Learning computationally efficient dictionaries and their implementation as fast transforms

Luc Le Magoarou, Rémi Gribonval

Supplementary material

A. Anonymous

1 Projection operator proof

We want to find the projection operator onto the following set:

$$\mathcal{E} := \{\mathbf{A} \in \mathbb{R}^{n \times n} : \|\mathbf{A}\|_0 \leq p, \|\mathbf{A}\|_F = 1\}, \tag{1}$$

with $p \in \mathbb{N}^*$. We are interested in the projection of some matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ onto the set $\mathcal{E}$, namely we want to find $\mathbf{S}^*$ such that:

$$\mathbf{S}^* = P_{\mathcal{E}}(\mathbf{S}) \in \arg\min_{\mathbf{U}} \{\|\mathbf{U} - \mathbf{S}\|_F^2 : \mathbf{U} \in \mathcal{E}\} \tag{2}$$

Proposition 1.1. *Projection operator formula.*

$$P_{\mathcal{E}}(\mathbf{S}) = B_1(T_p(\mathbf{S}))$$

where

$$T_p(\mathbf{S}) := \arg\min_{\mathbf{V}} \{\|\mathbf{V} - \mathbf{S}\|_F^2 : \|\mathbf{V}\|_0 \leq p\}$$

and

$$B_1(\mathbf{R}) := \arg\min_{\mathbf{W}} \{\|\mathbf{W} - \mathbf{R}\|_F^2 : \|\mathbf{W}\|_F = 1\}$$

Proof. Let us denote by P the set of indices corresponding to the p greatest entries (in absolute value) of $\mathbf{S}$, and by $\mathbf{S}_P$ the matrix we get by setting all the other entries of $\mathbf{S}$ to 0. Let us also introduce $\mathbf{S}_{\bar{P}} = \mathbf{S} - \mathbf{S}_P$, the matrix we get by keeping only the $n^2 - p$ littlest entries. We have:

$$\|\mathbf{S}\|_F^2 = \|\mathbf{S}_P\|_F^2 + \|\mathbf{S}_{\bar{P}}\|_F^2 \tag{3}$$

and

$$\mathbf{S}_P = \arg\min_{\mathbf{V}} \{\|\mathbf{V} - \mathbf{S}\|_F^2 : \|\mathbf{V}\|_0 \leq p\} = T_p(\mathbf{S}) \tag{4}$$

and

$$P = \arg\max_{J} \{\|\mathbf{S}_J\|_F : \mathrm{card}(J) \leq p\} \tag{5}$$

The operator B_1 is simply the projection onto the Euclidean sphere of radius 1, that is:

$$B_1(\mathbf{X}) = \frac{1}{\|\mathbf{X}\|_F}\mathbf{X}. \tag{6}$$

Let us now denote by I the support of any $\mathbf{U} \in \mathcal{E}$, we have:

$$\|\mathbf{U} - \mathbf{S}\|_F^2 = \|\mathbf{S}_{\bar{I}}\|_F^2 + \|\mathbf{U} - \mathbf{S}_I\|_F^2 \tag{7}$$

so the problem can be rewritten:

$$(\mathbf{S}^*, I^*) = \underset{(\mathbf{U},I)}{\arg\min}\{\|\mathbf{S}_{\bar{I}}\|_F^2 + \|\mathbf{U} - \mathbf{S}_I\|_F^2 : \|\mathbf{U}\|_F = 1, \mathrm{card}(I) \leq p\} \tag{8}$$

Isolating $\mathbf{U}$ we have:

$$\mathbf{S}^* = \underset{\mathbf{U}}{\arg\min}\{\|\mathbf{U} - \mathbf{S}_{I^*}\|_F^2 : \|\mathbf{U}\|_F \leq q\} = B_1(\mathbf{S}_{I^*}) \tag{9}$$

with:

$$\begin{aligned}
I^* = \quad & \underset{I}{\arg\min}\{\|\mathbf{S}_{\bar{I}}\|_F^2 + \inf_{\mathbf{U}}\{\|\mathbf{U} - \mathbf{S}_I\|_F^2 : \|\mathbf{U}\|_F = 1\} : \mathrm{card}(I) \leq p\} \\
I^* = \quad & \underset{I}{\arg\min}\{\|\mathbf{S}_{\bar{I}}\|_F^2 + \|B_1(\mathbf{S}_I) - \mathbf{S}_I\|_F^2 : \mathrm{card}(I) \leq p\} \\
I^* = \quad & \underset{I}{\arg\min}\{\|\mathbf{S}_{\bar{I}}\|_F^2 + (1 - \tfrac{1}{\|\mathbf{S}_I\|_F})^2 \|\mathbf{S}_I\|_F^2 : \mathrm{card}(I) \leq p\} \\
I^* = \quad & \underset{I}{\arg\min}\{\underbrace{\|\mathbf{S}_{\bar{I}}\|_F^2 + \|\mathbf{S}_I\|_F^2}_{\|\mathbf{S}\|_F^2} - 2\|\mathbf{S}_I\|_F : \mathrm{card}(I) \leq p\} \\
I^* = \quad & \underset{I}{\arg\max}\{\|\mathbf{S}_I\|_F : \mathrm{card}(I) \leq p\} \\
I^* = \quad & P
\end{aligned} \tag{10}$$

So we have: $\mathbf{S}^* = B_1(\mathbf{S}_P) = B_1(T_p(\mathbf{S}))$. □

2 Lipschitz moduli proof

Let us look at the Lipschitz moduli of the gradient of the smooth part of the objective:

$$\begin{aligned}
& \left\|\nabla_{\mathbf{S^i_j}} H(\mathbf{S}_1^{i+1} \ldots \mathbf{S}_1 \ldots \mathbf{S}_p^i, \lambda^i) - \nabla_{\mathbf{S^i_j}} H(\mathbf{S}_1^{i+1} \ldots \mathbf{S}_2 \ldots \mathbf{S}_p^i, \lambda^i)\right\|_F \\
= \quad & \left\|\lambda^i \mathbf{L}^T(\lambda^i \mathbf{L}\mathbf{S}_1\mathbf{R} - \mathbf{X})\mathbf{R}^T - \lambda^i \mathbf{L}^T(\lambda^i \mathbf{L}\mathbf{S}_2\mathbf{R} - \mathbf{X})\mathbf{R}^T\right\|_F \\
= \quad & (\lambda^i)^2 \left\|\mathbf{L}^T\mathbf{L}(\mathbf{S}_1 - \mathbf{S}_2)\mathbf{R}\mathbf{R}^T\right\|_F \\
= \quad & (\lambda^i)^2 \left\|(\mathbf{R}\mathbf{R}^T) \otimes (\mathbf{L}^T\mathbf{L}).\mathrm{vec}(\mathbf{S}_1 - \mathbf{S}_2)\right\|_2 \\
\leq \quad & (\lambda^i)^2 \left\|(\mathbf{R}\mathbf{R}^T) \otimes (\mathbf{L}^T\mathbf{L})\right\|_2 \|\mathbf{S}_1 - \mathbf{S}_2\|_F \\
= \quad & (\lambda^i)^2 \|\mathbf{R}\|_2^2 . \|\mathbf{L}\|_2^2 \|\mathbf{S}_1 - \mathbf{S}_2\|_F .
\end{aligned} \tag{11}$$

So we can say that the following quantity is a Lipschitz modulus: $L_j(\mathbf{L}, \mathbf{R}, \lambda^i) = (\lambda^i)^2 \|\mathbf{R}\|_2^2 . \|\mathbf{L}\|_2^2$.