Construction of Surrogates for Computer Experiments with Gradient Information

Fred J. Hickernell and Yiou Li

Applied Mathematics, Illinois Institute of Technology

Challenges

Realistic nuclear reactor simulations

- require a lot of time
- to compute one output, $y = f(\boldsymbol{x})$, from a large d-vector of inputs, $\boldsymbol{x}$.

We need cheap to evaluate surrogates, $\tilde{f} \approx f$, constructed from simulation data,

$$y_i = f(\boldsymbol{x}_i), \quad i = 1, \ldots, m.$$

Questions

- Can we avoid large m although d is large?
- What form should $\tilde{f}$ take?
- How do we choose the design $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$?

More Data from Gradients

Adjoint methods

- provide $\nabla f(\boldsymbol{x}_i)$ along with $f(\boldsymbol{x}_i)$ for ≈ 50–100% additional cost,
- yielding $d+1$ scalar data per data site, $\boldsymbol{x}_i$.
- Models with up to $m(d+1)$, not just m, free parameters can be fit with data from m sites.

References

Hickernell FJ, Li Y (2011) Designs for regression models with general data. in preparation

Li Y, Anitescu M, Roderick O, Hickernell FJ (2011) Orthogonal bases for polynomial regression with derivative information in uncertainty quantification. J Uncertain Quant To appear

Orthogonal Polynomials

Multiple regression models

with polynomial bases, $\boldsymbol{g}^T = (g_1, \ldots, g_p)$,

$$\begin{pmatrix} f(\boldsymbol{x}_i) \\ \nabla f(\boldsymbol{x}_i) \end{pmatrix} = \begin{pmatrix} \boldsymbol{g}^T(\boldsymbol{x}_i) \\ \nabla \boldsymbol{g}^T(\boldsymbol{x}_i) \end{pmatrix} \boldsymbol{\beta} + \varepsilon_i, \; i = 1, \ldots, m,$$

$$\varepsilon_i \text{ i.i.d.}, \quad \operatorname{cov}(\varepsilon_i) = \sigma^2 \operatorname{diag}(1, \lambda_1, \ldots, \lambda_d),$$

are fit to create the surrogate, $\tilde{f} = \boldsymbol{g}^T \hat{\boldsymbol{\beta}}$.
Reliable $\hat{\beta}$ need the information matrix,

$$\mathsf{M} = \frac{1}{m} \sum_{i=1}^{m} \left[\boldsymbol{g}(\boldsymbol{x}_i) \boldsymbol{g}^T(\boldsymbol{x}_i) + \sum_{\ell=1}^{d} \frac{1}{\lambda_\ell} \frac{\partial \boldsymbol{g}}{\partial x_\ell}(\boldsymbol{x}_i) \frac{\partial \boldsymbol{g}^T}{\partial x_\ell}(\boldsymbol{x}_i) \right],$$

