



MASTER 2 MICAS

Machine Learning Communications and Security

Course : MICAS931 - Introduction to Optimization.

School : Institut Polytechnique de Paris

Teacher : HADI GHAUCH

Subject : Homework 1 - Submit before 22/11/2020

Etudiant : Panongbene Jean Mouhamed Sawadogo.

Email : panongbene.sawadogo@telecom-paris.fr

SOMMAIRE :

Convex Optimization

Part I : Convex Analysis

Part II : Convergence of plain GD

Part II : Finding L and μ

Part IV : Duality and Optimality for Equality Constrained Quadratic Program

Convex Optimization

Part I : Convex Analysis

Consider a strongly convex function, $f(x)$, (with constant $\mu > 0$), defined over a convex set X . Recall that it fulfills the following :

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2, \forall (x_1, x_2) \in X \quad (1)$$

Prove all the following statements

- (1) is equivalent to a minimum positive curvature $\nabla^2 f(x) \succeq \mu I_d, \forall x \in X$

Answer : We assume that f is twice continuously differentiable.

According to the Taylor-Lagrange's formula, we have :

$$\forall x, h \in X, \exists \epsilon \text{ a function defined in } X, \lim_{h \rightarrow 0} \epsilon(h) = 0$$

$$f(a + h) = f(a) + \nabla f(a)^T h + \frac{1}{2} h^T \nabla^2 f(a) h + \|h\|_2^2 \epsilon(h)$$

Let $a, h \in X$, using Taylor-Lagrange's formula we have :

$$\begin{aligned} (1) &\iff f(a) + \nabla f(a)^T h + \frac{1}{2} h^T \nabla^2 f(a) h + \|h\|_2^2 \epsilon(h) \geq f(a) + \nabla f(a)^T h + \frac{\mu}{2} \|h\|_2^2 \\ &\iff \frac{1}{2} h^T \nabla^2 f(a) h + \|h\|_2^2 \epsilon(h) \geq \frac{\mu}{2} \|h\|_2^2 \\ &\iff h^T [\nabla^2 f(a) + 2\epsilon(h)I_d - \mu I_d] h \geq 0 \end{aligned}$$

Since $\epsilon(h)$ converges to 0 when h approaches 0, the previous expression is equivalent to :

$$\iff h^T [\nabla^2 f(a) - \mu I_d] h \geq 0$$

This expression is true for all h and a in X , so we can say that :

$$(1) \iff \forall a \in X, \nabla^2 f(a) \succeq \mu I_d$$

- (1) is equivalent to $(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2$

Answer : Let $x_1, x_2 \in X$, we have :

$$\begin{aligned}
(1) &\iff \begin{cases} f(x_1) \geq f(x_2) + \nabla f(x_2)^T (x_1 - x_2) + \frac{\mu}{2} \|x_1 - x_2\|_2^2 \\ f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2 \end{cases} \\
&\iff f(x_2) + f(x_1) \geq f(x_1) + f(x_2) + \nabla f(x_1)^T (x_2 - x_1) + \nabla f(x_2)^T (x_1 - x_2) + \mu \|x_2 - x_1\|_2^2 \\
&\iff \nabla f(x_2)^T (x_2 - x_1) - \nabla f(x_1)^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2 \\
&\iff (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2
\end{aligned}$$

So we can say that :

$$(1) \iff (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2$$

- (1) implies to $f(x) - f^* \leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2, \forall x \in X$

Answer : Let $x, y \in X$ we have :

$$\begin{aligned}
0 &\leq \frac{\mu}{2} \left\| y - x + \frac{\nabla f(x)}{\mu} \right\|_2^2 = \frac{\mu}{2} \left(y - x + \frac{\nabla f(x)}{\mu} \right)^T \left(y - x + \frac{\nabla f(x)}{\mu} \right) \\
&= \frac{\mu}{2} \|y - x\|_2^2 + \frac{1}{2} (y - x)^T \nabla f(x) + \frac{1}{2} \nabla f(x)^T (y - x) + \frac{\|\nabla f(x)\|_2^2}{2\mu} \\
&= \frac{\mu}{2} \|y - x\|_2^2 + \nabla f(x)^T (y - x) + \frac{\|\nabla f(x)\|_2^2}{2\mu}
\end{aligned}$$

