Elia Fantini - Sciper n. 336006

Math of Data: Homework 3

Exercise 1.1.a

Given $\mathcal{X} \in \mathbb{R}^{p \times m}$, we have that its indicator function $\delta_{\mathcal{X}}$ is the following:

$$\delta_{\mathcal{X}}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{X} \\ \infty & \text{if } X \notin \mathcal{X} \end{cases}$$

Since $\text{prox}_{\delta_{\mathcal{X}}}(Z) = \underset{X \in \mathbb{R}^{p \times m}}{\arg\min} \left\{ \delta_{\mathcal{X}}(X) + \frac{1}{2} \| X - Z \|_F^2 \right\}$,

due to the definition of $\delta_{\mathcal{X}}$ every $X \notin \mathcal{X}$ can not be a solution of this minimization problem. We can then rewrite $\text{prox}_{\delta_{\mathcal{X}}}$ as follows:

$$\text{prox}_{\delta_{\mathcal{X}}}(Z) = \underset{X \in \mathbb{R}^{p \times m}}{\arg\min} \left\{ \frac{1}{2} \| X - Z \|_F^2 : X \in \mathcal{X} \right\}$$

$$= \underset{X \in \mathcal{X}}{\arg\min} \left\{ \frac{1}{2} \| X - Z \|_F^2 \right\}$$

$$= \underset{X \in \mathcal{X}}{\arg\min} \left\{ \| X - Z \|_F^2 \right\} = \text{proj}_{\mathcal{X}}(Z)$$

$\frac{1}{2}$ is just a constant term, can be removed as it does not influence the solution.

Exercise 1.1.6

Since we have that $z^* = proj_{\mathcal{X}}(x) \iff \langle x - z^*, z - z^* \rangle \leq 0$ $\forall z \in \mathcal{X}$ and the projection of a point into $\mathcal{X} \in \mathcal{X}$, then we have that:

$$\langle x - proj_{\mathcal{X}}(x), proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq 0$$

and $\langle y - proj_{\mathcal{X}}(y), proj_{\mathcal{X}}(x) - proj_{\mathcal{X}}(y) \rangle \leq 0$

$\Rightarrow \langle proj_{\mathcal{X}}(y) - y, proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq 0$

If we add the two inequalities and use the property of inner product called additivity that says that $\langle x,z \rangle + \langle y,z \rangle = \langle x+y, z \rangle$ then:

$$\langle x - proj_{\mathcal{X}}(x), proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle + \langle proj_{\mathcal{X}}(y) - y, proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq 0$$

$\iff \langle x - proj_{\mathcal{X}}(x) + proj_{\mathcal{X}}(y) - y, proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq 0$

We apply the same property:

$\langle x - y, proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle + \langle proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x), proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq 0$

$\iff \langle x - y, proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \rangle \leq - \| proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \|_2^2 \cdot \cos(0)$

$\iff \| (x-y)^T (proj_{\mathcal{X}}(x) - proj_{\mathcal{X}}(y)) \|_2 \geq \| proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \|_2^2$

For Cauchy-Schwarz inequality we have that:

$$\| (x-y)^T (proj_{\mathcal{X}}(x) - proj_{\mathcal{X}}(y)) \|_2 \leq \| (x-y) \|_2 \| proj_{\mathcal{X}}(x) - proj_{\mathcal{X}}(y) \|_2$$

$\Rightarrow \| proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \|_2^2 \leq \| x - y \|_2 \cdot \| proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \|_2$

$\iff \| proj_{\mathcal{X}}(y) - proj_{\mathcal{X}}(x) \|_2 \leq \| y - x \|_2$

Exercise 1.1. c

Given a matrix A, its nuclear norm is defined as:

$\|A\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A)$, where A is a $m \times n$ matrix and

σ_i are its singular values.

If $Z = U \Sigma V^T$ is the singular value decomposition of Z, then Σ is diagonal with singular values that are all greater or equal to zero. Hence, given the L1-norm definition as $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ where x is a n-dimensional vector,

we can see that $\|\Sigma\|_* = \|\sigma\|_1$, with σ being the vector corresponding to the diagonal of Σ.

Then $\mathcal{X} = \{X : X \in \mathbb{R}^{p \times m}, \|X\|_* \leq k\} = \{X : X \in \mathbb{R}^{p \times m}, \|X\|_1 \leq k\}$

Since, given $\mathcal{X}$ and matrix $Z \in \mathbb{R}^{p \times m}$, we have that $\text{proj}_{\mathcal{X}}(Z) = \underset{X \in \mathcal{X}}{\text{argmin}} \|X - Z\|_F$ is the projection of Z into the nuclear norm ball $\mathcal{X}$,

then: $\text{proj}_{\mathcal{X}}(\Sigma_Z) = \Sigma_Z^{\ell_1}$, where $\Sigma_Z^{\ell_1}$ is the projection into the ℓ_1-norm ball of the same radius of $\mathcal{X}$.