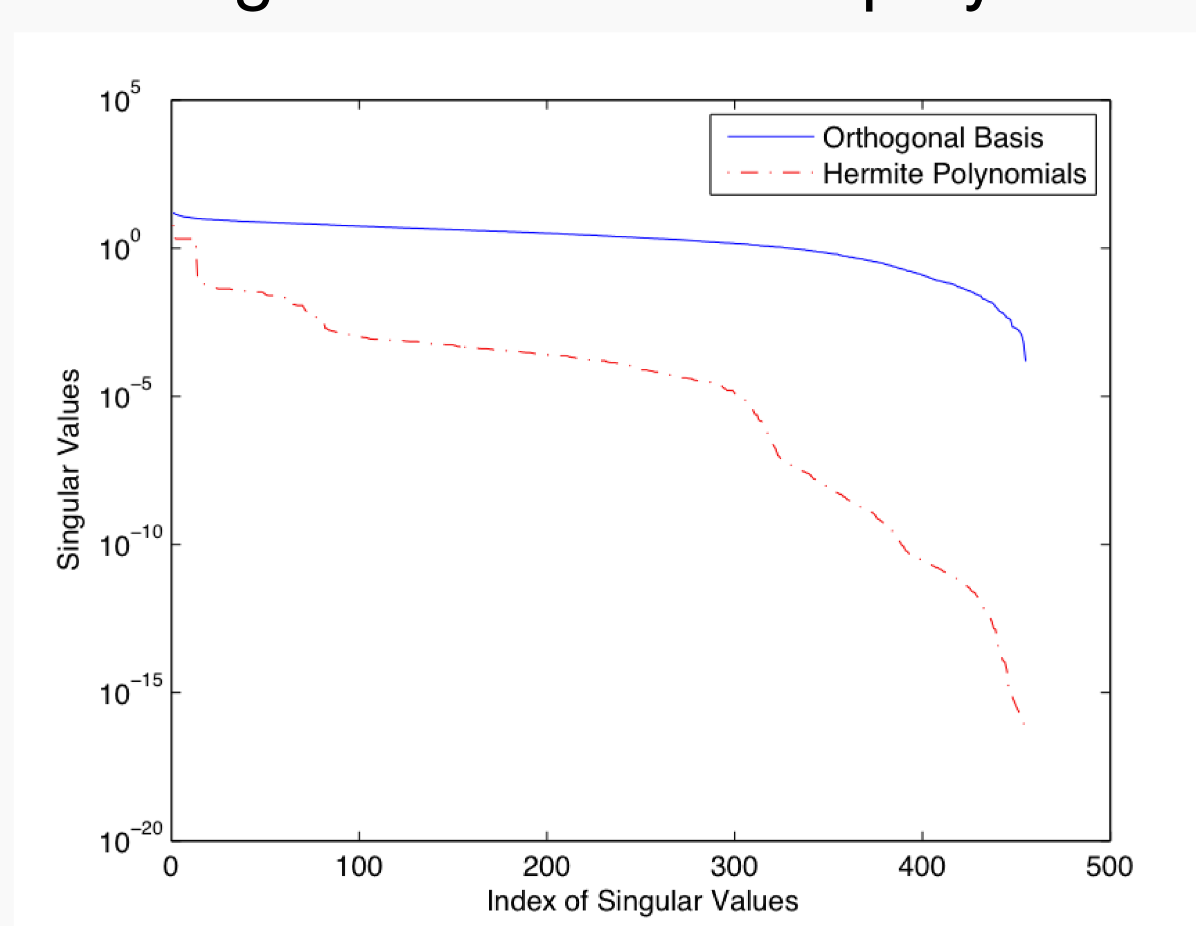
to have a small condition number, $\operatorname{cond}(M)$.

Finding: Orthogonality Including Gradients Improves $\operatorname{cond}(\mathsf{M})$ by 10^{10}

Li et al (2011) show that $\operatorname{cond}(\mathsf{M})$ is smaller when the g_j above are orthogonal under:

$$\langle f, g \rangle_{\mathcal{G}} = \langle f, g \rangle_{\mathcal{L}_2} + \sum_{\ell=1}^{d} \frac{1}{\lambda_\ell} \left\langle \frac{\partial f}{\partial x_\ell}, \frac{\partial g}{\partial x_\ell} \right\rangle_{\mathcal{L}_2},$$

i.e., not Legendre or Hermite polynomials.



The singular values for M plotted above come from a nuclear reactor core model.

The Design (Data Sites)

Discrepancy, $\|F - F_e\|_{\mathcal{M}}$

- measures the difference between a target probability distribution and the empirical distribution of the design,
- arises in the analysis of multivariate numerical integration.

Finding: Use Low Discrepancy Designs

- Hickernell and Li (2011) show that low discrepancy designs give M like that for continuous sampling.
- Low discrepancy designs are model agnostic.

Ongoing Work

Choosing Designs via Large Scale Optimization

- For $d = 1$ designs with optimal M are different when gradient information is used.
- Semi-definite programming may be used to find optimal designs for $d > 1$.

Acknowledgements

This work is supported in part by DE-SC0002100 and performed in collaboration with Mihai Anitescu and Oleg Roderick, both of Argonne National Laboratory.