Thus, using relation (1), we have :

$$\begin{aligned}
(1) &\implies 0 \leq \frac{\mu}{2} \|y - x\|_2^2 + \nabla f(x)^T (y - x) + \frac{\|\nabla f(x)\|_2^2}{2\mu} \leq f(y) - f(x) + \frac{\|\nabla f(x)\|_2^2}{2\mu} \\
&\implies f(x) - f(y) \leq \frac{\|\nabla f(x)\|_2^2}{2\mu} \\
&\implies f(x) - \inf_{y \in X} f(y) \leq \frac{\|\nabla f(x)\|_2^2}{2\mu}
\end{aligned}$$

So we have :

$$(1) \implies f(x) - f^* \leq \frac{\|\nabla f(x)\|_2^2}{2\mu}$$

- (1) implies to $f(x) + r(x)$ is strongly convex for any convex f and strongly convex r

Answer : Let f be a convex function and r a strongly convex. We assume f and r is continuously differentiable, then for $x, y \in X$ we have :

$$\begin{cases} f(y) \geq f(x) + \nabla f(x)^T (y - x) \\ r(y) \geq r(x) + \nabla r(x)^T (y - x) + \frac{\mu}{2} \|y - x\|_2^2 \end{cases}$$

$$\implies (f+r)(y) \geq (f+r)(x) + (\nabla f + \nabla r)(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

So, we can say that $f+r$ is strongly convex

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is L -smooth if it is differentiable and its gradient is L -Lipschitz-continuous (usually w.r.t. norm-2) :

$$\forall x_1, x_2 \in \mathbb{R}^d, \|\nabla f(x_2) - \nabla f(x_1)\|_2 \leq L \|x_2 - x_1\|_2, \quad (2)$$

Prove that (2) implies all the following statements (assume convexity if needed) :

$$(a) f(x_2) \leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2, \forall x_1, x_2 \in X$$

Answer : Let x_1 and $x_2 \in X$, we have :

$$f(x_2) = f(x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)), x_2 - x_1 \rangle dt$$

$$\implies f(x_2) = f(x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1) + \nabla f(x_1), x_2 - x_1 \rangle dt$$

$$\implies f(x_2) = f(x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle dt + \int_0^1 \langle \nabla f(x_1), x_2 - x_1 \rangle dt$$

$$\implies f(x_2) = f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle dt$$

Using Cauchy-Sawartz inequality, we have for all $t \in [0,1]$,

$$|\langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle| \leq \|\nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1)\|_2 \|x_2 - x_1\|_2$$

Using inequality (2), we have :

$$\|\nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1)\|_2 \leq L \|x_1 + t(x_2 - x_1) - x_1\|_2 = tL \|x_2 - x_1\|_2$$

$$\implies |\langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle| \leq tL \|x_2 - x_1\|_2^2$$

$$\implies \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle dt \leq \int_0^1 |\langle \nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1), x_2 - x_1 \rangle| dt \leq \int_0^1 tL \|x_2 - x_1\|_2^2 dt$$

$$\begin{aligned}
\nabla f(x_1), x_2 - x_1 > |dt| &\leq \int_0^1 tL\|x_2 - x_1\|_2^2 dt = \frac{L}{2}\|x_2 - x_1\|_2^2 \\
\implies f(x_2) &= f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \\
\nabla f(x_1), x_2 - x_1 > dt &\leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \int_0^1 \langle \nabla f(x_1 + t(x_2 - x_1)) - \\
\nabla f(x_1), x_2 - x_1 > |dt| &\leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{L}{2}\|x_2 - x_1\|_2^2 \\
\implies &\boxed{f(x_2) \leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{L}{2}\|x_2 - x_1\|_2^2}
\end{aligned}$$

$$(b) f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2, \forall x_1, x_2 \in X$$

Answer : Let $x_1, x_2 \in X$,

Let $y \in X$,

$$f(x_1) - f(x_2) = f(x_1) - f(y) + f(y) - f(x_2)$$

Using (a) we have ,

$$\begin{aligned}
f(y) - f(x_2) &\leq \nabla f(x_2)^T (y - x_2) + \frac{L}{2}\|y - x_2\|_2^2 \\
\implies f(x_1) - f(x_2) &\leq f(x_1) - f(y) + \nabla f(x_2)^T (y - x_2) + \frac{L}{2}\|y - x_2\|_2^2
\end{aligned}$$