From Mirsky's inequality we know that $\|X - Z\|_F \geq \|\Sigma_X - \Sigma_Z\|_F$ and as previously proved $\underset{\Sigma_X \in \mathcal{X}}{\text{argmin}} \|\Sigma_X - \Sigma_Z\|_F = \Sigma_Z^{\ell_1}$,

then $\|\Sigma_Z^{\ell_1} - \Sigma_Z\|_F$ is a lower bound for the projection argument $\|X - Z\|_F$.

Furthermore we have that: $\|U \Sigma_Z^{\ell_1} V^T - U \Sigma_Z V^T\|_F =$

$= \|U(\Sigma_Z^{\ell_1} - \Sigma_Z) V^T\|_F = \|\Sigma_Z^{\ell_1} - \Sigma_Z\|_F$ because U and V are unitary matrices by definition and orthogonal, and Frobenius norm is invariant of orthogonal matrices.

Then for $X = U \Sigma_Z^{\ell_1} V^T$ we reach the lower bound for $\|X - Z\|_F$, therefore $\text{proj}_{\mathcal{X}}(Z) = U \Sigma_Z^{\ell_1} V^T$

Exercise 1.2

As said in the hint, by definition $Kuv^T \in \mathcal{X}$, so we need to show
$\langle X, Z \rangle \geq \langle -Kuv^T, Z \rangle \quad \forall X \in \mathcal{X}$

We have that $\langle -Kuv^T, Z \rangle = Tr(Z^T(-Kuv^T)) = -K Tr(Z^T uv^T)$
Since u is the left singular vector corresponding to the largest singular value of Z, that we can call σ, then by definition $Z^T u = \sigma v$, hence:

$-K Tr(Z^T uv^T) = -K Tr(\sigma vv^T) = -K\sigma$, where v is an unit vector of the unitary matrix V, so $Tr(vv^T) = 1$.

By Holder's inequality we have that: $|\langle X, Z \rangle| \leq \|X\|_* \cdot \|Z\|_\infty$ with $q=1$ and $r=\infty$. From the inequality above we get the following lower bound: $\langle X, Z \rangle \geq -\|X\|_* \cdot \|Z\|_\infty$

Given that $X \in \mathcal{X}$, then $\|X\|_* \leq K \Rightarrow -\|X\|_* \geq -K$. Then, since the spectral norm of a matrix coincides with its largest singular value, we have that $\|Z\|_\infty = \sigma$.
Therefore:
$$\langle X, Z \rangle \geq -\|X\|_* \|Z\|_\infty \geq -K\sigma = \langle -Kuv^T, Z \rangle \quad \forall X \in \mathcal{X}$$

Exercise 1.3.a

In the following table are reported the execution times to compute the projection of the matrix onto the nuclear norm ball.

Table 1: Projection onto the nuclear norm ball of radius k=5000

Run #:	1	2	3	4	5	**Average**
100k data:	0.8051 sec	0.7532 sec	0.7312 sec	0.7292 sec	0.7492 sec	**0.7536**
1M data:	56.1070 sec	56.9551 sec	59.7932 sec	56.4497 sec	58.5015 sec	**57.5613**

Exercise 1.3.b

In the following table are reported the execution times to compute the linear minimization oracle of the nuclear norm ball.

Table 2: Computing LMO with k=5000

Run #:	1	2	3	4	5	**Average**
100k data:	0.0189 sec	0.0149 sec	0.0209 sec	0.0189 sec	0.0229 sec	**0.0193**
1M data:	0.2277 sec	0.2527 sec	0.2847 sec	0.2827 sec	0.2297 sec	**0.2555**

From the two tables we can see that generally LMO is much faster than computing projections, and it is also more scalable. In fact, projections require the singular value decomposition that has a complexity of $O(\min(m^2p, mp^2))$, making the computation time grow exponentially with the size of the matrix and, therefore, with the amount of data. On the other hand, LMO remains fast even with 1M data, with a computation time just 13.24x slower, on average.

