CPSC 540 — Assignment 1

Chaurette, Laurent
84060128

Knill, Stephanie
54882113

Vincart-Emard, Alexandre
85135127

1 Fundmanentals

Stephanie

2 Convex Functions

2.1 Minimizing Stricly Convex Quadratic Functions

1. $$f(w) = \frac{1}{2}\|w - v\|^2 \tag{1}$$

   Taking the gradient of $f(w)$, we get

   $$\begin{aligned}\nabla f(w) &= \nabla\left(\frac{1}{2}(w-v)^T(w-v)\right) \\ &= \frac{1}{2}\,\nabla\left(w^Tw - 2w^Tv + v^Tv\right) \\ &= w - v.\end{aligned} \tag{2}$$

   We now set the gradient to 0 to find the critical points

   $$\nabla f(w_{min}) = 0 = w_{min} - v, \tag{3}$$

   implies

   $$w_{min} = v. \tag{4}$$

2. $$f(w) = \frac{1}{2}\|Xw - y\|^2 + \frac{1}{2}w^T\Lambda w \tag{5}$$

   Taking the gradient of $f(w)$, we get

   $$\begin{aligned}\nabla f(w) &= \nabla\left(\frac{1}{2}\|Xw - y\|^2 + \frac{1}{2}w^T\Lambda w\right) \\ &= \frac{1}{2}\nabla\left[w^T\left(X^TX + \Lambda\right)w - 2w^TX^Ty + y^Ty\right] \\ &= \frac{1}{2}\left(2X^TX + \Lambda + \Lambda^T\right)w - X^Ty.\end{aligned} \tag{6}$$

   Setting the gradient to 0, we find the critical point

   $$0 = \frac{1}{2}\left(2X^TX + \Lambda + \Lambda^T\right)w_{min} - X^Ty \tag{7}$$

   which gives,

   $$\left(2X^TX + \Lambda + \Lambda^T\right)w_{min} = 2X^Ty \tag{8}$$

   and finally,

   $$w_{min} = 2\left(2X^TX + \Lambda + \Lambda^T\right)^{-1}X^Ty \tag{9}$$

3.
$$f(w) = \frac{1}{2}\sum_{i=1}^{n} v_i \left(w^T x_i - y_i\right)^2 + \frac{\lambda}{2}\|w - w^0\|^2 \tag{10}$$

Taking the gradient of $f(w)$, we get

$$\begin{aligned}
\nabla f(w) &= \nabla \left(\frac{1}{2}\sum_{i=1}^{n} v_i \left(w^T x_i - y_i\right)^2 + \frac{\lambda}{2}\|w - w^0\|^2\right) \\
&= \frac{1}{2}\nabla \left(\sum_{i=1}^{n} v_i \left(w^T x_i x_i^T w - 2w^T x_i y_i + y_i^2\right) + \lambda \left(w^T w - 2w^T w^0 + w^{0^T} w^0\right)\right) \\
&= \sum_{i=1}^{n} v_i \left(x_i x_i^T w - x_i y_i\right) + \lambda \left(w - w^0\right).
\end{aligned} \tag{11}$$

Setting the gradient to 0, we find the critical point

$$0 = \sum_{i=1}^{n} v_i \left(x_i x_i^T w_{min} - x_i y_i\right) + \lambda I \left(w_{min} - w^0\right), \tag{12}$$

where I is the $d \times d$ identity matrix. Thus,

$$\lambda I w^0 + \sum_{i=1}^{n} v_i x_i y_i = \left(\sum_{i=1}^{n} v_i x_i x_i^T + \lambda I\right) w_{min}. \tag{13}$$

We finally find

$$w_{min} = \left(\sum_{i=1}^{n} v_i x_i x_i^T + \lambda I\right)^{-1} \left(\lambda I w^0 + \sum_{i=1}^{n} v_i x_i y_i\right) \tag{14}$$

2.2 Proving Convexity

1.
$$f(w) = -\log(aw) \tag{15}$$

As the log function is twice differentiable, we will prove convexity by calculating the Hessian matrix

$$f''(w) = \frac{1}{w^2} > 0, \quad \forall\, w \in \mathcal{R}^+ \tag{16}$$

The hessian is therefore positive definite over the domain of $f(w)$ which means f is convex.

