Bayesian Inference with Optimal Maps
Tarek El Moselhy and Youssef Marzouk, Massachusetts Institute of Technology

Motivation
- Conditioning models on data, via statistical inference, is central to many engineering and science applications
- Bayesian approach: foundation for inverse problems, data assimilation, assessing uncertainty in computational predictions, decision-making under uncertainty
- Goal: develop computationally efficient approach to Bayesian inference in large-scale complex systems
- Would like algorithms that are theoretically sound, computationally efficient, and endowed with clear measures of accuracy and convergence
- Overcome bottlenecks associated with traditional Bayesian computation (e.g., MCMC)

Bayesian framework
- Model parameters represented by random variable \( x \); data \( d \)
- Bayes’ rule
- Likelihood function
- Prior density
- Posterior density
- Evidence or marginal likelihood

Challenges
- Extracting information from the posterior: moments, marginals, realizations, set probabilities
- Evidence/normalization is difficult to compute
- Posterior evaluations may be expensive: forward model (PDE) appears inside the likelihood
- Parameter \( x \) may be high-dimensional

State of the art
- Markov chain Monte Carlo (MCMC) simulation of the posterior: flexible and broadly applicable, but several drawbacks
- Variational Bayes: suitable for a special set of inference problems

Main idea: inference with maps
1. From prior to posterior knowledge
   - Prior knowledge
   - Data, forward model, error model [i.e., likelihood]
   - Posterior knowledge

2. As a map or random variable transformation
   - "Prior random variable"
   - \( Z = f(x) \)
   - "Posterior random variable"

3. Graphical representation
   - Prior density \( p(x) \)
   - Transform
   - Posterior density \( p(x|z) \)
   - Map pushes forward the prior measure to the posterior measure

Formulation
\[
\pi(z) = \frac{L(z|d)p(z)}{\beta} \quad Z = f(X) \quad q(x|f) = \frac{L(f(x)|d)p(f(x))}{\beta} \quad \log \frac{\partial f}{\partial x}
\]
a probability density for the prior random variable \( X \), parameterized by \( f \)

Optimization problems
- Map exists (under weak conditions) but is not unique
  - Example: linear Gaussian case, identity prior covariance, posterior \( X \sim N(\mu, \Sigma) \)
  - Any map of the form \( z = f(x) = \mu + Lx \), such that \( L^T \Sigma L = \Sigma \), is valid
- Two optimization formulations guarantee existence and uniqueness of a monotone map
  - Penalty from optimal transport theory [Caffarelli, McCann]
    \[
    \min \text{Var}(f(X); \mu) + \lambda \text{Var}(f(X) - X)^2
    \]
  - Knobloch-Rosenblatt transport, i.e., "triangular" construction
    \[
    \min \text{Var}(f(X); \mu) \quad \text{s.t.} \quad z = f(x, x_2, \ldots, x_n)
    \]
- These are stochastic optimization problems; use sample-average approximation with prior samples of \( X \)
- Represent \( f \) using an orthogonal polynomial expansion (e.g., Hermite chaos)
  \[
  f(x) = G^T \Psi(x)
  \]
Matrix of unknown coefficients

Numerical demonstrations

Example 1: Linear-Gaussian model
- \( d = Ax + e \)
  - \( A \in \mathbb{R}^{m \times n} \), \( e \in \mathbb{R}^n \)
  - \( A \) is randomly generated
  - Convergence in \( < 12 \) iterations

Example 2: Nonlinear reaction kinetics
- Five late-time observations of species \( A \)
- Infer parameters \( k_1 \) and \( k_2 \); truth is \( k_1 = 1 \), \( k_2 = 2 \)
- 5th order polynomial map

Example 3: Cascaded maps
- ODE model of a genetic toggle switch from [Gardner 2000]; infer six parameters
- Real experimental data: steady-state expression levels of one gene

Example 4: PDE-constrained inverse problem (139 dimensions)
- Elliptic PDE: infer heterogeneous diffusivity \( e \) from limited/noisy observations of pressure \( p \)
- 2-D spatial domain; \( e \) parameterized by Karhunen-Loeve expansion with 139 independent modes
- Optimization algorithm identifies polynomial map up to 5th order
- Substantial gains in both accuracy and speed over MCMC

Conclusions
- An optimal transport interpretation of Bayesian statistical inference
- Posterior expressed as pushforward measure of prior
- Map computed by solving an optimization problem
- Advantages: clear convergence criterion; arbitrary numbers of independent posterior samples; embarrassingly parallel; exploit gradient/Hessian information; marginal likelihood computed for free (model selection); posterior moments computed analytically; enables sequential uncertainty propagation and inference

Support from DOE Office of Science/ASCER Early Career Research Program

DOE Applied Mathematics Program Meeting, October 2011
ymarz@mit.edu