Exercise List: Proximal Operator.

Robert M. Gower.

September 26, 2017

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.

nuclear norm is the convex envelop of matrix rank !

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

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

Consequently

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

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 $\alpha^* = \text{prox}_{\lambda|\alpha|}(v)$.

Part III

If $\lambda < v$ show that the solution to the inclusion (6) is given by

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

Part IV

If $-\lambda < v < \lambda$ show that the solution to the inclusion (6) is given by

$$\alpha^* = 0.$$

Part V

Using the previous items, prove that

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

and that the equality (10) holds.

Answer (Ex. I) — The trick lies in noticing that

$$\begin{aligned}
\min_x \frac{1}{2}\|x - v\|_2^2 + \sum_{i=1}^d f_i(x_i) &= \min_x \frac{1}{2}\sum_{i=1}^d (x_i - v_i)^2 + \sum_{i=1}^d f_i(x_i) \\
&= \sum_{i=1}^d \min_{x_i} \frac{1}{2}(x_i - v_i)^2 + f_i(x_i).
\end{aligned}$$

Consequently

$$\begin{aligned}\arg\min_x \frac{1}{2}\|x-v\|_2^2 + f(x) &= \left(\arg\min_{x_1}\left\{\frac{1}{2}(x_1-v_1)^2 + f_1(x_1)\right\}, \ldots, \arg\min_{x_d}\left\{\frac{1}{2}(x_d-v_d)^2 + f_d(x_d)\right\}\right) \\ &= \left(\mathrm{prox}_{f_1}(v_1), \ldots, \mathrm{prox}_{f_d}(v_d)\right).\end{aligned}$$

Answer (Ex. II) — The solution to (5) must be such that

$$0 \in \partial\left(\frac{1}{2}(\alpha^* - v)^2 + \lambda|\alpha^*|\right) = \alpha^* - v + \lambda\partial|\alpha^*|.$$

Rearranging the above gives $\alpha^* \in v - \lambda\partial|\alpha^*|$.

Answer (Ex. III) — First note that

$$v - \lambda\partial|\alpha^*| \subset [v-\lambda,\, v+\lambda]. \tag{7}$$

If $\lambda < v$ then the above together with inclusion (6) shows that

$$\alpha^* \in \,]0,\, \infty[\quad \Rightarrow \quad \partial|\alpha^*| = 1.$$

Consequently (6) shows that

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

Answer (Ex. IV) — If $-\lambda \le v \le \lambda$ then due to (7) the solution to the inclusion (6) is bounded by

$$\alpha^* \in v - \lambda\partial|\alpha^*| \subset \begin{cases}\{v+\lambda\} & \text{if } \alpha^* < 0 \\ [v-\lambda,\, v+\lambda] & \text{if } \alpha^* = 0. \\ \{v-\lambda\} & \text{if } \alpha^* > 0\end{cases} \tag{8}$$

Now suppose that $\alpha^* < 0$. The above shows that

$$\alpha^* \in [0,\, 2\lambda]\,,$$

a contradiction. If $\alpha^* > 0$, then (8) shows that

$$\alpha^* \in [-2\lambda,\, 0]\,,$$

another contraction. Finally if $\alpha^* = 0$ then (8) offers no contradiction since it is equivalent to

$$\alpha^* \in [-2\lambda,\, 2\lambda]\,.$$

Consequently $-\lambda < v < \lambda \Rightarrow \alpha^* = 0$.

Answer (Ex. V) — Using analogous arguments to Ex III we can show that $v < -\lambda \Rightarrow \alpha^* = v + \lambda$. By combing this observation together the solutions of Ex III and IV we have that

$$\alpha^* = \begin{cases} v - \lambda & \text{if } \lambda < v \\ 0 & \text{if } -\lambda \leq v \leq \lambda \\ v + \lambda & \text{if } v < -\lambda. \end{cases}$$

Thus

$$\mathrm{prox}_{\lambda|\alpha|}(v) = \alpha^* \overset{(3)}{=} S_\lambda(v).$$

3 Singular Value Soft Thresholding

Consider the extension of proximal operators to matrices

$$\mathrm{prox}_F(A) \overset{\text{def}}{=} \arg\min_{X \in \mathbb{R}^{d \times d}} \frac{1}{2}\|X - A\|_F^2 + F(X). \tag{9}$$

We will now prove step by step that

$$\mathrm{prox}_{\lambda\|X\|_*}(A) = US_\lambda(\mathrm{diag}(\sigma(A)))V^\top, \tag{10}$$

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 [1].

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 (10) 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*

$$\left\langle UXV^\top, A \right\rangle \leq \langle \mathit{diag}(\sigma_i(X)), \mathit{diag}(\sigma_i(A)) \rangle\,, \tag{11}$$

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

it is also called Von Neumann's trace inequality

Answer (Ex. I) — By the definition of Frobenius norm we have that

$$\|OA\|_F^2 = \mathrm{Tr}\left(A^\top (O^\top O)A\right) = \mathrm{Tr}\left(A^\top A\right) = \|A\|_F^2.$$

For the nuclear norm, note that for any orthogonal matrices O and U we have that OU is an orthogonal matrix since

$$(OU)^\top OU = U^\top (O^\top O)U = U^\top U = I.$$

Thus is the SVD decomposition is given by $A = U\mathrm{diag}(\sigma_i(A))V^\top$ then $OA = OU\mathrm{diag}(\sigma_i(A))V^\top$ is the SVD decomposition of OA, that is, the matrix OA has the same singular values of A. Consequently we have that

$$\|OA\|_* = \|(OU)\mathrm{diag}(\sigma_i(A))V^\top\| = \sum_{i=1}^d \sigma_i(A) = \|A\|_*,$$

by definition of nuclear norm.

