Gradient-Based Simulation Optimization

Michael Fu

ISR students: Huashuai Qu, Marie Chau

THIRTY YEARS OF SYSTEMS RESEARCH EXCELLENCE

INSTITUTE FOR SYSTEMS RESEARCH
A. JAMES CLARK SCHOOL OF ENGINEERING

History: In the setting of (Monte Carlo) simulation for stochastic discrete-event systems, various direct gradient estimation techniques such as perturbation analysis and the likelihood ratio or score function method were invented in the 1970s, which enabled efficient gradient-based search for stochastic optimization problems.

Future Work:

- Incorporating DiGAR and GESK into sequential response surface methodology simulation optimization algorithms
- Applying STAR-SPSA to real-world problems
- Combining local search with global optimization

Direct Gradient Augmented Regression (DiGAR)

Traditional Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

DiGAR

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\ g_i &= \beta_1 + \epsilon_i' \end{aligned}$$

Theorem

Under certain assumptions,

$$\frac{\sigma_g^2}{\sigma^2} \le \frac{2\alpha}{1-\alpha} + \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \iff Var(\hat{\beta}_1^{DiGAR}) \le Var(\hat{\beta}_1^{standard}).$$

where $\hat{\beta}_1$ indicates the respective slope estimators.

Extensions: MLE and multivariate

Traditional

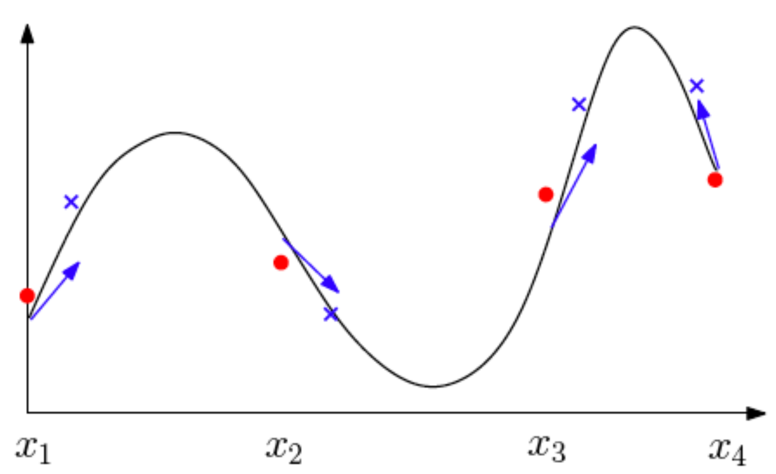
$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \end{aligned}$$

DiGAR

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \frac{1-\alpha}{\alpha}\bar{g}}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{1-\alpha}{\alpha}} \end{aligned}$$

M.C. Fu and H. Qu, "Augmented Regression With Direct Gradient Estimates," *INFORMS Journal on Computing*, Vol.26, No.3, 484–499, 2014.

Gradient-Enhanced Stochastic Kriging (GESK)



SK Model: $y_j(x_i) = \mathbf{f}(x_i)^T\beta + \mathrm{M}(x_i) + \epsilon_j(x_i)$

SKG Model: $g_j(x_i) = \nabla \mathbf{f}(x_i)'\beta + \nabla \mathrm{M}(x_i) + \delta_j(x_i)$

GESK Model:

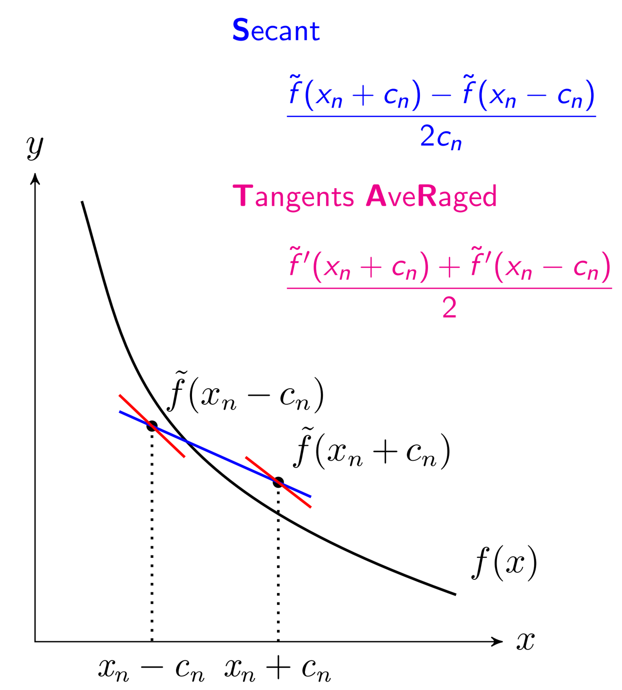
$$\begin{aligned} x_i^+ &= x_i + \Delta x_i \\ y_j^+(x_i) &= y_j(x_i) + g_j(x_i) \cdot \Delta x_i \\ &= \mathbf{f}(x_i^+)^T\beta + \mathrm{M}(x_i^+) + \epsilon_j(x_i^+) \end{aligned}$$

H. Qu and M.C. Fu, "Gradient Extrapolated Stochastic Kriging," *ACM Transactions on Modeling and Computer Simulation*, 24(4), 2014.

