Assessing Convergence in Gaussian Process Surrogate Model Optimization

Nick Grunloh
as advised by Dr. Herbie Lee

Applied Mathematics and Statistics
University of California, Santa Cruz

grunloh@soe.ucsc.edu
herbie@ucsc.edu

January 12, 2015

Optimization can be Tricky

$$\underset{\mathbf{x}}{\operatorname{argmin}} f(\mathbf{x})$$

Problems:

- $\mathbf{x}$ may be in many dimensions
- f may be poorly behaved
- Often no useful f' information
- f evaluations may be expensive

Example:
Computer simulation experiments

Newton-Raphson

Beyond Newton-Raphson: Pattern Search

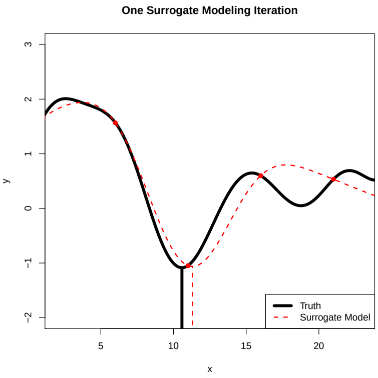
Beyond Newton-Raphson: Simulated Annealing

Beyond Newton-Raphson: Evolutionary Algorithms

Statistical Surrogate Modeling

Procedure:

1) Collect an initial set from the domain $\mathbf{x}$
2) Compute $f(\mathbf{x})$
3) Model f
4) Predict an optimal $\mathbf{x}^*$
4) Add $\mathbf{x}^*$ to $\mathbf{x}$
5) Check convergence
6) If converged exit. Otherwise go to 2).



Treed GP Surrogate Model

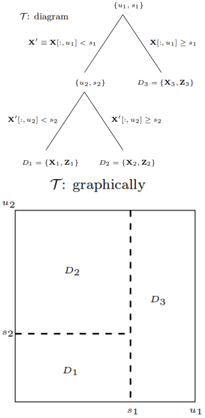
Partition f across R regions:

$$f_r \sim \mathrm{GP}\left(m_r(\mathbf{x}), C_r(\mathbf{x}, \mathbf{x}')\right)$$

$$\cup_{r=1}^{R} f_r = f$$

$$f \sim \mathcal{T}\mathrm{GP}$$

Hierarchical priors for $m_r(.)$ and $C_r(.)$ parameters to share across partitions.



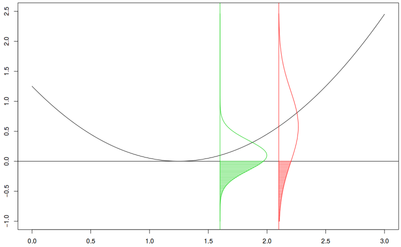
* R. B. Gramacy, (2007). *tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models*. Journal of Statistical Software, 19(9), 1-46.

Expected Improvement (EI)

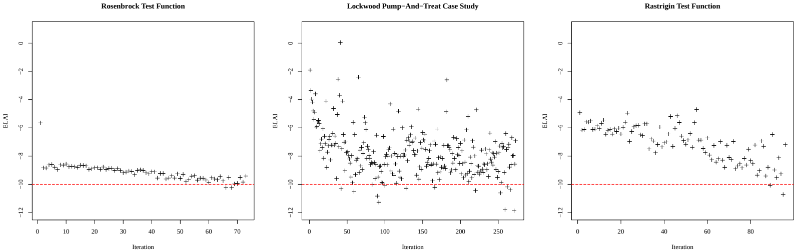
$$f_{min} = \min\left\{ f(\mathbf{x_1}), ..., f(\mathbf{x_N}) \right\}$$

$$\mathsf{I}(\mathbf{x}) = \max\left\{ \left(f_{min} - f(\mathbf{x})\right),\ 0 \right\}$$

