Optimal Transport Map

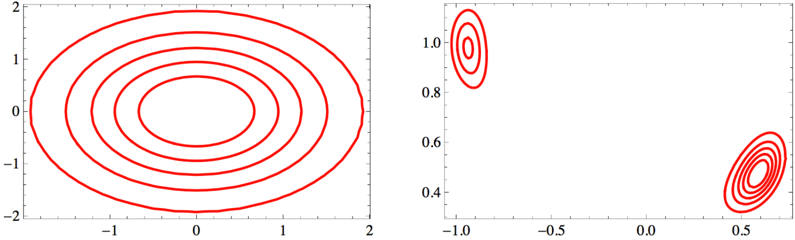
Institute for Computer and Sciences – Aerospace Software Engineering

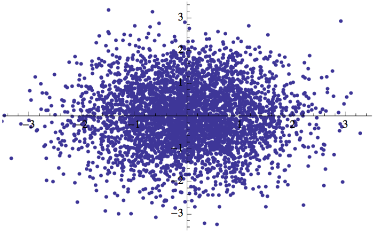
Outline

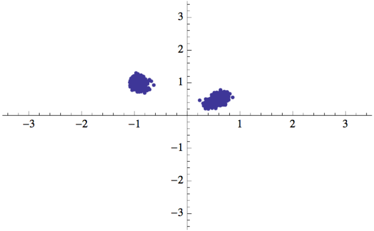
Introduction

Optimal Transport Map

Relation to the Particle Filter
- Relation to the particle flow filter
- Coupling Particle Filter

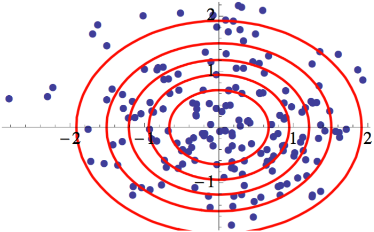

2
1
0
−1
−2
−1
0
1
2
1.0
0.8
0.6
0.4
−1.0
−0.5
0.0
0.5



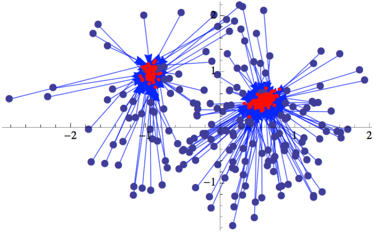


Ensembles of particles represent prior pdf and posterior pdf.
Compute expectations by summing over samples

How to update?



prior



posterior

how to update?

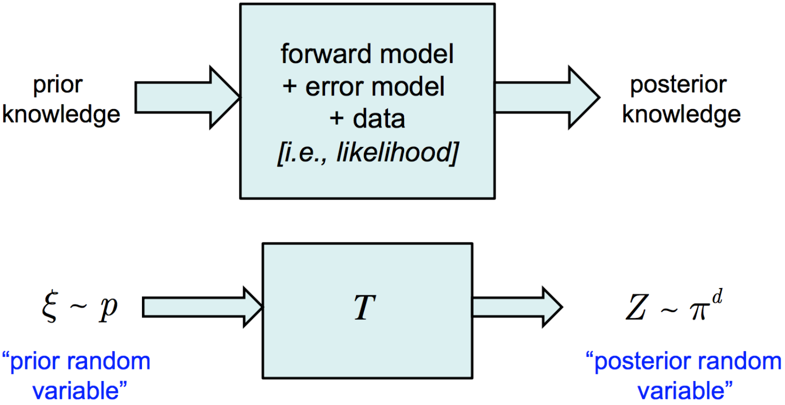
“These methods use . . . ‘samples’ that are drawn independently from the given initial distribution and assigned equal weights. . . . When observations become available, Bayes’ rule is applied either to individual samples . . . ” (Park and Xu, 2009)

Note: Bayes rule explains how to update probabilities, but not how to update samples.
(Particle filters do update the probability of each sample using Bayes rule.)

how to update?

