_p$ are linearly independent. $\exists \alpha_1, \ldots, \alpha_p$, not all zero, st $\sum_{k=1}^{p} \alpha_k A_k = 0$

Let $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ 0 \\ 0 \end{pmatrix}$, for any λ, $A(x + \lambda \alpha) = b$

$$A\alpha = \sum_{k=1}^{p} A_k \cdot \alpha_k = 0$$

Let j be selected such that $\left|\frac{x_j}{\alpha_j}\right| \leq \left|\frac{x_k}{\alpha_k}\right|$, $k = 1, \ldots, p$

Choose λ st $|\lambda| = \left|\frac{x_j}{\alpha_j}\right|$ and $x_j - \lambda \alpha_j = 0$

$x_k + \lambda \alpha_k \geqslant x_k - |\lambda \alpha_k| \geqslant x_k - \left|\frac{x_k}{\alpha_k}\right| \alpha_k \geqslant 0$

$x + \lambda \alpha$ has $p-1$ non-zero elements.

Proved.

Proof of 2

Let x be an optimal feasible solution $c^T(x + \lambda \alpha) = c^T x + \lambda c^T \alpha$

Since x is optimal, this implies $c^T \alpha = 0$

Linear Programming
Simplex Algorithm
Standard LP

$$\min c^T x$$

subject to $x \geq 0$, $Ax = b$, $A \in \mathbb{R}^{m \times n}$

$\text{Rank}(A) = m$,

Theorem. If LP has an optimal solution, it must have an optimal solution

Ex. $\min 3x_1 + 2x_2 + x_3 + 2x_4$

subject to $x_1, x_2, x_3, x_4 \geq 0$

$$x_1 + 2x_3 + 2x_4 = 4$$
$$x_2 - x_3 + 3x_4 = 3$$

possible solution : $x_1 = 4$, $x_2 = 3$, $x_3 = 0$, $x_4 = 0$, $c^T x = 18$

Let $x_4 = 0$, we get $\begin{cases} x_1 + 2x_3 = 4 \\ x_2 - x_3 = 3 \end{cases}$ plug in $c^T x = 18 + (-6+2+1)x_3$

Let $x_3 = 0$, $c^T x = 18 + (-6-6+2)x_4$ from $\begin{matrix} x_1 + 2x_4 = 4 \rightarrow x_4 = 2 \\ x_2 + 3x_4 = 3 \rightarrow x_4 = 1 \end{matrix}$

x_4 has priority to be changed

New solution :

$x_4 = 1$, $x_1 = 2$, $x_2 = 0$, $x_3 = 0$, $c^T x = 8$

$\begin{matrix} x_1 + 2x_3 + 2x_4 = 4 \\ x_2 - x_3 + 3x_4 = 3 \end{matrix} \Rightarrow \begin{matrix} (1-\frac{2}{3})x_3 - \frac{2}{3}x_2 + (2-\frac{2}{3})x_3 = 4-2 \\ \frac{1}{3}x_2 - \frac{1}{3}x_3 + x_4 = 1 \end{matrix} \Rightarrow \begin{matrix} x_1 - \frac{2}{3}x_2 + \frac{4}{3}x_3 = 2 \\ \frac{1}{3}x_2 - \frac{1}{3}x_3 + x_4 = 1 \end{matrix}$

Opt 1. Keep $x_3 = 0$
$$c^T x = 3x_1 + 2x_2 + 2x_4$$
$$= 3(2 + \frac{2}{3}x_2) + 2x_2 + 2(1 - \frac{1}{3}x_2)$$
$$= 8 + (2 + 2 - \frac{2}{3})x_2$$

$\begin{cases} x_1 - \frac{2}{3}x_2 = 2 \\ \frac{1}{3}x_2 + x_4 = 1 \end{cases}$

bad option.

Opt2 keep $x_2 = 0$

$$c^T x = 3x_1 + x_3 + 2x_4$$
$$= 3(2 - \tfrac{4}{3}x_3) + x_3 + 2(1 + \tfrac{1}{3}x_3)$$
$$= 8 + (-4 + 1 + \tfrac{2}{3})x_3$$

$$\begin{cases} x_1 + \frac{4}{3}x_3 = 2 \\ -\frac{1}{3}x_3 + x_4 = 1 \end{cases}$$

$\boxed{x_3 = \frac{6}{4}}$

no constrain

$x_1 = 0,\ x_2 = 0,\ x_3 = \frac{6}{4},\ x_4 = \frac{3}{2},\ c^T x = \frac{6}{4} + 3 < 8$

New solution $x_1 = 0,\ x_2 = 0,\ x_3 = \frac{3}{2},\ x_4 = \frac{3}{2}$

$$x_1 - \tfrac{2}{3}x_2 + \tfrac{4}{3}x_3 = 2 \qquad \Rightarrow \qquad \tfrac{3}{4}x_1 - \tfrac{1}{2}x_2 + x_3 = \tfrac{3}{2}$$
$$\tfrac{1}{3}x_2 - \tfrac{1}{3}x_3 + x_4 = 1 \qquad \qquad \tfrac{1}{4}x_1 - \tfrac{1}{6}x_2 + x_4 = 1 + \tfrac{1}{2}$$

Let $x_2 = 0$

$$c^T x = 3x_1 + x_3 + 2x_4 = 3x_1 + (\tfrac{3}{2} - \tfrac{3}{4}x_1) + 2(\tfrac{3}{2} - \tfrac{1}{4}x_1)$$
$$= \tfrac{3}{2} + 3 + (3 - \tfrac{3}{4} - \tfrac{2}{4})x_1 \qquad > 0$$

Let $x_1 = 0$

$$c^T x = 2x_2 + (\tfrac{3}{2} + \tfrac{1}{2}x_2) + 2(\tfrac{3}{2} + \tfrac{1}{6}) = \tfrac{3+3}{2} + (2 + \tfrac{1}{2} + \tfrac{1}{3})x_2 \qquad > 0$$

This says ↓ is the optimal solution

§ Basic Simplex Algorithm

Starting point Let $i_1, \dots, i_m$, such that $x_{i_k} \neq 0$, $k = 1, \dots, m$, $x_j = 0$, $j \neq i_1, \dots, i_m$

$$A = [A_1, \dots, A_n] \qquad A_{i_k} = e_k = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow k\text{th row}$$

By keeping all $x_j = 0$, for $j \neq i_1, \dots, i_m$ except j^*, then the constraints have the form

$$x_{i_k} + A_{kj^*} x_{j^*} = b_k, \quad k = 1, \dots, m.$$

and

$$c^T x = \sum_{k=1}^{m} c_{i_k} x_{i_k} + c_{j^*} x_{j^*} = \sum_{k=1}^{m} c_{i_k}(b_k - A_{kj^*} x_{j^*}) + c_{j^*} x_{j^*}$$
$$= \sum_{k=1}^{m} c_{i_k} b_k + \underbrace{\Big(c_{j^*} - \sum_{k=1}^{m} c_{i_k} A_{kj^*}\Big)}_{} x_{j^*}$$