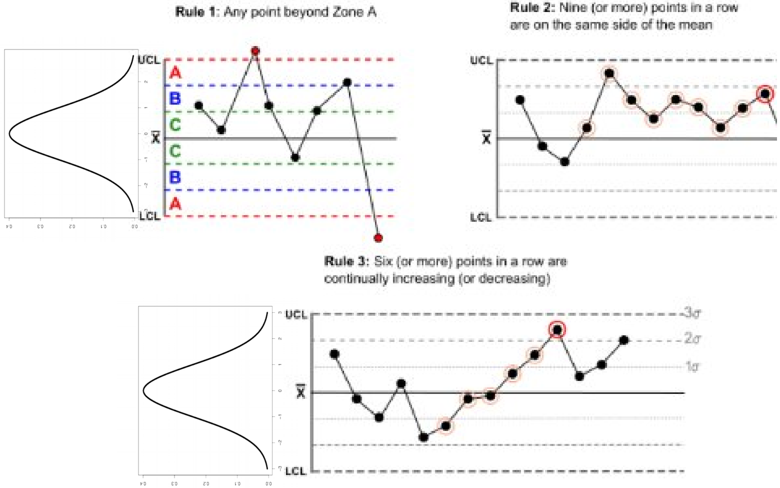
$$\mathsf{EI} = \mathbb{E}\left[\ \mathsf{I}(\mathbf{x})\ \right]$$



Convergence



Statistical Process Control (SPC)

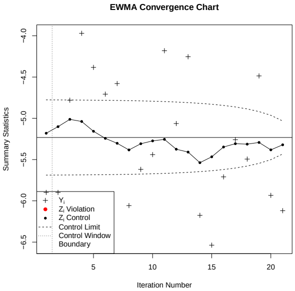


Convergence is Subtle

$$z_i = \lambda \bar{x}_i + (1-\lambda) z_{i-1}$$

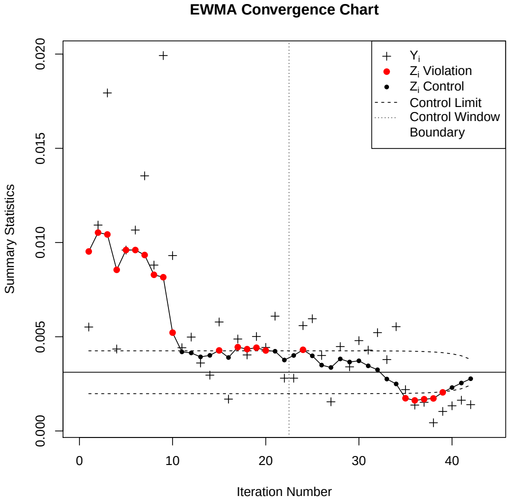
$$\sigma^2_{z_i} = \frac{\sigma^2_x}{n}\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]$$

$$\mathrm{CL}_i = \mu \pm c\ \sigma_{z_i}$$

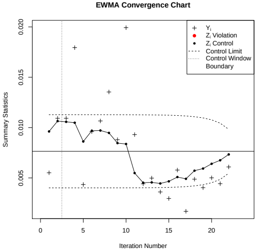


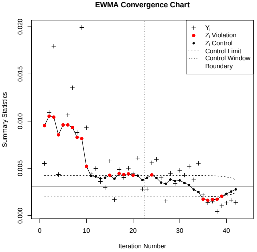
* L. Scrucca (2004). *qcc: an R package for quality control charting and statistical process control*. R News 4/1, 11-17.

