Multifidelity Optimization using Statistical Surrogate Modeling for Non-Hierarchical Information Sources (Lam, Allaire, Willcox)

Saul Toscano-Palmerin

School of Operations Research and Information Engineering, Cornell University

October 7, 2015

Introduction

- Goal: Designing and optimizing complex systems.
- The quantities of interest can be evaluated by different information sources (IS). They could be numerical models, experiments, or historical data.
- Each IS has an associated fidelity.

Method

- Adaptively build a multifidelity surrogate for multifidelity optimization.
- Definition of fidelity in the form of a variance metric.
- Gaussian processes are used to create an intermediate surrogate for each IS.
- A single multifidelity surrogate is constructed by fusing all the intermediate surrogates.
- Advantage: the multifidelity surrogate capability of integrating IS whose fidelity changes over the design space.
- Acquisition function: expected improvement criteria.

Multifidelity Surrogate: Intermediate GP

- For each IS f_m, construct an intermediate surrogate using a Gaussian process.

- Kernel: square exponential covariance function with additive Gaussian noise for each IS.

- The parameters $\sigma^2_{GP,m}(x)$, $\mu_m(x)$ of a GP are updated assuming homogeneous noise.

- The fidelity of an IS is its variance $\sigma^2_{f,m}(x)$. Those variances are provided as an input by the user.

Multifidelity Surrogate: Intermediate GP

- At each step, define a new surrogate σ_t^2 for each IS with parameters $\mu_{GP,m}(x), \sigma_{t,m}^2(x)$, where

$$\sigma_{t,m}^2(x) = \sigma_{f,m}^2(x) + \sigma_{GP,m}^2(x)$$

- This is different to our approach because we update the parameters assuming heterogeneous noise, however they add a known heterogeneous noise after they update the parameters of the GP.

Multifidelity Surrogate: Fusion of Information

- Following Winkler, they build a single multifidelity surrogate

$$S(x) \sim N\left(\bar{\mu}(x), \bar{\sigma}^2(x)\right)$$

where

$$\begin{aligned}
\bar{\mu}(x) &= \bar{\sigma}^2(x) \sum_i \frac{\mu_i(x)}{\sigma_{t,i}^2(x)} \\
\bar{\sigma}^2(x) &= \left(\sum_i \frac{1}{\sigma_{t,i}^2(x)} \right)^{-1}
\end{aligned}$$

The problem

- The problem is to solve

$$x^* = \text{arg min } \mu(x)$$

where

$$\begin{aligned}
\mu(x) &= \sigma^2(x) \sum_i \frac{f_i(x)}{\sigma^2_{f,i}(x)} \\
\sigma^2(x) &= \left(\sum_i \frac{1}{\sigma^2_{f,i}(x)} \right)^{-1}
\end{aligned}$$

- The solution given is the point that minimizes $\bar{\mu}(x)$.

Acquisition function

- At each iteration n of the optimization algorithm, a new design x_n is evaluated with IS number $m_n \in \{1, \ldots, M\}$. The new training point is added to the training set. The m_n^{th} intermediate surrogate and the multifidelity surrogate are then updated.

- They use EI:

$$EI(x) = \mathbb{E}\left[\max\left(y_{min} - Y_x\right), 0\right]$$

where $Y_x \sim N\left(\bar{\mu}(x), \hat{\sigma}^2(x)\right)$ where

$$\hat{\sigma}^2(x) = \left(\sum_i \frac{1}{\sigma^2_{GP,i}(x)}\right)^{-1}.$$

Acquisition function

- Note: they use $\hat{\sigma}^2(x)$ instead of $\bar{\sigma}^2(x)$ because they want to avoid picking designs that have already been evaluated with at least one IS.
- The next point chosen is $x_{n+1} = \arg\max EI(x)$.
- When the maximum EI found is lower than a threshold ϵ, they instead choose $x_{n+1} = \arg\min \bar{\mu}(x)$.

Choosing the Next Information Source to Query

- Heuristic motivation: large information gain for a low evaluation cost.
-

$$
\begin{aligned}
m_{n+1} &= \arg\min_m \frac{C_m(x_{n+1})}{\bar{\sigma}^2(x_{n+1}) - \tilde{\sigma}^2_m(x_{n+1})}, \\
\tilde{\sigma}^2_m(x_{n+1}) &= \left(\frac{1}{\sigma^2_{f,m}(x_{n+1})} + \sum_{i=1, i\neq m}^{M} \frac{1}{\sigma^2_{t,i}(x_{n+1})} \right)^{-1}
\end{aligned}
$$

where $C_m(x_{n+1})$ is the evaluation cost of the m^{th} IS at x_{n+1}.

- A better approach might be that in Moore et al., who choose the design and the IS at the same time by maximizing a value of information metric.