Introduction to optimisation

Lenie (H.M.) Goossens
S4349113

November 10, 2023

Exercise 1

Let $\phi : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be closed and $\mu-$ strongly convex. Use the Reciprocity formula to show why the Fenchel conjugate of ϕ must be $\frac{1}{\mu}$ strongly convex.
Solution:
Fencel conjugate: $\phi^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}$.
Completly forgot the Reciprocity formula.

In all the follows, we consider the problem $\mathcal{P}$ of minimizing a continuous and $\mu-$ strongly convex function $f : \mathbb{R}^N \to \mathbb{R}$ on $V = \{x \in \mathbb{R}^N : Ax = b\}$ where $A \in \mathbb{R}^{M \times N}$ and $b \in \mathbb{R}^M$

Exercise 2

Why can we assure that $\mathcal{P}$ has a unique solution?
Solution:
Since f is strongly convex, we see that the minimization problem of f, has a unique solution. And this minimization problem is given by $\mathcal{P}$.

Exercise 3

Use the first order optimality condition to show that $\hat{x}$ is a solution of $\mathcal{P}$ iff $A\hat{x} = b$ and there exists $\hat{y} \in \mathbb{R}^M$ s.t. $-A^T y \in \partial f(\hat{x})$. We say $(\hat{x}, \hat{y})$ is an optimal pair.
Since f is continuous, we have $\partial(f + \iota_V) = \partial f + \partial \iota_V = \partial f + \text{ran}(A^T)$. You do not need to prove this.
Solution:
For a general function $\beta : \mathbb{R}^N \to \mathbb{R}$ we see that if β is convex, then $\tilde{x}$ is a minimum if $0 \in \partial\beta(\tilde{x})$.
Since f is strictly convex, so $f + \iota_V$ is is strictly convex we see that there is a unique

minimizer, which we call $\hat{x}$, so we see that $0 \in \partial(f + \iota_V)(\hat{x})$.
Note that this means that $0 \in (\partial f + \text{ran}(A^T))(\hat{x})$. But we see that $\partial \iota_V(x) = \text{ran}(A^T)$ only if $x \in V$. Therefore we see that we must have $\hat{x} \in V$ so $A\hat{x} = b$.
Therefore $0 \in \partial f(\hat{x}) + \text{ran}(A^T)$. This means that $-\text{ran}(A^T) \in \partial f(\hat{x})$.
Note that $-A^T\hat{y} \in -\text{ran}(A^T)$ so therefore we get indeed that $-A^T\hat{y} \in \partial f(\hat{x})$.

In the rest of the exam, we shall establish the convergence of a sequence (x_k, y_k), constructed from an initial point $(x_0, y_0) \in \mathbb{R}^N \times \mathbb{R}^M$ by iterating

$$\begin{cases} x_{k+1} & = \arg\min\{L(x, y_k) : x \in \mathbb{R}^N\} \\ y_{k+1} & = y_k + \alpha(Ax_{k+1} - b) \end{cases}$$

with $\alpha > 0$ and $L(x, y) = f(x) + y \cdot (Ax - b) = f(x) + (A^T y) \cdot x - y \cdot b$ for each $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$.

Exercise 4

Why is x_{k+1} well defined?
Solution:
Since f is well-defined, and matrix multiplication is well defined, we see that $L(x, y)$ is well-defined. Since $\arg\min(K(..))$ is only well defined if the thing inside the min brackets is well defined (so $K(..)$), we see that $\arg\min\{L(x, y_k) : x \in \mathbb{R}^N\}$ is well defined.

Exercise 5

Write the optimality condition satisfied by x_{k+1} (this comes from the first subiteration).
Solution:
First note that since f is convex, we have that $L(x, y_k)$ is convex for a fixed y_k on the set V.
We see that if $\hat{x} \in V$ minimizes f, it also minimizes $L(x, y_k)$. Note that $L(x, y_k)$ is also differentiable at $\hat{x}$.
Therefore we see that since we want that $L(\hat{x}, y_k) \leq L(x, y_k)$ for all $x \in V$, we must have by Fermat's rule that $\nabla L(\hat{x}, y_k) \cdot (x - x_k) \geq 0, \forall x \in V$.
Note that $\nabla L(x, y_k) = \nabla f(x) + A^T y_k$. Therefore we must have that $(f(\hat{x}) + A^T y_k) \cdot (x - x_k) \geq 0$ for all $x \in V$.