If we assume f is convex, then we have

$$\begin{aligned}
f(x_1) - f(y) &\leq -\nabla f(x_1)^T (y - x_1) \\
\implies f(x_1) - f(x_2) &\leq -\nabla f(x_1)^T (y - x_1) + \nabla f(x_2)^T (y - x_2) + \frac{L}{2}\|y - x_2\|_2^2 \\
\implies f(x_1) - f(x_2) &\leq -\nabla f(x_1)^T (y - x_2) - \nabla f(x_1)^T (x_2 - x_1) + \nabla f(x_2)^T (y - x_2) \\
&\quad + \frac{L}{2}\|y - x_2\|_2^2 \\
\implies f(x_1) - f(x_2) &\leq -\nabla f(x_1)^T (x_2 - x_1) + (\nabla f(x_2) - \nabla f(x_1))^T (y - x_2) + \frac{L}{2}\|y - x_2\|_2^2
\end{aligned}$$

$$\text{Let } y = x_2 - \frac{1}{L} (\nabla f(x_2) - \nabla f(x_1)) \text{ then } y - x_2 = -\frac{1}{L} (\nabla f(x_2) - \nabla f(x_1))$$

So we have :

$$\begin{aligned}
f(x_1) - f(x_2) &\leq -\nabla f(x_1)^T (x_2 - x_1) - \frac{1}{L} (\nabla f(x_2) - \nabla f(x_1))^T (\nabla f(x_2) - \nabla f(x_1)) \\
&\quad + \frac{1}{L^2} * \frac{L}{2} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2
\end{aligned}$$

$$\implies f(x_1) - f(x_2) \leq -\nabla f(x_1)^T(x_2 - x_1) - \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2$$

$$\implies \boxed{f(x_1) + \nabla f(x_1)^T(x_2 - x_1) + \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \leq f(x_2)}$$

Part II : Convergence of plain GD

Consider the following optimization problem : $\min_{x \in \mathbb{R}^d} f(x)$ where f is strongly convex with constant μ , continuous with Lipschitz constant L . In other words, f is μ -strongly convex and L -smooth. Moreover, consider the Gradient descent (GD) with constant step-size

$$x_{k+1} = x_k - \alpha \nabla f(x_k), \alpha \geq 0 \quad (3)$$

We set the (constant) step-size as $\alpha = \frac{2}{L+\mu}$. This variant is known as plain/vanilla GD.

(1) Prove the convergence of GD with constant step size. Specifically, show that the iterations in (3) are such that :

$$\|x_k - x^*\|_2^2 \leq \left(1 - \frac{2}{1 + \frac{L}{\mu}}\right)^{2k} \|x_0 - x^*\|_2^2$$

Hints : you may use these observations.

From smoothness and vanishing gradient of the optimal point, conclude

$$f(x_k) - f(x^*) \leq \frac{L}{2} \|x_k - x^*\|_2^2 \quad (*)$$

Use the coercivity of the gradient :

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \frac{\mu L}{\mu + L} \|x - y\|_2^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|_2^2$$

Iterate over k and use (*) to obtain :

$$f(x_k) - f(x^*) \leq \frac{L}{2} \prod_{i \in [k]} \left(1 - 2\alpha \frac{\mu L}{\mu + L}\right) \|x_0 - x^*\|_2^2$$

Answer : We have

$$\begin{aligned}\|x_{k+1} - x^*\|_2^2 &= \|x_k - \alpha \nabla f(x_k) - x^*\|_2^2 = \|(x_k - x^*) - \alpha \nabla f(x_k)\|_2^2 \\ \implies \|x_{k+1} - x^*\|_2^2 &= \|x_k - x^*\|_2^2 - 2\alpha \langle \nabla f(x_k), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k)\|_2^2\end{aligned}$$

We have $\nabla f(x^*) = 0$ so, we can write :

$$\implies \|x_{k+1} - x^*\|_2^2 = \|x_k - x^*\|_2^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2$$

Using the coercivity of the gradient of f , we have :

