Solving Fredholm integral equations of the first kind via Wasserstein gradient flow

Arnaud -> Adam, Francesca

February 21, 2020

1 Fredholm integral equation of first kind

We want to solve the integral equation

$$\mu(y) = \int \rho\,(x)\, K(x, y)dx,$$

where ρ and μ are probability densities on $\mathbb{R}^n$ and $\mathbb{R}^m$ and K a Markov transition density, i.e. $\mu = \rho K$ in operator notation. The solution to this problem is not unique and we propose to regularize the problem using an entropy constraint; i.e. for a given $\lambda > 0$ we propose to minimize w.r.t. ρ

$$E(\rho) = \mathrm{KL}(\mu, \rho K) - \lambda \mathrm{Ent}(\rho)$$

where $\mathrm{KL}(\mu, \rho K)$ is the Kullback-Leibler divergence between μ and ρK and $\mathrm{Ent}(\rho) = -\int \rho \log \rho$ is the entropy of ρ. This requires solving a minimization problem in the space of probability measures. We are going to follow a Wasserstein gradient flow approach.

2 Gradient flow approach

We first need to compute

$$\lim_{\epsilon \to 0} \epsilon^{-1} \left(E(\rho + \epsilon\chi) - E(\rho)\right) = \int \frac{\delta E}{\delta \rho}\,(x)\, \chi\,(dx)$$

where χ is any signed measure such that $\rho + \epsilon\chi$ is a probability measure. We have

$$\begin{aligned} E(\rho) =& \mathrm{KL}(\mu, \rho K) - \lambda \mathrm{Ent}(\rho) \\ =& -\int \mu \log \rho K + \lambda \int \rho \log \rho + \int \mu \log \mu. \end{aligned}$$

It follows directly that

$$\frac{\delta \mathrm{Ent}\,(\rho)}{\delta \rho} = 1 + \log \rho.$$

and

$$\begin{aligned}\int \mu \log \left(\left(\rho + \epsilon \chi \right) K \right) - \int \mu \log \left(\rho K \right) &= \int \mu \left\{ \log \left(\rho K \right) + \log \left(1 + \frac{\epsilon \chi K}{\rho K} \right) \right\} - \int \mu \log \left(\rho K \right) \\ &= \int \mu \log \left(1 + \frac{\epsilon \chi K}{\rho K} \right) \\ &= \int \mu \left(\frac{\epsilon \chi K}{\rho K} + o \left(\frac{\epsilon \chi K}{\rho K} \right) \right) \\ &= \epsilon \int \mu \frac{\chi K}{\rho K} + o \left(\epsilon \int \mu \frac{\chi K}{\rho K} \right).\end{aligned}$$

We have

$$\int \mu \frac{\chi K}{\rho K} = \int \int \mu \left(dy \right) \frac{K(x,y)}{\rho K(y)} d\chi \left(x \right)$$

so

$$\frac{\delta E}{\delta \rho} \left(x \right) = \int \mu \left(dy \right) \frac{K(x,y)}{\rho K(y)}.$$

Hence, it follows that

$$\frac{\delta E}{\delta \rho} \left(x \right) = \int \mu \left(dy \right) \frac{K(x,y)}{\rho K(y)} + \lambda \left(1 + \log \rho \left(x \right) \right).$$

We can now compute the gradient of this functional derivative equation w.r.t. x

$$\nabla \frac{\delta E}{\delta \rho} \left(x \right) = \int \mu \left(dy \right) \frac{\nabla K(x,y)}{\rho K(y)} + \lambda \nabla \log \rho \left(x \right).$$

We now consider the following PDE

$$\partial_t \rho_t = \nabla \cdot \left(\rho_t \nabla \frac{\delta E}{\delta \rho_t} \right),$$

where $\nabla \cdot f = \sum_i \partial_i f_i$ is the divergence operator. The corresponding nonlinear ODE

$$dX_t = -\nabla \frac{\delta E}{\delta \rho_t} \left(X_t \right) dt, \quad X_0 \sim \rho_0 \tag{1}$$

is such that $Law(X_t) = \rho_t$. Then, by construction, one has

$$\frac{dE \left(\rho_t \right)}{dt} = - \int \left\| \nabla \frac{\delta E}{\delta \rho_t} \left(x \right) \right\|^2 \rho_t \left(dx \right).$$

The terminology nonlinear ODE is here used to indicate that the drift depends not only on X_t but on its distribution too.

Practically, what we would like to do is to simulate N particles $(X_t^1, ..., X_t^N)$ such that, at initialization, we sample iid particles $X_0^i \sim \rho_0$ and then implement numerically the N nonlinear ODEs

$$dX_t^i = - \int \mu \left(dy \right) \frac{\nabla K(X_t^i, y)}{\rho_t^N K(y)} dt - \lambda \nabla \log \left(\rho_t^N * H_\epsilon \left(X_t^i \right) \right) dt, \quad \rho_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}.$$

Approximating the first term on the r.h.s. is fine as practically $\mu \left(dy \right)$ is a discrete measure but approximating $\nabla \log \rho_t \left(x \right)$ from the empirical measure ρ_t^N is difficult and would require say convolution by some kernel H_ϵ. This is ugly and would be most likely highly inefficient. In the next section, we show how to address this issue.

3 Nonlinear SDE approach and numerical implementation

Let us rewrite the PDE

$$\begin{aligned}\partial_t \rho_t =& \nabla \cdot \left(\rho_t \nabla \frac{\delta E}{\delta \rho_t} \right) \\ =& \nabla \cdot \left(\rho_t \int \mu\left(dy\right) \frac{\nabla K(x,y)}{\rho_t K(y)} \right) + \lambda \nabla \cdot \left(\rho_t \nabla \log \rho_t \right).\end{aligned}$$

However, we have

$$\nabla \cdot \left(\rho_t \nabla \log \rho_t \right) = \nabla \cdot \nabla \rho_t = \triangle \rho_t,$$

where $\triangle f = \sum_i \partial_i^2 f_i$ is the Laplacian. So we can consider the following non-linear SDE (McKean-Vlasov)

$$dX_t = -\int \mu\left(dy\right) \frac{\nabla K(X_t,y)}{\rho_t K(y)} dt + \sqrt{2\lambda} dW_t, \quad X_0 \sim \rho_0, \tag{2}$$

where W_t is a standard $n-$dimensional Brownian motion. The SDE (2) has the same marginal distributions as the nonlinear ODE (1). So to solve the minimization problem of interest, we will simulate in practice N particles $(X_t^1, ..., X_t^N)$ such that, at initialization, we sample iid particles $X_0^i \sim \rho_0$ and then they evolve according to the non-linear (McKean-Vlasov) SDE

$$dX_t^i = -\int \mu\left(dy\right) \frac{\nabla K(X_t^i,y)}{\rho_t^N K(y)} dt + \sqrt{2\lambda} dW_t^i, \quad \rho_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}.$$

We will also need to further discretize in time these SDEs obviously.