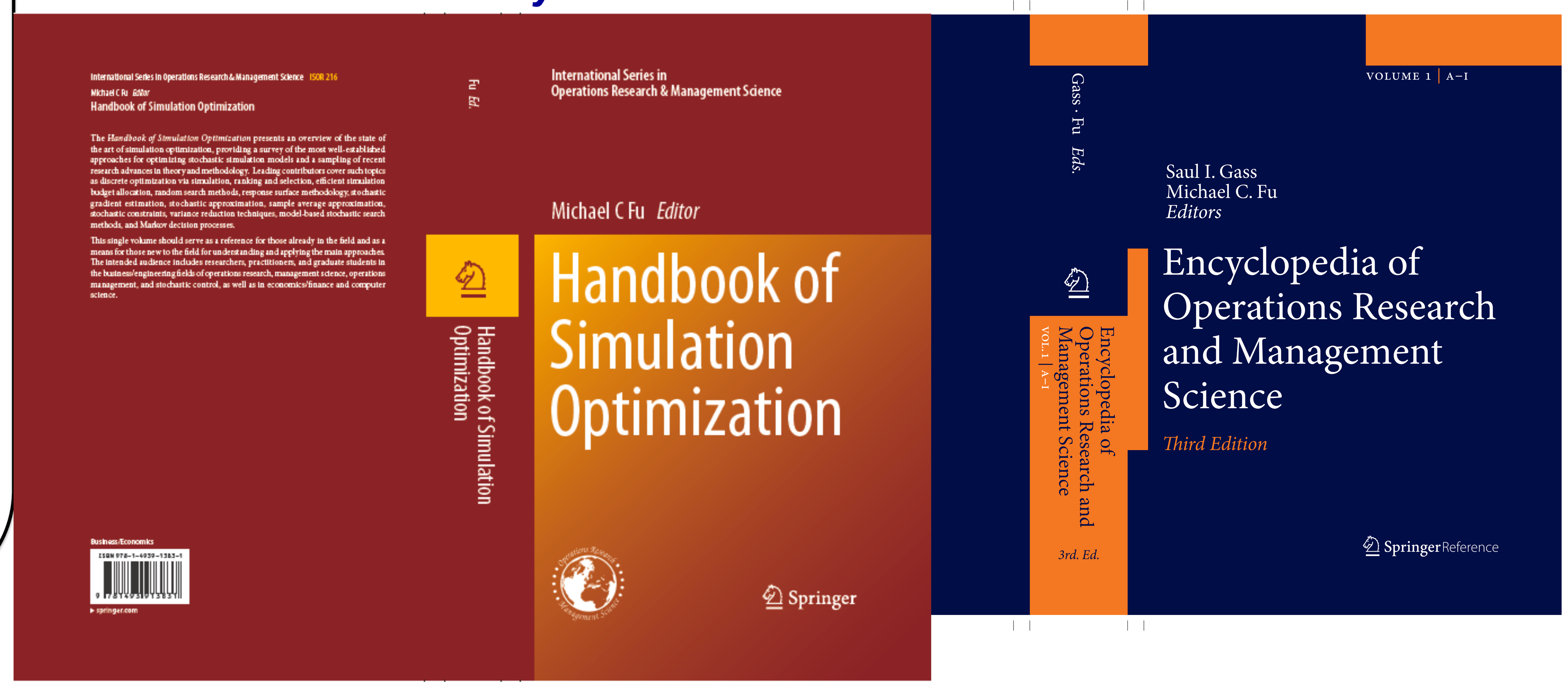
Secant Tangents AveRaged (STAR) Stochastic Approximation

- 1st stochastic approximation to COMBINE direct and indirect gradient estimates
- convex combination, weights chose in an "optimal" manner
- provably convergent (in mean square and almost sure)
- extended to high dimensions via simultaneous perturbation stochastic approximation
- variance of gradient provably lower than original gradients used in Robbins-Monro or Kiefer-Wolfowitz stochastic approximation algorithms

- M. Chau, M.C. Fu, and H. Qu, "Multivariate Stochastic Approximation Using a Secant- Tangents AveRaged (STAR) Gradient," *Operations Research*, under revision.
- M. Chau, H. Qu, and M.C. Fu, "A New Hybrid Stochastic Approximation Algorithm," Proceedings of the 2014 Workshop on Discrete Event Systems, 241–246, 2014.



Recently Published Books in Related Areas



Main Sources of Support: NSF-DMI/CMII, NSF-EECS, AFOSR