Dictionary Learning through SCA

Leonardo Di Nino

May 2024

The dictionary learning problem can be formulated as:

$$\begin{cases} \min\limits_{D,X} \frac{1}{2}||Y - DX||_F^2 + \lambda ||X||_1 \\ \qquad ||d_i||_2 \leq 1, i = 1, ..., m \end{cases}$$

given a matrix of observed signals $Y \in \mathbb{R}^{s \times t}$ and fixed a dimension m for the number of atoms in the dictionary such that $D \in \mathbb{R}^{s \times m}$ and $X \in \mathbb{R}^{m \times t}$: we want to learn both the basis and the sparse representation of these observed signals.

Unfortunately, the problem is not convex, but it is block wise convex with respect to both the decision variables, so that we can define the following convex surrogates for a SCA approach:

$$\widetilde{F}_1(D|X^k) = \frac{1}{2}||Y - DX^k||_F^2 \tag{1}$$

$$\widetilde{F}_2(X|D^k) = \frac{1}{2}||Y - D^k X||_F^2 + \lambda ||X||_1 \tag{2}$$

The two related problems can be easily solved in closed form.

Minimizing equation (1) leads to the following projected gradient procedure:

$$D^{k+1} = \Pi_{\mathcal{D}}[D^k - \alpha \frac{\partial}{\partial D}(\frac{1}{2}||Y - DX^t||_F^2)] \tag{3}$$

where $\frac{\partial}{\partial D}(\frac{1}{2}||Y - DX^t||_F^2) = -YX^T - DXX^T$ and $\Pi_{\mathcal{D}}$ is the projection over $\mathcal{D} = \{D \in \mathbb{R}^{s \times m} : ||[D]_{:,i}||_2 \leq 1, i = 1, ..., m\}$ which has the following closed form:

$$\Pi_{\mathcal{D}}(D) = \frac{[D]_{:,i}}{\max(1, ||[D]_{:,i}||_2)} \tag{4}$$

Instead, minimization of equation (2) can be tackled using any sparse recovery algorithm on each column x_j of X.