Global, Robust, Multi-Objective Optimization of Stellarators

David Bindel

29 March 2019

Department of Computer Science
Cornell University

What Makes a Good Stellarator?

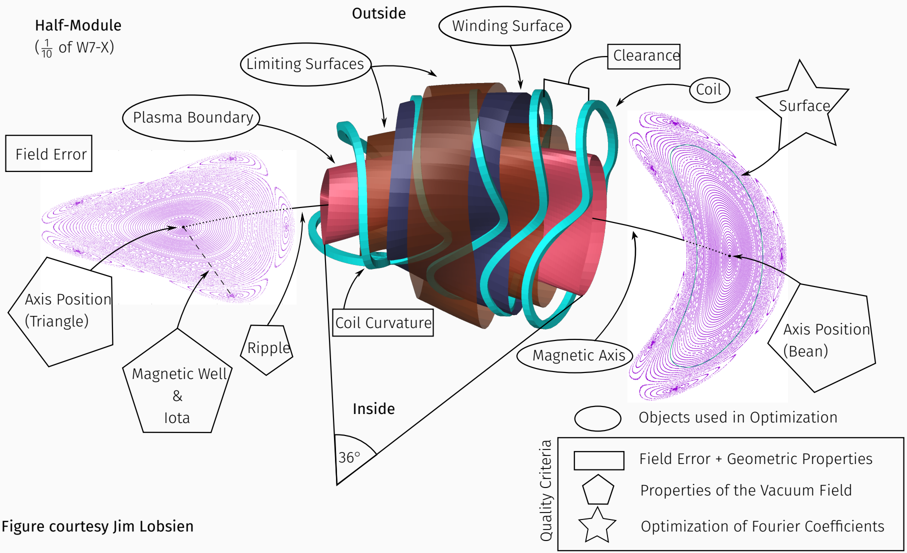
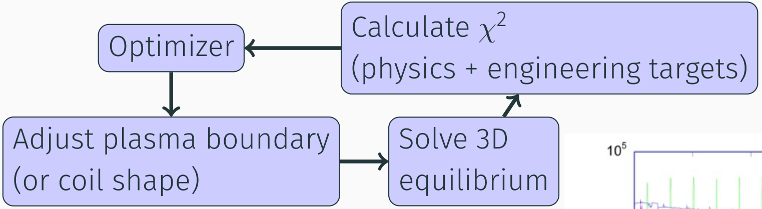
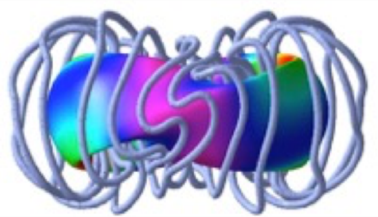


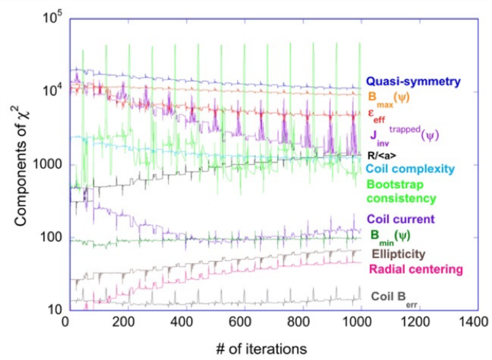
Figure courtesy Jim Lobsien

How Do We Optimize? (STELLOPT Approach)





$$r(\phi,\theta) + iz(\phi,\theta) = \sum \alpha_{m,n} e^{i(m\phi - n\theta)}$$



STELLOPT Approach

Goal: Design MHD equilibrium (coil opt often separate)

- Possible parameters for boundary: $C \subset \mathbb{R}^n$
- Physics/engineering properties: $F : C \to \mathbb{R}^m$
- Target vector: $F^* \in \mathbb{R}^m$

Minimize χ^2 objective over C:

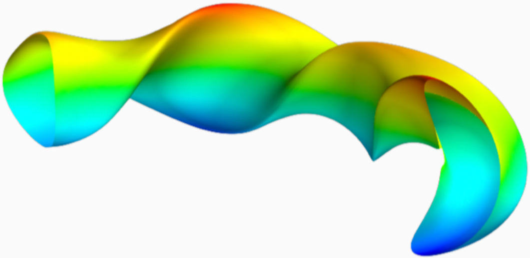
$$\chi^s(x) = \sum_{k=1}^{m} \frac{J_k(x)}{\sigma_k^2}, \quad J_k(x) = (F_k(x) - F_k^*)^2$$