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \frac{\mu L}{\mu + L} \|x - y\|_2^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|_2^2 \quad \forall x, y \in X$$

we have :

$$\begin{aligned}\|x_{k+1} - x^*\|_2^2 &\leq \|x_k - x^*\|_2^2 - 2\alpha \left(\frac{\mu L}{\mu + L} \|x_k - x^*\|_2^2 + \frac{1}{\mu + L} \|\nabla f(x_k) - \nabla f(x^*)\|_2^2 \right) + \\ &\quad \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2 \\ \implies \|x_{k+1} - x^*\|_2^2 &\leq \left(1 - 2\alpha \frac{\mu L}{\mu + L} \right) \|x_k - x^*\|_2^2 + \left(\alpha^2 - \frac{2\alpha}{\mu + L} \right) \|\nabla f(x_k) - \nabla f(x^*)\|_2^2\end{aligned}$$

We have :

$$\alpha^2 - \frac{2\alpha}{\mu + L} = 0$$

so, we have :

$$\begin{aligned}\|x_{k+1} - x^*\|_2^2 &\leq \left(1 - 2\alpha \frac{\mu L}{\mu + L} \right) \|x_k - x^*\|_2^2 \\ \|x_{k+1} - x^*\|_2^2 &\leq \left(\frac{L - \mu}{\mu + L} \right)^2 \|x_k - x^*\|_2^2 \\ \implies \|x_{k+1} - x^*\|_2^2 &\leq (1 - \alpha \mu)^2 \|x_k - x^*\|_2^2\end{aligned}$$

When we iterate over k on the right side, we get :

$$\implies \|x_{k+1} - x^*\|_2^2 \leq (1 - \alpha \mu)^{2(k+1)} \|x_0 - x^*\|_2^2$$

we can conclude that :

$$\|x_k - x^*\|_2^2 \leq \left(1 - \frac{2}{1 + \frac{L}{\mu}}\right)^{2k} \|x_0 - x^*\|_2^2$$

2) What is the convergence rate of this variant, in the $\mathcal{O}(-)$ sense?

Answer : The rate of convergence of this variant is $(c)^k = \mathcal{O}((c)^{2k})$ where $c = 1 - \frac{2}{1 + \frac{L}{\mu}}$

3) Recall : μ and L are upper and lower bounds on the largest and smallest eigenvalues of the Hessian of a μ -strongly convex and L -smooth, respectively

a) What happens to the convergence rate (question 1) when $L/\mu \rightarrow 1$? Explain to which scenario does this correspond. Discuss the practical implications of this scenario for GD.

Answer : When $\frac{L}{\mu} \rightarrow 1$, the convergence constant $c = 1 - \frac{2}{1 + \frac{L}{\mu}} \rightarrow 0$ so the convergence rate $(c)^k$ decreases quickly : We have a fastest convergence for GD.

b) What happens to the convergence rate (question 1) when $L/\mu \rightarrow \infty$? Explain to which scenario does this correspond. Discuss the practical implications of this scenario for GD.

Answer : When $\frac{L}{\mu} \rightarrow \infty$, the convergence constant $c = 1 - \frac{2}{1 + \frac{L}{\mu}} \rightarrow 1$ so the convergence rate $(c)^k$ decreases slowly : We have a slowest convergence for GD.

Part III : Finding L and μ

Consider a linear ridge regression : $\min_w f(w) = \frac{1}{N} \sum_{i \in [N]} f_i(w) + \lambda \|w\|_2^2$ where

the loss for sample i is given by : $f_i(w) = (y_i - w^T x_i)^2$. Use the Bodyfat dataset (available in the ML toolbox in MATLAB)

a) Is f Lipschitz continuous? If so, find a small Lipschitz constant L ?

Answer : We have for all w :

$$f(w) = \frac{1}{N} \sum_{i \in [N]} f_i(w) + \lambda \|w\|_2^2$$

$$\Rightarrow f(w) = \frac{1}{N} \sum_{i \in [N]} (f_i(w) + \lambda \|w\|_2^2)$$

$$\text{Let } g_i(w) = f_i(w) + \lambda \|w\|_2^2$$

then we have :

$$f(w) = \frac{1}{N} \sum_{i \in [N]} g_i(w)$$

NB : we consider the functions g_i thus defined in the other questions

$$\text{Let } B > 0 \text{ and } \mathcal{H} = \{w \in \mathbb{R}^d : \|w\|_2 \leq B\}$$

