Applying Bayesian Estimation to Noisy Simulation Optimization

Geng Deng Michael C. Ferris

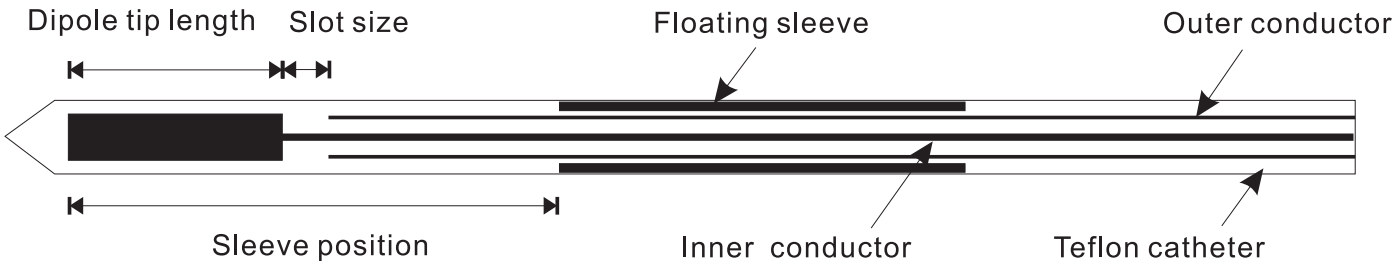
University of Wisconsin-Madison

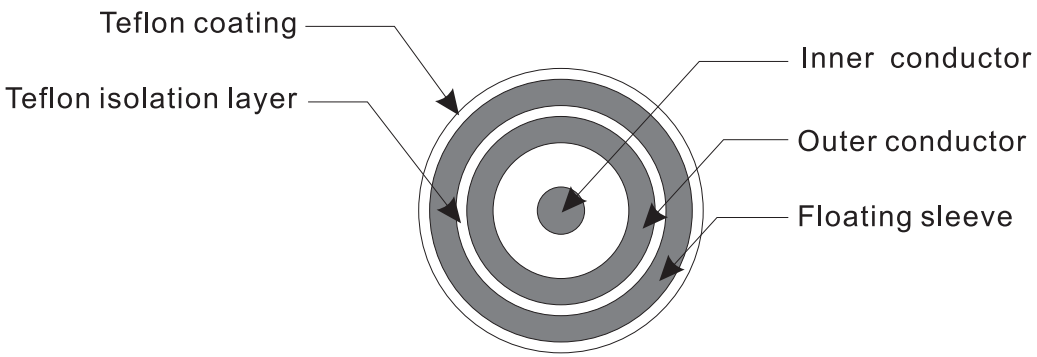
INFORMS Annual Meeting Pittsburgh 2006

Simulation-based optimization problem

- Computer simulations are used as substitute to evaluate complex real systems.
- Simulations are widely applied in manufacturing, supply chain management, medical treatment and many other fields.
- The goal: Optimization finds the best values of the decision variables (design parameters or controls) that minimize some performance measure of the simulation.

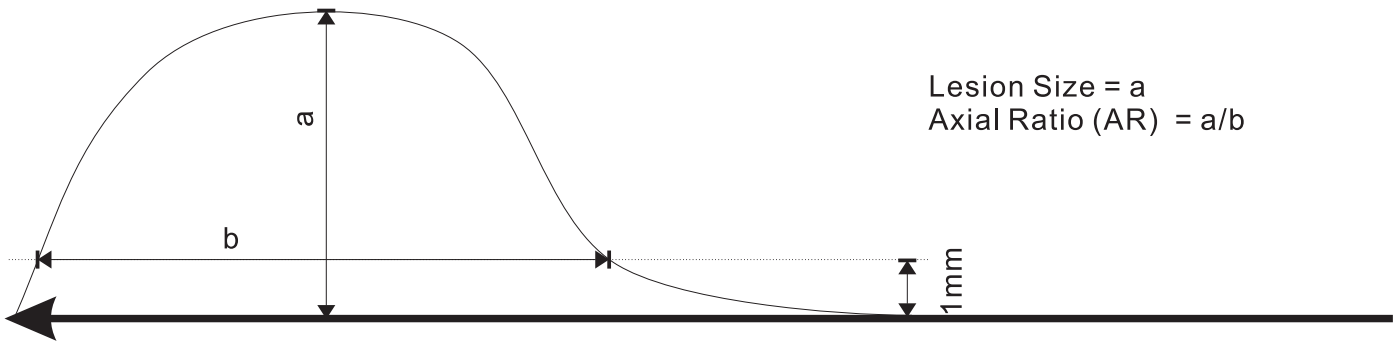
Design a coaxial antenna for hepatic tumor ablation





