

Introduction

It is rare for the quantity of interest and the quantity measured to perfectly match: often the signal $s \in \mathcal{R}^{m \times n}$ of interest is actually mixed with itself. We will set up a certain number of recording devices, generally n and from our recorded signals try to recover the sources. If we assume the sources are linearly mixed our problem is then:

$$x = As$$

where $x \in \mathcal{R}^{k \times n}$, $A \in \mathcal{R}^{k \times m}$. While not always true or needed we will assume that A is square and invertible and call $A^{-1} = W$.

We will see that the sole hypothesis of the independence of the s_i is enough to find a solution to the problem.

Because of the way the problem is posed one cannot hope to recover the real s as for example a scaling or permutation of A leaves the problem unchanged and we therefore cannot hope to recover the magnitudes (and therefore signs) of the s_i or their order.

ICA as an eigenvalue problem

We want to optimize a measure of independence of our estimated sources. Most approaches try to optimize an approximation of the mutual information. The measure chosen here [1] is for two variables:

$$\rho_{\mathcal{F}} = \max_{f_1, f_2 \in \mathcal{F}} \text{corr}(f_1(x_1), f_2(x_2))$$

By exploiting the kernel trick we can obtain

$$\rho_{\mathcal{F}} = \max_{f_1, f_2 \in \mathcal{F}} \text{corr}(\langle \Phi_1(x_1), f_1 \rangle, \langle \Phi_2(x_2), f_2 \rangle)$$

We recognize a CCA problem and if we note K_1 and K_2 the respective Gram matrices (assuming they are centred) we get the problem:

$$\begin{pmatrix} 0 & K_1 K_2 \\ K_2 K_1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \rho \begin{pmatrix} K_1^2 & 0 \\ 0 & K_2^2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

This problem is unfortunately not well posed and will always be equal to 0 for most kernels, we therefore adopt a regularized version as an estimator:

$$\rho_{\mathcal{F}} = \max_{f_1, f_2 \in \mathcal{F}} \frac{\text{cov}(f_1(x_1), f_2(x_2))}{(\text{var} f_1(x_1) + \kappa \|f_1\|_{\mathcal{F}}^2)^{1/2} (\text{var} f_2(x_2) + \kappa \|f_2\|_{\mathcal{F}}^2)^{1/2}}$$

The problem estimated at the first order is then:

$$\begin{pmatrix} 0 & K_1 K_2 \\ K_2 K_1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \rho \begin{pmatrix} (K_1 + \frac{N\kappa}{2}\mathbb{I})^2 & 0 \\ 0 & (K_2 + \frac{N\kappa}{2}\mathbb{I})^2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

Of course we are interested in solving the m -variables problems, we can easily extend CCA to m variables. We can then transform that problem in the problem of finding the eigenvalues of:

$$\tilde{\mathcal{K}}_{\kappa} = \begin{pmatrix} \mathbb{I} & r_{\kappa}(K_1)r_{\kappa}(K_2) & \cdots & r_{\kappa}(K_1)r_{\kappa}(K_m) \\ r_{\kappa}(K_2)r_{\kappa}(K_1) & \mathbb{I} & \cdots & r_{\kappa}(K_2)r_{\kappa}(K_m) \\ \vdots & \vdots & \ddots & \vdots \\ r_{\kappa}(K_m)r_{\kappa}(K_1) & r_{\kappa}(K_m)r_{\kappa}(K_2) & \cdots & \mathbb{I} \end{pmatrix}$$

$$r_{\kappa}(K_i) = K_i(K_i + \frac{N\kappa}{2}\mathbb{I})^{-1}$$

We then optimize:

$$J(W) = -\frac{1}{2} \det \tilde{\mathcal{K}}_{\kappa}$$

Reducing Complexity

If we decompose the K_i as

$$K_i = G_i G_i^T = U_i \Lambda_i U_i^T$$

with Λ_i diagonal then if R_i is Λ_i regularized by $\lambda \rightarrow \frac{\lambda}{\lambda + N\kappa/2}$ we have

$$\tilde{\mathcal{K}}_{\kappa} = (\mathcal{U}\mathcal{V}) \begin{pmatrix} \mathcal{R}_{\kappa} & 0 \\ 0 & \mathbb{I} \end{pmatrix} (\mathcal{U}\mathcal{V})^T$$

with

$$\mathcal{R}_{\kappa} = \begin{pmatrix} \mathbb{I} & R_1 U_1^T U_2 R_2 & \cdots & R_1 U_1^T U_m R_m \\ R_2 U_2^T U_1 R_1 & \mathbb{I} & \cdots & R_2 U_2^T U_m R_m \\ \vdots & \vdots & \ddots & \vdots \\ R_m U_m^T U_1 R_1 & R_m U_m^T U_2 R_2 & \cdots & \mathbb{I} \end{pmatrix}$$

And therefore

$$\det \tilde{\mathcal{K}}_{\kappa} = \det \mathcal{R}_{\kappa}$$

We will therefore make heavy use of incomplete Choleski decompositions to reduce the computational complexity from $\mathcal{O}(m^2 n^3)$ to $\mathcal{O}(m^2 M^2 n)$ with $M \ll n$.