Continue with that example

$3x_1 + 2x_2 + x_3 + 2x_4$

$$\begin{cases} x_1 + 2x_3 + 2x_4 = 4 \\ x_2 - x_3 + 3x_4 = 3 \end{cases}$$

$$c = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 3 \end{bmatrix} \quad b = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \quad \begin{matrix} i_1 = 1 \\ i_2 = 2 \end{matrix}$$

$\mathrm{Rank}(A) = m$, $\exists\, i_1, \dots, i_m$ s.t. $A_{i_k} = e_k$, $k = 1, \dots, m \Rightarrow x_{i_k} = b_k$, $x_j = 0$, $j \neq i_1, \dots, i_m$

Algorithm consists of finding a new basic solution

step 1) For each $j \neq i_1, \dots, i_m$

$$x_{i_k} + \sum_{j=1}^{n} A_{kj} x_j = b_k, \quad k = 1, \dots, m$$

New cost

$$c^T x = \sum_{\substack{j=1 \\ j \neq i_1, \dots, i_m}}^{n} c_j x_j + \sum_{k=1}^{m} c_{i_k} \Big(b_k - \sum_{\substack{j=1 \\ j \neq i_1, \dots, i_m}}^{n} A_{kj} x_j \Big)$$

$$= \sum_{\substack{j=1 \\ j \neq i_1, \dots, i_m}}^{n} \underbrace{\Big(c_j - \sum_{k=1}^{m} c_{i_k} A_{kj} \Big)}_{r_j} x_j + \boxed{\sum_{k=1}^{m} c_{i_k} b_k}$$

Evaluate $\boxed{r_j = c_j - \sum_{k=1}^{m} c_{i_k} A_{kj}}$

if all r_j are positive or 0, the current basis solution is optimal.

step 2) Let j^* be such that, $r_{j^*} < 0$ and $r_{j^*} \leq r_j$, $\forall j \neq i_1, \dots, i_m$.

We want to keep all other $x_j = 0$ $j \neq j^*$ and find the largest value for x_{j^*}

$$x_{i_k} + A_{kj^*} x_{j^*} > b_k, \quad k = 1, \dots, m$$

If $A_{kj^*} > 0$, $x_{j^*} \leq \frac{b_k}{A_{kj^*}}$

If for all k, $A_{kj^*} \leq 0$, then the problem is unbounded. i.e. $\min c^T x = -\infty$

Otherwise, $x_{j^*} = \min \frac{b_k}{A_{kj^*}}$, $A_{kj^*} > 0$

New basis solution, has the property that

$$x_{i_{k^*}} = 0, \quad x_{j^*} = \frac{b_{k^*}}{A_{k^*, j^*}}.$$

step 3) Using Gaussian elimination technique to transform the constraints in

$$\hat{A} x = \hat{b} \quad \text{s.t.} \quad \hat{A}_{j^*} = e_{k^*}$$

Divide k^*-the row of matrix A by A_{k^*, j^*}

$\Rightarrow$ Subtract from row k, the A_{kj^*} multiple of the new row k^*

Ex $\min\ 3x_1 + 2x_2 - x_3 + 2x_4$

subject to $x_1 - 2x_3 + 2x_4 = 4$

$x_2 - x_3 + 3x_4 = 3$

		x_1	x_{i_1}		x_n	b
c_i		c_1		0	c_n	
c_{i_1}		A_{11}			A_{1n}	b_1
c_{i_n}		A_{m1}			$A_{m,n}$	b_m
r			////			$\sum_{k=1}^{m} c_{i_k} b_k$

	x_1^*	x_2^*	x_3	x_4	b
c_i	3	2	-1	2	
$c_{i_1} = 3$	1	0	-2	2	4
$c_{i_2} = 2$	0	1	-1	3	3
r	////	////	$-1-(-2\times 3-2)$	$2+2\times 3-2\times 3$	18

$j^* = 4,\ k^* = 2$

	x_1^*	x_2	x_3	x_4^*	b
c_i	3	2	-1	2	
$c_{i_1} = 3$	1	$-\frac{2}{3}$	$-1+\frac{2}{3}$	0	2
$c_{i_2} = 2$	0	$\frac{1}{3}$	$-\frac{1}{3}$	1	1
r	////			////	8

Simplex Method

$\min c^T x$

subject to $x \geq 0,\ Ax = b$ $\quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$

1) $b \geq 0$

2) $\exists\ i_1, \ldots, i_n$ $\quad A_{i_k} = e_k$

Transforming General LP into standard form to start Simplex Algorithm.

1) Making RHS of inequality constraints to be non-negative.

2) Change inequalities to equalities.

$$\begin{matrix} A_1 x \geq b_1 \\ A_2 x \leq b_2 \end{matrix} \Rightarrow \begin{matrix} A_1 x + y = b_1 & y \geq 0 \\ A_2 x - y = b_2 & y \geq 0 \end{matrix}$$

3). Change unsigned variable to non-negative variable

$x = x^+ - x^-$, $x^+, x^- \geq 0$

4) Change absolute value to non-negative variable

$x = x^+ - x^-$, $x^+, x^- \geq 0$ replace $|x|$ by $x^+ + x^-$

5) Adding auxiliary variables and solve the LP in 2 phases

$$\min \ c^T x$$
subject to $x \geq 0$
$$Ax = b \geq 0$$

a) $\min \sum_{k=1}^{m} y_k$

subject to $x \geq 0$, $y \geq 0$

$Ax + y = b$

$\Rightarrow$ ① unbounded X ② has a solution $y \neq 0$ ③ $y = 0$ is an optimal solution

b) Once we have $\hat{x}$ with m basic component, we can solve the original problem

Ex $\max \ 2x_1 + 5x_2 \iff \min -2x_1 - 5x_2$

subject to $x_1, x_2 \geq 0$ $y_1 \geq 0$ $y_2 \geq 0$ $y_3 \geq 0$

$$\begin{matrix} x_1 \leq 4 \\ x_2 \leq 6 \\ x_1 + x_2 \leq 8 \end{matrix} \Rightarrow \begin{matrix} x_1 + y_1 = 4 \\ x_2 + y_2 = 6 \\ x_1 + x_2 + y_3 = 8 \end{matrix}$$

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$$

	x_1	x_2	y_1	y_2	y_3	b
c	-2	-5	0	0	0	
0	1	0	1	0	0	4
0	0	1	0	1	0	6
0	1	1	0	0	1	8
r	-2	-5	///	///	///	0

Duality in Linear Programming

Standard Primal Problem (Symmetric form)

$\min \ c^T x$

subject to $x \geq 0$

$Ax \geq b, \ A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m$

Dual Problem

$\max \ b^T \lambda$

subject to $\lambda \geq 0$

$A^T \lambda \leq c$

Ex $\min \ 3x_1 + 2x_2 - x_3$

subject to $x_1, x_2, x_3 \geq 0$

$2x_1 - 2x_2 + 3x_3 \leq 1$

$x_1 + x_2 - 5x_3 \geq 3$

Standard Primal Problem

$c = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}, \ A = \begin{pmatrix} -2 & 2 & -3 \\ 1 & 1 & -5 \end{pmatrix} \quad b = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$