Simulation of the electromagnetic radiation profile

Finite element models (MultiPhysics v3.2) are used to generate the electromagnetic (EM) radiation fields in liver given a particular design



Metric	Measure of	Goal
Lesion radius	Size of lesion in radial direction	Maximize
Axial ratio	Proximity of lesion shape to a sphere	Fit to 0.5
S_{11}	Tail reflection of antenna	Minimize

A general problem formulation

- We formulate the simulation-based optimization problem as

$$\min_{x \in \mathcal{S}} F(x) = \mathbb{E}_{\omega}[f(x, \omega(x))], \tag{1}$$

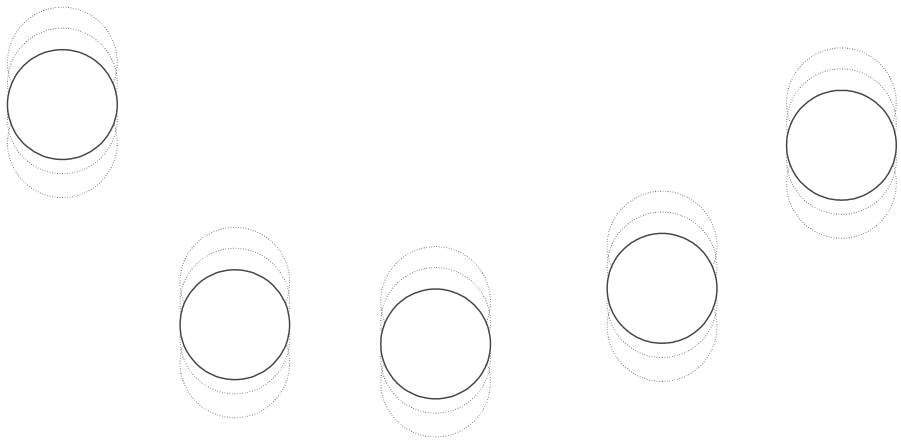
where $\omega(x)$ is a random factor arising in the simulation process.
The sample response function $f(x, \omega)$

  - Typically does not have a closed form, thus cannot provide gradient or Hessian information
  - Is normally computationally expensive
  - Is affected by uncertain factors in simulation

The underlying objective function $F(x)$ has to be estimated.

The discrete optimization case

- For example, test elasticity of a set of balls. Here $\mathcal{S} = \{1, 2, 3, 4, 5\}$ represents a set of 5 balls.



- Objective: Choose the ball with the largest expected bounce height $F(x_i)$. $f(x_i, \omega_j)$ corresponds to a single measurement in an experiment.

How to select the best system

- Choose the maximum sample mean

$$\arg\max_{i \in \mathcal{S}} \bar{\mu}_i := \frac{1}{N_i} \sum_{j=1}^{N_i} f(x_i, \omega_j), \tag{2}$$

  where N_i is the number of experiments.
- Select the best system with high accuracy, while controlling the total amount of simulation runs.
- Two approaches
  - Ranking and selection
    S.-H. Kim and B. L. Nelson, "Selecting the Best System: Theory and Methods."
  - Bayesian approach
    S. E. Chick, and K. Inoue, "New Two-stage and Sequential Procedures for Selecting the Best Simulated System."
    H.-C. Chen, C.-H. Chen, and E. Yucesan, "An Asymptotic Allocation for Simultaneous Simulation Experiments."

