Sequential Experimental Designs for Kriging Model

Yongdao Zhou

June 17, 2026

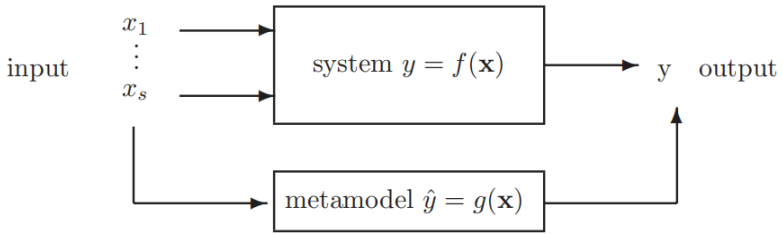
Outline

Outline

Introduction

Computer experiments have been widely used in industry, aerospace and other fields due to their high efficiency.

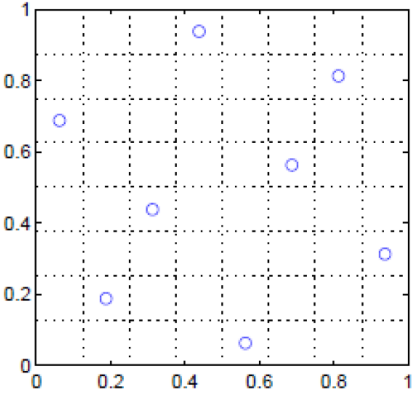
- Computer experiments: samples with the same input setting will produce identical outputs.
- Scientists and engineers make use of computer simulations to explore the complex relationship between inputs and outputs.



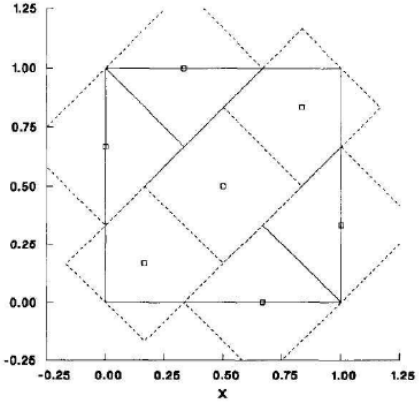
Introduction

- One of the goals of computer experiments is to find an approximate model that is much simpler than the true (but complicated) model by simulating the behavior on a computer.
- The good properties of Kriging model:
  - **Exact interpolation**: Perfectly fits all observed points.
  - **Uncertainty quantification**: Provides not only a prediction but also a predictive variance at any unobserved point.
- Use space-filling design to gather the most information, so that the model can ultimately achieve the goal of global fitting.
  - Latin hypercube design
  - Maximin distance design
  - Uniform design

Space-filling design

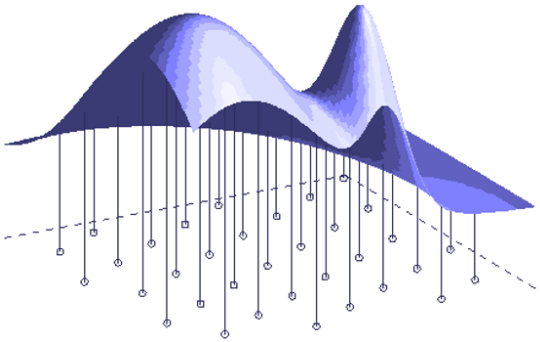


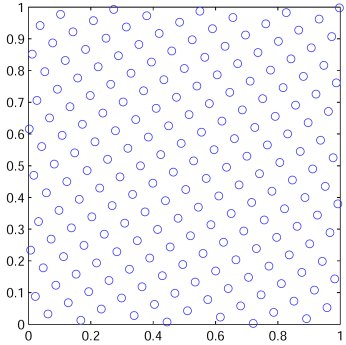
Latin hypercube design



maximin distance design

Uniform design





Theory of uniform design:

- Uniformity criteria (Fang and Wang (1994), Zhou and Xu (2014, JASA)
- Construction methods (Zhou and Xu (2015, Biometrika), Zhou and Xu (2017, JASA), Fang, Liu, Qin and Zhou (2018, Springer))

Kriging model

Kriging model is a Gaussian process regression:

$$y(\boldsymbol{x}) = f(\boldsymbol{x})^T \beta + z(\boldsymbol{x}),$$

where $z(\boldsymbol{x})$ is the zero-mean stationary Gaussian process with the covariance structure

$$Cov(Z(\boldsymbol{x}), Z(\boldsymbol{y})) = \tau^2 k(\boldsymbol{x}, \boldsymbol{y}),$$

and τ^2 represents the variance.

- Gaussian separable kernel:

$$k(\boldsymbol{x}, \boldsymbol{y}; \theta) = \exp\left\{ -\sum_{t=1}^{m} \theta_t |x_t - y_t|^2 \right\} + g\delta(\boldsymbol{x}, \boldsymbol{y}).$$

- Matérn correlation function:

$$k(\boldsymbol{x}, \boldsymbol{y}; \nu, \ell) = \frac{1}{\Gamma(\nu) 2^{\nu-1}} (2\sqrt{\nu}\ell \|\boldsymbol{x} - \boldsymbol{y}\|)^{\nu} K_{\nu}(2\sqrt{\nu}\ell \|\boldsymbol{x} - \boldsymbol{y}\|).$$

Kriging model

According to the observation points $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ and their observation values $\boldsymbol{Y}_{\boldsymbol{X}} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))$, we can compute the prediction and its variance at an unobserved point $\boldsymbol{x}$.

- Prediction: $\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) = \boldsymbol{r}_{\boldsymbol{X}}^T(\boldsymbol{x})\boldsymbol{K}^{-1}(\boldsymbol{X})\boldsymbol{Y}_{\boldsymbol{X}}$;
- Variance: $\widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x}) = \tau^2 \left(k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{r}_{\boldsymbol{X}}^T(\boldsymbol{x})\boldsymbol{K}^{-1}(\boldsymbol{X})\boldsymbol{r}_{\boldsymbol{X}}(\boldsymbol{x})\right)$.

where $\boldsymbol{r}_{\boldsymbol{X}}(\boldsymbol{x}) = (k(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}, \boldsymbol{x}_n))^T$ and $\boldsymbol{K}(\boldsymbol{X}) = (k(\boldsymbol{x}_i, \boldsymbol{x}_j))_{n\times n}$.

Introduction

However, single-stage experiments (i.e., all experimental points are predetermined in a single step) have the following limitations.

- Relies entirely on a priori assumptions about the experimental domain and the true model.
- Cannot dynamically adjust observation points based on existing information.
- May lead to poor model fitting if key regions are omitted.

Therefore, a sequential experimental design strategy is necessary.

Sequential design framework

Algorithm 1 Sequential design for Kriging model

1: Select a space-filling design as the initial design $\boldsymbol{X}_0$.
2: $i \leftarrow 0$.
3: **while** termination condition is not satisfied **do**
4: $\quad \boldsymbol{Y}_{\boldsymbol{X}_i} \leftarrow$ the observed values of $\boldsymbol{X}_i$.
5: $\quad$ Select $\boldsymbol{B}_{i+1}$ according to the criterion ϕ.
6: $\quad \boldsymbol{X}_{i+1} \leftarrow (\boldsymbol{X}_i, \boldsymbol{B}_{i+1})$.
7: $\quad i \leftarrow i + 1$.
8: **end while**
9: Fit the Kriging model using the final $\boldsymbol{X}_i$ and $\boldsymbol{Y}_{\boldsymbol{X}_i}$.

