1 Fundamentals of unconstrained optimization

1. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, write A as a sum of rank 1 matrices. Use the Eigenvalue decomposition of A.

2. Let $a \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Define $f_1(x) = a^\top x$ and $f_2(x) = \frac{1}{2}x^\top Ax$. Compute ∇f_1, $\nabla^2 f_1$, ∇f_2, and $\nabla^2 f_2$.

3. Let f be a strictly convex function. Show that if x^* is a local minimizer then it is a global minimizer.

4. Show that $-\nabla f$ is a descent direction.

5. Let $f(x,y) = (x+y^2)^2$. Show that $p = [-1,1]^\top$ is a descent direction at $[1,0]^\top$. State the steepest descent direction and Newton's direction. What is the optimum of this function? For Newton's method how many steps do we need to reach such optimum?

2 Newton and Quasi-Newton methods

1. State Newton's update formula.
2. State two differences between Newton's method and steepest descent (other than the formulas).
3. Let $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ be a constant matrix (i.e. independent of x). Let $d \in \mathbb{R}^n$ such that it solves the following system:

$$0 = \nabla f(y) + \nabla^2 f(y) d.$$

   Prove that $y + d$ is a stationary point of f.

4. Let:

$$B_k = (I - \rho\gamma s^\top)B_{k-1}(I - \rho s\gamma^\top) + \rho\gamma\gamma^\top,$$

   with $\rho = 1/(\gamma^\top s)$, γ, s fixed vectors.

   (a) Which update formula is this?

   (b) What does B_k approximates? What is the difference between this method and Newton's method?

   (c) Prove that $B_k s = \gamma$

   (d) Prove that if B_{k-1} is spd and $s^\top\gamma > 0$ then B_k is also spd. What is the practical importance of this?

   (e) Prove that if $B_{k-1}s = \gamma$ then $B_k = B_{k-1}$. What does this mean?

3 KKT Conditions

1. State the KKT conditions for the following optimization problem:

$$
\begin{aligned}
\min\ & f_0(x) & & \\
\text{s.t. }\ & f_k(x) = 0 & & k = 1, ..., m \\
& g_j(x) \leq 0 & & j = 1, ..., r
\end{aligned}
$$

2. Solve the following optimization problem using the KKT conditions:

$$
\begin{aligned}
\min\ & -xy \\
\text{s.t. }\ & x + y = 10
\end{aligned}
$$

3. Using the KKT conditions, find the point on the circle $x^2 + y^2 = 80$ that is closest to $(1, 2)$. You are going to find to KKT points, what is the relation between that other non optimal point and $(1, 2)$?

4 Duality

1. What is strong duality? What is weak duality?
2. State the definition of the Lagrange dual function
3. How does the Lagrange dual function relate to the dual LP problem?
4. Find the Lagrange dual function of the following problem:

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

5. Consider the following LP in standard form:

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

   Compute the Lagrangua dual function. What can you tell about the dual LP problem from this function?

5 Quadratic programming

1. Consider the following optimization problem:

$$\min q(x) = \frac{1}{2}x^\top Qx + c^\top x \tag{1}$$
$$\text{s.t. } Ax = 0, \tag{2}$$

where $Q \in \mathbb{R}^{n\times n}$ is spd, $A \in \mathbb{R}^{m\times n}$ with $rank(A) = m$.

(a) Let x^* be the optimum, prove that $q(x^*) \leq 0$.

(b) Prove that $q(x^*) = 0 \iff x^* = 0$.

2. Consider the problem of finding the point on a hyperplane H that has the minimum distance to a fixed point x_0. This hyperplane is defined as:

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

where x_0, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ are fixed and $rank(A) = m$.

(a) Write this as a constrained optimization problem.

(b) Write down the Lagrange function for this problem

(c) Deduce that the solution is given by:

$$x^* = x_0 + A^\top(AA^\top)(b - Ax_0)$$
$$\lambda^* = -(AA^\top)^{-1}(b - Ax_0).$$

Let $A = QDQ^T$

$Q = [q_1 \cdots q_n]$ $\quad Q^T = \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$

$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

Then:

$$A = QDQ^T = [q_1 \cdots q_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = [q_1 \cdots q_n] \begin{bmatrix} \lambda_1 q_1^T \\ \vdots \\ \lambda_n q_n^T \end{bmatrix} = \sum_{k=1}^{n} \lambda_k q_1 q_1^T$$

