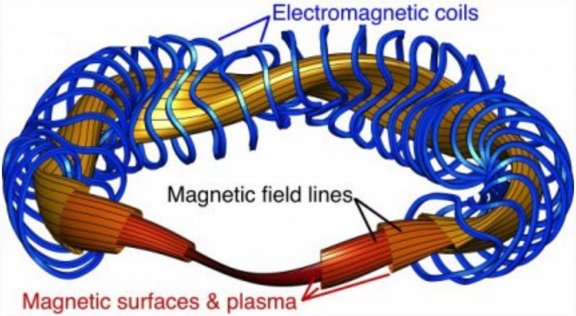
Bayesian Optimization Under Uncertainty with Local Refinement

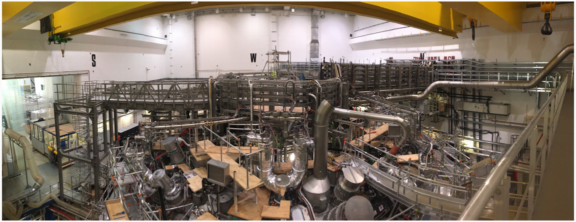
David Bindel

24 Jun 2024

Department of Computer Science
Cornell University

Stellarator Concept and Practice



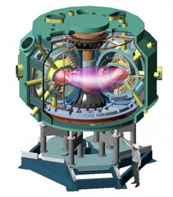


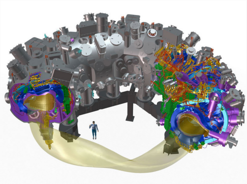
Optimization Under Uncertainty

Low construction tolerances:

- NCSX: 0.08%
- Wendelstein 7-X: 0.1% – 0.17%

Higher tolerances as coil opt goal!

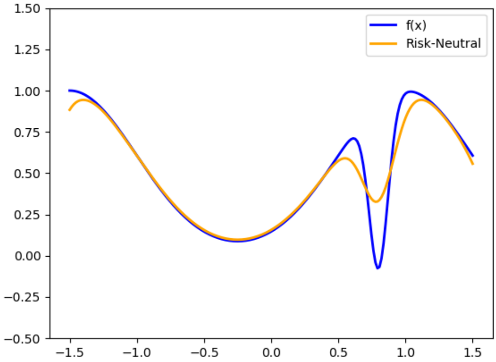




Also want tolerance to

- Changes to control parameters
- Uncertainty in physics or model

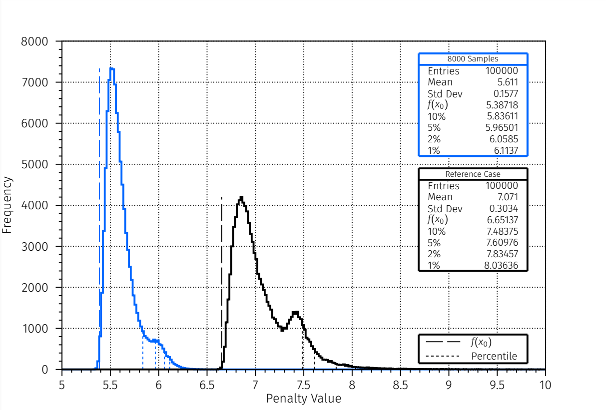
Risk-neutral OUU



Want efficient OUU in ~ 200 dimensions

$$\min_{x \in \Omega} \mathbb{E}_U[f(x - U)]$$

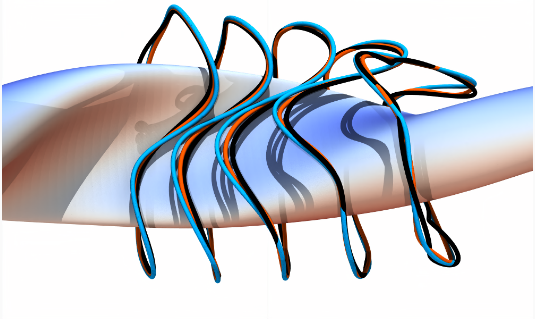
Prior: Monte Carlo Approach



Robustness & mean perf greatly improved (w/ $\sim 10^8$ evals)

J.-F. Lobsien, M. Drevlak, T. Kruger, S. Lazerson, C. Zhu, T. S. Pedersen, *Improved performance of stellarator coil design optimization*, Journal of Plasma Physics, 2020.