Let $w \in \mathcal{H}$ then we have :

$$g_i(w) = (y_i - w^T x_i)^2 + \lambda \|w\|_2^2$$

$$\Rightarrow \nabla g_i(w) = -2x_i (y_i - w^T x_i) + 2\lambda w$$

$$\Rightarrow \nabla g_i(w) = -2x_i y_i + 2(x_i x_i^T + \lambda I_d) w$$

$$\Rightarrow \|\nabla g_i(w)\|_2 = \|-2x_i y_i + 2(x_i x_i^T + \lambda I_d) w\|_2 \leq 2\|x_i y_i\|_2 + 2\|(x_i x_i^T + \lambda I_d)\|_2 B$$

$$\Rightarrow \|\nabla f_i(w)\|_2 = \left\| \frac{1}{N} \sum_{i \in [N]} \nabla g_i(w) \right\|_2 \leq \frac{1}{N} \sum_{i \in [N]} \|\nabla g_i(w)\|_2 \leq \frac{2}{N} \sum_{i \in [N]} \|x_i y_i\|_2 +$$

$$\frac{2B}{N} \sum_{i \in [N]} \|(x_i x_i^T + \lambda I_d)\|_2$$

$$\Rightarrow \|\nabla f_i(w)\|_2 \leq \frac{2}{N} \sum_{i \in [N]} \|x_i y_i\|_2 + \frac{2B}{N} \sum_{i \in [N]} \|(x_i x_i^T + \lambda I_d)\|_2$$

So we can say that f is L -lipschitz and we can take

$$L = \frac{2}{N} \sum_{i \in [N]} \|x_i y_i\|_2 + \frac{2B}{N} \sum_{i \in [N]} \|(x_i x_i^T + \lambda I_d)\|_2$$

as lipschitz constant

b) Is f strongly convex? If so, find a large μ ?

Answer : Consider this function : $h(w) = \lambda \|w\|_2^2$ then :

$$\nabla h(w) = 2\lambda w$$

$$\implies \nabla^2 h(w) = 2\lambda I_d$$

$$\implies \nabla^2 h(w) \succ 2\lambda I_d$$

So, the h function is 2λ -strongly convex (**)

For $i \in [N]$, we have

$$f_i(w) = (y_i - w^T x_i)^2$$

$$\implies \nabla f_i(w) = -2x_i (y_i - w^T x_i)$$

$$\implies \nabla^2 f_i(w) = 2x_i x_i^T$$

$$\implies \nabla^2 f_i(w) \succcurlyeq 0$$

so we can say that the function f_i is convex (***)

(**) and (***) $\implies g_i(w) = f_i(w) + h(w)$ is 2λ -strongly convex

So, $f(w) = \frac{1}{N} \sum_{i \in [N]} g_i(w)$ is 2λ -strongly convex and we can choose

$\mu = 2\lambda$

c) There is a simple way (trick) to find L and μ , for the optimization problem considered here.

Express, mathematically, the steps for doing so, and derive the expressions for L , μ , and their ratio L/μ

What inherent properties of the dataset impact (and determine) the ratio L/μ ?

Answer : We can find the values of L and μ by calculating the eigenvalues of the Hessian matrix of f . Indeed μ and L are respectively the min and the max of the eigenvalues of the Hessian matrix of f .

step 1 : We compute the matrix $\nabla^2 f(w) = \frac{2}{N} \sum_{i \in [N]} x_i x_i^T + 2\lambda I_d$

step 2 : Compute the eigenvalues $\{\lambda_i\}_{i \in [N]}$ of $\nabla^2 f(w)$

step 3 : compute the values of L and μ

$$L = \max \{\lambda_i\}_{i \in [N]}$$

$$\mu = \min \{\lambda_i\}_{i \in [N]}$$

To calculate μ and L , we need to determine the eigenvalues of $\nabla^2 f(w) = \frac{2}{N} \sum_{i \in [N]} x_i x_i^T + 2\lambda I_d$.

So, the larger our dataset, i.e. N is big, more we need times to calculate $\frac{2}{N} \sum_{i \in [N]} x_i x_i^T$.