Solve via Levenberg-Marquardt, GA, differential evolution
(avoids gradient information apart from finite differences)

Challenges

1. Costly and "black box" physics computations
   - Each step: MHD equilibrium solve, transport, coil design, ...
   - Several times per step for finite-difference gradients
2. Managing tradeoffs
   - How do we choose the weights in the χ^2 measure? By gut?
   - Varying the weights does not expose tradeoffs sensibly
3. Dealing with uncertainties
   - What you simulate $\neq$ what you build!
4. Global search
   - How to avoid getting stuck in local minima?

Challenge 1: Costly Physics Constraints



Beltrami field (Taylor state):

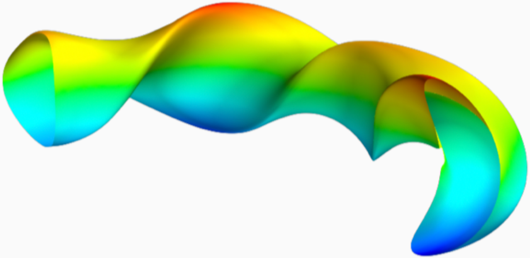
$$\nabla \times B = \lambda B \text{ on } \Omega$$
$$B \cdot n = 0 \text{ on } \partial\Omega$$
$$\nabla \cdot B = 0$$

+ Flux conditions for well-posedness

What goes into the optimization objective and constraints?

- Costly physics solves (MHD equilibrium, transport, ...)
  - PH: "The equations are all first order, and should not be taken too literally."
  - One approach: work with simpler/cheaper proxies
  - Does this actually get us what we want?
- Derivatives require PDE sensitivity / adjoints

Physics-Constrained Optimization



Beltrami field (Taylor state):

$$\nabla \times B = \lambda B \text{ on } \Omega$$

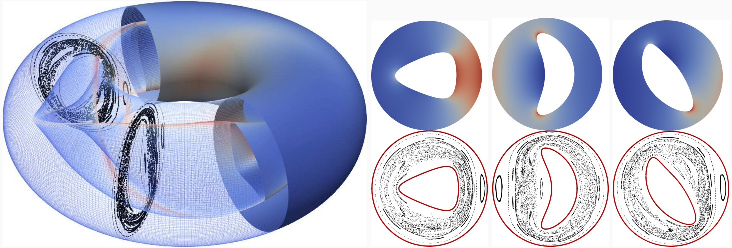
$$B \cdot n = 0 \text{ on } \partial\Omega$$

$$\nabla \cdot B = 0$$

+ Flux conditions for well-posedness

Key: Exploit PDE properties

- PDE-constrained: Solves are part of the optimization
- PDE structure influences objective landscape
- PDE properties: compactness, smoothing, near/far fields
- Opportunities for dimension reduction in optimization



High-order BIE solvers for

- Taylor states in stellarators [Malhotra, Cerfon, Imbert-Gérard, O'Neil, 2019]
- Taylor states in toroidal geometries [O'Neil, Cerfon, 2018]
- Laplace-Beltrami on genus 1 surfaces [Imbert-Gérard, Greengard, 2017]

See also talk of Stuart Hudson, poster by Dhairya Malhhotra, others here

Adjoint-Based Vacuum-Field Optimization

Single-stage optimization of coil shapes and vacuum-field properties:

- Targets: rotational transform, ripple, coil length, magnetic axis length
- Constraints: Magnetic axis is generated by coils

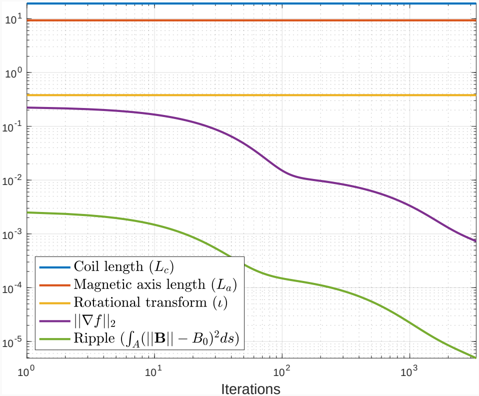
With adjoint solves, not a problem to have many geometric parameters:

N_p	102	192	282	372	462	552
Finite differences	84	222	411	664	1057	1473
Adjoint approach	4	11	26	48	83	116

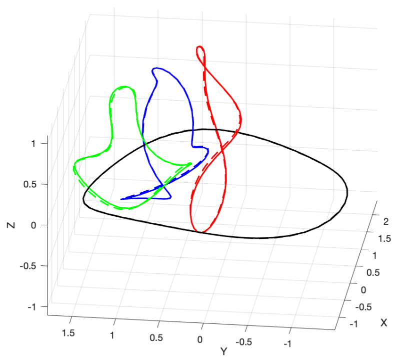
Timings on a modern laptop.

[Giuliani, Cerfon, Landreman, Stadler]

Example: Optimization of Ripple in NCSX Coils



(a) Convergence curve.



(b) Coils before and after optimization.

[Giuliani, Cerfon, Landreman, Stadler]

Challenge 2: Multi-Objective Optimization

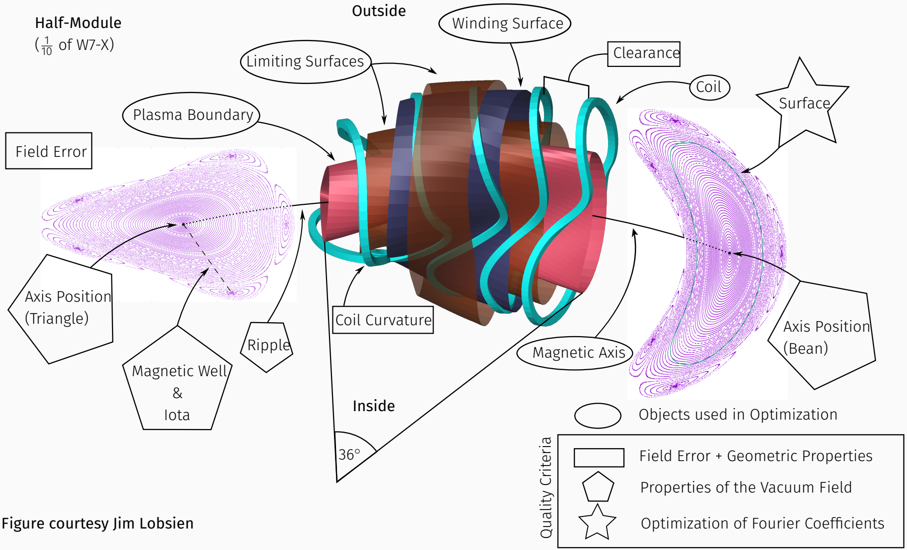


Figure courtesy Jim Lobsien

Stellarator Quality Measures

What makes an "optimal" stellarator?

- Approximates field symmetries (which measures?)
- Satisfies macroscopic and local stability
- Divertor fields for particle and heat exhaust
- Minimizes collisional and energetic particle transport
- Minimizes turbulent transport
- Satisfies basic engineering constraints (cost, size, etc)

Each objective involves different approximations, uncertainties, and computational costs.

Incompleteness of χ-square Combination

Structural Optimization 14, 63–69 © Springer-Verlag 1997

A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems

I. Das and J.E. Dennis

Department of Computational and Applied Mathematics, Rice University of Houston, TX 77251-1892, USA

June 4, 2015 Matt Landreman

Some optimal solutions to a smooth multi-objective problem cannot be found by minimizing a total χ^2

Exploring the Pareto Frontier

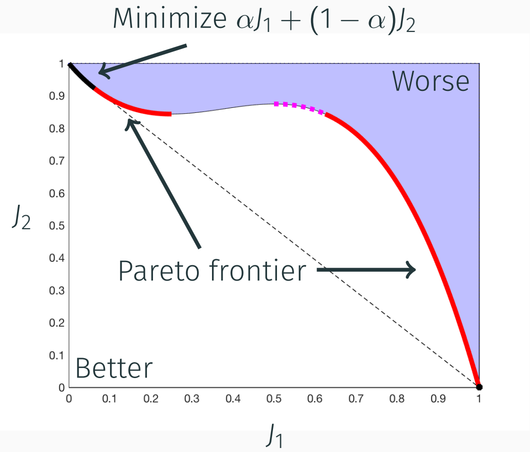
x **dominates** y if

$$\forall k, J_k(x) \leq J_k(y)$$

and not all strict.