Optimization on the Stiefel Manifold

Our problem is

$$\min_W J(W) \\ \text{s.t } WW^T = \mathbb{I}$$

The set $\{W \mid WW^T = \mathbb{I}\}$ has a particular geometry: it is Riemannian manifolds and most common optimization procedures can be performed on it [2]. The simplest implementation is steepest descent along geodesics in the direction of the gradient using the following: if $W, H \in \mathcal{R}^{m \times n}$ s.t $W^T W = \mathbb{I}$ and $A = W^T H$ skew-symmetric then the geodesic on the Stiefel manifold emanating from W in direction H is given by the curve

$$\begin{aligned} W(t) &= WM(t) + QN(t) \\ QR &= (\mathbb{I} - WW^T)H \\ \text{where} \\ \text{and} \end{aligned} \quad \begin{pmatrix} M(t) \\ N(t) \end{pmatrix} = \exp \left(t \begin{pmatrix} A & -R^T \\ R & 0 \end{pmatrix} \begin{pmatrix} \mathbb{I}_n \\ 0 \end{pmatrix} \right)$$

Given the cost of the gradient computations a natural extension is to perform conjugate gradient on the manifold, see [2] for the procedure. Clever use of the block structure of $\tilde{\mathcal{K}}_{\kappa}$ gives an efficient $\mathcal{O}(m^2)$ algorithm to compute first order differences.

Unmixing images



Figure 1: 3 images mixed linearly then unmixed

Independent images basis

This time the quantity of interest is W , the unmixing matrix. s representing the random coefficients in the ICA basis of natural images. In this setting the columns of A form the basis and the columns of W the detectors.

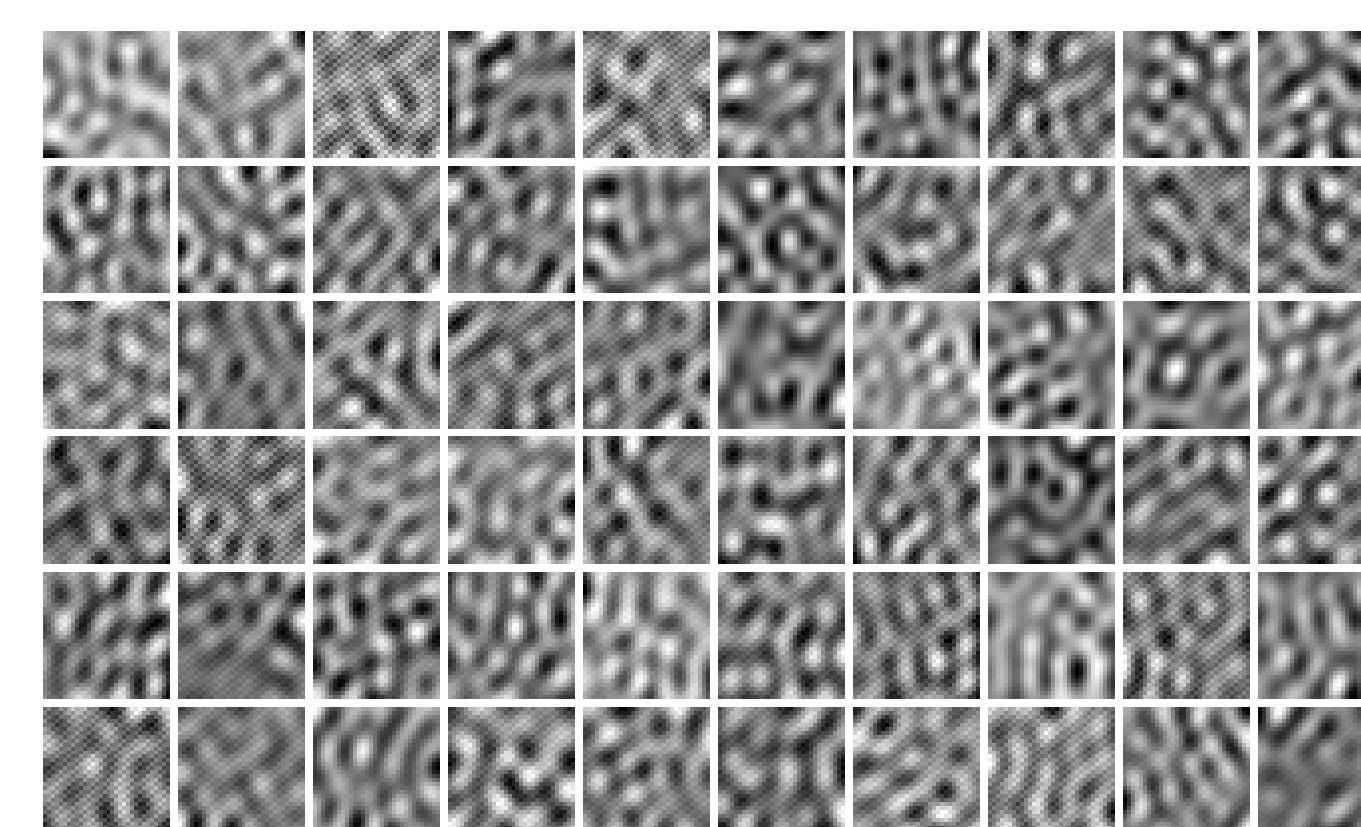


Figure 2: ICA basis of the 3 previous images

Conclusion

When seen as a semi-parametric problem ICA becomes a pure optimization problem and improvements can be achieved with better optimization techniques that exploits the geometry [3]. Using the geometry of the manifold and incomplete Cholesky decomposition we transform Kernel-ICA into a tractable problem. ICA has proven to be invaluable to many real world problems such as EEGs processing and also can bring insights to many unsupervised learning problems such as compressed sensing of images.

References

- [1] F. R. Bach and M. I. Jordan.
Kernel Independent Component Analysis.
Journal of Machine Learning Research, 3:1–48, 2002.
- [2] A. Edelman, T. A. Arias, and S. T. Smith.
The Geometry of Algorithms with Orthogonality Constraints.
SIAM Journal on Matrix Analysis and Applications, 20(2):303–353, 1998.
- [3] H. Shen, S. Jegelka, and A. Gretton.
Fast Kernel-Based Independent Component Analysis.
IEEE Transactions on Signal Processing, 57, 2009.