Existing sequential design criteria

For a given criterion ϕ, the next point $\widetilde{\boldsymbol{x}}$ to be added is the point that maximizes $\phi(\boldsymbol{x})$ in the experimental region $\mathcal{X}$, that is,

$$\widetilde{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \phi(\boldsymbol{x}).$$

- ϕ_{MD} is the negative of mixture discrepancy (MD)[1];
- $\phi_s(\boldsymbol{x}) = \widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x})$;
- $\phi_{EI_0}(\boldsymbol{x}) = (\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*))^2 + \widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x})$;
- $\phi_{EI_1}(\boldsymbol{x}) = \left(\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*) - \nabla^T f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)\right)^2 + \widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x})$;
- $\phi_{EI_2}(\boldsymbol{x}) = \left(\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*) - \nabla^T f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)\right)^2 \times d(\boldsymbol{x},\boldsymbol{x}^*) + \widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x})$.

[1] Zhou, Fang and Ning (2013, J Complexity)

Introduction

Existing sequential designs are mainly divided into two categories.

- Pointwise sequential design: observe **one point** at a time.
  - Advantages: more sufficient information utilization and less likely to fall into local optima.
  - Disadvantages: more iterations and higher resource consumption.
  - limitations of current researches: most existing methods lack theoretical guarantees.
- Batch sequential design: observe **multiple points** at a time.
  - Advantages: significantly reduces the number of experimental runs and improves resource utilization efficiency.
  - Disadvantages: prone to falling into local optima.
  - limitations of current researches:
    - Observation-based methods are relatively scarce.
    - There is no framework that bridges pointwise sequential design and batch sequential design.

Introduction

Then, we consider the following improvement.

- Two novel pointwise sequential criteria are proposed, along with corresponding theoretical results.
- A general batch sequential design framework is presented, which can extend any pointwise sequential criterion to batch scenarios.
- The effectiveness and robustness of the proposed methods are verified through simulations on multiple test functions.

Outline

Gradient-based criterion: motivation

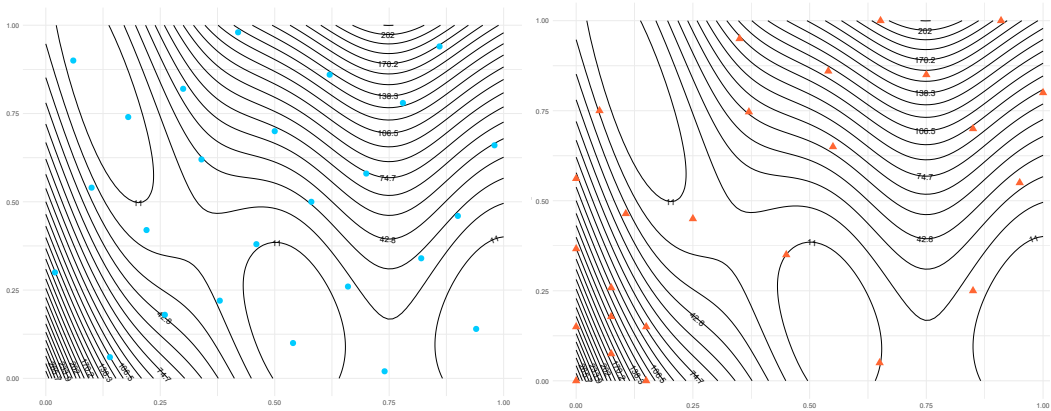
- Consider the Branin function as an example:

$$f(x_1, x_2) = \left(15x_2 - \frac{5.1}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 10.$$

- The kernel function is the Gaussian separable kernel.
- Two methods: uniform design based on mixture discrepancy and gradient-based sequential design.
- Evaluation metrics: root mean square error (RMSE).

$$\mathrm{RMSE}(\boldsymbol{X}) = \sqrt{\frac{\sum_{\boldsymbol{x} \in \boldsymbol{X}_{test}} \left(f(\boldsymbol{x}) - \hat{y}(\boldsymbol{x})\right)^2}{n}}.$$

Gradient-based criterion: motivation



(a) Uniform design

(b) Gradient-based sequential design.

