From Hardness to Efficiency in Structured Sparse Matrix Factorization

- **Quoc-Tung Le** (Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France)
- Elisa Riccietti (Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France)
- Rémi Gribonval (Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France)

Keywords: Optimization, Matrix Factorization, Sparsity, NP-hardness.

Abstract: Approximating a dense matrix with a product of sparse factors is a fundamental problem for many signal processing and machine learning tasks. Such a problem can be formulated as the following optimization problem [1]:

$$\begin{aligned} &\underset{\mathbf{X}^J,\ldots,\mathbf{X}^1}{\text{Minimize}} \|\mathbf{Z} - \mathbf{X}^J \ldots \mathbf{X}^1\|_F^2, \\ &\text{Subject to} : \mathbf{X}^\ell \in \Sigma_\ell, \forall 1 \leq \ell \leq J \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\Sigma_\ell, \ell = 1, \ldots, J$, are sets of sparse matrices that constraint the structure of the factors $\mathbf{X}^\ell$. Typical choices of Σ_ℓ can be the set of k-sparse matrices $\Sigma_k = \{X \mid \|X\|_0 \leq k\}$ ($\|\cdot\|_0$ is the number of nonzero coefficients of the matrix) and its variants (k-row/column sparse matrices).

In general, Problem (1) can be decomposed into two subproblems: finding the positions of the non-zero coefficients in the sparse factors $\mathbf{X}^i$, and determining their values. While the first step is usually seen as the most challenging one due to its combinatorial nature, this work focuses on the second step, referred to as sparse matrix approximation with fixed support (FSMF) and we analyse it in-depth the case two factors ($J = 2$).

First, we show the NP-hardness of FSMF in the case where there are only two factors. This is proved by reducing FSMF to the problem of low rank matrix recovery, which is shown to be NP-hard in [2]. While this result contrasts with the theory established for coefficient recovery with a fixed support in the classical sparse recovery problem (that can be trivially addressed by least squares), it is in line with the known hardness of related matrix/tensor factorization with additional constraints or different losses.

Second, we also present a non-trivial family of support constraints making the FSMF problem practically tractable with a gradient-free algorithm, of fixed bounded complexity, that decomposes the problem into "blocks" that can be exploited to build up an optimal solution using blockwise SVDs.

Finally, we investigate the landscape of FSMF optimization formulation, proving the absence of spurious local valleys [3] and spurious local minima [4], whose presence could prevent local optimization methods from achieving global optimality under our hypothesis of tractability. Other challenges for the general FSMF will also be discussed.

References:

[1] Le Magoarou, L. and Gribonval, R. Flexible Multi-layer Sparse Approximations of Matrices and Applications *IEEE Journal of Selected Topics in Signal Processing*, Volume 10, page 688 - 700, 2016.

[2] Gillis, N. and Glineur, F. Low-Rank Matrix Approximation with Weights or Missing Data Is NP-Hard SIAM Journal on Matrix Analysis and Applications, Volumn 32

[3] Venturi, L. and Bandeira, A. S. and Bruna, J. Spurious Valleys in One-hidden-layer Neural Network Optimization Landscapes *Journal of Machine Learning Research*, Volume 20, page 688 - 700, 2020.

[4] Kawaguchi, K. Deep Learning without Poor Local Minima Advances in Neural Information Processing Systems 29, page 586-594, 2016