Peyresq 2019 - Riemannian optimization - PA Absil

30 June – 6 July 2019
Project

1 Context

The cocktail party problem consists in recovering statistically independent source signals from linear combinations of these sources (see figure 1).

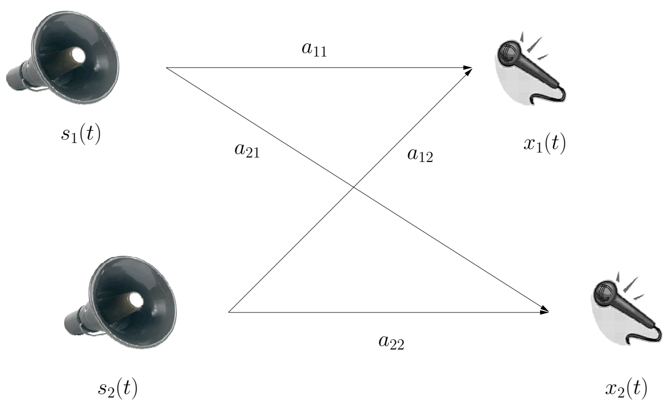

Figure 1: Cocktail party problem: case of two recorded signals generated by two source signals.

The problem can be modeled as

$$x(t, \omega) = As(t, \omega), \quad t \in \mathcal{T}, \omega \in \Omega$$

where $A \in \mathbb{R}^{n \times n}$ is a constant unknown matrix termed *mixing matrix*, $\mathcal{T} \subset \mathbb{R}$ is the time set, Ω is the sample space of a probability space, $x : \mathcal{T} \times \Omega \to \mathbb{R}^n$ is an observed stochastic process, and $s : \mathcal{T} \times \Omega \to \mathbb{R}^n$ is a hidden stochastic process whose components $s_1, \ldots, s_n$ are statistically independent.

However, in practical applications, we only have a sampled signal $(x(t_1), \ldots, x(t_N))$ with values in $\mathbb{R}^n$, and we do not know the probability distribution of x, but we assume that x comes from a realization $(s(t_1), \ldots, s(t_N))$ of the random process s according to a linear law

$$x(t_i) = As(t_i), \quad i = 1, \ldots, N.$$

By defining matrices

$$S := \begin{bmatrix} s_1(t_1) & s_1(t_2) & \cdots & s_1(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ s_n(t_1) & s_n(t_2) & \cdots & s_n(t_N) \end{bmatrix}$$

and

$$X := \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_N) \end{bmatrix},$$

we obtain the form

$$X = AS,$$

where X is known (observed) and A and S are unknown. The goal is to recover A and S as faithfully as possible, knowning that the rows of S are realizations of statistically independent random processes.

A popular approach is to look for an *unmixing matrix* $W \in \mathbb{R}^{n\times n}$ such that the rows of matrix

$$Z = W^T X \tag{1}$$

are "as independent as possible". To this end, we make use of a *contrast function* $\gamma_X : \mathbb{R}^{n\times n} \to \mathbb{R} : W \mapsto \gamma_X(W)$ which measures the "dependence" of the rows of Z. The problem then reduces to computing a matrix W that minimizes the cost function γ_X. This approach is termed Independent Component Analysis (ICA).

Most ICA algorithms minimize a contrast whose domain of definition is restricted to the orthogonal group O_n (instead of $\mathbb{R}^{n\times n}$). This restriction is made possible by the whitening technique, as follows. Let us suppose that the rows of X are linearly independent (a reasonable assumption). Let $W = U\Sigma V$ be a singular value decomposition of W. Equation (1) becomes

$$Z = V^T \Sigma U^T X.$$

Whitening consists in choosing matrices Σ and U such that the correlation matrix of $\Sigma U^T x$ is a multiple of the identity, i.e., $(\Sigma U^T X)(\Sigma U^T X)^T = \sigma I$. It then remains to choose $V \in O_n$ that minimizes $\gamma_{\tilde{X}}(V)$, where $\tilde{X} := \Sigma U^T X$. From now on, we assume that X has been whitenend (i.e., $XX^T = \sigma I$) and that it remains to find $V \in O_n$ that minimizes the interdependence between the rows of $Z = V^T X$.