Dual Problem

$\max \ -\lambda_1 + 3\lambda_2$

subject to $-2\lambda_1 + \lambda_2 \leq 3$

$\lambda_1 + \lambda_2 \leq 2$

$-3\lambda_1 - 5\lambda_2 \leq -1$

Ex $\min\ x_1 + 2x_2$ Standard Primal Problem

subject to $x_1, x_2 \geq 0$

$x_1 + 5x_2 = 10$ $\Rightarrow$ $x_1 + 5x_2 \geq 10$

$x_1 - 2x_2 \geq 1$ $\quad -x_1 - 5x_2 \geq -10$

$x_1 - 2x_2 \geq 1$

Dual of Dual LP

1st Dual $\rightarrow$ $\min\ -b^T\lambda$

subject to $\lambda \geq 0$

$-A^T\lambda \geq -C$

$\xrightarrow{\text{Dual again}}$ $\max\ -C^T x$ subject to $x \geq 0$, $-Ax \leq -b$ $\Leftrightarrow$ $\min C^T x$ subject to $x \geq 0$, $Ax \geq b$

Dual of Dual is itself

Consider a LP

$\min C^T x$

subject to $x \geq 0$

$Ax = b,\ A \in \mathbb{R}^{m \times n}$

$\downarrow$ Standard Primal Problem in Asymmetric form.

$\Leftrightarrow$ $\min C^T x$

$x \geq 0$

$Ax \geq b$

$-Ax \leq -b$

$\begin{bmatrix} A \\ -A \end{bmatrix} x \geq \begin{bmatrix} b \\ -b \end{bmatrix}$

Dual Problem: $\lambda^+, \lambda^- \in \mathbb{R}^m$

$\max\ b^T\lambda^+ - b^T\lambda^-$

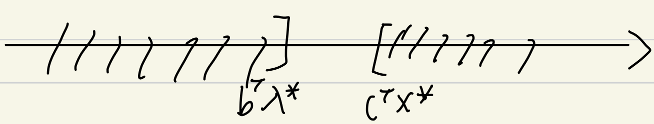
$\lambda^+, \lambda^- \geq 0$

$A\lambda^+ - A\lambda^- \leq C$

$\boxed{\Rightarrow \max\ b^T\lambda \quad \lambda \in \mathbb{R}^m \quad \text{subject to } A^T\lambda \leq C}$

Duality Property

Let x be a feasible solution for the primal problem and λ be a feasible solution for the dual problem, we have $\boxed{b^T\lambda \leq C^T x}$



Terminology: Let p^* be the optimal cost of the primal problem

$$p^* = \begin{cases} -\infty, & \text{if primal problem is unbounded in cost} \\ p^* & \text{optimal cost} \\ +\infty & \text{if primal problem is infeasible, ie admissable set is empty.} \end{cases}$$

Let d^* be the optimal cost for the dual problem, then

$$d^* = \begin{cases} -\infty & \text{if the dual problem is infeasible} \\ d^* & \text{optimal} \\ +\infty & \text{if the dual problem is unbounded.} \end{cases}$$

Ex

Primal Problem

$\min x$

subject to $x \geq 0$

$x + 1 \leq 0$

(infeasible)

$\Rightarrow$

standard form

$\min x$

subject to $x \geq 0$

$-x \geq 1$

Dual Problem

$\max \lambda$ $\quad \lambda \geq 0$

subject to $-\lambda \leq 1$

$d^* = \lambda = +\infty$

Ex

$\min x$

subject to $x \geq 0$

$A = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$0x \geq 1$

$x \geq -1$

Dual Problem

$\max \lambda_1 - \lambda_2$

subject $\lambda_1, \lambda_2 \geq 0$

$\lambda_1 0 + \lambda_2 \leq -1$

$d^* = -\infty$

Simplex Method example

maximize $2x_1 + 5x_2$

subject to
$$x_1 \le 4$$
$$x_2 \le 6$$
$$x_1 + x_2 \le 8$$
$$x_1, x_2 \ge 0$$

minimize $-2x_1 - 5x_2$

subject to
$$x_1 + x_3 = 4$$
$$x_2 + x_4 = 6$$
$$x_1 + x_2 + x_5 = 8$$
$$x_1, x_2, x_3, x_4, x_5 \ge 0$$

basic

$$\begin{array}{cccccc} & & 1 & 1 & 1 & \\ a_1 & a_2 & a_3 & a_4 & a_5 & b \\ 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1 & 0 & 6 \\ 1 & 1 & 0 & 0 & 1 & 8 \end{array} \qquad \begin{bmatrix} 0 \\ 0 \\ 4 \\ 6 \\ 8 \end{bmatrix}$$

$$r_1 = c_1 - z_1$$
$$= -2 - (c_3 y_{11} + c_4 y_{21} + c_5 y_{31})$$
$$= -2$$
$$r_2 = c_2 - z_2$$
$$= -5 - (c_3 y_{12} + c_4 y_{22} + c_5 y_{32}) = -5$$

Choose a_2

$$\frac{y_{2,0}}{y_{2,2}} = \frac{6}{1} = 6$$
$$\frac{y_{3,0}}{y_{3,3}} = \frac{8}{1} = 8$$

basic

$$\begin{array}{cccccc} 1 & 1 & & & 1 & \\ a_1 & a_2 & a_3 & a_4 & a_5 & b \\ 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1 & 0 & 6 \\ 1 & 0 & 0 & -1 & 1 & 2 \end{array} \qquad \begin{bmatrix} 0 \\ 6 \\ 4 \\ 0 \\ 2 \end{bmatrix}$$

$$r_1 = c_1 - z_1 = c_1 - (c_3 y_{11} + c_2 y_{21} + c_5 y_{31})$$
$$= -2 - (0(1) + 5(0) + 0(1)) = -2$$
$$r_4 = c_4 - z_4 = c_4 - (c_3 y_{14} + c_2 y_{24} + c_5 y_{34})$$
$$= 0 - (0(0) + (-5)(1) + 0(-1)) = 5$$

$$\frac{y_{10}}{y_{11}} = \frac{4}{1}, \quad \frac{y_{30}}{y_{31}} = 2$$

basic

$$\begin{array}{cccccc} 1 & 1 & 1 & & & \\ a_1 & a_2 & a_3 & a_4 & a_5 & b \\ 0 & 0 & 1 & 1 & -1 & 4 \\ 0 & 1 & 0 & 1 & 0 & 6 \\ 1 & 0 & 0 & -1 & 1 & 2 \end{array} \qquad \begin{bmatrix} 2 \\ 6 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

$$r_4 = c_4 - z_4$$
$$= -(c_3 y_{14} + c_2 y_{24} + c_1 y_{34})$$
$$= -(0(1) + (-5)(1) + (-2)(-1)) = 3$$
$$r_5 = c_5 - z_5$$
$$= 2$$

$$f(x) = -34$$

another way

	a_1	a_2	a_3	a_4	a_5	b
	1	0	1	0	0	4
	0	1	0	1	0	6
	1	1	0	0	1	8
C^T	-2	-5	0	0	0	0

$$\begin{bmatrix} 0 \\ 0 \\ 4 \\ 6 \\ 8 \end{bmatrix} \checkmark$$

$\frac{y_{2,0}}{y_{2,2}} = \frac{6}{1}$ $\quad \frac{y_{3,0}}{y_{3,2}} = \frac{8}{1} > 6$

	a_1	a_2	a_3	a_4	a_5	b	
	1	0	1	0	0	4	
	0	1	0	1	0	6	
	1	0	0	-1	1	2	✓
C^T	-2	0	0	5	0	30	