Best points called **Pareto optimal** (non-dominated, non-inferior, efficient)

Pareto frontier is an $(m-1)$-manifold with corners, generally.



Minimizing $\sum_k \alpha_k J_k$ only explores convex hull of Pareto frontier!
Other methods sample / approximate the full frontier.

Challenge 3: Optimization Under Uncertainty

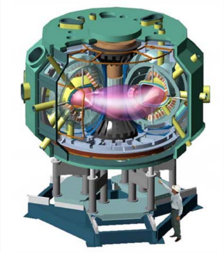
Low construction tolerances:

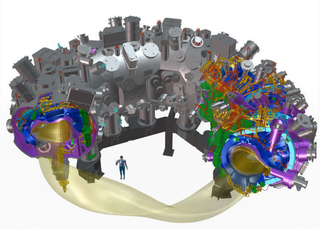
- NCSX: 0.08%
- Wendelstein 7-X: 0.1% – 0.17%

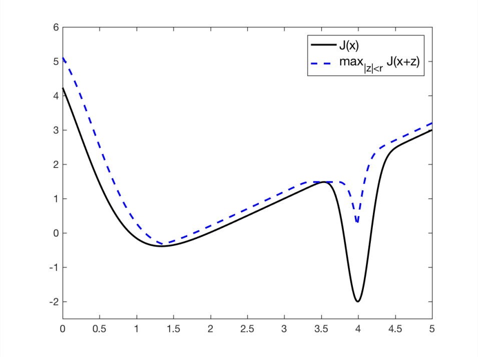
Want: higher tolerances as coil optimization goal!

Also want tolerance to

- Changes to control parameters in operation
- Uncertainty in physics or model parameters







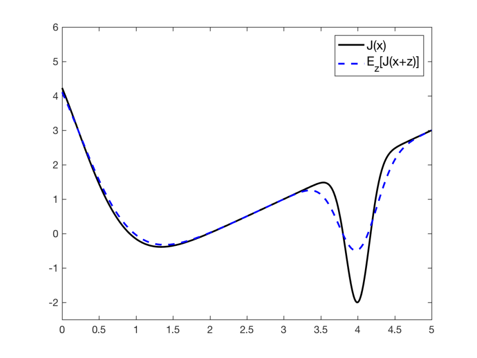
Robust optimization idea:

- Characterize uncertainty region
- Optimize for worst-case

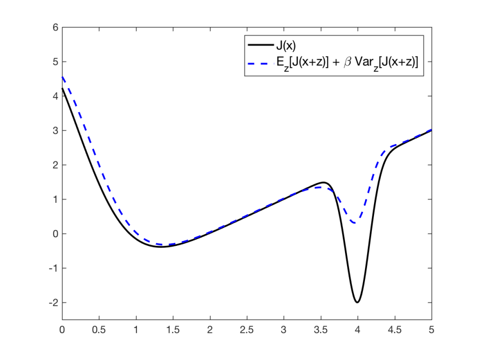
Drawbacks:

- May be pessimistic
- Need an inner optimization
- Non-smooth outer objective

Optimization Under Uncertainty



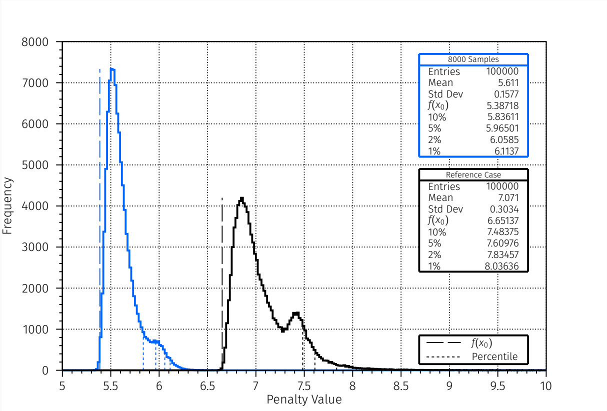
Risk-Neutral Objective



Risk-Averse Objective

- Requires distributional assumption for uncertainty
- Inner computation of moments (MC or quadrature)
- Outer objective becomes smoother