Gradient-based sequential design is dense in regions with large gradients and sparse in regions with small gradients, and maintain overall uniformity across the entire design space

Gradient-based criterion: motivation

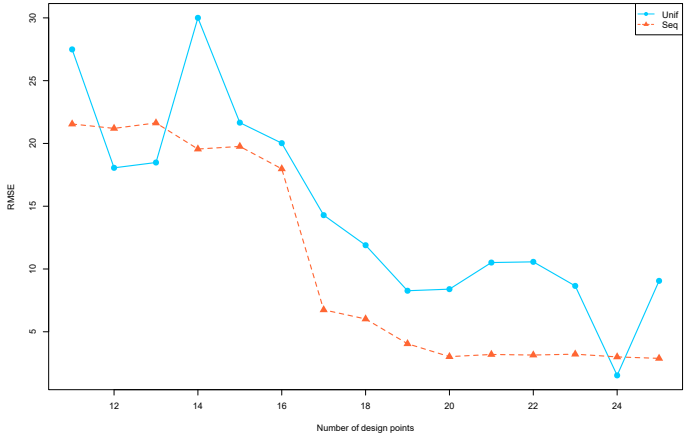


Figure: The RMSE of the uniform design and the gradient-based sequential design with different number of design points.

Gradient-based criterion: theoretical results

The following theorem shows that the upper bound of the pointwise error is controlled by ϕ_{gra}.

Theorem

Let $f(\boldsymbol{x})$ be the smooth true model and $\hat{y}_{\boldsymbol{X}}(\boldsymbol{x})$ be the predicted value of the Kriging model at point $\boldsymbol{x}$ based on $\boldsymbol{X}$. Then for any $\boldsymbol{x} \in \mathcal{X}$, we have

$$|f(\boldsymbol{x}) - \hat{y}_{\boldsymbol{X}}(\boldsymbol{x})| \leq \|\nabla f(\boldsymbol{t})\| \times d(\boldsymbol{x}, \boldsymbol{x}^*) + |\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*)|,$$

where $\boldsymbol{x}^* = \arg\min_{\boldsymbol{y} \in \boldsymbol{X}} d(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{t} \in \{\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{x}^* \,|\, 0 \leq \alpha \leq 1\}$.

Gradient-based criterion: expression

Gradient-based criterion:

$$\phi_{gra}(\boldsymbol{x}) = \sqrt{E(\|\nabla f(\boldsymbol{x})\|^2)} \times d(\boldsymbol{x}, \boldsymbol{x}^*) + |\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*)|,$$

where $\boldsymbol{x}^*$ is the point in the existing design that is closest to $\boldsymbol{x}$.

- $\sqrt{E(\|\nabla f(\boldsymbol{x})\|^2)}$: add points where gradient is large.
- $d(\boldsymbol{x}, \boldsymbol{x}^*)$: keep all points uniformly distributed.
- $|\hat{y}_{\boldsymbol{X}}(\boldsymbol{x}) - f(\boldsymbol{x}^*)|$: add points where variation is large.

Variance-based criterion: motivation

- Sacks and Schiller (1988) proposed a criterion that selects the point minimizing the maximum prediction variance $\widehat{s^2_{\boldsymbol{X}}}(\boldsymbol{x})$.
- Relying solely on the prediction variance of design point estimates is insufficient to fully characterize the overall estimation error.
- The estimation error is also governed by the variance of the gradient estimate.

Variance-based criterion: theoretical results

Theorem

Let $\mathcal{X} \subseteq \mathbb{R}^m$ be open, $\mathcal{N}_k(\mathcal{X})$ be the native Hilbert space corresponding to the symmetric positive definite kernel $k(\boldsymbol{x}, \boldsymbol{y}) \in C^{2m}(\mathcal{X}, \mathcal{X}) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and f be the smooth true function. Denote the i-th diagonal element of $\nabla^2 k(\boldsymbol{x}, \boldsymbol{x}) - \nabla r_{\boldsymbol{X}}^T(\boldsymbol{x}) \boldsymbol{K}^{-1}(\boldsymbol{X}) \nabla r_{\boldsymbol{X}}(\boldsymbol{x})$ as $g_i(\boldsymbol{x})$. Then we have