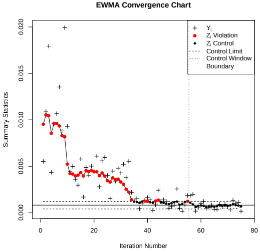
Build a Convergence Chart



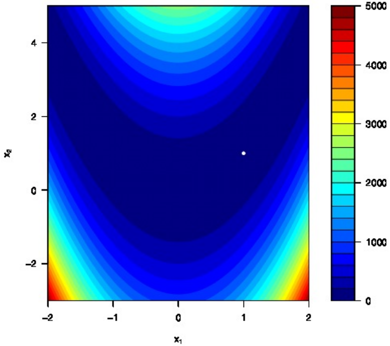
Interpret a Convergence Chart

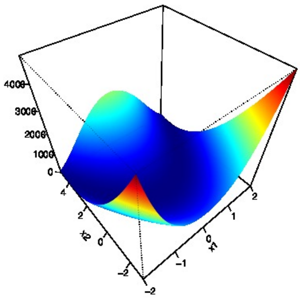






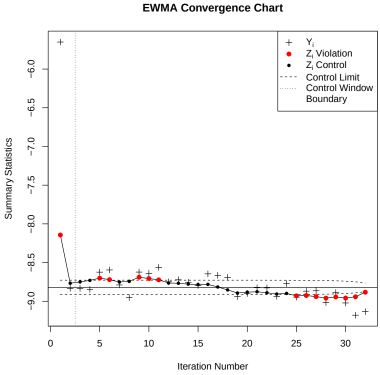
Rosenbrock Test Function

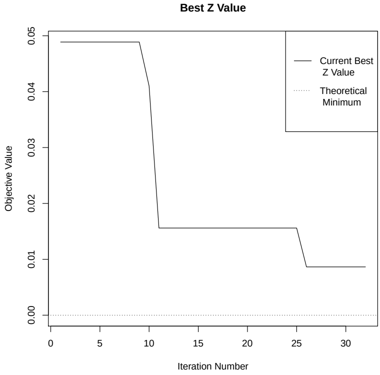




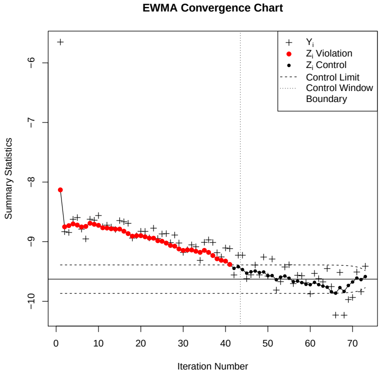
$$
\begin{aligned}
f(x_1, x_2) &= 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 \\
\text{Minimum} &: \quad f(1,1) = 0
\end{aligned}
$$

