Continuous optimization, an introduction

Exercices oraux

Exercice 1 - (Gradient à pas fixe)

Dans un espace de Hilbert $\mathcal{X}$, on considère $f : \mathcal{X} \to \mathbb{R}$, convexe, telle que ∇f est L-Lipschitzien.

On considère un point x et le point $x - d$, où d est un direction à trouver.

1. Montrer que

$$f(x-d) \leq f(x) - \nabla f(x) \cdot d + \frac{L}{2}\|d\|^2. \tag{1}$$

Minimiser cette expression par rapport à d. On définit l'algorithme de descente comme un algorithme qui à x associe $x + d$. .

2. Soit $\bar{x} \in \mathcal{X}$ et $\hat{x} = \bar{x} - \bar{d}$ la direction trouvée précédemment. Montrer que pour tout x

$$f(x) + \frac{L}{2}\|x - \bar{x}\|^2 \geq f(\hat{x}) + \frac{L}{2}\|x - \hat{x}\|^2. \tag{2}$$

3. On introduit l'algorithme suivant (de Nesterov, 1983): étant donné $x^0 = x^{-1} = \bar{x}^{-1} \in \mathcal{X}$, $t_0 = 0$, on définit pour $k \geq 0$:

- $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$;
- $\bar{x}^k = x^k + \frac{t_k - 1}{t_{k+1}}(x^k - x^{k-1})$
- $x^{k+1} = \bar{x}^k - \frac{1}{L}\nabla f(x^k)$.

Vérifier que $t_k \geq (k+1)/2$ si $k \geq 1$. En prenant $x = ((t_{k+1}-1)x^k + x^*)/t_{k+1}$, $\hat{x} = x^{k+1}$ et $\bar{x} = \bar{x}^k$ dans (2), montrer par récurrence que pour tout $k \geq 1$,

$$f(x^k) - f(x^*) \leq L\frac{\|x^* - x^0\|^2}{2t_k^2} \tag{3}$$

Conclure.

Exercice 2

Calculer les transformées de Legendre (conjuguées convexes) des fonctions

1. $x \mapsto |x|^3/3$;
2. $x \mapsto 3x$;
3. $x \mapsto \langle Ax, x\rangle /2$ où $x \in \mathbb{R}^N$ et A est une matrice définie positive;
4. $x \mapsto -\sqrt{x}$ si $x \geq 0$, $+\infty$ si $x < 0$.

Exercice 3: normes duales

Show that if $\|x\|$ is a norm and $\|y\|^\circ = \sup_{\|x\|\le 1} \langle x, y\rangle$ is the *polar* or *dual* norm, then

$$\|\cdot\|^*(y) = \delta_{B_{\|\cdot\|^\circ}(0,1)}(y) = \begin{cases} 0 & \text{if } \|y\|^\circ \le 1, \\ +\infty & \text{else.} \end{cases}$$

Hint: write $\sup_x \langle x, y\rangle - \|x\|$ as $\sup_{t>0}\left(\sup_{\|x\|\le t} \langle x, y\rangle\right) - t$.

What is $\|\cdot\|^{\circ\circ}$?

Exercice 4 (*Schatten norms*)

Let $X \in \mathbb{R}^{n\times p}$ be a matrix.

a. Show that $X^T X$ and XX^T are a symmetric $p\times p$ and $n \times n$ (respectively) matrix and that they have the same nonzero eigenvalues $(\lambda_1,\dots,\lambda_k)$ ($k \le \min\{p, n\}$). The values $\mu_i = \sqrt{\lambda_i}$ are the "singular values" of X.

b. Show that if $(e_1,\dots,e_p)$ is an orthonormal basis of eigenvectors of $X^T X$ (associated to the eigenvalues λ_i, or 0 if $i > k$), then $(Xe_i)_i$ are orthogonal. Show that one can write, for $\mu_i > 0$, $Xe_i = \mu_i f_i$ where f_i are also orthonormal. Completing f_i into an orthonormal basis of $\mathbb{R}^n$, deduce that

$$X = \sum_{i=1}^{k} \mu_i f_i \otimes e_i = VD\,{}^tU$$

where U is the column vectors $(e_i)_{i=1}^p$, V the column vectors $(f_i)_{i=1}^n$, D is the $n\times p$ matrix with $D_{ii} = \mu_i$, $i = 1,\dots,k$, $D_{ij} = 0$ for all other entries (just evaluate $Xx = X(\sum_{i=1}^p \langle x, e_i\rangle e_i)$, etc.) What type of matrices are the matrices U, V? This is called the "singular value decomposition" (SVD) of X (one usually orders the μ_i by nonincreasing values).