We propose to use a contrast function of "SOBI" type [BAMCM97] which measures the "diagonality" of lagged covariance matrices of z. Concretely, for a lag d, we define the (symmetric) covariance matrix $R(d) \in \mathbb{R}^{n\times n}$ whose element (i, j) is given by

$$(R_X(d))(i,j) = \sum_{k=1}^{N-d} (X(i,k)X(j,k+d) + X(i,k+d)X(j,k)).$$

Since $Z = V^T X$, we get

$$R_Z(d) = V^T R_X(d) V. \tag{2}$$

We choose a collection of lags $d_1, \ldots, d_K$; from matrix X (known), we build the covariance matrices

$$C_k = R_X(d_k), \quad k = 1, \ldots, K,$$

and we consider the contrast

$$f : O_n \to \mathbb{R} : V \mapsto f(V) = \sum_{k=1}^{K} \|\mathrm{off}(V^T C_k V)\|_F^2 \tag{3}$$

where $\mathrm{off}(M)$ stands for M with the diagonal elements set to zero, and $\|\cdot\|_F$ denotes the Frobenius norm. This contrast can be thought of as a "measure of diagonality" of matrices $R_Z(d_k)$, $k = 1, \ldots, K$, and thus gives an account of the "level of dependence" of the rows of Z.

2 Questions

1. Show that it is possible to choose matrices Σ diagonal and U orthogonal such that $(\Sigma U^T X)(\Sigma U^T X)^T$ is a multiple of the identity.
2. Show that matrices $R_X(d)$ are symmetric. Are they always positive definite? Prove equation (2).
   Suggestion: Write the expression of $R_X(d)$ in matrix form.
3. Prove that the orthogonal group $O_n = \{V \in \mathbb{R}^{n\times n} : V^T V = I\}$ is a submanifold of $\mathbb{R}^{n\times n}$.
4. Give an expression for $T_V O_n$.
5. Give an expression for the Riemannian metric induced on O_n by the canonical metric $\langle Z_1, Z_2\rangle = \mathrm{trace}(Z_1^T Z_2)$ of $\mathbb{R}^{n\times n}$.
6. Give an expression for $\mathrm{grad}\, f(V)$, where f is the cost function on O_n defined in (3).
   Remarks: $\|M\|_F^2 = \mathrm{trace}(M^T M)$. For all matrices A and B of compatible dimensions, $\mathrm{trace}(\mathrm{off}(A)\mathrm{off}(B)) = \mathrm{trace}(\mathrm{off}(A)B) = \mathrm{trace}(A\mathrm{off}(B))$; $\mathrm{trace}(AB) = \mathrm{trace}(BA)$; $\mathrm{trace}(A^T) = \mathrm{trace}(A)$.
7. Write an expression for $\mathrm{Hess}\, f(V)[Z]$ where $Z \in T_V O_n$.
8. Write a steepest-descent method for f and use it in the provided Matlab template.
9. Likewise with a conjugate gradient method.

3 Complementary questions

The following, more difficult questions involve quotient manifolds.

Let $\mathbb{R}_*^{n\times p}$, $p \leq n$, denote the set of all full rank matrices of size $n \times p$. Consider the cost function

$$\overline{f} : \mathbb{R}_*^{n\times p} \to \mathbb{R} : Y \mapsto \sum_{k=1}^{K} (\log\det \mathrm{ddiag}(Y^T C_k Y) - \log\det(Y^T C_k Y)), \tag{4}$$

where $C_1, \ldots, C_K$ are given symmetric positive-definite matrices.

The Hadamard inequality guarantees that $\det(M) \leq \prod_i m_{ii}$ for all symmetric positive-definite matrices M, and that the equality holds if and only if M is diagonal. It follows that $\overline{f}(V)$ is nonnegative for all V and is zero if and only if all the matrices $V^T C_k V$ are diagonal.

Let $\mathcal{D}_*^p$ be the set of nonsingular diagonal matrices. Let $Y\mathcal{D}_*^p$ denote the set $\{YD : D \in \mathcal{D}_*^p\}$ and let $\mathbb{R}_*^{n\times p}/\mathcal{D}_*^p$ denote the quotient space $\{Y\mathcal{D}_*^p : Y \in \mathbb{R}_*^{n\times p}\}$.

1. Prove that the function $\overline{f}$ defined in (4) satisfies the invariance property
$$\overline{f}(YD) = \overline{f}(Y)$$
for all $D \in \mathcal{D}_*^p$.

2. Choose a Riemannian metric on $\mathbb{R}_*^{n\times p}$ that turns $\mathbb{R}_*^{n\times p}/\mathcal{D}_*^p$ into a Riemannian quotient manifold.

3. Let f be the projection of $\overline{f}$ onto the quotient $\mathbb{R}_*^{n\times p}/\mathcal{D}_*^p$, i.e., $\overline{f} = f \circ \pi$ where π is the natural projection associated with the quotient. Write the gradient and the Hesian of f by means of horizontal lifts.

References

[BAMCM97] Adel Belouchrani, Karim Abed-Meraim, Jean-François Cardoso, and Eric Moulines, *A blind source separation technique using second-order statistics*, IEEE Trans. Signal Process. **45** (1997), no. 2, 434–444.