$f_1(x) = a^T x$

$f_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{k=1}^{n} a_i x_i$

$\frac{\delta f_1}{\delta x_k} = a_k \Rightarrow \nabla f_1(x) = a$

$\Rightarrow \nabla^2 f_1(x) = 0 \in \mathbb{R}^{n \times n}$

$f_2(x) = \frac{1}{2} x^T A x = \frac{1}{2} x^T \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$f_2(x) = \frac{1}{2} x^T \begin{bmatrix} \sum_{k=1}^{n} a_{1k} x_k \\ \vdots \\ \sum_{k=1}^{n} a_{nk} x_k \end{bmatrix} = \frac{1}{2} [x_1 \cdots x_n] \begin{bmatrix} \sum_{k=1}^{n} a_{1k} x_k \\ \vdots \\ \sum_{k=1}^{n} a_{nk} x_k \end{bmatrix}$

$= \frac{1}{2} \sum_{i=1}^{n} x_i \sum_{k=1}^{n} a_{ik} x_k$

$\Rightarrow \frac{\delta f_2}{\delta x_j} = \frac{1}{2} \left(\sum_{k=1}^{n} a_{jk} x_k + \sum_{k=1}^{n} x_k a_{kj} \right)$

$= \frac{1}{2} \left(2 \sum_{k=1}^{n} a_{kj} x_k \right) = \sum_{k=1}^{n} a_{kj} x_k$

$\Rightarrow \nabla f_2 = \begin{bmatrix} \sum_{k=1}^{n} a_{k1} x_k \\ \sum_{k=1}^{n} a_{k2} x_k \\ \vdots \\ \sum_{k=1}^{n} a_{kn} x_k \end{bmatrix} = Ax$

Then:

$\frac{\delta^2 f_2}{\delta x_i \delta x_j} = a_{ij} \Rightarrow \nabla^2 f = A$

Let x^* be a local minimum, this means that for a ball of radius $\varepsilon > 0$ around x^*, $f(x^*) \leq f(x)$, $x \in B_\varepsilon(x^*)$.

We know that the function is convex, then for $t \leq \varepsilon$:

$$f(x^*) \leq f(tx + (1-t)x^*)$$

$$\leq tf(x) + (1-t)f(x^*)$$

Now suppose that there is another x such that $f(x) < f(x^*)$, then:

$$f(x^*) \leq f(tx + (1-t)x^*) \leq tf(x) + (1-t)f(x^*) < tf(x^*) + (1-t)f(x^*) = f(x^*).$$

This is a contradiction, thus x^* must be a global minimum.

$$f(x,y) = (x+y^2)^2$$

$$\Rightarrow \nabla f(x,y) = \begin{bmatrix} 2(x+y^2) \\ 4(x+y^2)y \end{bmatrix}$$

At $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ we have that $\nabla f(1,0) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

For the direction $p = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$,

$$p^T \nabla f(1,0) = [-1, 1] \begin{bmatrix} 2 \\ 0 \end{bmatrix} = -2 < 0$$

So p is a descent direction.

The steepest descent direction is $-\nabla f(1,0) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$.

For Newton's method we need to calculate the Hessian:

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 4y \\ 4y & 4(x+y^2)+8y^2 \end{bmatrix} \Rightarrow \nabla^2 f(1,0) = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

$$\Rightarrow \left(\nabla^2 f(1,0)\right)^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}$$

Then Newton's direction is:

$$d_N = -\left(\nabla^2 f(1,0)\right)^{-1} \nabla f(1,0) = -\begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Let $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ be a constant matrix (i.e. independent of x). Let $d \in \mathbb{R}^n$ such that

$$0 = \nabla f(y) + \nabla^2 f(y) d$$

Prove that $y+d$ is a stationary point

Using Taylor's series:

$$f(y+d) = f(y) + d^T \nabla f(y)$$

Then:

$$\nabla f(y+d) = \nabla f(y) + \nabla^2 f(y) d = 0$$

Let $B_k = (I - \rho \gamma s^T) B_{k-1} (I - \rho s \gamma^T) + \rho \gamma \gamma^T$ with $\rho = \frac{1}{\gamma^T s}$

with $s, \gamma \in \mathbb{R}^n$ fixed vectors such that $s^T\gamma \neq 0$.

