Adaptive Bayesian Quadrature for Approximate Inference

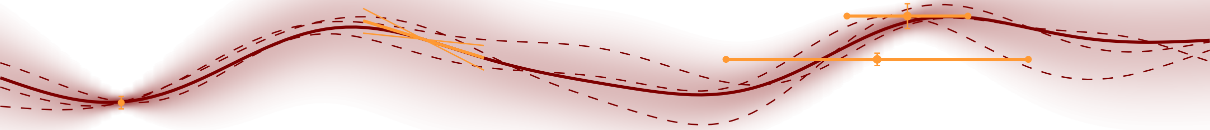
Alexandra Gessner
April 12, 2018

MAX-PLANCK-GESELLSCHAFT

Research Group for Probabilistic Numerics
Max Planck Institute for Intelligent Systems
Tübingen, Germany

erc

imprs-is



Bayesian quadrature for inference

Goal

Use Bayesian quadrature to integrate **positive** function $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ against a prior π

$$Z = \int f(\mathbf{x}) \, \mathrm{d}\pi(\mathbf{x})$$

in a potentially **high-dimensional space**

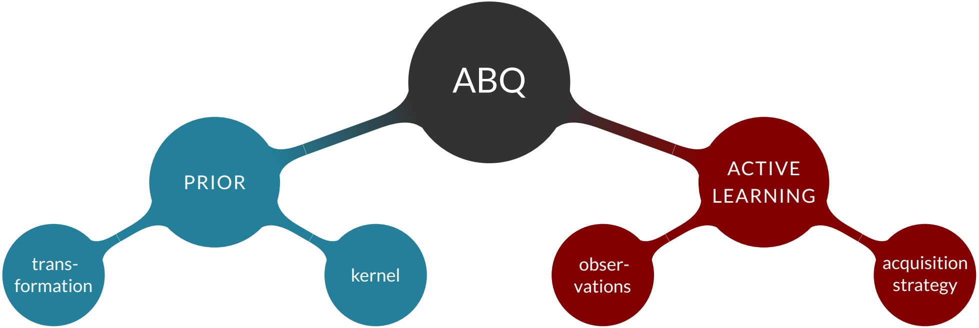
Optimal nodes in standard BQ

$$X = \underset{X \in \mathbb{R}^d}{\arg\min} \, \mathbb{V}_{\mathcal{D}}[Z] = \underset{X \in \mathbb{R}^d}{\arg\max} \iint k(\mathbf{x}, X) K_{XX}^{-1} k(X, \mathbf{x}') \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}'$$

Optimal acquisition rule for BQ is independent of observed values f
but for constrained functions, an active evaluation strategy might be desirable.

Adaptive BQ for inference

Ingredients



$\mathcal{T} : f \to g$

$g \sim \mathcal{GP}(0, k)$

approximate f by GP

"compatible"

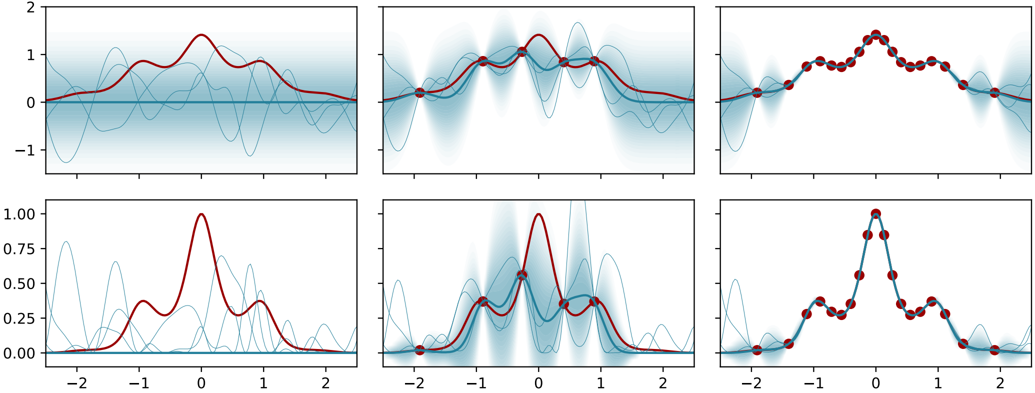
e.g. function evaluations $\{f(\mathbf{x}_i)\}_{i=1}^N$ or linear observations

"information theoretically sensible"
e.g. uncertainty sampling

EXAMPLE: Warped Sequential Active Bayesian Integration

WSABI [Gunter et al. 2014]

transformation	kernel	observations	acquisition
$f = \frac{1}{2} g^2$	SE kernel	function evaluations $\{f(\mathbf{x}_i)\}_{i=1}^N$	uncertainty sampling



ABQ prospects

PRIOR		ACTIVE LEARNING	
$f = \frac{1}{2} g^2$	+ many kernels that allow closed form integration – cannot handle high dynamic range	$\{f(\mathbf{x}_i, (\mathbf{x}_i))\}_{i=1}^{N}$	– bad coverage in high-dim. space
$f = \exp(g)$	+ high dynamic range – no closed form integration	Linear obs. $\mathcal{L}[f(\mathbf{x})]$	+ better coverage? – no longer linear due to transform