- $|f(\boldsymbol{x}) - \hat{y}_{\boldsymbol{X}}(\boldsymbol{x})| \leq \sqrt{k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{r}_{\boldsymbol{X}}^T(\boldsymbol{x}) \boldsymbol{K}^{-1}(\boldsymbol{X}) \boldsymbol{r}_{\boldsymbol{X}}(\boldsymbol{x})} \times |f|_{\mathcal{N}_k(\mathcal{X})}$,
- $|f(\boldsymbol{x}) - \hat{y}_{\boldsymbol{X}}(\boldsymbol{x})| \leq \sqrt{\sum_{i=1}^m g_i(\boldsymbol{t})} \times d(\boldsymbol{x}, \boldsymbol{x}^*) \times |f|_{\mathcal{N}_k(\mathcal{X})}$,

where $\boldsymbol{t} \in \{\alpha \boldsymbol{x} + (1 - \alpha) \boldsymbol{x}^* \mid 0 \leq \alpha \leq 1\}$ and $|f|_{\mathcal{N}_k(\mathcal{X})}$ is the norm of function f in the native space $\mathcal{N}_k(\mathcal{X})$.

Variance-based criterion: expression

Thus, we present the specific expression for the variance-based criterion as

$$\phi_{var}(\boldsymbol{x}) = \min\left(\sqrt{k(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{r}_{\boldsymbol{X}}^T(\boldsymbol{x})\boldsymbol{K}^{-1}(\boldsymbol{X})\boldsymbol{r}_{\boldsymbol{X}}(\boldsymbol{x})}, \sqrt{\sum_{i=1}^{m} g_i(\boldsymbol{x})} \times d(\boldsymbol{x},\boldsymbol{x}^*)\right).$$

When the number of current design points is n and the dimension is m, the computational complexity of both ϕ_{gra} and ϕ_{var} is $O(n^3 + mn^2)$.

Outline

Cluster-based top-b sequential design: motivation

- Simply selecting the b points that maximize $\phi(\boldsymbol{x})$ would inevitably cluster around the global maximum point, leading to the inherent problem of point clustering.
- In fact, each iteration yields more information beyond the maximum point.
- The current local maximum could become the global maximum in the next iteration.

Idea: Apply clustering to select multiple local maximum points in each iteration.

Cluster-based top-b sequential design: algorithm

Algorithm 2 The selection of $\boldsymbol{B}_i$ in top-b sequential design

Input: $\boldsymbol{D}_{all}$, $\phi(\boldsymbol{x})$, b, α, β
Sort the points in $\boldsymbol{D}_{all}$ by their ϕ values in descending order and denoted them as $\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n_{all})}$
$Cluster = \{\{\boldsymbol{x}_{(1)}\}\}$
for $i = 2$ **to** n_{all} **do**
    if the number of clusters in $Cluster$ is b **then**
        break
    end if
    for each cluster C in $Cluster$ **do**
        if the number of points in C is greater than 1 **then**
            $\bar{d} \leftarrow$ the average distance of each point in C to the cluster centroid
            $d \leftarrow$ the distance between point $\boldsymbol{x}$ and the centroid
            if $d < \beta\bar{d}$ **then**
                Assign $\boldsymbol{x}_{(i)}$ to cluster C
                break
            end if
        end if
        if the number of points in C is 1 **then**
            Denote the point in C as $\boldsymbol{x}_C$
            $num \leftarrow$ the number of points in the cube formed by $\boldsymbol{x}_C$ and $\boldsymbol{x}$
            if $num \leq \alpha$ **then**
                Assign $\boldsymbol{x}_{(i)}$ to cluster C
                break
            end if
        end if
    end for
    Add $\boldsymbol{x}_{(i)}$ as a new cluster to $Cluster$.
end for
Output: the first point of each cluster in $Cluster$