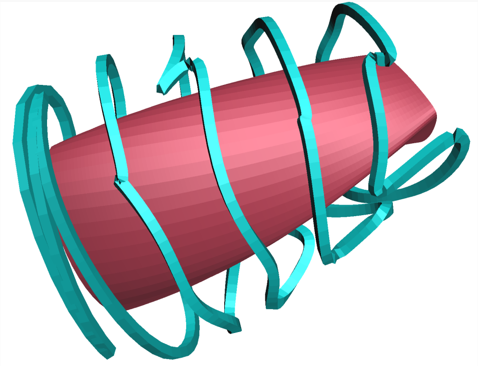
Monte Carlo Approach



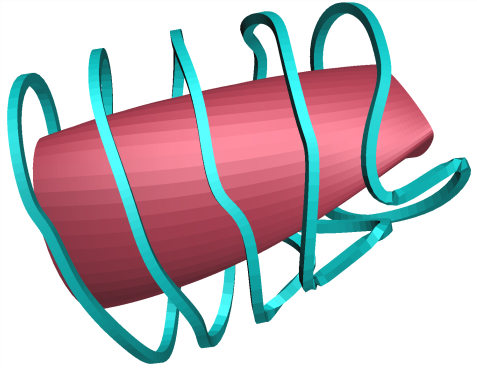
Robustness & average performance significantly improved

[Lobsien, Drevlak, Sunn Pedersen]

Example

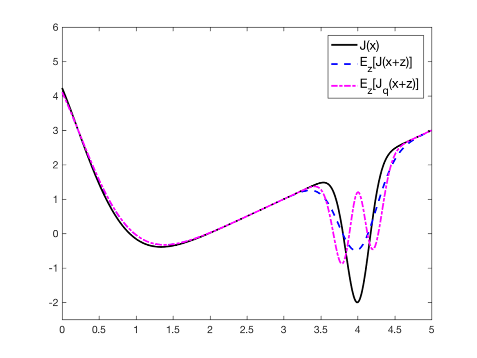


Classic Coil Optimization
(Deviation 0 mm, Penalty = 4.19)

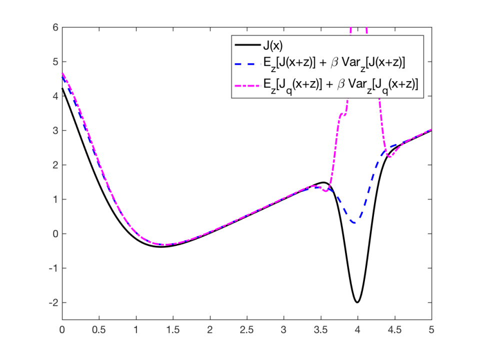


Stochastic Optimization
(Deviation 2.5 mm, Penalty = 2.24)

Efficient Optimization Under Uncertainty



Risk-Neutral Objective



Risk-Averse Objective

$$J_q(x+z) = J(x) + J'(x)z + \frac{1}{2}z^T H_J(x) z, \quad z \sim N(0, C)$$

Use quadratic approximation to compute robust or uncertain objectives

[c.f. Alexanderian, Petra, Ghattas, Stadler, 2017]

Efficient Optimization Under Uncertainty

- Consider objective $J(x, z)$ where x is control and z uncertain
- Model z as multivariate Gaussian
- Use local quadratic approximation in stochastic variables
  - Require $\partial J/\partial z$ and action of Hessian $\partial^2 J/\partial z^2$ on vectors
  - Assume Hessian is (approximately) low rank — dimension reduction
  - Scaling with low intrinsic dimension vs. number of parameters
- Beyond Gaussian: use approximation as a control variate

Lots of remaining challenges (high nonlinearity, turbulence, etc)

Challenge 4: Global Optimization

- Global optimization is hard!
  - Especially in high-dimensional spaces
  - Effective solvers are tailored to structure (e.g. convexity)
  - More general methods are often heuristic
- Want algorithms that balance
  - **Exploration**: Evaluating novel designs with unknown properties
  - **Exploitation**: Refining known designs from previously explored regions
- Global model-based techniques help (with the right models!)