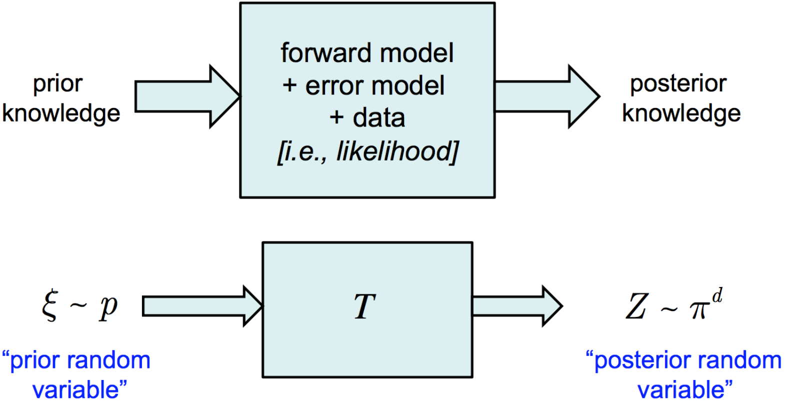
- If prior is Gaussian and posterior is Gaussian, then any linear transformation of variables that obtains the correct posterior mean and covariance is OK.
- when the observation operator is nonlinear?
    - Randomized maximum likelihood (Kitanidis, 1995; Oliver et al., 1996)
    - **Optimal map (El Moselhy and Marzouk, 2012)**
    - Implicit filters (Chorin et al., 2010; Morzfeld et al., 2012)
    - Ensemble-based iterative filters/smoothers

A different viewpoint



$$Z = T(\xi)$$

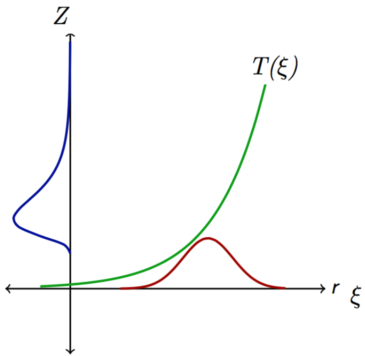
Optimal transport



$$Z = T(\xi)$$

Optimal transport

Deterministic coupling of two random variables, $\xi \sim \mu, Z \sim v$



Monge problem: $\min_T \int c(\xi, T(\xi)) d\mu(\xi)$, where $T^\sharp \mu = v$

Optimal transport

Let π be the target density and p be the density of a reference random variable ξ. Given the ability to

- sample ξ
- evaluate π up to a normalizing constant

one can compute numerical approximations to the optimal map. The reference distribution could be the prior, or it could be something else easy to sample from

Constructing a map

Let $\mathcal{T}$ be an appropriate set of diffeomorhpisms. Then, any global minimizer of the optimization problem:

$$\begin{aligned} min \quad & D_{KL}(T_\sharp \eta || \pi) \\ s.t. \quad & det \nabla T > 0 \\ & T \in \mathcal{T} \end{aligned} \tag{1}$$

In fact, any global minimizer of (1) achieves the minimum cost $D_{KL}(T_\sharp \eta || \pi) = 0$ and implies that $T_\sharp \eta = \pi$. The constraint det $\nabla T > 0$ ensures that the pushforward density $T_\sharp \eta$ is strictly postive on the support of the target.

Various types of transport

- **Optimal transport:** Monge problem:
  $\min_T \int c(\xi, T(\xi)) d\mu(\xi)$, where $T^\sharp \mu = v$(1781)
  many nice properties, but numericallt challengin in general cases
- **Knothe-Rosenblatt rearrangement**

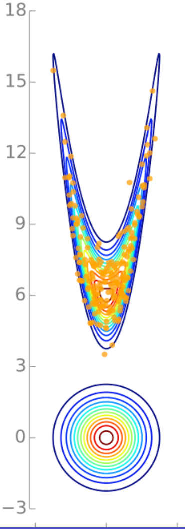
$$T(x) = \begin{bmatrix} T^1(x_1) \\ T^2(x_1, x_2) \\ \vdots \\ T^n(x_1, x_2, \ldots, x_n) \end{bmatrix}$$

  - Exists and is unique (up to ordering) under mild conditions
  - Jacobian determinant easy to evaluate

Other possible transports:

Other possible transports:

- Stein variational gradient descent [Liu & Wang 2016]
- Normalizing flows [Rezende & Mohamed 2015]
- Particle flows [Heng *et al.* 2015; Doucet, Daum. . .]
- Approximations of the optimal transport [Tabak 2013–16]



Potential advantages

- Move samples; don't just reweigh them
- Use optimization to enhance integration
- Independent, unweighted, and cheap samples from the target (or close to it
- Clear convergence criterion, even with unnormalized target density
- high-dimensional sampling

Relation with PFF

"We emphasize that our flow is not KnotheRosenblatt transport, but rather it is a completely different algorithm for particle flow."