Exercise 1.4.a

Given the objective function $f(X) = \frac{1}{2} \| A(X) - b \|_2^2$, where A is a linear operator, let's compute the gradient:

$$\nabla f(X) = 2 \cdot \frac{1}{2} \cdot A^T (A(x) - b) = A^T(A(x) - b)$$

Then: $\| \nabla f(X) - \nabla f(Y) \|_2 = \| A^T(A(x) - b) - A^T(A(Y) - b) \|_2$

$$= \| A^T A(x) - A^T b - A^T A(Y) + A^T b \|_2$$

$$= \| A^T A (X - Y) \|_2$$

For Cauchy-Schwartz inequality $\| A^T A (X-Y) \|_2 \leq \| A^T A \|_{2 \to 2} \cdot \| X - Y \|_2$

A function f is called Lipschitz continous if

$$\| f(X) - f(Y) \|_2 \leq L \cdot \| X - Y \|_2 \quad \forall X, Y \in A,$$

where $f : A \to B$.

Since we proved that $\| \nabla f(X) - \nabla f(Y) \|_2 \leq \| A^T A \|_{2 \to 2} \cdot \| X - Y \|_2$ then ∇f is Lipschitz continuous.

Exercise 1.4.b

Starting from the blurred image in Figure 1 and using a rough estimate for the support, a 17x17 pixels box at the center of the domain shown in Figure 2, we could already obtain good results. Computing the LMO of the nuclear norm ball with radius k=1000, the plate was easy to guess after just 15 iterations, and after 30 it was clearly readable that the blurred plate is “J209 LTL”. Figures 3 and 4 show the results.

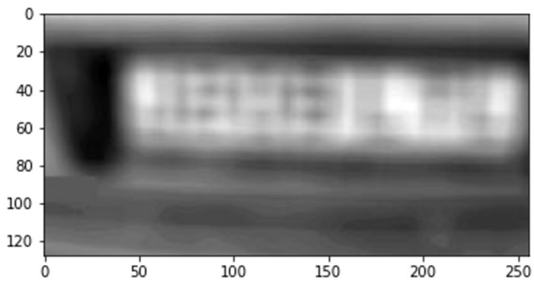

Figure 1 Blurred plate

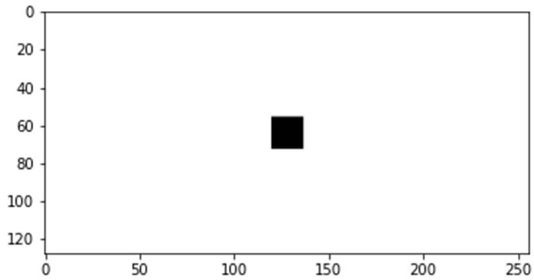

Figure 2 Support estimate

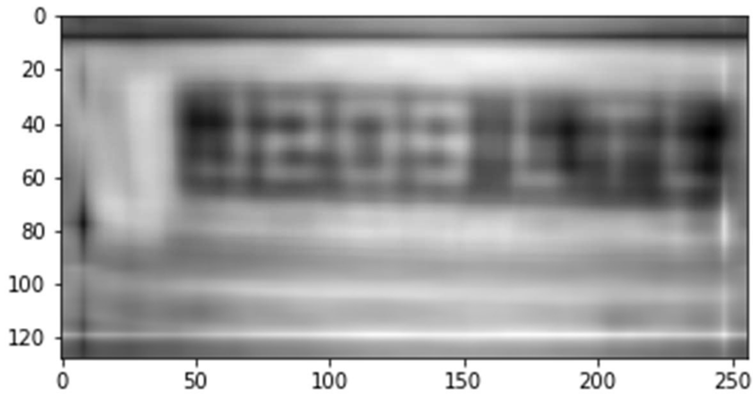

Figure 3 Result after 33 iterations

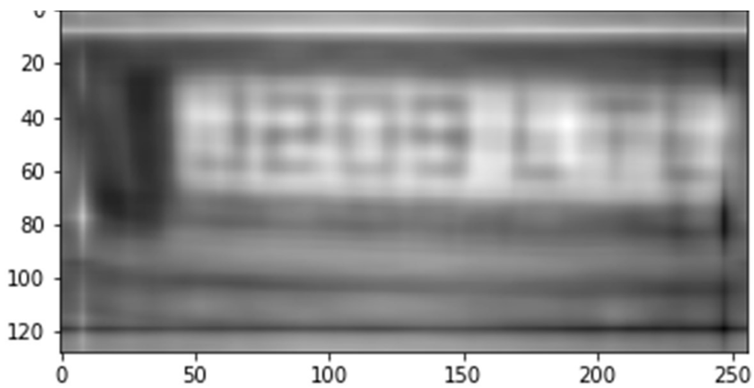

Figure 4 Result after 34 iterations