Cluster-based top-b sequential design: algorithm

- $\boldsymbol{D}_{all}$: uniform design for discretizing the design space.
- α and β govern the cluster assignment of candidate points.
  - Small value: produces overly small clusters and leads to point clustering.
  - Large value: may result in excessively large clusters that hinder local optimum selection.
  - The sensitivity analysis shows that $\alpha = 10$ and $\beta = 5$ perform favorably.
- For the choice of the parameter b, it generally depends on the practical scenario.
  - Large b: reduces the number of iterations and thus saves computational resources.
  - Small b: allows more precise selection of observation points and improves model accuracy.
  - In general, values of b ranging from 2 to 5 are suitable.

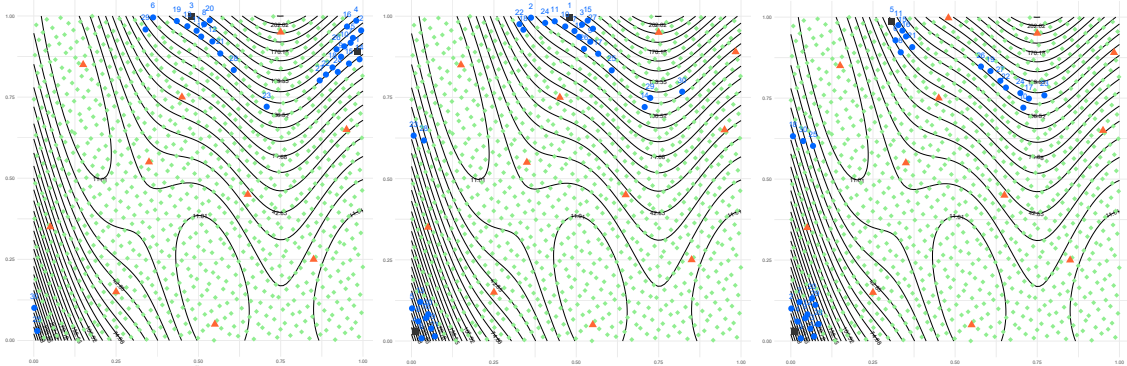
Cluster-based top-b sequential design: example

- The true response model is the Branin function.
- Discretize $\mathcal{X} = [0, 1]^2$ using a uniform design $\boldsymbol{D}_{all}$ with 1000 points, and select $\boldsymbol{B}_i$ from $\boldsymbol{D}_{all}$.

The following three figures show the pointwise sequential designs corresponding to the cases where the numbers of points are 10, 11, and 12, respectively.

- Red triangles: current design $\boldsymbol{X}_i$.
- Green diamonds: points in $\boldsymbol{D}_{all}$.
- Blue circles: top 30 points in $\boldsymbol{D}_{all}$ that maximize $\phi_{gra}(\boldsymbol{x})$.
- Black squares: the points selected by clustering $\boldsymbol{D}_{all}$ and choosing the maximum ϕ_{gra} point in the first two clusters (Algorithm 2).

Cluster-based top-b sequential design: example



Cluster-based top-b sequential design: example

- The points maximizing ϕ_{gra} tend to exhibit a clustering pattern.
- The point labeled 3 in the first figure becomes the new maximizer of $\phi_{gra}(\boldsymbol{x})$ in the second figure, and so on.
- Thus, we can directly select both the point labeled 1 and the point labeled 3 for simultaneous observation in the first iteration.
- This batch design selection strategy not only reduces the cost of additional iterative observations but also yields outcomes that are comparable to those of the pointwise sequential design.