- Particle flow inspired by Knothe-Rosenblatt transport for nonlinear filters
- Particle flow and Monge-Kantorovich transport

Relation with PFF

The particle flow filter is similar with the Monge-Kantorovich optimal transport(MKT) algorithms. But they differ in computational complexity and the overall intendend purpose.(Fred Daum & Jim Huang)

methods	PFF	MTK
purpose	fix the particle degenercy	transform particles
computational	fast	slow
dimension	high	low
approach	log-homotopy	homotopy

Relation with PFF

In contrast to MKT, for nonlinear filters, we are not interested in transporting physical objects or minimizing anything related to work or effort or energy or anything physical, but rather we merely want to move our particles from the prior density to the posteriori density with a small computational complexity. This is exactly what Knothe-Rosenblatt transport (KRT) tries to do!

Unfortunately, despite the reduced computational complexity of KRT, it is still too much for high dimensional nonlinear filter problems.

Coupling Particle filter

Although the optimal mass transport problem has many desirable properties it also has drawbacks. Monge's formulation is a nonconvex optimization problem and the Kantorovich formulation results in large-scale optimization problems. A recent development to address this computational problem builds on adding an entropic barrier term and solving the resulting optimization problem using the so called Sinkhorn iterations.

- Sen, D., Thiery, A. and Jasra, A. (2016) On coupling particle filter trajectories. arXiv preprint arXiv:1606.01016.
- Jacob P E, Lindsten F, Schön T B. Coupling of Particle Filters[J]. arXiv preprint arXiv:1606.01156, 2016.

Coupling Particle filter

Within particle filters, random variables are used to initialize, to resample and to propagate the particles.

Particle filters are randomized algorithms which can be written as a deterministic function of some random variables and a parameter value. However, they are discontinuous functions of their inputs, due to the resampling steps.

Our proposed strategy relies on common random numbers for the initialization and propagation steps, while the resampling step is performed jointly for a pair of particle systems, using ideas inspired by maximal couplings and optimal transport ideas.

Coupling Particle filter

A coupling of particle filters, given two parameter values θ and $\tilde{\theta}$, refers to a pair of particle systems, denoted by $(w_t^k, x_t^k)_{k=1}^N$ and $(\tilde{w}_t^k, \tilde{x}_t^k)_{k=1}^N$:

- marginally,each system has the same distribution as if it was generated by a particle filter
- the two systems are in some sense correlated

In the case of particle filters, the goal is to introduce postive correlations between likelihood estimators $\hat{p}(y_{1:t}|\theta)$ and $\hat{p}(y_{1:t}|\tilde{\theta})$, which improves the performance of score estimators and of MH schemes.

Coupling Particle filter

We consider the problem of jointly resampling (w, x) and $(\tilde{w}, \tilde{x})$. A joint distribution on $\{1, ..., N\}^2$ is characterized by a matrix P with non-negative entries P^{ij}, for $i, j \in \{1, ..., N\}$, that sum to one. The value P^{ij} represents the probability of sampling the pair (i, j).

The choice $P = w\tilde{w}^T$ corresponds to an independent coupling of w and $\tilde{w}$. Sampling from this matrix P is done by sampling a with probabilities w and $\tilde{a}$ with probabilities $\tilde{w}$.

Transport resampling

we want to choose $P \in \mathcal{J}(w, \tilde{w})$, the resampled particles are as similar as possible between the two systems.

The expected distance between the resampled particles x^a and $\tilde{x}^{\tilde{a}}$, conditional upon (w, x) and $(\tilde{w}, \tilde{x})$, is given by $\sum_{i=1}^{N} \sum_{j=1^N} P^{ij} d(x^i, \tilde{x}^j)$.
Denote by D the distance matrix with entries $D^{ij} = d(x^i, \tilde{x}^j)$. The optimal transport problem considers a matrix P^* that minimizes the expected distance over all $P \in \mathcal{J}(w, \tilde{w})$.