2.
$$f(w) = \frac{1}{2}w^T Aw + b^T w + \gamma \tag{17}$$

Once again, we will use the Hessian matrix criteria,

$$\nabla^2 f(w) = \frac{1}{2}\left(A + A^T\right) \succeq 0. \tag{18}$$

As A is positive semi-definite, its transpose is also positive semi-definite and so is the sum $A + A^T$. The Hessian is therefore positive semi-definite, which implies f(w) is convex.

3.
$$f(w) = \|w\|_p \tag{19}$$

Here, we will choose two points, $(v, f(v))$ and $(w, f(w))$ and show that the line always lies above the function itself. The line joining the two points can be written parametrically as

$$\theta\left(\sum_{i=1}^{n} \mid w_i \mid^p\right)^{1/p} + (1-\theta)\left(\sum_{i=1}^{n} \mid v_i \mid^p\right)^{1/p}, \quad 0 \leq \theta \leq 1 \tag{20}$$

while the function in between those two points can be expressed as

$$\left(\sum_{i=1}^{n} \mid \theta w_i + (1-\theta)v_i \mid^p\right)^{1/p}. \tag{21}$$

However, the triangle inequality gives us directly that

$$\begin{aligned} f(\theta w + (1-\theta)v) &= \left(\sum_{i=1}^{n} \mid \theta w_i + (1-\theta)v_i \mid^p\right)^{1/p} \\ &\leq \theta\left(\sum_{i=1}^{n} \mid w_i \mid^p\right)^{1/p} + (1-\theta)\left(\sum_{i=1}^{n} \mid v_i \mid^p\right)^{1/p} \\ &= \theta f(w) + (1-\theta)f(v), \end{aligned} \tag{22}$$

showing that the line joining two points is always above the function itself and therefore $f(w)$ is convex.

4.
$$f(w) = \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i w^T x_i\right)\right) \tag{23}$$

Once again, we will compute the Hessian matrix. First of all, we find the gradient

$$\nabla f(w) = \sum_{i=1}^{n} \frac{-y_i x_i}{1 + \exp\left(-y_i w^T x_i\right)}. \tag{24}$$

The Hessian is therefore

$$\nabla^2 f(w) = \sum_{i=1}^{n} \frac{y_i^2 x_i x_i^T}{\left[1 + \exp\left(-y_i w^T x_i\right)\right]^2}, \tag{25}$$

which is a real symmetric matrix. This implies the Hessian is positive semi-definite and the function $f(w)$ is convex.

5.
$$f(w) = \|Xw - y\|_p + \lambda \|Aw\|_q \tag{26}$$

The first term on the right hand side of equation (27) is simply the affine composition of the $l_p - norm$ which we have showed is convex in 3. That term is therefore convex.

The second term of the expression is the scalar product of the affine composition of the $l_q - norm$ and is also convex.

The sum of two convex functions is convex and therefore $f(w)$ is convex.

6.
$$f(w) = \sum_{i=1}^{N} \max\{0, \mid w^T x_i - y_i \mid -\epsilon\} + \frac{\lambda}{2}\|w\|_2^2 \tag{27}$$

The second term on the right hand side is convex as it is the scalar product of the $l_2 - norm$ which is convex.

As of the first term on the right hand side, the function $g(w) = 0$ is convex. The function $h(w) = w^T x_i - y_i$ is convex as it is a linear function. taking the absolute value of $h(w)$ can be written as $|h(w)| = \max\{-h(w), h(w)\}$ and the maximum of two convex functions is convex. Substracting a constant ϵ remains convex as the addition of convex functions is convex. The maximum of $g(w)$ and $|h(w)| - \epsilon$ is therefore still convex and so is summing over i as it is the sum of convex functions.

As both terms on the right hand side are convex and $f(w)$ is the sum of those two terms, $f(w)$ is thus convex.

7.
$$f(w) = \max_{ijk}\{|x_i| + |x_j| + |x_k|\} \tag{28}$$

Once again, the absolute value $|x_i| = \max\{x_i, -x_i\}$ is convex as it is the maximum of two convex functions. The sum $|x_i| + |x_j| + |x_k|$ is convex as it is the sum of convex functions. The maximum over ijk preserves convexity and therefore $f(w)$ is convex.

3 Numerical Optimization

Alex