Optimization homework

Pierre Ablin

Instructions

Please read carefuly these instructions

You should answer the questions below. You should send me your results by the **18/10/2021**, 23:59 at my email adress pierreablin@gmail.com . The title of your email should be "Optimization homework ..." where "..." is your full name. You should submit your work individualy. You can either send a scan of a handwriten sheet, or a numeric document.

1 Convexity

1.1

Let f a convex function, and w_1, w_2 such that for all $w \in \mathbb{R}^d$, $f(w_1) \leq f(w)$ and $f(w_2) \leq f(w)$. Demonstrate that for all $\lambda \in [0, 1]$ and for all $w \in \mathbb{R}^p$ we have $f(\lambda w_1 + (1-\lambda) w_2) \leq f(w)$.

1.2

Let f a strongly convex function. Show that f only admits one minimizer.

Hint: you can use the strong convexity property:

$$\forall x, y \in \mathbb{R}^d, \;\; \lambda \in [0, 1], \;\; f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\lambda(1-\lambda)\mu}{2}\|x - y\|^2$$

1.3

Let f a convex, differentiable function. Demonstrate that $\nabla f(x^*) = 0$ if, and only if, for all $x \in \mathbb{R}^p$, $f(x) \leq f(x^*)$.

2 Gradient descent

Let f a L-smooth function. We let $w \in \mathbb{R}^p$, $\eta > 0$ a step size, and y the output of **one** gradient descent step:

$$y = w - \eta \nabla f(w)$$

2.1

Show that when $\eta \in [0, \frac{2}{L}]$, we have

$$f(y) \leq f(w).$$

2.2

Can we have a larger interval ?

Hint: you can try to find a L-smooth function such that $f(y) > f(w)$ whenever $\eta > \frac{2}{L}$.

2.3

We assume that f is μ-strongly convex. Demonstrate that it verifies the Polyak-Lojasciewicz inequality:

$$\forall w \in \mathbb{R}^p, f(w) - f^* \leq \frac{1}{2\mu} \|\nabla f(w)\|^2$$

3 Logistic regression

We define $\phi(x) = \log(1 + \exp(-x))$ for $x \in \mathbb{R}$. We let $X = [x_1, \ldots, x_n]^\top \in \mathbb{R}^{n \times p}$ and $y_1, \ldots, y_n = \pm 1$.

We define the logistic loss function

$$f(w) = \frac{1}{n} \sum_{i=1}^{n} \phi(y_i \langle w, x_i \rangle)$$

3.1

Show that ϕ is convex, and demonstrate that for all $x \in \mathbb{R}$, we have $\phi''(x) \leq \frac{1}{4}$.

3.2

Show that the Hessian of f can be written as

$$\nabla^2 f(w) = \frac{1}{n} X^\top D(w) X,$$

where $D(w)$ is the diagonal matrix of size $n \times n$ with coefficients $D(w)_{ii} = \phi''(y_i \langle w, x_i \rangle)$.

3.3

Show that f is convex, and find a smoothness constant of f.

Hint: you can use the fact that for two matrices A, B, $\sigma_{\max}(AB) \leq \sigma_{\max}(A)\sigma_{\max}(B)$.

4 Proximal operations

4.1

Demonstrate that the prox of the absolute value is

$$\operatorname{prox}_{u|\cdot|}(x) = \begin{cases} x - u \text{ if } x > u \\ x + u \text{ if } x < -u \\ 0 \text{ otherwise} \end{cases}$$

4.2

Compute the proximal operator of the squared ℓ_2 norm

$$\operatorname{prox}_{u\|\cdot\|_2^2}(x) = ?$$

4.3

Let C a closed, convex set. Define I_C its indicator function:

$$I_C(x) = \begin{cases} 0 \text{ if } x \in C \\ +\infty \text{ otherwise} \end{cases}$$

Give the expression of $\operatorname{prox}_{I_C}(u)$.

4.4

Assume that we want to solve $\min_{w \in C} f(w)$, where f is a L-smooth function.

Can you rewrite the problem as an unconstrained minimization problem using the indicator function? Write the associated proximal gradient algorithm update. Explain in your own words the meaning of these iterations.

4.5

We consider the Lasso problem

$$\min_w \frac{1}{2}\|Xw - y\|^2 + \lambda\|w\|_1$$

Demonstrate that $w = 0$ is a solution of the problem when $\lambda \geq \|X^\top y\|_\infty$.

5 Quasi-Newton methods

We consider A a positive definite symetric matrix. We let f a twice differentiable function. We consider the variable metric gradient descent iterations

$$x_{n+1} = x_n - \alpha A \nabla f(x_n).$$

For $C \in \mathbb{R}^{p\times p}$ an invertible matrix, we make the change of variable $x = Cy$, and define the function $g(y) = f(Cy)$.

5.1

Compute the gradient and Hessian of g.

5.2

We consider a gradient descent on g:

$$y_{n+1} = y_n - \alpha \nabla g(y_n).$$

What are the corresponding iterations for $x_n = Cy_n$? For which matrices C do they correspond to the variable metric gradient descent iterations?