Global Optimization of Expensive Black-box function

Abstract: Number of **function evaluation** is limited by time and cost. Address this problem by **fitting response surface to data** collected by evaluating objective and **constraint function** at few point. The key is balancing between **exploit the approximating surface** (sampling it where it is minimized) and **improve the approximation** (by sampling where prediction error is high)

The response surface methodology based on **modeling the objective and constraint function with stochastic process,** which is approach to **approximate function** in 3 different literatures: mathematical geology (not use in this paper), global optimization ( called Bayesian global optimization or “random function approach”) and statistic (approximating integral and other hard-to-compute functionals of functions)[1]. This approach have 3 advantages: **require the fewest function evaluations** (method can be able to see obvious trends and jump to conclusion instead of move step-by-step along trajectory), **credible stopping rule** based on the expected improvement from further searching (statistical model provide confidence interval on function’s value at unsampled point) and **fast approximation that can be used to identify important variable** (not only estimate the optimal point)

The stochastic process model: Linear regression is the simplest example for response surface. The observations are treated as generated from the model

It rises 2 problem. **Practical problem** is we don’t know what function forms for regression terms. **Conceptual problem** is it is false for the assumption of independent errors (any lack of fit will be modeled as errors) and this is addressed by stochastic process. The correlation of errors depend on the **distance** between corresponding points. We don’t use **Euclidean distance** because it weights all variables equally. Instead we use this

Where measure the importance or activity of the variable and is related to smoothness of the function in coordinate direction h ( close to 1 mean less smoothness). It turns out the model we use in stochastic process: , is mean of stochastic process and is whose estimations must be combined with estimate of correlation parameter in order to make prediction. We have correlation matrix as:

[1] when input space, **R,** of the function itself consists of functions, the functions of the elements of R are called **functionals**