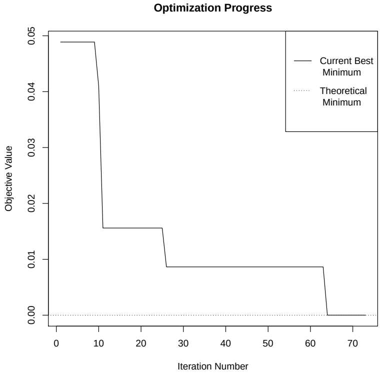
Rosenbrock Pre-Convergence



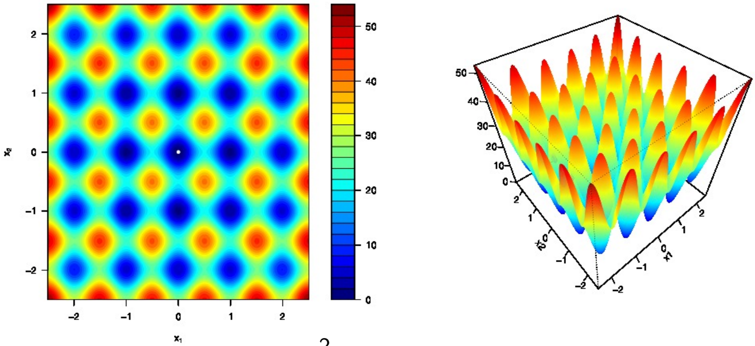


Rosenbrock Convergence





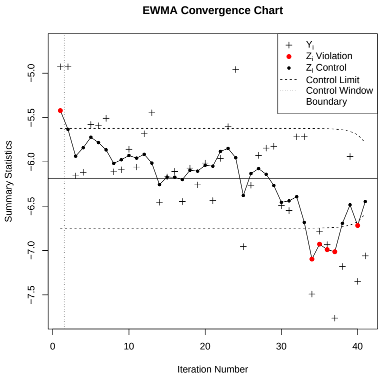
Rastrigin Test Function

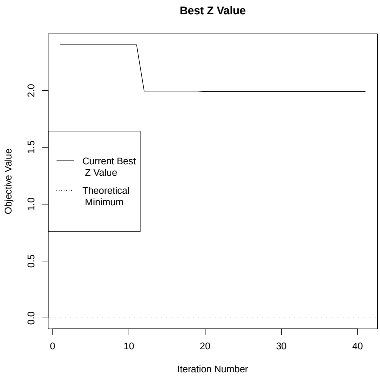


$$f(x_1, x_2) = \sum_{i=1}^{2} \left[x_i^2 - 10\cos(2\pi x_i)\right] + 2(10)$$

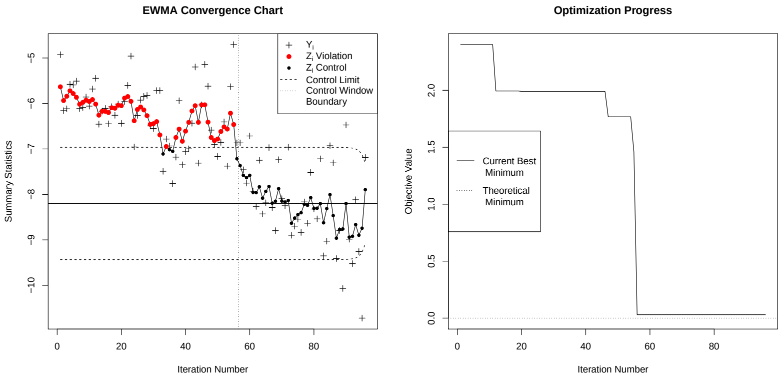
$$\text{Minimum} \quad : \quad f(0,0) = 0$$