$\frac{4}{1}$ v.s. $\frac{2}{1}$

	a_1	a_2	a_3	a_4	a_5	b
	0	0	1	1	-1	2
	0	1	0	1	0	6
	1	0	0	-1	1	2
C^T	0	0	0	3	2	34

$\frac{4}{1}$ v.s. $\frac{2}{1}$

Duality in LP

Theorem (weak duality). Let x, λ be feasible solutions for a primal LP and its dual, respectively. Then, $c^T x \geq b^T \lambda$

Proof: symmetric form of duality $x \geq 0$, $\lambda \geq 0$, $Ax \geq b$, $A^T \lambda \leq c$

$$c^T x \geq \lambda^T A x \geq \lambda^T b = b^T \lambda$$

Theorem: Let x, λ be feasible solution for primal & its dual LP, if

$$c^T x = \lambda^T b$$

then x, λ must be optimal solutions for the primal & dual problem

Theorem If x^* is an optimal solution for the primal LP, then the dual problem also has an optimal solution.

Proof. Consider x^* is a solution of asymmetric form of primal LP,

$x \geq 0$, $Ax = b$ Assuming x^* is basic $x^* = \begin{pmatrix} x_b^* \\ 0 \end{pmatrix}$

Let $A = [A_b, A_n]$, $A_b \in R^{m \times m}$, $A_b \in R^{m \times n}$, $A_n \in R^{n-m}$

A_b is full rank, i.e., A_b is invertible

$$Ax = b \iff [I_m, A_b^{-1} A_n] \begin{bmatrix} x_b^* \\ 0 \end{bmatrix} = A_b^{-1} b = x_b^*$$

$$c^T x^* = c_b^T x_b^* \qquad c = \begin{pmatrix} c_b \\ c_n \end{pmatrix}$$

$$= \underbrace{c_b^T \cdot A_b^{-1}}_{\lambda} b$$

$$\lambda^* = (A_b^{-1})^T c_b \qquad A^T = \begin{bmatrix} A_b^T \\ A_n^T \end{bmatrix} \quad A^T \lambda^* = \begin{bmatrix} A_b^T (A_b^{-1})^T c_b \\ A_n^T (A_b^{-1}) c_b \end{bmatrix} = \begin{bmatrix} c_b \\ (A_b^{-1} A_n)^T c_b \end{bmatrix}$$

$$\leq \begin{pmatrix} c_b \\ c_n \end{pmatrix}$$

$$r = c_n - A_b^{-1} A_n c_b \geq 0$$

<u>Theorem</u> Let $x \cdot \lambda$ be feasible solutions for primal & dual L.P. problems, they are optimal iff

$$(c^T - \lambda^T A) x = 0 \quad \text{and} \quad \lambda^T (Ax - b) = 0$$

<u>Proof</u>: Let primal LP be assymmetric form, so $\lambda^T (Ax - b) = 0$

$(c^T - \lambda^T A) x = 0$ implies $c^T x - \lambda^T A x = 0 \Rightarrow c^T x - \lambda^T b = 0$

Note, for symmetric form of primal LP problem

$Ax - b = y \geq 0$ $\quad$ y is the slack

$\lambda^T y = 0$, if we know an optimal solution for the dual problem if $\lambda_i > 0$, $y_i = 0$

Let x^*, λ^* be optimal solutions for primal & dual LP problems. We want to change the lower bound b for the primal problem to $b + \Delta b$. Let $x^* + \Delta x$ be the new optimal solution. We want to estimate $c^T (x^* + \Delta x)$

$$\geq (b + \Delta b)^T \lambda^* \quad \Rightarrow \quad c^T \Delta x \geq \lambda^{*T} \Delta b$$

General Constrained Minimization

$$\min f_0(x), \quad x \in \mathbb{R}^n$$

subject to $f_i(x) \le 0, \quad i = 1, \dots, m$

$h_i(x) = 0, \quad i = 1, \dots, p$

Ex $f_0(x) = \frac{1}{2} x^T A x - b^T x$

subject to $Bx = C$, $B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^m$

Observation: The admissible set is an affine subset $S = x_0 + V$, $Bx_0 = C$, $V \in Null(B)$

$\dim(Null(B)) < n$

$\min f(x_0 + v) = \frac{1}{2}(x_0 + v)^T A (x_0 + v) = \frac{1}{2} v^T A v + x_0^T B v + \frac{1}{2} x_0^T A x_0$

subject to $v \in Null(B)$. Let $v_1, \dots, v_p$ be a basis for $Null(B)$, then the problem can be formulated as an unconstrained optimization problem.

Equality Constraints

Optimality Condition: Let x be a point satisfying the equality constraints such that $\{\nabla h_i(x), i = 1, \dots, p\}$ is a set of linear independent vectors, then if x is a local min for f_0, among the admissible solutions, there exists $\mu_1, \dots \mu_p$ s.t. $\nabla f_0(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) = 0$ $\iff$ FONC

Special case for equality constraints.

$h_i(x) = 0, \quad i = 1, \dots, p$

Let $S = \{x : h_i(x) = 0, \quad i = 1, \dots, p\}$

Let $x_0 \in S$, we define $T(x_0)$ as the collection of vectors, with the property that

$v \in T(x_0)$, then $\exists\, x_k \in S$ $\lim_{k \to \infty} \frac{x_k - x_0}{\|x_k - x_0\|} = v$ (tangent vector)

$$0 = h_i(x_k) - h_i(x_0) \cong \nabla h_i(x_0)^T (x_k - x_0)$$

$$\lim_{k \to \infty} \nabla h_i(x_0)^T \frac{x_k - x_0}{\|x_k - x_0\|} = 0 \Rightarrow \nabla h_i(x_0)^T \cdot v = 0, \; v \in T(x_k), \; i = 1, \dots, p$$

Let $N(x_0)$ to be the normal vectors of S at x_0. We define

$N(x_0) = T(x_0)^\perp$ that is, $u \in N(x_0)$ if and only if for any $v \in T(x_0)$, $u^T v = 0$

$\{\nabla h_i(x_0), i = 1, \dots, p\} \subseteq N(x_0)$

When does $N(x_0) = sp\{\nabla h_i(x_0)\}$?

FONC: $\forall v \in T(x_0)$, $\nabla f_0(x_0)^T v = 0$.