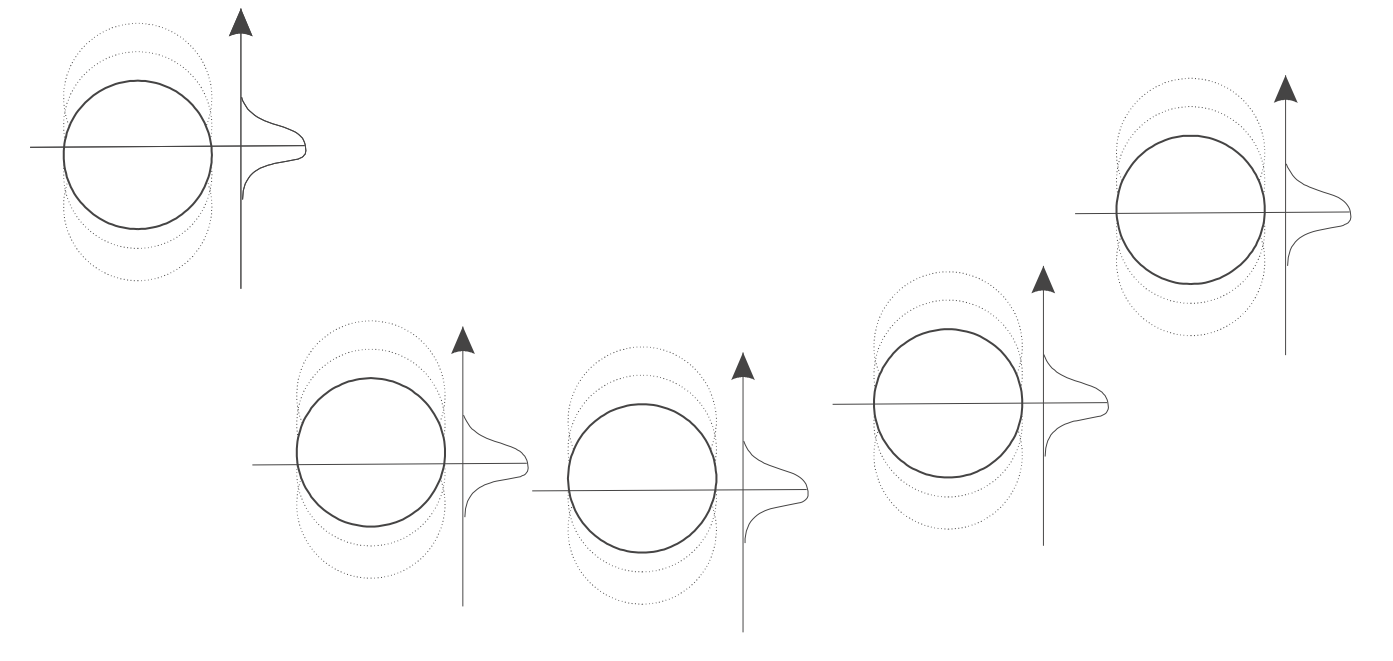
Bayesian approach

- Denote the mean of the simulation output for each system as $\mu_i = F(x_i) = \mathbb{E}_\omega[f(x_i, \omega)]$
- In Bayesian perspective, the means are considered as Gaussian random variables whose posterior distributions can be estimated as

$$\mu_i | X \sim N(\bar{\mu}_i, \hat{\sigma}_i^2 / N_i) \tag{3}$$

where $\bar{\mu}_i$ is sample mean and $\hat{\sigma}_i^2$ is sample variance. The above formulation is one type of posterior distributions.

Posterior distributions facilitate comparison



Now it is easy to compute the probability of correct selection (PCS).