Rastrigin Pre-Convergence

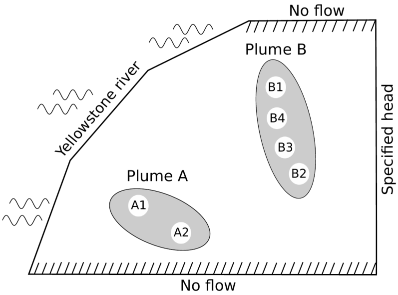


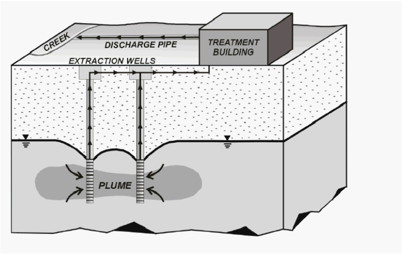


Rastrigin Convergence



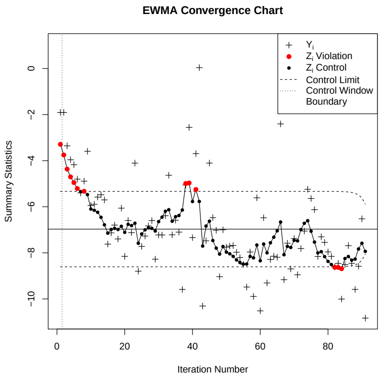
Lockwood Pump-and-Treat Case Study

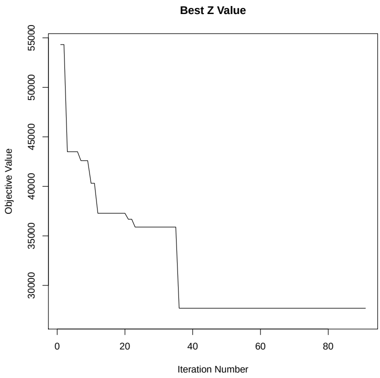




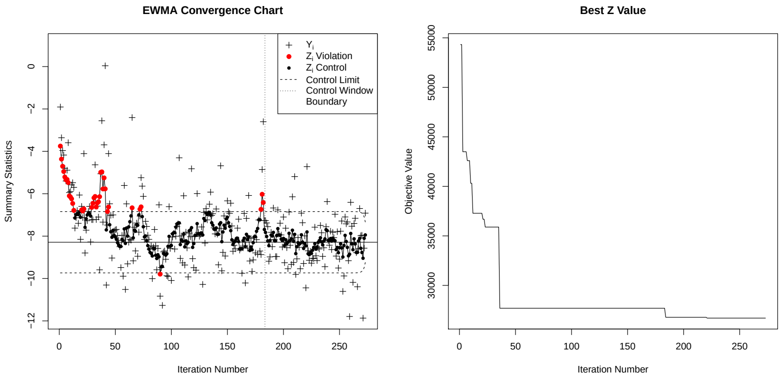
$$f(\mathbf{x}) = \sum_{i=1}^{6} x_i + 2\left[c_A(\mathbf{x}) + c_B(\mathbf{x})\right] + 20000\left[\mathbb{1}_{c_A(\mathbf{x})>0} + \mathbb{1}_{c_B(\mathbf{x})>0}\right]$$

Lockwood Pre-Convergence





Lockwood Convergence



Conclusions

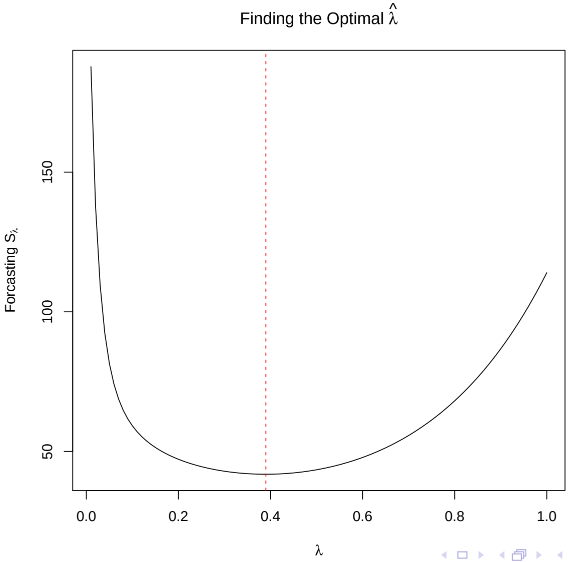
- A method for considering stochastic convergence criterion.
- Convergence, in this sense, describes when surrogate modeling algorithms have exhausted their searching potential.
- I want an objective way to choose the control limit size w.

The Deets

$$Z(\mathbf{x}) = m(\mathbf{x}, \boldsymbol{\beta}) + \epsilon(\mathbf{x}) + \eta(\mathbf{x}).$$

$$\begin{aligned}
\mathbf{Z} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{K} &\sim N_n\left(\mathbf{F}\boldsymbol{\beta},\ \sigma^2\mathbf{K}\right) & \sigma^2 &\sim IG\left(\frac{\alpha_\sigma}{2},\ \frac{\beta_\sigma}{2}\right) \\
\boldsymbol{\beta} \mid \sigma^2, \tau^2, \boldsymbol{\beta}_0, \mathbf{W} &\sim N_m\left(\boldsymbol{\beta}_0,\ \sigma^2\tau^2\mathbf{W}\right) & \tau^2 &\sim IG\left(\frac{\alpha_\tau}{2},\ \frac{\beta_\tau}{2}\right) \\
\boldsymbol{\beta}_0 &\sim N_m\left(\boldsymbol{\mu},\ \mathbf{B}\right) & \mathbf{W} &\sim IW\left(\rho\mathbf{V},\ \rho\right).
\end{aligned}$$

$$K(\mathbf{x}_j, \mathbf{x}_k | d) = \exp\left\{-\frac{||\mathbf{x}_j - \mathbf{x}_k||^p}{d}\right\} + g\delta_{j,k}$$


Finding the Optimal $\hat{\lambda}$
Forcasting S_λ
50
100
150
0.0
0.2
0.4
0.6
0.8
1.0
λ