$\Rightarrow \nabla f_0(x_0) \in N(x_0)$

If $N(x_0) = sp\{\nabla h_i(x_0)\}$, then FONC says if x_0 is a local min, then $\exists\, \mu_1, \mu_p$ s.t.

$$\nabla f(x_0) = \sum_{i=1}^{p} \mu_i \nabla h_i(x_0)$$

<u>Ex</u> $\min \frac{1}{2} x^T A x - b^T x$

subject to $B^T x = c, \quad B \in \mathbb{R}^{p \times n}, \quad c \in \mathbb{R}^p$

$B = [B_1, \dots, B_p] \quad B^T x = c \iff c_k = B_k^T x, \quad B_k^T x - c_k = h_k(x), \quad \nabla h_k(x) = B_k$

$f_0(x) = \frac{1}{2} x^T A x - b^T x$

$\nabla f_0(x) = Ax - b$

If x_0 is a local min of f_0, then $Ax_0 - b = \underbrace{\sum_{k=1}^{p} B_k \mu_k}_{\text{linear combination}} = B\mu, \ \mu = \mathbb{R}^n$

$\Rightarrow$ solve $Ax_0 - B\mu = b$ (n equations)

$B^T x_0 = c$ (p equations)

$$\begin{bmatrix} A & -B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ \mu \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}$$

symmetric

Ex $\min c^T x$

subject to $\frac{1}{2} x^T A x - b^T x + d = 0 = h(x)$

$n = 2$

level curve of $f_0(x) = c^T x$

x is a candidate for local min if

$\exists \mu$ s.t. $\nabla f_0(x) + \mu \nabla h(x) = 0$

$\nabla f_0(x) = c$

$\nabla h(x) = Ax - b$

$c + \mu(Ax - b) = 0$

$h(x) = 0$

$c + \mu A x - \mu b = 0$

$\mu A x = \mu b - c$

$A x = \frac{1}{\mu}(\mu b - c)$

$x = \frac{1}{\mu} A^{-1}(\mu b - c)$

$h(x) = \frac{1}{2}\left(\frac{1}{\mu} A^{-1}(\mu b - c)\right)^T A \left(\frac{1}{\mu} A^{-1}(\mu b - c)\right) - b^T\left(\frac{1}{\mu} A^{-1}(\mu b - c)\right) + d = 0$

$\frac{\mu^2}{2} b^T A^{-1} b - \cancel{\mu c^T A^{-1} b} + \frac{1}{2} c^T A c - \mu^2 b^T A^{-1} b + \cancel{\mu c^T A^{-1} b} + \mu^2 d = 0$

$-\frac{\mu^2}{2} b^T A^{-1} b + \frac{1}{2} c^T A c + \mu^2 d = 0$

$\mu^2 \left(\frac{1}{2} b^T A^{-1} b - d\right) = \frac{1}{2} c^T A c$

Computationally, we need to find solutions to system of nonlinear equations

$\nabla f_0(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) = 0$ n equations

$h_i(x) = 0$ p equations

Newton's method to solve?

$G(x;\mu) : \mathbb{R}^n \times \mathbb{R}^p \longmapsto \mathbb{R}^n \times \mathbb{R}^p$

$$G(x;\mu) = \begin{pmatrix} \nabla f_0(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) \\ h_1(x) \\ \vdots \\ h_p(x) \end{pmatrix}$$

Newton's Method

Start with x_0, $G(x_k + \Delta x, \mu^{(k)} + \Delta \mu) \simeq G(x_k, \mu^{(k)}) + \nabla G(x_k, \mu^{(k)}) \begin{pmatrix} \Delta x \\ \Delta \mu \end{pmatrix} + O(\ )$

Gradient   Jacobian

$$\begin{pmatrix} x^{k+1} \\ \mu^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \mu^k \end{pmatrix} - \nabla G^{-1} G(x^k; \mu^k)$$

Special Case

$\min \ f_0(x)$

subject $Ax - b = 0$

$$G(x,\mu) = \begin{pmatrix} \nabla f(x) + A^T \mu \\ Ax - b \end{pmatrix}$$

Jacobian

$$\nabla G(x,\mu) = \begin{array}{c} \begin{array}{cc} n\times n & n\times p \end{array} \\ \begin{bmatrix} H_f(x) & A^T \\ A & 0 \end{bmatrix} \\ \begin{array}{cc} p\times n & \end{array} \end{array}$$

Inequality Constraints

$\min \ f_0(x)$

subject to $f_i(x) \le 0, \ i = 1, \dots, m$

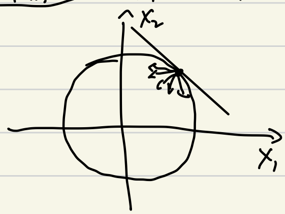
Let x satisfy the constraints

If $f_i(x) < 0$ for a given i, the constraint i is said to be inactive at x

Only active constraints limit the admissible direction. In particular, a direction d is admissible if and only if $\nabla f_i(x)^T d \le 0$. for all i s.t. $f_i(x) = 0$

Definition : The Tangent cone of the admissible set at x is defined as the collection of vectors d such that $\nabla f_i(x)^T d \le 0$

Definition : A set K of vectors is said to be a cone if $\forall x \in K$, and $\lambda \ge 0$
$\lambda x \in K$.



FONC If x is a local min of f_0, then $\nabla f_0(x)^T d \ge 0$, $\forall d \in T(x_0)$

Constrained Nonlinear Optimization characterization of solutions

Equality Constraints

$$\min f_0(x), \ x \in \mathbb{R}^n$$

subject to $h_i(x) = 0, \ i = 1, \dots, p$

FONC . If x^* is a solution, then $\exists \mu_i \ \ i = 1, \dots, p$, such that

$$\nabla f_0(x^*) = \sum_{i=1}^{p} \mu_i \nabla h_i(x^*)$$

SONC If x^* satisfies FONC, then for any vector $d \in T(x^*)$

$$d^T H_{f_0}(x^*) d \ge 0 . \ T(x^*) = (sp\{\nabla h_i(x^*)\})^\perp$$

Inequality Constraints

If x^* is a local minimum of f_0 satisfying $f_i(x) \le 0, \ i = 1, \dots, m$

Then $\nabla f_0(x^*) = -\sum_{i=1}^{m} \lambda_i \nabla f_i(x^*)$, for $\lambda_i \ge 0$.

Definition: A set K is called a cone if $\forall x \in K$, $\forall \alpha > 0$, $\alpha x \in K$

Definition: Let S be a set of vectors, a cone generated by S is defined as

$$K = \{\alpha x,\ x \in S,\ \alpha \geq 0\} = co(S)$$

Let $\{v_1, v_2, \dots, v_m\}$, $co\{v_1, v_2, \dots, v_m\}$

Definition: The dual cone of a cone K is defined by $K^* = \{y,\ y^T x \geq 0,\ \forall x \in K\}$

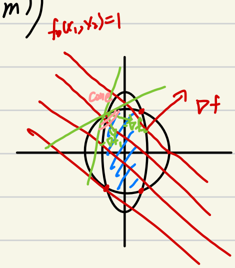
1) Show K^* is a cone
2) Equivalent statement for FONC of inequality constraints is

$$\nabla f_0(x^*) \in \left(co(-\nabla f_i(x^*),\ i = 1, \dots, m)\right)^*$$

Ex, $\min\ 3x_1 + 2x_2$

subject to $x_1^2 + x_2^2 - 1 \leq 0$ $\quad f_1(x)$

$x_1^2 + \frac{x_2^2}{4} - \frac{1}{2} \leq 0$ $\quad f_2(x)$



Complete Nonlinear Constrained Optimization Problem

$$\min\ f_0(x)$$

subject to $f_i(x) \leq 0,\ i = 1, \dots, m$

$h_i(x) = 0,\ i = 1, \dots, p$

Theorem (Karush-Kuhn-Tucker, KKT condition)

If a point x^* is a local minimum of f_0 under the constraints, there must be a vector $\lambda \in \mathbb{R}^m$, $\lambda \geq 0$, a vector $\mu \in \mathbb{R}^p$ such that

$$\nabla f_0(x^*) + \sum_{1}^{m} \lambda_i \nabla f_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) = 0$$

and $\lambda_i f_i(x^*) = 0$, $\forall i = 1, \dots, m$, $h_i(x^*) = 0$, $i = 1, \dots, p$.