Compute the PCS

- Pairwise comparison

$$PCS = Pr(\mu_1 \geq \mu_2) \sim Pr(\mu_1 \geq \mu_2 | X) = Pr(\mu_1 | X - \mu_2 | X \geq 0). \quad (4)$$

- Multiple comparisons (Bonferroni inequality):

$$\begin{aligned} PCS &= Pr(\mu_b - \mu_i \geq 0, i = \{1, 2, \cdots, K\} \setminus \{b\}) \\ &\sim 1 - \sum_{i=1, i \neq b}^{K} Pr(\mu_b - \mu_i < 0) \end{aligned} . \quad (5)$$

Summary of the Bayesian approach

- Once the PCS is determined, future work is to choose the suitable sample number of each system N_i such that the best system is selected with desired accuracy

$$PCS \geq 1 - \alpha.$$

- Bayesian approach
  - Utilizes both mean and variance information
  - Simple and direct to implement
  - Without using indifference-zone parameter δ

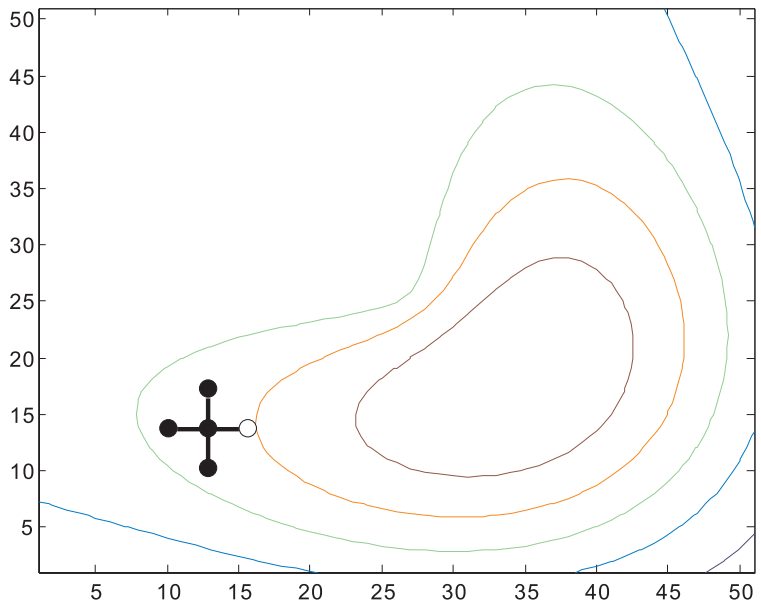
Unconstrained continuous optimization case

$\mathcal{S} = \mathbb{R}^n$

- Basic approach: reduce function uncertainty by averaging multiple samples per point, which is similar to the discrete case.
- Potential difficulty:
  efficiency of algorithm VS number of simulation runs
- We apply Bayesian approach to determine appropriate number of samples per point, while simultaneously enhancing the algorithm efficiency
- Guarantee the global convergence of the algorithm

A noisy extension of compass search

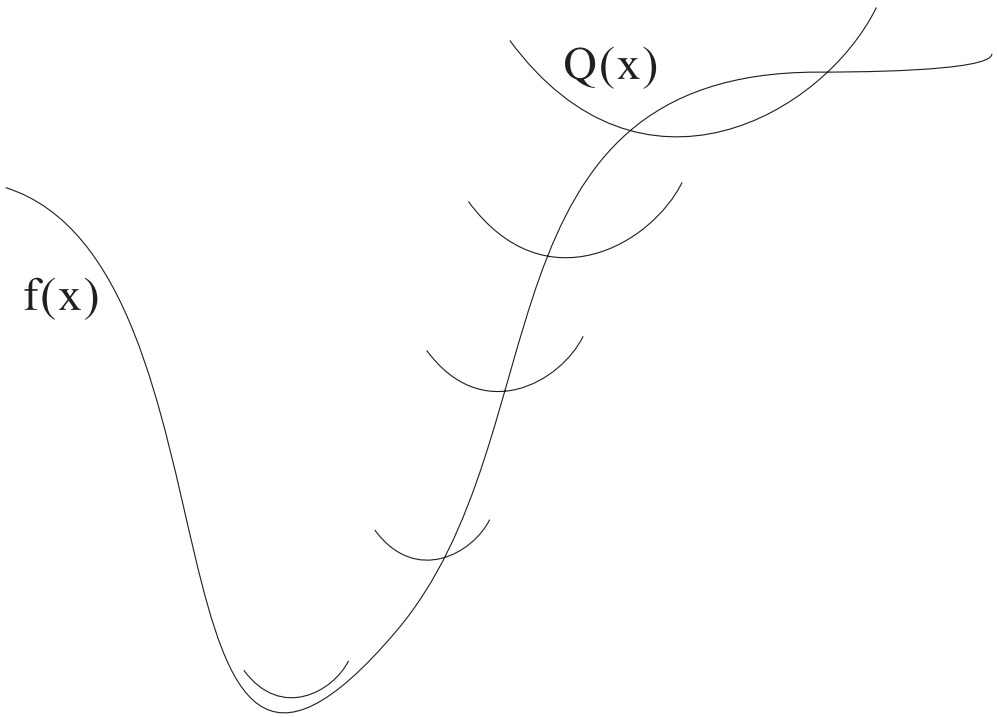
- Direct search methods do not attempt to make gradient estimates.
- Compass search is one type of the direct search methods.



