Learning computationally efficient dictionaries and their implementation as fast transforms
Luc Le Magoarou, Rémi Gribonval

To cite this version:
Luc Le Magoarou, Rémi Gribonval. Learning computationally efficient dictionaries and their implementation as fast transforms. 2015. hal-01010577v3

HAL Id: hal-01010577
https://inria.hal.science/hal-01010577v3
Preprint submitted on 26 Feb 2015
Supplementary material

A. Anonymous

1 Projection operator proof

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

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

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

\[ S^* = P_\mathcal{E}(S) \in \arg \min_U \{\| U - S \|_F^2 : U \in \mathcal{E}\} \quad (2) \]

**Proposition 1.1. Projection operator formula.**

\[ P_{\mathcal{E}}(S) = B_1(T_p(S)) \]

where

\[ T_p(S) := \arg \min_V \{\| V - S \|_F^2 : \| V \|_0 \leq p\} \]

and

\[ B_1(R) := \arg \min_W \{\| W - R \|_F^2 : \| W \|_F = 1\} \]

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

\[ \| S \|_F^2 = \| S_P \|_F^2 + \| S_{\bar{P}} \|_F^2 \quad (3) \]

and

\[ S_P = \arg \min_V \{\| V - S \|_F^2 : \| V \|_0 \leq p\} = T_p(S) \quad (4) \]

and

\[ P = \arg \max_J \{\| S_J \|_F : \text{card}(J) \leq p\} \quad (5) \]

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

\[ B_1(X) = \frac{1}{\| X \|_F} X. \quad (6) \]
Let us now denote by $I$ the support of any $U \in E$, we have:
\[
\|U - S\|_F^2 = \|S_I\|_F^2 + \|U - S_I\|_F^2
\]  
so the problem can be rewritten:
\[
(S^*, I^*) = \arg \min_{(U, I)} \{ \|S_I\|_F^2 + \|U - S_I\|_F^2 : \|U\|_F = 1, \text{card}(I) \leq p \}
\]  
Isolating $U$ we have:
\[
S^* = \arg \min_{U} \{ \|U - S_I^*\|_F^2 : \|U\|_F \leq q \} = B_1(S_{I^*})
\]  
with:
\[
I^* = \arg \min_{I} \{ \|S_I\|_F^2 + \inf_{U} \{ \|U - S_I\|_F^2 : \|U\|_F = 1 \} : \text{card}(I) \leq p \}
\]
\[
I^* = \arg \min_{I} \{ \|S_I\|_F^2 + \|B_1(S_I) - S_I\|_F^2 : \text{card}(I) \leq p \}
\]
\[
I^* = \arg \min_{I} \{ \|S_I\|_F^2 + (1 - \frac{1}{\|S_I\|_F})^2 \|S_I\|_F^2 : \text{card}(I) \leq p \}
\]
\[
I^* = \arg \min_{I} \{ \|S_I\|_F^2 + \frac{1}{\|S_I\|_F^2} \|S_I\|_F^2 : \text{card}(I) \leq p \}
\]
\[
I^* = \arg \max_{I} \{ \|S_I\|_F : \text{card}(I) \leq p \}
\]
\[
I^* = P
\]
So we have: $S^* = B_1(S_P) = B_1(T_P(S))$.

2 Lipschitz moduli proof

Let us look at the Lipschitz moduli of the gradient of the smooth part of the objective:
\[
\left\| \nabla_{S_I} H(S_1^{i+1} \ldots S_1 \ldots S_{i_r}, \lambda^i) - \nabla_{S_I} H(S_1^{i+1} \ldots S_2 \ldots S_{i_r}, \lambda^i) \right\|_F
\]
\[
= \left\| \lambda^iL^T(\lambda^iLS_1R - X)R^T - \lambda^iL^T(\lambda^iLS_2R - X)R^T \right\|_F
\]
\[
= (\lambda^i)^2 \left\| L^T(LS_1 - S_2)RR^T \right\|_F
\]
\[
= (\lambda^i)^2 \left\| (RR^T) \otimes (LT) \right\|_F
\]
\[
\leq (\lambda^i)^2 \left\| (RR^T) \otimes (LT) \right\|_2 \|S_1 - S_2\|_F
\]
\[
= (\lambda^i)^2 \left\| R \right\|_2^2 \cdot \left\| L \right\|_2 \cdot \left\| S_1 - S_2 \right\|_F
\]
So we can say that the following quantity is a Lipschitz modulus: $L_j(L, R, \lambda^i) = (\lambda^i)^2 \left\| R \right\|_2^2 \cdot \left\| L \right\|_2$. 