a) Prove that

$$B_k s = \gamma$$

$$\left[(I - \rho \gamma s^T) B_{k-1} (I - \rho s \gamma^T) + \rho \gamma \gamma^T\right] s$$

$$= (I - \rho \gamma s^T) B_{k-1} (I - \rho s \gamma^T) s + \rho \gamma \gamma^T s$$

$$= (I - \rho \gamma s^T) B_{k-1} (s - \rho s \gamma^T s) + \gamma$$

$$= (I - \rho \gamma s^T) B_{k-1} (0) + \gamma = \gamma$$

b) If B_{k-1} is symmetric positive definite and $s^T\gamma > 0$ then B_k is spd

Note that if B_{k-1} is symmetric then B_{k-1} is also symmetric. Write $B_{k-1} = QDQ^T$ (its Eigenvalue factorization). Since all its Eigenvalues are positive, then set $\sqrt{B_{k-1}} = QD^{1/2}Q^T$, where $D^{1/2} = \text{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\}$. Then for any $x \in \mathbb{R}^n$:

$$x^T B_k x = x^T (I - \rho \gamma s^T) B_{k-1}^{1/2} B_{k-1}^{1/2} (I - \rho s \gamma^T) x + \rho\, x^T \gamma \gamma^T x$$

$$= \left(B^{1/2}(I-\rho s\gamma)x\right)^T\left(B^{1/2}(I-\rho s\gamma)x\right) + \rho(\gamma^T x)^2$$

$$= \|B^{1/2}(I-\rho s\gamma)x\|_2^2 + \rho(\gamma^T x)^2 \geq 0$$ since all these terms are non negative.

c) $B_{k-1} s = \gamma \Rightarrow B_k = B_{k-1}$

$$B_k = (I-\rho\gamma s^T)B_{k-1}(I-\rho s\gamma^T) + \rho\gamma\gamma^T$$

$$= (I-\rho\gamma s^T)(B_{k-1} - \rho B_{k-1} s\gamma^T) + \rho\gamma\gamma^T$$

$$= B_{k-1} - \rho B_{k-1} s\gamma^T - \rho\gamma s^T B + \rho^2 \gamma s^T B_{k-1} s\gamma^T + \rho\gamma\gamma^T$$

$$= B_{k-1} - \rho\gamma\gamma^T - \rho\gamma\gamma^T + \rho^2\gamma s^T\gamma\gamma^T + \rho\gamma\gamma^T$$

$$= B_{k-1} - \rho\gamma\gamma^T + \rho\gamma\gamma^T = B_{k-1}$$

$$\min \quad f(x,y) = -xy$$
$$\text{s.t.} \quad x + y = 10$$

$$\mathcal{L}(x,y,\lambda) = -xy + \lambda(x+y-10)$$

$$\frac{\delta \mathcal{L}}{\delta x} = -y + \lambda = 0 \iff \lambda = y$$
$$\Rightarrow x = y$$
$$\frac{\delta \mathcal{L}}{\delta y} = -x + \lambda = 0 \iff \lambda = x$$

Then

$$x + y = 10 \iff 2x = 10 \Rightarrow x = 5 \Rightarrow y = 5$$

Thus $f(x^*, y^*) = f(5,5) = -25$

Find the points on the circle $x^2 + y^2 = 80$ that are the closest to $(1,2)$.

$$\text{Min} \quad (x-1)^2 + (y-2)^2$$
$$\text{s.t.} \quad x^2 + y^2 = 80$$

$$\mathcal{L}(x,y,\lambda) = (x-1)^2 + (y-2)^2 + \lambda(x^2 + y^2 - 80)$$

$$\frac{\delta \mathcal{L}}{\delta x} = 2(x-1) + 2\lambda x = 0 \iff x - 1 + \lambda x = 0 \Rightarrow \lambda = \frac{1-x}{x}$$

$$\frac{\delta \mathcal{L}}{\delta y} = 2(y-2) + 2\lambda y = 0 \iff y - 2 + \lambda y = 0 \Rightarrow \lambda = \frac{2-y}{y}$$

$$\Rightarrow y - xy = 2x - xy \Rightarrow y = 2x$$

since $x^2 + y^2 = 80$ then

$$x^2 + (2x)^2 = 80 \Rightarrow x^2 + 4x^2 = 80 \Rightarrow x^2 = 16 \Rightarrow x = \pm 4$$

Then we have two KKT points

$$x_1 = 4 \qquad x_2 = -4$$
$$y_1 = 8 \qquad y_2 = -8$$

We evaluate on f:

$$f(x_1, y_1) = (4-1)^2 + (8-2)^2 = 9 + 36 = 45$$

$$f(x_2, y_2) = (-4-1)^2 + (-8-2)^2 = 25 + 100 = 125$$

We have that

$$(x^*, y^*) = (4, 8)$$

$$\min \; x^T x$$
$$\text{s.t.} \quad Ax = b$$

To compute the dual problem:

$$\mathcal{L}(x, \lambda) = x^T x + \lambda^T (Ax - b)$$

$$\nabla_x \mathcal{L} = 2x + A^T \lambda = 0 \iff x = -\left(\tfrac{1}{2}\right) A^T \lambda$$