Outline

Test functions

- f_1: the Branin function.
- $f_2(x_1, x_2) = \left(1 - \exp\left(-\frac{0.5}{x_2}\right)\right) \frac{2300x_1^3+1900x_1^2+2092x_1+60}{100x_1^3+500x_1^2+4x_1+20}$.
- $f_3(x_1, x_2) = \log(2 + \sin(2\pi x_1)) \cos(2\pi x_2^2)$.
- Hartman function of dimension three:
  $f_4(x_1, x_2, x_3) = -\sum_{i=1}^{4} c_i \exp\left\{-\sum_{j=1}^{3} \alpha_{ij}(x_j - p_{ij})^2\right\}$.
- Ackley function of dimension five:

$$f_5(\boldsymbol{x}) = -20 \exp\left(-0.2\sqrt{\frac{1}{5}\sum_{i=1}^{5} x_i^2}\right) - \exp\left(\frac{1}{5}\sum_{i=1}^{5} \cos(2\pi x_i)\right) + 20 + e.$$

- Zakharov function: $f_6(\boldsymbol{x}) = \sum_{i=1}^{m} x_i^2 + \left(\sum_{i=1}^{m} 0.5 i x_i\right)^2 + \left(\sum_{i=1}^{m} 0.5 i x_i\right)^4$.
- Rosenbrock function: $f_7(\boldsymbol{x}) = \sum_{i=1}^{m-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]$.

Comparison results: MAE of pointwise sequential designs

We compare our methods with 5 existing criteria on 7 test functions with dimensions ranging from 2 to 8 under maximum absolute error (MAE).

Table 1: The MAE of different pointwise sequential designs.

Test function	m	ϕ_{EI_0}	ϕ_{EI_1}	ϕ_{EI_2}	ϕ_s	ϕ_{MD}	ϕ_{gra}	ϕ_{var}
f_1	2	49.61	9.85	3.07	3.79	21.30	26.02	**2.05**
f_2	2	2.41	0.99	0.82	1.24	3.26	**0.75**	0.97
f_3	2	0.84	0.54	0.26	0.25	0.37	0.99	**0.21**
f_4	3	1.00	1.17	1.27	1.27	0.96	0.80	**0.80**
f_5	5	14.49	13.94	14.05	14.65	13.91	14.05	**13.69**
f_6	3	1.90	2.09	1.15	0.35	1.17	2.60	**0.28**
f_6	5	235.12	298.32	**112.92**	213.95	426.67	228.48	186.25
f_6	8	14565	14359	10203	25987	25124	**9478**	11158
f_7	3	6.92	2.95	2.21	1.44	6.86	6.23	**1.41**
f_7	5	91.53	163.05	153.94	61.67	147.37	104.37	**58.59**
f_7	8	235.79	248.43	204.63	179.14	278.62	255.74	**177.21**

Comparison results: RMSE of pointwise sequential designs

Table 2: The RMSE of different pointwise sequential designs.

Test function	m	ϕ_{EI_0}	ϕ_{EI_1}	ϕ_{EI_2}	ϕ_s	ϕ_{MD}	ϕ_{gra}	ϕ_{var}
f_1	2	4.49	0.72	0.72	1.11	1.70	**0.52**	0.65
f_2	2	0.43	0.26	0.25	0.35	0.30	**0.21**	0.23
f_3	2	0.11	0.09	0.07	0.07	0.08	**0.06**	0.06
f_4	3	0.20	0.21	0.23	0.26	0.20	**0.17**	0.31
f_5	5	0.91	0.81	0.82	0.77	0.81	0.81	**0.77**
f_6	3	0.18	0.15	0.10	0.08	0.15	0.22	**0.06**
f_6	5	24.05	25.32	25.06	28.83	24.92	**23.11**	35.49
f_6	8	1637.46	1611.18	**1162.48**	2103.61	1349.59	1207.69	1487.18
f_7	3	0.76	0.42	0.30	0.28	0.59	**0.24**	0.30
f_7	5	16.06	19.92	15.19	**13.18**	17.71	15.68	13.25
f_7	8	37.11	42.03	35.40	30.97	47.88	**29.61**	30.34

Comparison results: MAE of batch sequential designs

Table 3: The MAE of different batch sequential designs.