Our Approach: fast TuRBO-ADAM



Black: ref; red: TuRBO-ADAM 10mm; blue: TuRBO-ADAM 20mm.

Evaluate objective with FOCUS from PPPL.

- Global search with modified TuRBO
- Local refinement with ADAM with control variate

Costs about 0.01% the evaluation budget.

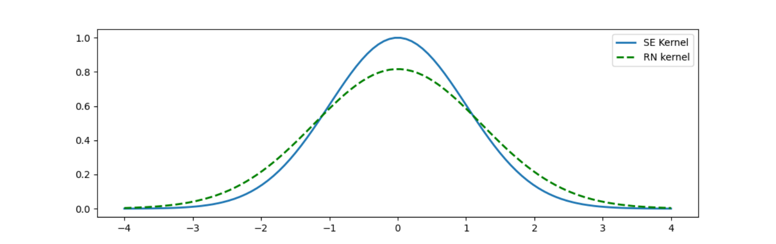
(D)TuRBO: Trust-Region BO

For high-d: combine local BO with multi-start strategy

- Rough global sampling at M points
- Local GP models and trust-region around each point
- Thompson sampling to choose which local model (and trust region) to refine next

(Eriksson, Pearce, Gardner, Turner, Poloczek, 2019)

(D)TuRBO + OUU



- TuRBO builds GP models for $f(x)$ (nominal objective)
- Simple transform from GP for $f(x)$ to GP for $E_U[f(x+U)]$ (Beland and Nair, 2017) — use regularized kernel

Problem: TuRBO explores a lot — want more refinement

Adam + Control Variates

- Regular Adam: stochastic gradient algorithm with "adaptive momentum" for step size control. Use directions

$$g(x) = \nabla f(x + U)$$

  for a random draw U (can also do mini-batch).
- Variance reduction with control variates (Wang, Chen, Smola, Xing, 2013)

$$\begin{aligned} g(x) &= \nabla f(x + U) + \alpha(\hat{g}(x) - E[\hat{g}(x)]) \\ \hat{g}(x) &= \nabla f(x) + HU. \end{aligned}$$

- True Hessian not avail, so set H to be an approximate Hessian (BFGS approximation via gradients from Adam).

Additional Information

Multi-output GPs model $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^k$

- Model covariance over space and across outputs.
- Example: function values + derivatives

$$\mu^{\nabla}(\mathrm{x}) = \begin{bmatrix} \mu(x) \\ \nabla_x \mu(x) \end{bmatrix}, \quad k^{\nabla}(x, x') = \begin{bmatrix} k(x, x') & (\nabla_{x'} k(x, x'))^T \\ \nabla_x k(x, x') & \nabla^2 k(x, x') \end{bmatrix}$$

- Can also model multi-fidelity sims, etc

Pro: FOCUS provides gradients, easy to incorporate!
Con: Matrix dimensions scale like $n(d+1)$
(Partial) Fix: Variational inference (Bindel, Gardner, Huang, Padidar, Zhu, NeurIPS 2021)

For more

S. Glas, M. Padidar, A. Kellison, and D. Bindel,
"Global Stochastic Optimization of Stellarator Coil Configurations,"
Journal of Plasma Physics, vol. 88, no. 2, Apr. 2022.
arXiv: 2110.07464

`https://hiddensymmetries.princeton.edu/`
`https://hifistell.plasma.princeton.edu/`