- Determine the next iterate by direct comparisons. (Selecting the best system!)

A noisy extension of the UOBYQA algorithm

The base derivative free optimization algorithm: The UOBYQA (Unconstrained Optimization BY Quadratic Approximation) algorithm is based on a trust region method. It constructs a series of local quadratic approximation models of the underlying function.



Quadratic model construction and solve trust region subproblem

For iteration $k = 1, 2, \ldots,$

- $\ldots$
- construct a quadratic model via interpolation

$$Q(x,\omega) = f(x_k,\omega) + g_Q^T(\omega)(x - x_k) + \frac{1}{2}(x - x_k)^T G_Q(\omega)(x - x_k) \tag{6}$$

  The model is unstable interpolating noisy data

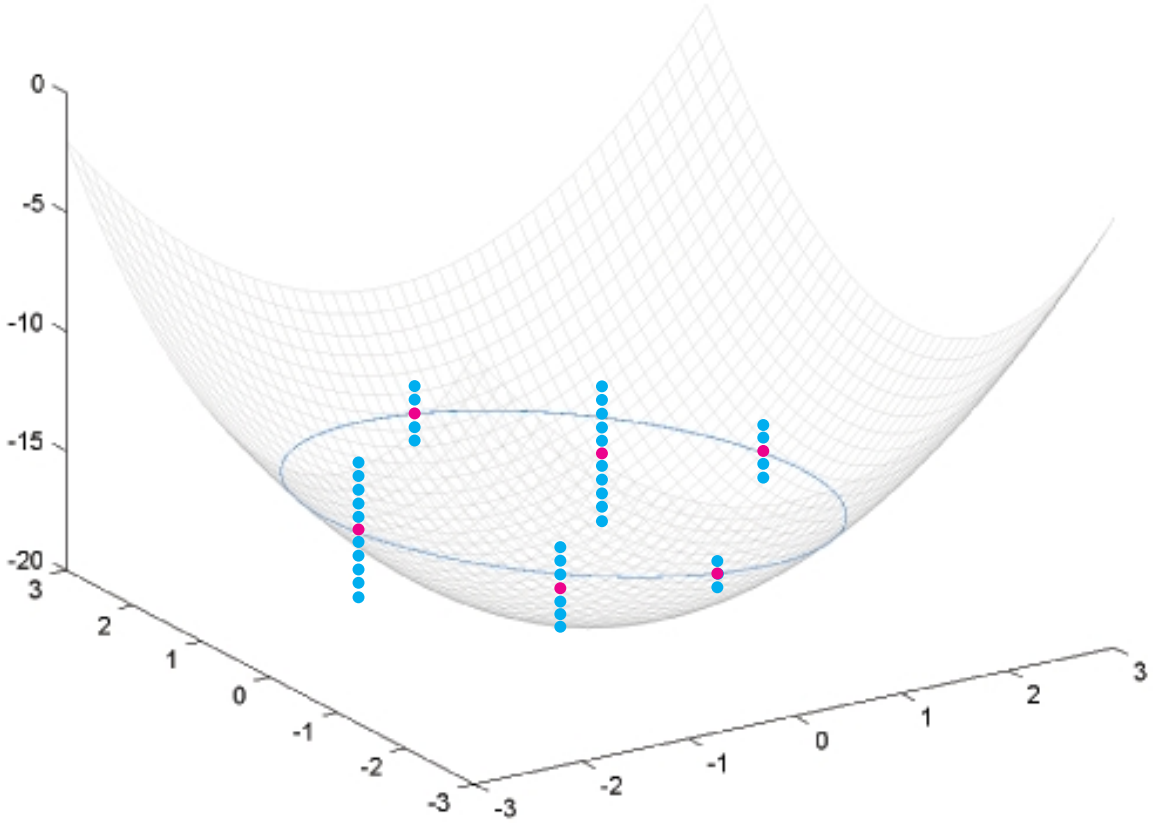
- Solve trust region subproblem

$$\begin{aligned} s_k(\omega) = \arg\min_s \quad & Q(x_k + s, \omega) \\ s.t. \quad & \|s\|_2 \leq \Delta_k \end{aligned} \tag{7}$$

  The solution is thus unstable

- $\ldots$

Why is the quadratic model unstable?



How to stabilize the quadratic model?

Let $\mathcal{I} = \{y^1, y^2, \ldots, y^L\}$ be the interpolation set.

- Quadratic interpolation model is a linear combination of Lagrange functions:

$$Q(x, \omega) = \sum_{j=1}^{L} f(y^j, \omega) l_j(x). \tag{8}$$

- Each piece $l_j(x)$ is a quadratic polynomial, satisfying

