# Exercise List: Proximal Operator.

Robert M. Gower and Francis Bach

April 19, 2019

## 1 Introduction

This is an exercise in deducing closed form expressions for proximal operators. In the first part we will show how to deduce that the proximal operator of the L1 norm is the soft-thresholding operator. In the second part we will show the equivalence between the proximal operator of the matrix nuclear norm and the singular value soft-thresholding operator.

First some necessary notation.

**Notation:** For every $x, y, \in \mathbb{R}^n$ let $\langle x, y\rangle \overset{\text{def}}{=} x^\top y$ and let $\|x\|_2 = \sqrt{\langle x, x\rangle}$. Let $\sigma(A) = [\sigma_1(A), \ldots, \sigma_n(A)]$ be the singular values of $A$.

Let $\|A\|_F^2 \overset{\text{def}}{=} \mathrm{Tr}\left(A^\top A\right) = \sum_{ij} A_{ij}^2$ denote the Frobenius norm of $A$ and let $\|A\|_* = \sum_i \sigma_i(A)$ be the nuclear norm.

## 2 Soft Thresholding

Let $f : x \in \mathbb{R}^d \to f(x)$ be a convex function. Consider the proximal operator

$$\mathrm{prox}_f(v) \overset{\text{def}}{=} \arg\min_x \frac{1}{2}\|x - v\|_2^2 + f(x). \tag{1}$$

**Ex. 1 —** In this exercise we will show step-by-step that the proximal operator of the L1 norm is the soft thresholding operator, that is

$$\mathrm{prox}_{\lambda\|w\|_1}(v) = \left(S_\lambda(v_1), \ldots, S_\lambda(v_n)\right), \tag{2}$$

where

$$S_\lambda(v) = \begin{cases} v - \lambda & \text{if } \lambda < v \\ 0 & \text{if } -\lambda \leq v \leq \lambda \\ v + \lambda & \text{if } v < -\lambda. \end{cases} \tag{3}$$

*Part I*

Show that if $f(x)$ is separable, that is, if $f(x) = \sum_{i=1}^{d} f_i(x_i)$ then

$$\mathrm{prox}_f(v) = (\mathrm{prox}_{f_1}(v_1), \ldots, \mathrm{prox}_{f_d}(v_d)). \tag{4}$$

Consequently

$$\mathrm{prox}_{\lambda\|w\|_1}(v) = (\mathrm{prox}_{\lambda|w_1|}(v_1), \ldots, \mathrm{prox}_{\lambda|w_d|}(v_d)).$$

*Part II*

Show that if

$$\alpha^* = \arg\min_{\alpha} \frac{1}{2}(\alpha - v)^2 + \lambda|\alpha| \tag{5}$$

then

$$\alpha^* \in v - \lambda\partial|\alpha^*|. \tag{6}$$

Note that by definition $\alpha^* = \mathrm{prox}_{\lambda|\alpha|}(v)$.

*Part III*

If $\lambda < v$ show that the solution to the inclusion (**??**) is given by

$$\alpha^* = v - \lambda.$$

*Part IV*

If $-\lambda < v < \lambda$ show that the solution to the inclusion (**??**) is given by

$$\alpha^* = 0.$$

*Part V*

Using the previous items, prove that

$$\mathrm{prox}_{\lambda|\alpha|}(v) = S_\lambda(v)$$

and that the equality (**??**) holds.

*Part VI*

Show that the soft-threshold operator can be written in a more compact way as

$$S_\lambda(v) = sign(v)(|v| - \lambda)_+, \tag{7}$$

where $(\alpha)_+ \overset{\text{def}}{=} \max_{\{} 0, \alpha\}$ and $sign(v)$ is the sign function given by $sign(v) = 1_{v\geq 0} - 1_{v<0}$. This can be implemented efficiently in python using numpy and

```
prox_L1(v,lmbda)  =   np.sign(v) * np.maximum(np.abs(v) - lmbda, 0.)
```

The above code also works when $v$ is a vector!

## 3 Singular Value Soft Thresholding

Consider the extension of proximal operators to matrices

$$\mathrm{prox}_F(A) \stackrel{\text{def}}{=} \arg\min_{X\in\mathbb{R}^{d\times d}} \frac{1}{2}\|X-A\|_F^2 + F(X). \tag{10}$$

We will now prove step by step that

$$\mathrm{prox}_{\lambda\|X\|_*}(A) = US_\lambda(\mathrm{diag}(\sigma(A)))V^\top, \tag{11}$$

where $\|X\|_* = \sum_{i=1}^d \sigma_i(X)$ and $A = U\mathrm{diag}(\sigma_i(A))V^\top$ is the singular value decomposition of $A$.

This proximal operator forms the basis of the celebrated algorithm for solving the matrix completion problem [**CaiCandes:2010**].

**Ex. 2** — *Part I*

Show that the nuclear and the Frobenius norm are invariant under rotations. That is, for any matrix $A$ and orthogonal matrices $O$ and $Q$ we have that

$$\|A\|_F^2 = \|OA\|_F^2 = \|AQ\|_F^2$$

and

$$\|A\|_* = \|OA\|_* = \|AQ\|_*.$$

*Part II*

**(Level HARD):** Prove that (**??**) holds. You may use the following Theorem by Von Neumann

**Theorem 3.1 (Von Neumann 1937)** *For any matrices $X$ and $A$ of the same dimensions and orthogonal matrices $U$ and $V$, we have that*

$$\left\langle UXV^\top, A\right\rangle \leq \langle diag(\sigma_i(X)), diag(\sigma_i(A))\rangle\,, \tag{12}$$

*where $diag(\sigma_i(A))$ is a diagonal matrix with the singulars values of $A$ on the diagonal.*