$f_i(x^*) \leq 0$, $i = 1, \dots, m$.

We define a Lagrange function for this general constrained minimization problem as

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

$$x \in D = Dom(f_0) \cap \bigcap_{i=1}^{m} Dom(f_i(x)) \bigcap_{i=1}^{p} Dom(h_i(x))$$

$$(*) \iff \nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

$$\nabla_\mu L(x^*, \lambda^*, \mu^*) = 0$$

$$\nabla_\lambda L(x^*, \lambda^*, \mu^*) = \begin{pmatrix} f_1(x^*) \\ f_m(x^*) \end{pmatrix}$$

$$\lambda^T \nabla_\lambda L(x^*, \lambda^*, \mu^*) = 0$$

Let a function $g : \mathbb{R}^m \times \mathbb{R}^p$ be defined by

$$g(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$$

g is called the Lagrange dual of the minimization problem.

$g(\lambda, \mu)$ can take value $-\infty$

Suppose p^* is the minimal value of f_0 in the admissible set. Then

$$g(\lambda, \mu) \le p^* \le f_0(x), \ x \text{ admissible}$$

Notes by Dongwei Zhang when I was absent

11/5

Nonlinear - Constrained Minimization.

Primal - $\min f_0(x)$

subject $f_i(x) \leq 0$, $i = 1, \dots m$.

$h_i(x) = 0$, $i = 1 \dots p$

Lagrange functional. $L: \mathcal{D} \times R_+^m \times R^p$ $\mathcal{D} = \mathrm{Dom}(f_0)$

$\cap \bigcap_{i=1}^{m} \mathrm{Dom}(f_i) \cap \bigcap_{i=1}^{p} \mathrm{Dom}(h_i)$

$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$

Lagrange dual

$$g(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x \lambda \mu).$$

Ex LP $\min C^T x$. $\quad x \in R^n$ $A \in R^{m \times n}$ $b \in R^m$.

subj to $x \geq 0$

$Ax \geq b$

$-x \leq 0$

$b - Ax \leq 0$

$L(x, \lambda_1, \lambda_2) = C^T x - \lambda_1 x + \lambda_2 (b - Ax)$

$\lambda_1 \in R^n$, $\lambda_2 \in R^m$ $\mathcal{D} = R^n \times R_+^{m+n}$.

$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \in R_+^{m+n}$. $\quad g(\lambda_1, \lambda_2) = \begin{cases} -\infty & C - \lambda_1 - A^T\lambda_2 \neq 0 \\ \lambda_2^T b & \text{otherwise.} \end{cases}$

Dual $\max \lambda_2^T b$,

subj to $\lambda_1 \lambda_2 \geq 0$

$C - A^T \lambda_2 = \lambda_1$

$C - A^T \lambda_2 \geq 0$

$\max \lambda_2^T b$

s.t. $\lambda_2 \geq 0$

$A^T y$

Ex $\min \frac{1}{2} x^T A x - b^T x$. $\quad b \in R^n$, $A \in R^{n \times n}$, $A > 0$

subj to $Bx = C$ $\quad B \in R^{p \times n}$ $C \in R^p$.

Ex. $\min \frac{1}{2} x^T A x - b^T x$ $\quad b \in R^n$, $A \in R^{n \times n}$ $A > 0$

subj to. $Bx = C$. $\quad B \in R^{p \times n}$, $C \in R^p$

$L(x, \mu) = \frac{1}{2} x^T A x - b^T x + \mu^T (Bx - c)$ $\quad \mu \in R^p$

$= \frac{1}{2} (x - A^{-1}(b - B^T\mu))^T A (x - A^{-1}(b - B^T\mu)) - \frac{1}{2}(b - B^T\mu)^T A^{-1}$

$-\mu^T C$ $\quad g(\mu) = -\frac{1}{2}(b - B^T\mu)^T A^{-1}(b - B^T\mu) - \mu^T C$

$g(\mu) = -\frac{1}{2}(b - B^T\mu)^T A^{-1}(b - B^T\mu) = \mu^T C = -\frac{1}{2}\mu^T B A^{-1} B^T \mu$

$+ b^T A^{-1} B^T \mu - \frac{1}{2} b^T A^{-1} b - C^T \mu.$

We assume B is full rank. $\text{rank}(B) = p \leq n$

In the case $D = BA^{-1}B^T$ is positive definite

$g(\mu) = -\frac{1}{2}\mu^T D \mu - C^T\mu - \frac{1}{2} b^T A^{-1} b = \frac{1}{2}(\mu + D^{-1}C)^T D(\mu + D^{-1}C)$

$+ \frac{1}{2} C^T D^{-1} C - \frac{1}{2} b^T A^{-1} b.$ Dual Problem:

$\max g(\mu) \Rightarrow$ optimal solution $\mu^* = -D^{-1}C$

$x^* = A^{-1}(b - B^T\mu^*) = A^{-1}b - A^{-1}B^T\mu^*$

Ex. $\min \frac{1}{2}x^T A x - b^T x$ $\quad b \in \mathbb{R}^n,\ A \in \mathbb{R}^{n \times n},\ A > 0$

subj to $\boxed{Bx = C}$

$B \in \mathbb{R}^{p \times n},\ C \in \mathbb{R}^p$

$Bx^* = BA^{-1}b - BA^{-1}B^T(-(BA^{-1}B^T)^{-1}(C - BA^{-1}b))$

$= BA^{-1}b + C - BA^{-1}b = C$

Ex $\min C^T x$: subj to . $\frac{1}{2}x^T A x - b^T x \leq r^2$

$\min 3x_1 + 2x_2 - 4x_3$, subj to $x_1^2 + \frac{x_2^2}{4} + \frac{x_3^2}{9} + x_1 - x_2 \leq 25$

Duality

Primal	Dual
minimize $C^T x$	minimize $\lambda^T b$
subj to $Ax \geq b$	subj to $\lambda^T A \leq C^T$
$x \geq 0$	$\lambda \geq 0$

symmetric

Primal

$\min C^T x$ subj to $Ax = b$ $\quad Ax \geq b$ $\quad \min C^T x$

$x \geq 0$ $\quad Ax = b \Leftrightarrow \begin{cases} Ax \geq b \\ -Ax \geq -b \end{cases}$ $\quad \Rightarrow$ subject to $\begin{bmatrix} A \\ -A \end{bmatrix} x \geq \begin{bmatrix} b \\ -b \end{bmatrix}$

$x \geq 0$

Constrained Nonlinear Minimization.