$$l_j(y^i) = \delta_{ij}, i = 1, 2, \cdots, L.$$

- The coefficients of l_j are uniquely determined, regardless of the random objective function.

Bayesian estimation of coefficients c_Q, g_Q, G_Q

In Bayesian approach, the mean of function output $\mu(y^j) := \mathbb{E}_\omega f(y^j, \omega)$ is considered as a random variable:
Normal posterior distributions:

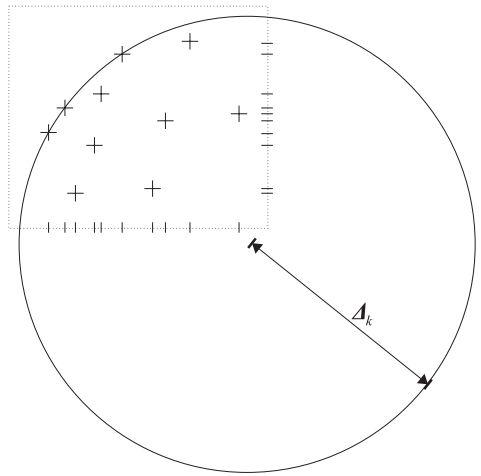
$$\mu(y^j)|X \sim N(\bar{\mu}(y^j), \hat{\sigma}^2(y^j)/N_j). \tag{9}$$

Thus the coefficients of the quadratic model are estimated as:

$$\begin{array}{rcl} g_Q|X & = & \sum_{j=1}^{L}(\mu(y^j)|X)g_j, \\ G_Q|X & = & \sum_{j=1}^{L}(\mu(y^j)|X)G_j. \end{array} \tag{10}$$

- g_j, G_j are coefficients of Lagrange functions l_j.
- g_j, G_j are deterministic and determined by points y^j.

Constraining the variance of coefficients



- Trial solutions are generated within a trust region. The standard deviation of the solutions are constrained.

$$\max_{i=1}^{n} std([s^{*(1)}(i), s^{*(2)}(i), \cdots, s^{*(M)}(i)]) \leq \beta\Delta_k. \qquad (11)$$

Other approaches to constrain the variance of coefficients

- Test the sufficient reduction criterion

$$Pr\left(Q_k(x_k) - Q_k(x_k + s^*) \geq \kappa_{mdc}\|g_k^\infty\| \min\left[\frac{\|g_k^\infty\|}{\kappa_{Qh}}, \Delta_k\right]\right) \geq 1-\alpha \tag{12}$$