Then the dual function is:

$$g(\lambda) = -\tfrac{1}{2} \lambda^T A A^T \lambda - b^T \lambda$$

$$\min \; c^T x$$
$$\text{s.t.} \quad Ax = b$$
$$x \geq 0$$

$$\mathcal{L}(x, \lambda, s) = c^T x + \lambda^T (Ax - b) - s^T x$$
$$= -b^T \lambda + (c + A^T \lambda - s)^T x$$

Thus:

$$\nabla_x \mathcal{L} = c + A^T \lambda - s = 0$$

Then:

$$g(\lambda, s) = \begin{cases} -b^T \lambda & \text{if } c + A^T \lambda - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual:

$$\max \; -b^T\lambda \qquad \rightsquigarrow \qquad \min \; -b^T\lambda$$
$$\text{s.t. } -A^T\lambda \leq c \qquad\qquad \text{s.t. } A^T\lambda \leq c$$

Prove that the optimal x^* of

$$\min \; q(x) = \frac{1}{2}x^TQx + c^Tx \qquad (1)$$
$$\text{s.t. } Ax = 0$$

where $Q \in \mathbb{R}^{n\times n}$ spd, $A \in \mathbb{R}^{m\times n}$ with full rank ($\text{rank}(A) = m$) satisfies $q(x^*) \leq 0$ and $q(x^*) = 0 \iff x^* = 0$.

Let $Z \in \mathbb{R}^{n\times(n-m)}$ be the matrix formed such that its columns form a basis for the null space of A. Then for $Ax = 0$ there is $y \in \mathbb{R}^{n-m}$ such that

writting x in the Z basis

$$Ax = 0 \Rightarrow x = Zy$$

Thus we can change (1) to an unconstrained minimization problem.

$$q(y) = \frac{1}{2}y^TZ^TQZy + c^TZy.$$

From the original problem:

$$\nabla q = Qx + c$$

Substituting:

$$\nabla q = QZy + c = 0 \iff Zy = -Q^{-1}c$$
$$y^TZ^T = -c^TQ^{-1}$$

Thus:

$$q(y) = \frac{1}{2}(-c^TQ^{-1})Q(-Q^{-1}c) + c^T(-Q^{-1}c)$$
$$= \frac{1}{2}c^TQ^{-1}c - c^TQ^{-1}c = -\frac{1}{2}c^TQ^{-1}c \leq 0$$

because Q^{-1} is also spd.

We check that $g(x^*)=0 \iff x^*=0$

"$\Rightarrow$" we have that $0=-\frac{1}{2}c^TQ^{-1}c \iff c=0$

$\iff y^TZ=0 \iff x=0$

"$\Leftarrow$" If $x^*=0$ then by substituting:

$$g(x^*)=\frac{1}{2}x^{*T}Qx^*+c^Tx^*=0$$

Consider the problem of finding the shortest distance of a fixed point x_0 to the hyperplane

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

with $A\in\mathbb{R}^{m\times n}$ and $\operatorname{rank}(A)=m$

$$\min \quad \frac{1}{2}(x-x_0)^T(x-x_0)$$
$$\text{s.t.} \quad Ax=b$$

Deduce that the solution is

$$x^*=x_0+A^T(AA^T)^{-1}(b-Ax_0)$$

$$\lambda^*=-(AA^T)^{-1}(b-Ax_0)$$

The Lagrange function is

$$\mathcal{L}(x,\lambda)=\frac{1}{2}(x-x_0)^T(x-x_0)+\lambda^T(Ax-b)$$

$$\nabla_x\mathcal{L}=x-x_0+A^T\lambda=0 \iff x=x_0-A^T\lambda$$

But we also have the KKT conditions:

$Ax=b$

$A(x_0-A^T\lambda)=b$

$$\Rightarrow A x_0 - A A^T \lambda = b$$

$$\Rightarrow -A A^T \lambda = b - A x_0$$

$$\Rightarrow \lambda^* = -(A A^T)^{-1}(b - A x_0)$$

Then

$$x^* = x_0 - A^T\left(-(A A^T)^{-1}(b - A x_0)\right)$$

$$= x_0 + A^T (A A^T)^{-1}(b - A x_0)$$