$$\begin{cases} \min & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i=1,\dots,m \\ & h_i(x) = 0, \quad i=1,\dots,p \end{cases}$$

Convex Optimization

$$\left[\begin{array}{l} f_0, f_1, \dots, f_m \text{ convex} \\ Ax - b = 0 \end{array} \right.$$

Convex Optimization

$$\min \; f_0(x)$$

subject to $f_i(x) \le 0, \; i=1,\dots,m$

$Ax - b = 0, \; A \in \mathbb{R}^{p \times n}, \; b \in \mathbb{R}^p$

$f_0, f_1, \dots f_m$ are convex

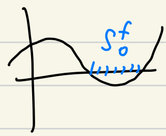
Convex function

$$f(tx + (1-t)y) \le t f(x) + (1-t) f(y)$$

Concave function : $-f$ is convex

Definition: Let f be a function from $\mathbb{R}^n \mapsto \mathbb{R}^1$. A sublevel set S_α^f is defined as

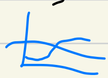
$$S_\alpha^f = \{ x \in \mathbb{R}^n, f(x) \le \alpha \}$$



Lemma 1. If f is convex, then for any α, S_α^f is convex

Proof : Let $x, y \in S_\alpha^f$, for any $t \in [0,1]$, $f(tx + (1-t)y) \le t \underbrace{f(x)}_{\le \alpha} + (1-t) \underbrace{f(y)}_{\le \alpha} \le \alpha$

The converse is false

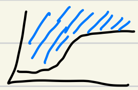


- Graph of a function $f: \mathbb{R}^n \mapsto \mathbb{R}^1$ is a subset in $\mathbb{R}^{n+1}$

$$G(f) = \{ (x,t), \; x \in \mathbb{R}^n, \; t = f(x) \}$$

- Epigraph of a function $f : \mathbb{R}^n \mapsto \mathbb{R}^1$ is a subset in $\mathbb{R}^{n+1}$ defined by

$$\operatorname{epi}(f) = \{ (x,t), \; x \in \mathbb{R}^n, \; t \ge f(x) \}$$



above the line

- Lemma: A function f is convex iff its epigraph is convex
- Proof : f is convex, let $(x,t), (y,s) \in \operatorname{epi}(f)$

$$\theta(x,t) + (1-\theta)(y,s) \in \operatorname{epi}(f)$$

show: $f(\theta x + (1-\theta) y) \le \theta t + (1-\theta) s.$

$$\begin{aligned} f(\theta x + (1-\theta) y) &\le \theta f(x) + (1-\theta) f(y) \\ &\le \theta t + (1-\theta) s \end{aligned}$$

Conversely, if epi(f) is convex, show f must be convex.

take $x, y \in \mathbb{R}^n$, $\forall \theta \in [0,1]$

Show $f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$

$(x, f(x)), (y, f(y)) \in epi$

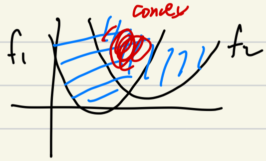
$\theta(x, f(x)) + (1-\theta)(y, f(y)) \in epi$

$(\theta x + (1-\theta) y, \theta f(x) + (1-\theta) f(y)) \in epi$

Lemma: Let f_α be a family of convex functions for $\alpha \in \mathcal{A}$. Then

$$f_{max}(x) = \sup_{\alpha \in \mathcal{A}} f_\alpha(x) \quad \text{is convex}$$

proof: Intersection of epigraph is convex



Corollary: Let $g(\lambda, \mu)$ be the lagrange dual function for a general constrained minimization problem, i.e.

$\min f_0(x)$

subject to $f_i(x) \le 0, \quad i = 1, \dots, m$

$h_i(x) = 0, \quad i = 1, \dots, p$

Then $g(\lambda, \mu)$ is <u>concave</u>

<u>Proof</u>: $g(\lambda, \mu) = \inf_x f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)$

$$= -\sup_{x \in \mathcal{D}} \Big(-\underline{f_0(x)} - \sum_{i=1}^m \underbrace{\lambda_i}_{\text{constant}} f_i(x) - \sum_{i=1}^p \mu_i \underline{h_i(x)} \Big) \rightarrow \text{affine (linear)} \Rightarrow \text{convex and concave}$$

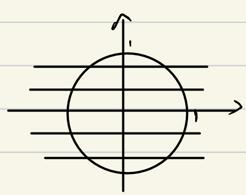
Slate Condition: We assume that $\exists\, \hat{x} \in \operatorname{int} \mathcal{D}$ (interior), s.t. $f_i(\hat{x}) < 0$, $i = 1, \dots, m$

If f_i is not affine, then the strong duality holds, ie

$\max g(\lambda, \mu) = \min f_0(x)$ subject to constraints

<u>Ex</u> $\min x_2$

subject to $x_1^2 + x_2^2 - 1 = 0$



$\Downarrow$

Solve: $L(x,\mu) = x_2 + \mu(x_1^2 + x_2^2 - 1)$ $\qquad$ Do this for final.

$$g(\mu) = \begin{cases} -\infty, & \mu \le 0 \\ -\frac{1}{4\mu} - \mu, & \mu > 0 \end{cases} \qquad \Downarrow$$

$$\underbrace{x_2 + \mu x_1^2 + \mu x_2^2}_{\mu(x_2^2 + \frac{1}{\mu}x_2)} - \mu$$

$$= \mu x_1^2 + \mu(x_2 + \tfrac{1}{2\mu})^2 - \frac{1}{4\mu} - \mu$$

Dual Problem

$\max \; -\frac{1}{4\mu} - \mu$

subject to $\mu > 0$

$$g'(\mu) = \frac{1}{4\mu^2} - 1 = \frac{4\mu^2 - 1}{4\mu^2}, \quad \mu^* = \frac{1}{2}, \quad g(\tfrac{1}{2}) = -\tfrac{1}{2} - \tfrac{1}{2} = -1$$

$$\boxed{d^* = -1}$$

Computational Method for Convex Optimization Problem.

Sometimes g is not differentiable

Equality Constraints

$\min \; f(x)$

subject to $\boxed{Ax - b = 0}$ equality constraint

$$L(x,\mu) = f(x) + \mu^T(Ax - b)$$

$$\nabla_x L(x,\mu) = \nabla f(x) + A^T\mu \qquad \underset{\text{Hessian}}{\nabla^2 L(x,\mu)} = \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$

$$\nabla_\mu L(x,\mu) = Ax - b$$

Newton's Method

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} - \left(\nabla^2 L(x,\mu)\right)^{-1} \begin{pmatrix} \nabla_x f \\ \nabla_\mu f \end{pmatrix}$$

hard to do

solutions may not satisfy constraints

$\therefore$ Need to assume/choose x_0 satisfies constraints.