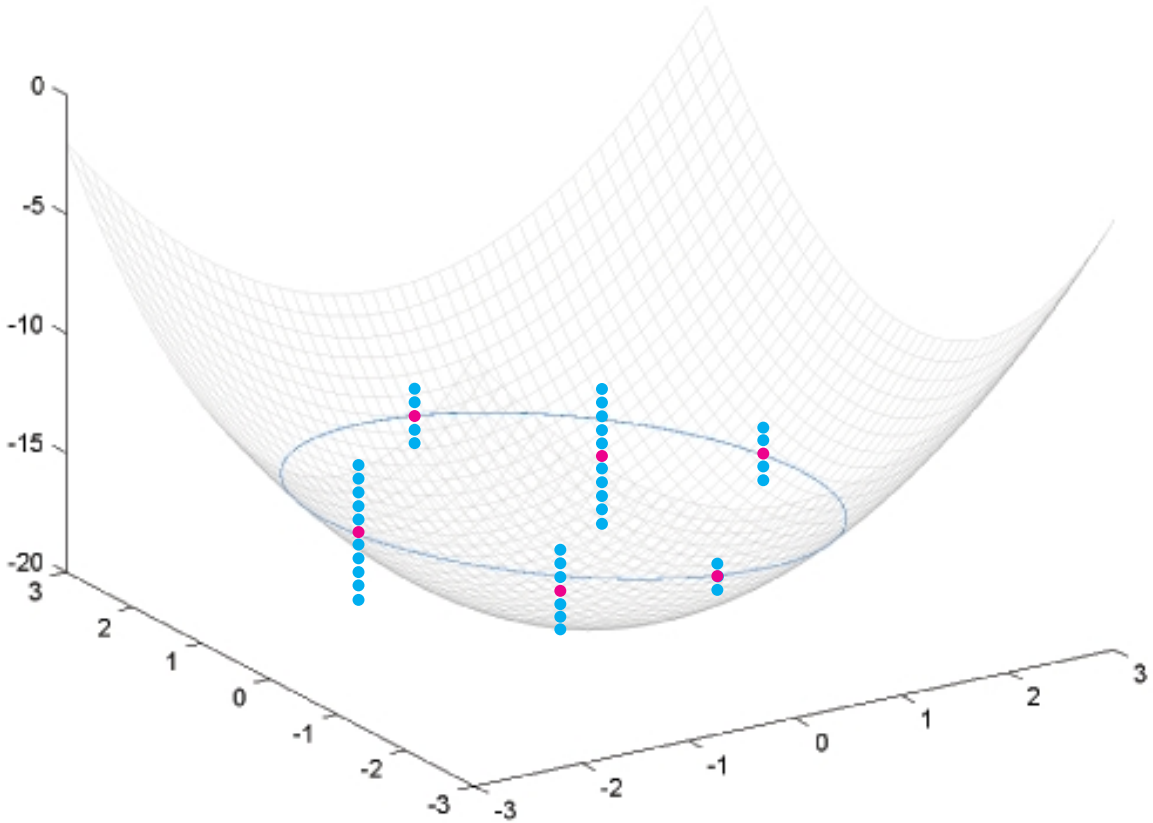
- Quantify variance of individual coefficient in Q:

$$\frac{std(g_Q(i'))}{E[g_Q(i')]} \leq \beta, i' = 1, \cdots, n \tag{13}$$

$$\frac{std(G_Q(i', j'))}{E[G_Q(i', j')]} \leq \beta, i', j' = 1, \cdots, n \tag{14}$$

Optimally allocating computing resources

Select appropriate N_j for the point y^j in the interpolation set



A numerical test

Table: The performance of Noisy UOBYQA for the Rosenbrock function, with $n = 2$ and $\sigma^2 = 0.01$.

Iteration (k)	FN	$F(x_k)$	Δ_k
1	1	404	2
20	78	3.56	9.8×10^{-1}
40	140	0.75	1.2×10^{-1}
60	580	0.10	4.5×10^{-2}
80	786	0.0017	5.2×10^{-3}
100	1254	0.0019	2.8×10^{-4}
120	2003	0.0016	1.1×10^{-4}
✓ Stops here with the termination criterion $\Delta_k \leq 10^{-4}$			

Thank you!