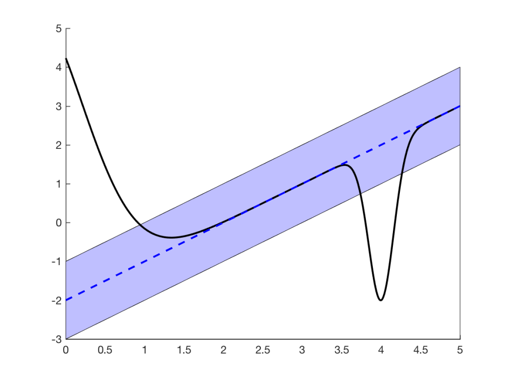
Surrogates

Surrogates (aka *response surfaces*) approximate costly functions

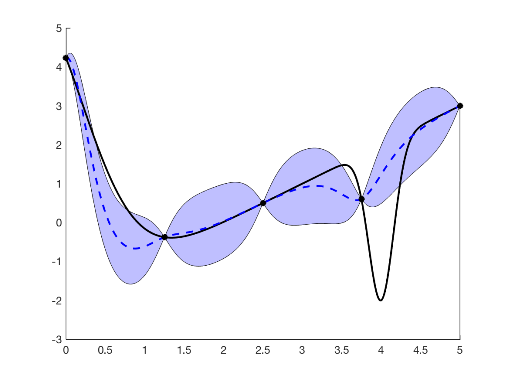
- Different variants
  - Fixed physics-based approximations
  - Parametric: polynomial, ridge, NN
  - Non-parametric: kernel methods
- Incorporate function values, gradients, bounds, ...
- May also estimate uncertainty (e.g. Gaussian process models)

Bayesian optimization uses GP mean and variance to guide sampling.