c. One defines the p-Schatten norm of X, $p \in [1,\infty]$, as $\|X\|_p^p = \sum_{i=1}^k \mu_i^p$, $\|X\|_\infty = \max_i \mu_i$. Show that

$$\|X\|_2^2 = \sum_{i,j} x_{i,j}^2 = \mathrm{Tr}({}^tX\,X); \quad \|X\|_\infty = \sup_{\|x\|\le 1} \|Xx\|.$$

(where in the latter $\|x\|$ is the 2-norm). $\|\cdot\|_\infty$ is called the *spectral* norm or *operator* norm.

d. [Exercice **3.** is necessary for this question.] Why do we have that

$$\{X : \|X\|_1 \le 1\} = \mathrm{conv}\{f\otimes e : f \in \mathbb{R}^n, e \in \mathbb{R}^p, \|f\| \le 1, \|e\| \le 1\}\,?$$

Deduce that

$$\|X\|_\infty = \sup_{\{\|Y\|_1\le 1\}} \langle Y, X\rangle$$

where we use the Frobenius (or Hilbert-Schmidt) scalar product $\langle Y, X\rangle = \sum_{i,j} Y_{i,j}X_{i,j} = \mathrm{Tr}({}^tY X)$. Deduce that

$$\|X\|_1 = \sup_{\{\|Y\|_\infty\le 1\}} \langle Y, X\rangle\,.$$

(One can also show that $\|\cdot\|_p^\circ = \|\cdot\|_{p'}$, $1/p + 1/p' = 1$.)

e. We want to compute

$$\bar{Y} = \arg\min_{\|X\|_\infty \le 1} \|X - Y\|_2^2 = \mathrm{prox}_{\delta_{\{\|X\|_\infty \le 1\}}}(Y). \qquad (P_\infty)$$

Show first that it is equivalent to estimate $\min_{\|X\|_\infty \le 1} \|X - D\|_2^2$ where $Y = VD\,{}^tU$ is the SVD decomposition of Y. Show that the matrix X which optimizes this last problem is diagonal, and satisfies $X_{i,i} = \max\{D_{i,i}, 1\}$. Deduce the solution $\bar{Y}$ of (P_∞). Deduce the proximity operator $\mathrm{prox}_{\tau\|\cdot\|_1}$.

f. A company rents movies and has a file of clients $X_{i,j} \in \{-1, 0, 1\}$ which states for each client $i = 1, \dots, p$ whether he/she has already rented the film $j = 1, \dots, n$ (otherwise $X_{i,j} = 0$) and has liked it ($X_{i,j} = 1$), or not ($X_{i,j} = -1$). It wants to determine a matrix of "tastes" for all the clients $Y \in \{-1, 1\}^{p \times n}$. Assuming that the clients can be grouped into few categories, this matrix should have low rank. One could look therefore for an approximation of Y by minimising

$$\min_Y \|Y\|_1 + \frac{\lambda}{2} \sum_{i,j : X_{i,j} \neq 0} (X_{i,j} - Y_{i,j})^2 + \frac{\varepsilon}{2} \sum_{i,j : X_{i,j} = 0} Y_{i,j}^2$$

where $\lambda >> \varepsilon > 0$ are parameters.

Design an iterative algorithm to solve this problem.

Exercice 5 (projection sur le simplexe)

On considère dans $\mathbb{R}^N$ la fonction

$$f : x \mapsto \max_{i=1,\dots,N} x_i.$$

1. Evaluer la transformée de Legendre de f. On commencera par remarquer qu'il ságit de la fonction caractéristique d'un convexe qu'on cherchera ensuite à déterminer.

2. Soit $\tau > 0$. On veut calculer le "prox" de f, soit

$$\arg\min_x \frac{\|x - \bar{x}\|^2}{2\tau} + \max x_i.$$

Trouver un algorithme pour calculer x.

3. En déduire un algorithme de projection sur le simplexe unité ($\Sigma = \{v \in \mathbb{R}^N : v_i \geq 0, \sum_i v_i = 1\}$).