Exercise 6

Show that the dual $\mathcal{D}$ of the problem $\mathcal{P}$ is $\min\{h(y) : y \in \mathbb{R}^M\}$ where
$h(y) = f^*(-A^T y) + b \cdot y$

Solution:

Note that we can write $\mathcal{P}$ as $f(x)+g(x)$ where $g(x) = \iota_V(x)$ and $f(x)$ as in the equation. Therefore we see that the dual is given by

$$\inf_{y\in\mathbb{R}^M}\{f^*(-P^T y) + g^*(y)\} \quad P \in \mathbb{R}^{M\times N}$$

Take $P = A$, so then we see that the dual is given by

$$\inf_{y\in\mathbb{R}^M}\{f^*(-A^T y) + g^*(y)\}$$

If we can show that $g^*(y) = b \cdot y$, we see that we are done.
Since $g(x) = \iota_V(x)$ we see that $g^*(x) = \sup\{x \cdot z | z \in V\}$. Note that if $z \in V$ then $Az = b$. So we have that $g^*(x) = \sup\{x \cdot z | Az = b\}$.

Exercise 7

Compute ∇h and verify that h is $l-$ smooth with $l = \frac{\|A\|^2}{\mu}$.

Solution:

We see that h is $l-$ smooth if there exists $l > 0$ s.t. $\|\nabla h(z) - \nabla h(y)\| \leq l\|z - y\|, \forall z, y \in V$. Now let $k(y) = -A^T y$, so we see that $h(y) = f^*(k(y)) + b \cdot y$. Therefore we see that

$$\begin{aligned}
\nabla h(y) &= \nabla f^*(y)|_{k(y)} \cdot \nabla k(y) + b \cdot \nabla y \\
&= \nabla f^*(y)|_{k(y)} \cdot -A^T + b \\
&= -\nabla f^*(y)|_{k(y)} \cdot A^T + b
\end{aligned}$$

Therefore we see that

$$\begin{aligned}
\|\nabla h(z) - \nabla h(y)\| &= \|-\nabla f^*(x)|_{k(z)} \cdot A^T + b - (-\nabla f^*(x)|_{k(y)} \cdot A^T + b)\| \\
&= \|-A^T(\nabla f^*(x)|_{k(z)} - \nabla f^*(x)|_{k(y)})\| \\
&\leq \|A\| \cdot \|\nabla f^*(x)|_{k(z)} - \nabla f^*(x)|_{k(y)}\|
\end{aligned}$$

Now use question 1. we see that f is μ smooth, so f^* is $\frac{1}{\mu}-$ smooth, so we get

$$\begin{aligned}
\|\nabla f^*(x)|_{k(z)} - \nabla f^*(x)|_{k(y)}\| &\leq \frac{1}{\mu}\|k(z) - k(y)\| \\
&= \frac{1}{\mu}\|-A^T z + A^T y\| \leq \frac{\|A\|}{\mu}\|z - y\|
\end{aligned}$$

If we substiute this back we get indeed

$$\|\nabla h(z) - \nabla h(y)\| \leq \frac{\|A\|^2}{\mu}\|z - y\|$$

Exercise 8

Show that the sequence (y_k) satisfies $y_{k+1} = y_k - \alpha\nabla h(y_k)$
Solution:
So it is enough to show that $-\alpha\nabla h(y_k) = \alpha(Ax_{k+1} - b)$ so $\nabla h(y_k) = b - Ax_{k+1}$
We see that $\nabla h(y) = -\nabla f^*(z)|_{z=-A^Ty} \cdot A^T + b$. Therefore we get that $\nabla h(y_k) = -\nabla f^*(z)|_{z=-A^Ty_k} \cdot A^T + b$. Therefore we see that it is enough to show that $\nabla f^*(z)|_{z=-A^Ty_k} \cdot A^T = Ax_{k+1}$. But since I do not see how to work with $\nabla f^*(z)$ I do not see how to continue.

Exercise 9

For which values of α can we guarantee that the sequence (x_k, y_k) converges to an optimal pair $(\hat{x}, \hat{y})$ as $k \to \infty$. Express the result in terms of μ and $\|A\|$.
Solution:
We see that $(x_k, y_k) \to (\hat{x}, \hat{y})$ as $k \to \infty$ if $\|(\hat{x} - x_k) + (\hat{y} - y_k)\| \xrightarrow{k\to\infty} 0$.
Furhtermore if $y_k \to \hat{y}$ as $k \to \infty$ then we see that $\|y_{k+1} - y_k\| \to 0$ as $k \to \infty$. So $\|-\alpha\nabla h(y_k)\| \to 0$ as $k \to \infty$.

$$\|\nabla h(y_{k+1}) - \nabla h(y_k)\| \leq \frac{\|A\|^2}{\mu}\|y_{k+1} - y_k\| = \frac{\|A\|^2}{\mu}\|\alpha(Ax_{k+1} - b)\| = \frac{\|A\|^2|a|^2}{\mu}\|Ax_{k+1} - b\|$$

Now use that we want that $x_{k+1} \to \hat{x}$ so $Ax_{k+1} - b \to 0$. So we have convergence if $\frac{\|A\|^2|\alpha|^2}{\mu} < 1$. Therefore wemust have that $\alpha < \frac{\sqrt{\mu}}{\|A\|}$.

Exercise 10

What can you say about the convergence rate of this algorithm?
Solution:
The larger we make $\|A\|$, the faster this function converges. Note that μ is fixed for a certain f.