Kernel-Based Surrogates

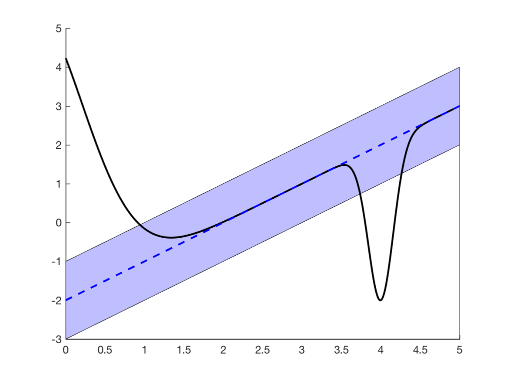


Prior

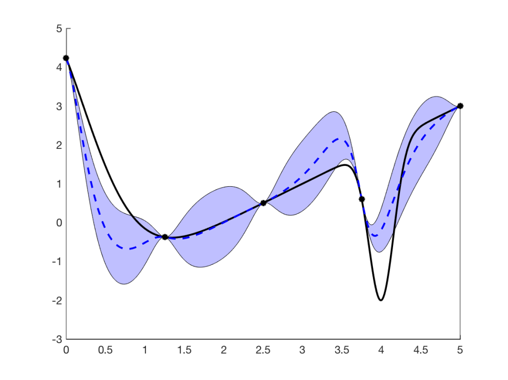


Function data

Kernel-Based Surrogates

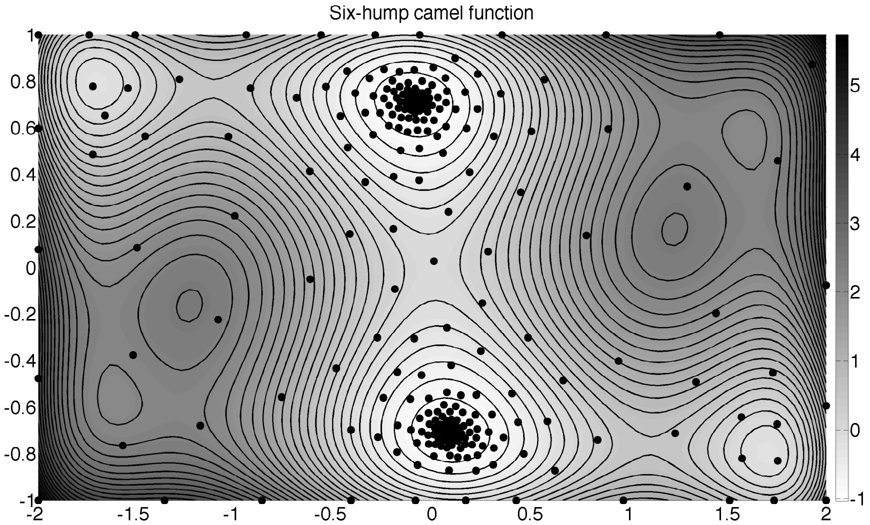


Prior



Hermite data

Exploration vs Exploitation (Eriksson and B)



Surrogate Optimization

- Example: Single objective Bayesian optimization
  - Sample objective and fit a GP model
  - Use acquisition function to guide further sampling (EI, PI, UCB, KG)
- Active work on recent variants for
  - Pareto (ParEGO [Knowles 2004], GPareto [Binois, Picheny, 2018])
  - Multi-fidelity optimization [e.g. March, Willcox, Wang, 2011]
  - Incorporating gradients [Wu, Poloczek, Wilson, Frazier, 2018]
  - Simultaneous dimension reduction [Eriksson, Dong, Lee, B, 2018]
  - Objectives with quadrature [Toscano-Palmerin, Frazier, 2018]
- Several options in PySOT toolkit [Eriksson, Shoemaker, B]

Surrogates with Side Information

- Problem: Need predictions from *limited data*
- Shape surrogate to have known structure (*inductive bias*)
  - Meaningful mean fields
  - Structured kernels (symmetry, regularity, dimension reduction, etc)
  - Tails that capture known singularities and other features
- Alternative: Jointly predict $J_{\mathrm{costly}}(x)$ and $J_{\mathrm{corr}}(x)$
  - Kernel captures correlation betwen functions and across space
  - Basic idea is old: e.g. *co-kriging* in geostatistics
  - Use in computational science and engineering is active research [Peherstorfer, Willcox, Gunzburger, many others]

General Formulation

$$
\begin{aligned}
\min_{\text{coils}} \quad & \mathbb{E}_z[J_{\text{int}}(B,z)], \mathbb{E}_z[J_{\text{qs}}(B,z)], \mathbb{E}_z[J(B,q,z)], \ldots, \mathcal{R}(B,q,z,\ldots) \\
\text{s.t.} \quad & \text{Manufacturing and physics constraints} \\
& \text{PDEs relating coils to magnetic field } B \\
& \text{Physics of particle or heat transport } q
\end{aligned}
$$

- Optimize integrability (J_{int}), quasi-symmetry (J_{qs}), etc
- Take into account uncertain parameters z
- Include risk aversion objective $\mathcal{R}$
- Find Pareto points vs using scalarized objective

General Formulation

$$\begin{aligned} \min_{\text{coils}} \; & \mathbb{E}_z[J_{\text{int}}(B,z)], \mathbb{E}_z[J_{\text{qs}}(B,z)], \mathbb{E}_z[J(B,q,z)], \ldots, \mathcal{R}(B,q,z,\ldots) \\ \text{s.t.} \; & \text{Manufacturing and physics constraints} \\ & \text{PDEs relating coils to magnetic field } B \\ & \text{Physics of particle or heat transport } q \end{aligned}$$

Costs beyond deterministic PDE solves

- Stochastic objectives require many deterministic solves each
- Pareto frontier is an $(m-1)$-dimensional manifold with corners
- Non-convex global optimization requires a lot of searching

Common issue: **curse of dimensionality** — dimension reduction a common theme

Summary

I was tense, I was nervous, I guess it just wasn't my night.
Art Fleming gave the answers; oh, but I couldn't get the questions right.
— *Weird Al Yankovic, "I Lost on Jeopardy"*

Stellarator optimization is hard. Challenges include:

1. Costly and "black box" physics computations
2. Managing tradeoffs
3. Dealing with uncertainties
4. Global search

Many challenges/opportunities in the formulation – not unique to stellarators!