Conditional Bayesian Quadrature

Zonghao Chen[1,*] Masha Naslidnyk[1,*] Arthur Gretton[1] François-Xavier Briol[1]

[1] University College London @Hudson19990518

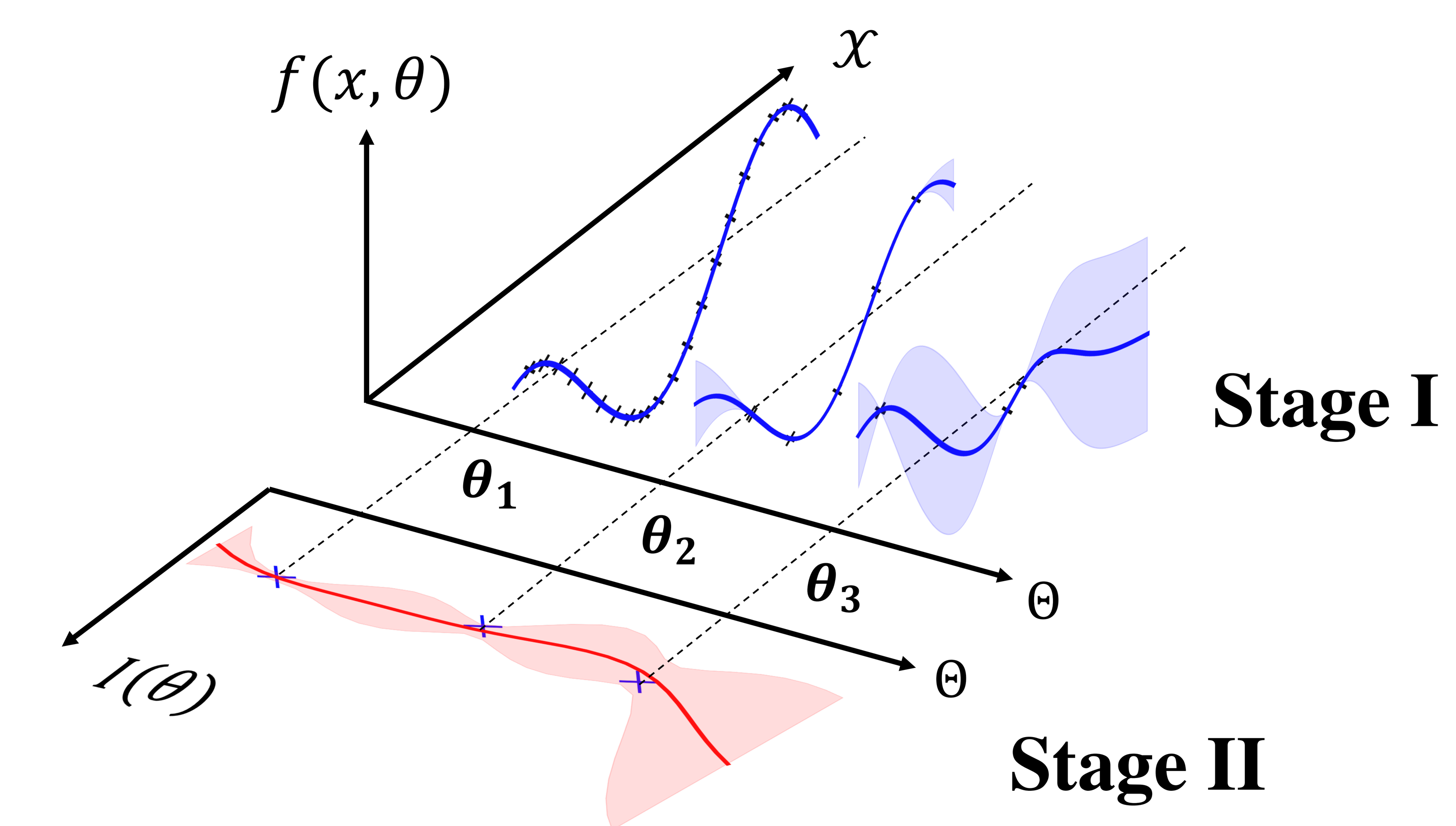
TL;DR

We propose conditional Bayesian quadrature (CBQ): a numerical algorithm for conditional/parametric expectations.