Thus, we can say that the size of our data set (N) impacts the calculation of the ratio $\frac{L}{\mu}$

d) What can you say about the ratio L/μ for the Bodyfat dataset? is it a 'good' or 'bad' setup for a plain GD method

Answer : The bodyfat dataset contains 252 items, so the number of items is relatively small. So this is a good dataset for GD method

Part IV : Duality and Optimality for Equality Constrained Quadratic Program

Consider the following Equality Constrained Quadratic Program (ECQP).

$$ECQP : x^* := \begin{cases} \arg \min_{x \in \mathbb{R}^d} f_0(x) = x^T S x \\ s.t. \quad Ax = b \end{cases}$$

where $S \in \mathbb{R}^{d \times d} \succ 0$ is a positive definite matrix, $A \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$

1) Derive the Lagrangian $\mathcal{L}()$, the Lagrange function $g()$, and the dual problem, that correspond to the above ECQP

Answer : We have

$$\begin{aligned} \mathcal{L}(x, v) &= f_0(x) + v^T (Ax - b) \\ \implies \boxed{\mathcal{L}(x, v) &= x^T S x + v^T (Ax - b)} \end{aligned}$$

Nothing that for all v , the function $x \rightarrow \mathcal{L}(x, v)$ is strongly convex and continuously differentiable. So, we have :

$$\hat{x} = \arg \min_{x \in \mathbb{R}^d} \mathcal{L}(x, v) \in \{x \in \mathbb{R}^d \mid \nabla \mathcal{L}_x(x, v) = 0\}$$

We have :

$$\begin{aligned} \nabla \mathcal{L}_x(x, v) &= \nabla_x (x^T S x + v^T (Ax - b)) \\ \implies \nabla \mathcal{L}_x(x, v) &= 2Sx + A^T v \\ \implies \nabla \mathcal{L}_x(x, v) = 0 &\iff 2Sx + A^T v = 0 \\ \implies x &= -(2S)^{-1} A^T v \text{ So we have} \end{aligned}$$

$$\boxed{\hat{x} = -(2S)^{-1} A^T v}$$

We have $g(v) = \inf_{x \in \mathbb{R}^d} \mathcal{L}(x, v) = \mathcal{L}(\hat{x}, v)$

$$\begin{aligned} \implies g(v) &= \mathcal{L}(-(2S)^{-1} A^T v, v) \\ \implies g(v) &= [-(2S)^{-1} A^T v]^T S [-(2S)^{-1} A^T v] + v^T (A [-(2S)^{-1} A^T v] - b) \\ \implies g(v) &= \frac{1}{4} [A^T v]^T S^{-1} [A^T v] - \frac{1}{2} [A^T v]^T S^{-1} [A^T v] - v^T b \end{aligned}$$

$$g(v) = -\frac{1}{4} [A^T v]^T S^{-1} [A^T v] - v^T b$$

The Lagrangian dual problem D is define by :

$$\max_v g(v) = \min_v -g(v) = \min_v \left\{ \frac{1}{4} [A^T v]^T S^{-1} [A^T v] + v^T b \right\}$$

As $S \succ 0$ then $S^{-1} \succ 0$, \implies the dual problem is strongly convex QP in v . So we have :

$$-\nabla g(v) = \frac{1}{2} A S^{-1} A^T v + b$$

$$\implies -\nabla g(v) = 0 \iff \frac{1}{2} A S^{-1} A^T v + b = 0$$

$$\implies -\nabla g(v) = 0 \iff v = -2 (A S^{-1} A^T)^{-1} b$$

So we have :

$$\hat{v} = \arg \max_v g(v) = -2 (A S^{-1} A^T)^{-1} b$$

$$\implies \hat{x} = S^{-1} A^T (A S^{-1} A^T)^{-1} b$$

$$\implies \hat{x} = (S^{-1} A^T)^{-1} A (S^{-1} A^T)^{-1} b$$

2) Derive the KKT conditions that correspond to the above ECQP. Use these KKT condition to derive a closed-form analytical solution, x^* , as a function of optimal dual variables, μ^* ?

Answer : We have :

1- The ECQP problem is strongly convex

2- $\nabla \mathcal{L}(\hat{x}, \hat{v}) = 0$

3- The constrain function ($Ax=b$) is affine

So using the KKT condition, we can say : the close form analytical solution

of ECQP problem is given by :

$$\hat{x} = (S^{-1}A^T) A (S^{-1}A^T) b$$

we have also :

$$\hat{v} = \arg \max_v g(v) = -2 (AS^{-1}A^T) b$$