Answer (Ex. II) — Substituting $A = U\mathrm{diag}(\sigma_i(A))V^\top$ gives

$$\mathrm{prox}_{\lambda\|X\|_*}(A) = \arg\min_{X\in\mathbb{R}^{d\times d}} \frac{1}{2}\|U(U^\top XV - \mathrm{diag}(\sigma_i(A)))V^\top\|_F^2 + \lambda\|X\|_*.$$

Changing variable name

$$\bar{X} = U^\top XV, \tag{12}$$

and noting that the Frobenius norm and the nuclear norm are invariant to orthogonal transforms we have that

$$\min_{X\in\mathbb{R}^{d\times d}} \frac{1}{2}\|U(\bar{X} - \mathrm{diag}(\sigma_i(A)))V^\top\|_F^2 + \lambda\|U\bar{X}V^\top\|_* = \min_{X\in\mathbb{R}^{d\times d}} \frac{1}{2}\|\bar{X} - \mathrm{diag}(\sigma_i(A))\|_F^2 + \lambda\|\bar{X}\|_*. \tag{13}$$

I now claim that the solution $\bar{X}$ to the above must be a diagonal matrix. This is where Von Neumann's theorem comes into play. To see this let $U_x\mathrm{diag}(\sigma_i(X))V_x^\top$ be the SVD decomposition of X. Thus

$$\begin{aligned}
\|\bar{X} - \mathrm{diag}(\sigma_i(A))\|_F^2 &= \|U_x\mathrm{diag}(\sigma_i(X))V_x^\top - \mathrm{diag}(\sigma_i(A))\|_F^2 \\
&= \|\mathrm{diag}(\sigma_i(X))\|_F^2 + \|\mathrm{diag}(\sigma_i(A))\|_F^2 - 2\left\langle U_x\mathrm{diag}(\sigma_i(X))V_x^\top, \mathrm{diag}(\sigma_i(A))\right\rangle \\
&\overset{\text{Theorem 3.1}}{\geq} \|\mathrm{diag}(\sigma_i(X))\|_F^2 + \|\mathrm{diag}(\sigma_i(A))\|_F^2 - 2\left\langle \mathrm{diag}(\sigma_i(X)), \mathrm{diag}(\sigma_i(A))\right\rangle \\
&= \|\mathrm{diag}(\sigma_i(X)) - \mathrm{diag}(\sigma_i(A))\|_F^2.
\end{aligned}$$

Consequently

$$\min_{\bar{X}} \frac{1}{2}\|\bar{X} - \mathrm{diag}(\sigma_i(A))\|_F^2 + \lambda\|\bar{X}\|_* \quad \geq \quad \min_{\bar{X}} \frac{1}{2}\|\mathrm{diag}(\sigma_i(\bar{X})) - \mathrm{diag}(\sigma_i(A))\|_F^2 + \lambda\|\mathrm{diag}(\sigma_i(\bar{X}))\|_*,$$

where we used the invariance of the nuclear norm under orthogonal transformations. This proves that the solution $\bar{X} = \mathrm{diag}(\bar{X}_{11}, \ldots, \bar{X}_{dd})$ will be a diagonal matrix. From now on we assume that $\bar{X} = \mathrm{diag}(\bar{X}_{ii})$ is a diagonal matrix. Let $\bar{x} = (\bar{X}_{11}, \ldots, \bar{X}_{dd})$ be the vectorization of $\bar{X}$. Thus $\|\bar{X}\|_* = \|\bar{x}\|_1$ and $\|\bar{X}\|_F^2 = \|\bar{X}\|_2^2$. Let $\sigma(A) = [\sigma_1(A), \ldots, \sigma_d(A)] \in \mathbb{R}^d$. Finally we have that (13) becomes

$$\min_{X \in \mathbb{R}^{d\times d}} \frac{1}{2}\|\bar{X} - \mathrm{diag}(\sigma_i(A))\|_F^2 + \lambda\|\bar{X}\|_* = \min_{\bar{x} \in \mathbb{R}^d} \frac{1}{2}\|\bar{x} - \sigma(A)\|_2^2 + \lambda\|\bar{x}\|_1.$$

Consequently, taking the minimum argument we have that

$$S_\lambda(\mathrm{diag}(\sigma(A))) = \arg \min_{\bar{X} \text{ is diag}} \frac{1}{2}\|\bar{X} - \mathrm{diag}(\sigma_i(A))\|_F^2 + \|\bar{X}\|_*,$$

where $S_\lambda(\mathrm{diag}(\sigma(A))) := \mathrm{diag}(S_\lambda(\sigma_i(A)))$. To conclude, note that our original argument is $UXV^\top = \bar{X}$ due to (12). Thus finally

$$\mathrm{prox}_{\lambda\|X\|_*}(A) = \arg\min_X \|X - A\|_F^2 + \lambda\|X\|_* = US_\lambda(\mathrm{diag}(\sigma(A)))V^\top. \qquad \blacksquare$$

References

[1] J.-F. Cai, E. J. Candès, and Z. Shen. “A Singular Value Thresholding Algorithm for Matrix Completion”. In: *SIAM J. on Optimization* 20.4 (Mar. 2010), pp. 1956–1982.