Continuous optimization, an introduction

Exercises (22th Nov. 2016)

1. We recall that for a convex function $f : X \to \mathbb{R}$,

$$\mathrm{prox}_{\tau f}(x) = \arg\min_y f(y) + \frac{1}{2\tau}\|y - x\|^2.$$

Evaluate $\mathrm{prox}_{\tau f}(x)$ for $\tau > 0$, and

- $x \in \mathbb{R}^N$, $f(x) = \frac{1}{2}\|x - x^0\|^2$: what does Moreau's identity give in this case, for $x^0 = 0$?;
- $x \in \mathbb{R}^N$, $f(x) = \delta_{\{\|x\| \le 1\}}$[1];
- $x \in \mathbb{R}^N$, $f(x) = \langle p, x\rangle - \sum_i g_i \log x_i$ ($g_i > 0$) if $x_i > 0$ for all $i = 1, \dots, n$ and $f(x) = +\infty$ else;
- $x \in \mathbb{R}^N$, $f(x) = \delta_{\{|x_i| \le 1 : i=1,\dots,N\}}(x) + \varepsilon\|x\|^2/2$.

2. Evaluate the convex conjugate of:

- $f(x) = \delta_{\{|x_i| \le 1 : i=1,\dots,N\}}(x) + \varepsilon\|x\|^2/2$, $x \in \mathbb{R}^N$;
- $f(x) = \sum_{i=1}^N x_i \log x_i$, $x \in \mathbb{R}^N$, where $x \mapsto x \log x$ is $+\infty$ for $x < 0$ and extended by continuity (that is, with the value 0) in $x = 0$;
- $f(x) = \sqrt{1 + \|x\|_2^2}$, $x \in \mathbb{R}^N$.

In each case of the three cases above, describe ∂f and ∂f^*.

3. 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 $\|\cdot\|^{\circ\circ}$?

4. (*Schatten norms*) Let $X \in \mathbb{R}^{n \times p}$ be a matrix.

a. Show that $X^T X$ and $X X^T$ 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

$$X = \sum_{i=1}^k \mu_i f_i \otimes e_i = V D\, {}^t U$$

[1] $\delta_C(x) = 0$ if $x \in C$, $+\infty$ if $x \notin C$

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

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

(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 $\langle Y, X \rangle = \sum_{i,j} Y_{i,j} X_{i,j} = \mathrm{Tr}({}^t Y X)$. 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 $Y = V D\, {}^t U$ 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.