Note: (x^*, μ^*) is a saddle point for $L(x,\mu)$, ie $L(x^*,\mu^*) = \min_x L(x,\mu^*)$

$$= \max_\mu L(x^*,\mu)$$

So no gradient descent

Special case: $f(x) = \frac{1}{2}x^TQx - c^Tx$

$$\nabla^2 L(x,\mu) = \begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix}, \quad w = \begin{pmatrix} y \\ \lambda \end{pmatrix},$$

$$w^T\nabla^2 L(x,\mu)w = y^TQy + 2\lambda^TAy = (\underbrace{y + Q^{-1}A^T\lambda}_{z})^T Q(\underbrace{y + Q^{-1}A^{-1}\lambda}_{z}) - \lambda^TAQ^{-1}A^T\lambda$$

$$= z^TQz - \lambda^TAQ^{-1}A^T\lambda$$

$$\begin{bmatrix} Q & 0 \\ 0 & AQ^{-1}A^T \end{bmatrix}$$

At each step k, if $\nabla L(x_k, \mu_k)$ is sufficiently small in norm, the algorithm will stop.

Suppose x_0 satisfies $Ax_0 - b = 0$

$$\begin{pmatrix} \Delta x \\ \Delta \mu \end{pmatrix} = -\left(\nabla^2 L(x_k, \mu_k)\right)^{-1} \nabla L(x_k, \mu_k).$$

$$\nabla^2 L(x_k, \mu_k) \begin{pmatrix} \Delta x \\ \Delta \mu \end{pmatrix} = -\nabla L(x_k, \mu_k)$$

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \mu \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) + A^T \mu_k \\ Ax_k - b \end{pmatrix}$$

$$\begin{cases} \nabla^2 f(x) \Delta x + A^T \Delta \mu = -(\nabla f(x_k) + A^T \mu_k) \\ A \Delta x = -(A x_k - b) \end{cases}$$

$$\begin{cases} \Delta x + (\nabla^2 f(x_k))^{-1} A^T \Delta \mu = -(\nabla^2 f(x_k))^{-1} (\nabla f(x_k) + A^T \mu_k) \\ A \Delta x + A (\nabla^2 f(x_k))^{-1} A^T \Delta \mu = -A (\nabla^2 f(x_k))^{-1} (\nabla f(x_k) + A^T \mu_k) \end{cases}$$

$$A (\nabla^2 f(x_k))^{-1} A^T \Delta \mu = -A (\nabla^2 f(x_k))^{-1} (\nabla f(x_k) + A^T \mu_k) + (A x_k - b)$$

If $A(\nabla^2 f(x_k))^{-1} A^T$ is invertible, this gives us $\Delta \mu$

To find Δx, we need to find $\Delta \mu$ plug into this

$A(x_k + \Delta x) = b$

Quasi Newton's Method // In case we can't compute $\nabla^2 f(x)$

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} - \begin{pmatrix} H_k & A^T \\ A & 0 \end{pmatrix}^{-1} \nabla L(x_k, \mu_k)$$

$$\min f_0(x)$$

subject to $f_i(x) \leq 0$, $i = 1, \dots, m$

Penalty Approach

$$\min f_0(x) + \sum x_i I_-(f_i(x)), \quad I_-(t) = \begin{cases} 0 & t \leq 0 \\ +\infty & t > 0 \end{cases}$$

subj to $Ax - b = 0$

Idea: Replace I_- with $\hat{I}_-(s) = \begin{cases} -\log(-s), & s < 0 \\ +\infty, & s \geq 0 \end{cases}$

$\Rightarrow \min f_0(x) + \sum_{i=1}^{m} \frac{1}{t} \hat{I}_-(f_i(x))$

$\Rightarrow \min t f_0(x) + \sum_{i=1}^{m} -\log(-f_i(x))$

Let $x^*(t)$ be the optimal solution to the problem t. Then we have $Ax^*(t) - b = 0$

$$t \nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{f_i(x^*(t))} \nabla f(x^*(t)) + \frac{\hat{\mu}^T}{t} A = 0$$

$f_i(x^*(t)) \le 0, \; i = 1, \ldots, m$

We define $\lambda_i^* = -\frac{1}{t f_i(x^*(t))}$, $\mu^* = \frac{\hat{\mu}}{t}$

Consider the Lagrange function

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \mu^T \overbrace{(Ax - b)}^{0}$$

$$L(x^*(t), \lambda^*(t), \mu^*(t)) = f_0(x^*(t)) - \sum_{i=1}^{m} \frac{1}{t f_i(x^*(t))} f_i(x^*(t))$$

$$= f_0(x^*(t)) - \frac{m}{t}$$

Since $\nabla_x L(x^*(t), \lambda^*(t), \mu^*(t)) = 0$

$$\therefore \boxed{= g(\lambda^*(t), \mu^*(t)) \le p^*}$$

$$p^* + \frac{m}{t} \ge f_0(x^*(t)) \ge p^*$$

$$f(x) = t f_0(x) + \sum_{i=1}^{m} -\log(-f_i(x))$$

$$\nabla f(x) = t \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{f_i(x)} \nabla f_i(x)$$

$$\nabla^2 f(x) = t \nabla^2 f_0(x) + \sum_{i=1}^{m} \left[-\frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \frac{1}{f_i(x)} \nabla^2 f_i(x) \right]$$

need to take t large enough so that $\nabla^2 f(x)$ is positive definite.

In summary, we can use Newton's method as follows

1) Start with x_0 such that $f_i(x_0) < 0$, $i = 1, \ldots, m$, $\mu_0 \in \mathbb{R}^p$

2) "Select sufficient large t". At each step k.

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} - \begin{bmatrix} \nabla^2 F(x_k; t) & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{pmatrix} \nabla F(x_k; t) + A^T \mu_k \\ Ax_k - b \end{pmatrix}$$

for each t, we would like to have

$$\left.\begin{array}{l} x_k \to x^*(t) \\ \mu_k \to \mu^*(t) \end{array}\right\} p^* + \frac{m}{t} \ge f_0(x^*(t)) \ge p^*$$

to guarantee convergence, we typically

1) $\|\nabla^2 L(x, \mu) - \nabla^2 L(\hat{x}, \hat{\mu})\| \le K \left\| \begin{pmatrix} x \\ \mu \end{pmatrix} - \begin{pmatrix} \hat{x} \\ \hat{\mu} \end{pmatrix} \right\|$ (Lipshitz continuous)

2) $\|\nabla^2 L(x, \mu)^{-1}\| \le C$

since $\nabla^2 L(x, \mu) = \begin{bmatrix} \nabla^2 f(x; t) & A^T \\ A & 0 \end{bmatrix}$, $\therefore \nabla^2 L(x, \mu) - \nabla^2 L(\hat{x}, \hat{\mu})$

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

$$\begin{bmatrix} \nabla f(x) & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = (\overbrace{u + Q^{-1} A^T u}^{z}) Q (\overbrace{u + Q^{-1} A^T u}^{z}) - v^T A Q^{-1} A^T v$$

$$= \begin{pmatrix} z \\ v \end{pmatrix}^T \begin{bmatrix} Q & 0 \\ 0 & K \end{bmatrix} \begin{pmatrix} z \\ v \end{pmatrix} \quad K = AQ^{-1} A^T$$

$Q^{-1} = (\nabla f^2(x))^{-1}$

$K^{-1} = [A \nabla f^2(x) A^T]^{-1}$

Global Optimization
- Genetic Algorithm
- Naive Bayes Algorithm

Nelder-Mead Algorithm
- Often used in chemistry.