Off-the-Grid Wasserstein Group Lasso

P. Catala [1] , V. Duval [2], G. Peyré [1]

Abstract. In this contribution, we propose a new off-the-grid (i.e. without spatial discretization) solver for a sparse regularization method for multi-canal inverse problems, which integrates a Wasserstein distance between the recovered measures. This solver uses a semidefinite programming (SDP) relaxation based on Lasserre's hierarchy.

The goal is to estimate two (hopefully sparse, *i.e.* sums of Diracs) positive Radon measures (μ_0, ν_0) on the torus $\mathbb{T}^d$ from low-resolution noisy measurements of the form $u = \mathcal{F}\mu_0 + w$ and $v = \mathcal{F}\nu_0 + \varepsilon$, where $\mathcal{F}$ is a linear operator and w, ε account for some unknown noises. We assume that $\mathcal{F}$ is a convolution with a low-pass filter, so that without loss of generality, we may consider a Fourier transform $(\mathcal{F}\mu)_k = \int_{\mathbb{T}^d} e^{-2\imath\pi\langle k,x\rangle}\mathrm{d}\mu(x)$, for $k \in \Omega_c \overset{\text{def.}}{=} [\![-f_c, f_c]\!]^d$. In practice, the sought after sources (μ, ν) are linked to physical or biological phenomena, and one wishes to constrain the relative positions of the spikes in both measures. The classical approach (often called the group-Lasso problem) is to use a vectorial total variation norm, which imposes that μ and ν share the same support. This constraint is often too strong, and following [1] we propose to relax this assumption by rather penalizing their respective Wasserstein "distance". This distance is defined, for some cost $C(x, y)$, by $W_C(\mu, \nu) = \min_{\gamma_1=\mu,\gamma_2=\nu} C(x, y)\mathrm{d}\gamma(x, y)$ where (γ_1, γ_2) are the two marginals of the transport plan γ, which is a positive measure over the product space $\mathbb{T}^d \times \mathbb{T}^d$. The sources are thus estimated by solving the following infinite dimensional optimization problem

$$\min_{\mu,\nu\in\mathcal{M}_+(\mathbb{T}^d)} \frac{1}{2}\|u - \mathcal{F}\mu\|^2 + \frac{1}{2}\|v - \mathcal{F}\nu\|^2 + \lambda|\mu|(\mathbb{T}^d) + \lambda|\nu|(\mathbb{T}^d) + \tau W_C(\mu, \nu). \tag{1}$$

Instead of considering a fixed spatial discretization, we rather search for the Fourier moments of the transport plan γ. For $\ell \geq f_c$, we consider $z_{(s,t)} \overset{\text{def.}}{=} \int_{\mathbb{T}^d\times\mathbb{T}^d} e^{-2\imath\pi\langle s,x\rangle} e^{-2\imath\pi\langle t,y\rangle}\mathrm{d}\gamma(x, y)$ for $s, t \in [\![-\ell, \ell]\!]^d$, and $R_\ell(z) = (z_{(s-s',t-t')})_{s,s',t,t'\in[\![0,\ell]\!]^d}$. We denote by $\hat{C}$ the Fourier coefficients of C. Let also $z_1 = z_{(\cdot,0)}$, $z_2 = z_{(0,\cdot)}$ and $z_0 = z_{(0,0)}$. Then the sequence of SDP problems

$$\min_{z\in\mathbb{C}^{|\Omega_c|}} \frac{1}{2}\|u - z_1\|^2 + \frac{1}{2}\|v - z_2\|^2 + 2\lambda z_0 + \tau\langle\hat{C}, z\rangle \quad \text{s.t} \quad R_\ell(z) \succeq 0 \tag{$\mathcal{P}_\ell$}$$

can be shown to define a "Lasserre" hierarchy of increasingly tighter relaxations of (1), and in many cases, one can recover the support of a solution of (1) from a solution of $(\mathcal{P}_\ell)$. Note that when $\lambda = 0$ and as $\tau \to 0$, solving (1) defines a low-frequency approximation of the celebrated optimal transport problem. In this talk, I will outline the theoretical aspects of the relaxation $(\mathcal{P}_\ell)$, explain efficient tailored numerical solvers and showcase numerical illustrations.

References

[1] Wasserstein regularization for sparse multi-task regression. Janati et al. AISTATS 2019.

[1] DMA, Ecole Normale Supérieure, CNRS, UMR 8553, PSL

[2] INRIA Paris – CEREMADE, Université Paris-Dauphine, CNRS, UMR 7534, PSL