$$I(\theta) = \mathbb{E}_{X\sim\mathbb{P}_\theta}[f(X,\theta)] = \int_{\mathcal{X}} f(x,\theta)\mathbb{P}_\theta(\mathrm{d}x)$$

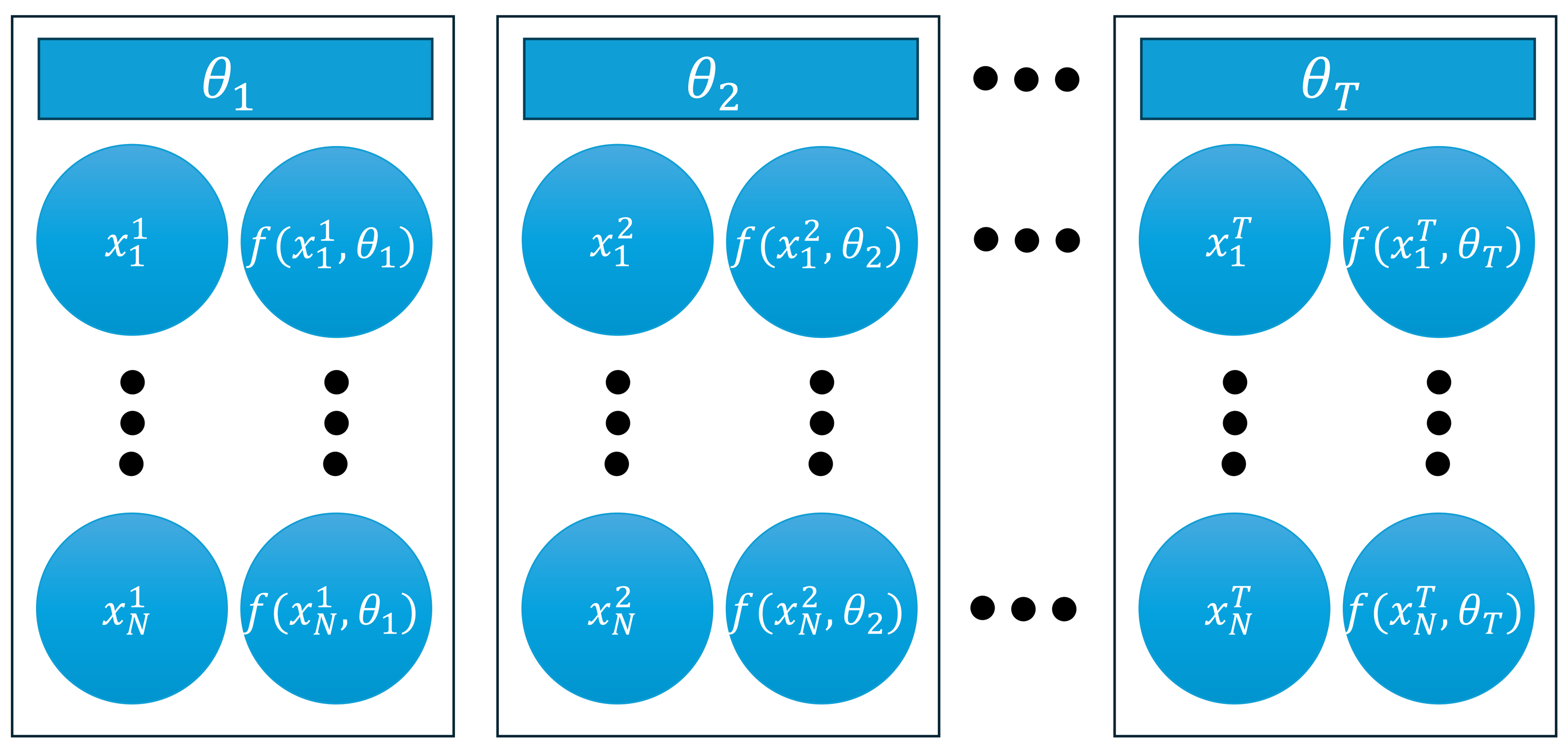
Our Contributions:

1. CBQ incorporates prior smoothness information
   - $f(x,\cdot)$ lies in the Sobolev space $\mathcal{W}_2^{s_f}(\mathcal{X})$.
   - $f(\cdot,\theta)$ lies in the Sobolev space $\mathcal{W}_2^{s_I}(\Theta)$.
2. CBQ has fast rate of convergence.
$$\left\|\hat{I}_{\mathrm{CBQ}} - I\right\|_{L^2(\Theta)} = \mathcal{O}\left(N^{-\frac{s_f}{d}}\right) + \mathcal{O}\left(T^{-\frac{1}{4}}\right)$$
3. CBQ gives finite-sample Bayesian uncertainty.



Full paper & code:
https://github.com/hudsonchen/CBQ

Observed samples



Conditional Bayesian Quadrature

Bayesian Quadrature

A probabilistic numerical integration algorithm based on Gaussian process regression.

$$\hat{I}_{\mathrm{BQ}} = \mathbb{E}_X[k_{\mathcal{X}}(X, x_{1:N})]k_{\mathcal{X}}(x_{1:N}, x_{1:N})^{-1} f(x_{1:N})$$

Conditional Bayesian Quadrature

- **Stage I**: Bayesian quadrature is employed to obtain BQ posterior means $\hat{I}_{\mathrm{BQ}}(\theta_1), \cdots, \hat{I}_{\mathrm{BQ}}(\theta_T)$ and posterior variances $\sigma^2_{\mathrm{BQ}}(\theta_1), \cdots, \sigma^2_{\mathrm{BQ}}(\theta_T)$.
- **Stage II**: Heteroscedastic GP regression is performed over the outputs from Stage I to give a GP posterior mean: $\hat{I}_{\mathrm{CBQ}}(\theta)$ and covariance $\sigma^2_{\mathrm{CBQ}}(\theta)$.

CBQ Estimator:

$$\hat{I}_{\mathrm{CBQ}}(\theta) := k_\Theta(\theta, \theta_{1:T})^\top \big(k_\Theta(\theta_{1:T}, \theta_{1:T}) + \sigma^2_{\mathrm{BQ}}(\theta_{1:T})\big)^{-1} \hat{I}_{\mathrm{BQ}}(\theta_{1:T})$$

Main Theorem

Suppose the following assumptions hold:

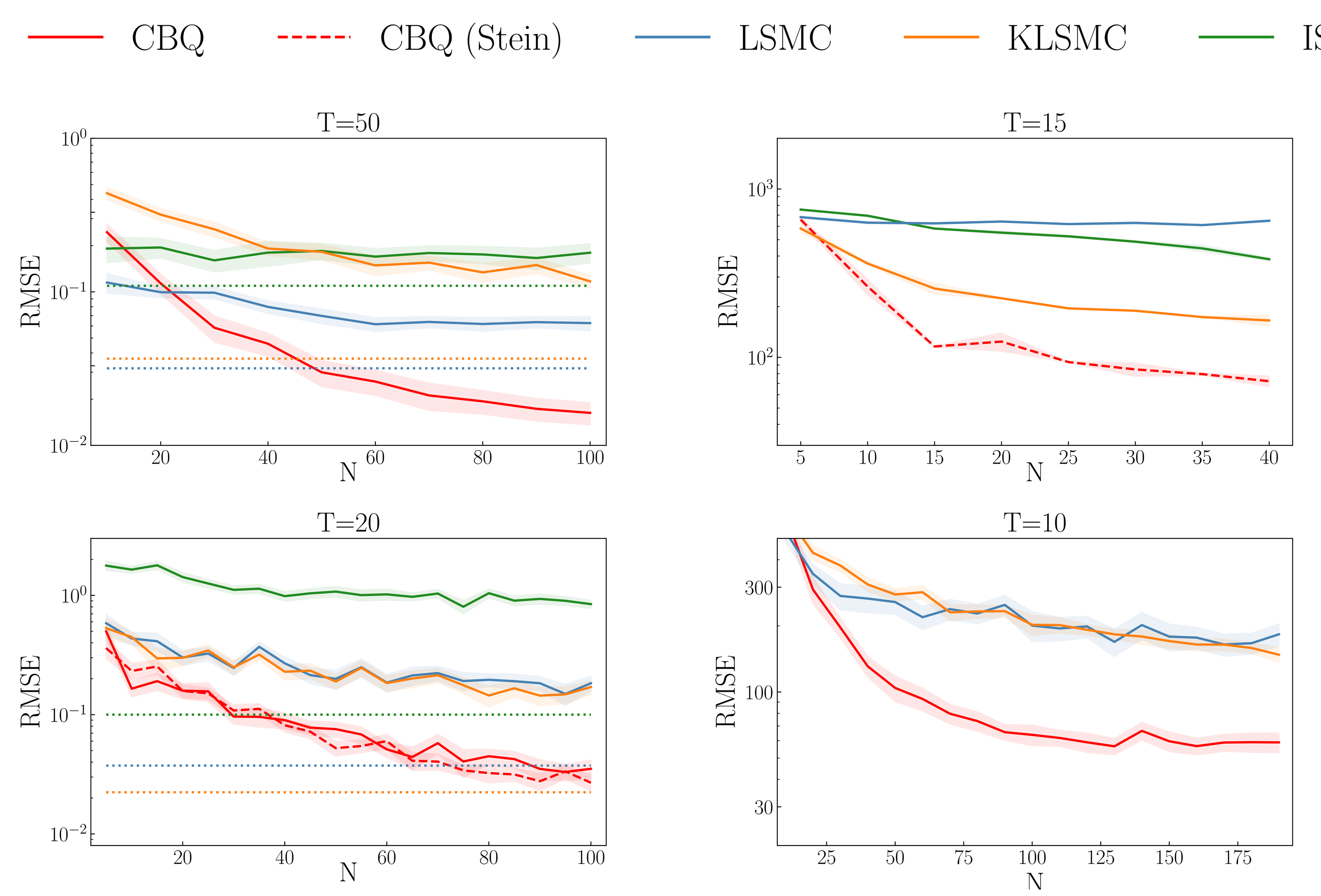
1. Compact domains $\mathcal{X} \subset \mathbb{R}^d$ and $\Theta \subset \mathbb{R}^p$.
2. The kernels $k_{\mathcal{X}}$ and k_Θ are Matérn kernels of smoothness $s_{\mathcal{X}} > d/2$ and $s_\Theta > p/2$.
3. The function $x \mapsto f(x,\theta)$ is of smoothness at least $s_{\mathcal{X}}$, $\theta \mapsto I(\theta)$ is of smoothness at least s_Θ.
4. Other regularity assumptions.

For sufficiently large N, T, with probability $\geq 1-\delta$,

$$\|\hat{I}_{\mathrm{CBQ}} - I\|_{L^2(\Theta)} \leq C_0(\delta)N^{-\frac{s_{\mathcal{X}}}{d}+\varepsilon} + C_1(\delta)T^{-\frac{1}{4}}.$$

Empirical Evaluations

- Bayesian sensitivity analysis.
- Susceptible-Infectious-Recovered (SIR) model
- Option pricing in mathematical finance.
- Uncertainty decision making.



Faster rate in N!