Test function	m	b	ϕ_{EI_0}	ϕ_{EI_1}	ϕ_{EI_2}	ϕ_s	ϕ_{MD}	ϕ_{gra}	ϕ_{var}
f_1	2	2	4.73	5.63	2.56	3.41	11.41	4.92	**1.64**
f_1	2	3	4.88	1.94	1.28	0.77	4.18	4.49	**0.72**
f_1	2	5	5.39	0.36	0.30	**0.08**	2.66	8.00	0.11
f_2	2	2	1.95	1.08	1.03	0.85	4.32	0.90	**0.83**
f_2	2	3	1.33	0.99	0.42	0.39	1.39	1.66	**0.38**
f_2	2	5	0.56	0.50	0.32	0.37	0.66	0.69	**0.21**
f_3	2	2	0.62	0.65	0.30	0.25	1.09	0.68	**0.21**
f_3	2	3	0.31	0.15	0.23	**0.09**	0.25	0.41	0.11
f_3	2	5	0.18	0.07	0.10	0.06	0.31	0.12	**0.05**
f_4	3	2	1.27	1.19	**0.84**	1.28	1.11	1.21	1.49
f_4	3	3	1.20	1.19	0.87	1.00	0.90	1.18	**0.84**
f_4	3	5	1.00	0.94	0.87	**0.54**	0.82	1.18	1.01
f_5	5	2	13.30	12.66	12.75	13.15	12.78	12.71	**11.15**
f_5	5	3	**11.38**	12.86	12.78	12.68	12.68	12.13	12.89
f_5	5	5	12.61	12.92	12.68	12.79	12.55	12.11	**11.87**

Comparison results: RMSE of batch sequential designs

Table 4: The RMSE of different batch sequential designs.

Test function	m	b	ϕ_{EI_0}	ϕ_{EI_1}	ϕ_{EI_2}	ϕ_s	ϕ_{MD}	ϕ_{gra}	ϕ_{var}
f_1	2	2	0.61	1.26	0.50	7.92	1.62	**0.40**	0.43
f_1	2	3	0.45	0.24	0.18	0.23	0.42	**0.14**	0.18
f_1	2	5	0.39	0.03	0.03	0.02	0.15	0.07	**0.02**
f_2	2	2	0.33	0.23	0.25	0.26	0.28	**0.23**	0.26
f_2	2	3	0.25	0.18	0.14	0.12	0.24	0.33	**0.12**
f_2	2	5	0.15	0.10	0.06	0.06	0.08	**0.05**	0.06
f_3	2	2	0.11	0.09	0.08	0.08	0.11	**0.06**	0.07
f_3	2	3	0.083	0.042	0.028	0.029	0.032	0.070	**0.028**
f_3	2	5	0.026	0.011	0.009	0.014	0.019	0.029	**0.009**
f_4	3	2	0.21	0.21	0.26	0.27	0.21	**0.20**	0.29
f_4	3	3	0.21	0.23	0.14	0.20	0.14	**0.12**	0.15
f_4	3	5	0.20	0.13	0.16	0.10	0.11	**0.10**	0.14
f_5	5	2	1.28	0.83	**0.81**	0.84	0.81	0.82	0.84
f_5	5	3	0.90	0.81	0.81	0.75	0.77	**0.73**	0.79
f_5	5	5	0.74	0.73	0.75	0.66	0.67	**0.63**	0.69

Comparison results

- For the pointwise sequential designs, the best-performing design in most cases is one of the two sequential designs proposed.
- The batch design performs comparably in most cases, and it is slightly outperformed by the pointwise design in a few instances.

How to choose the criterion?

- For the objective of minimizing RMSE: gradient-based criterion.
- For the objective of minimizing MAE: variance-based criterion.

Outline

Concluding remarks

- We introduce two novel criteria: gradient-based criterion and variance-based criterion.
- We put forward a framework termed the cluster-based top-b sequential design approach, which is applicable to any pointwise sequential design criterion.
- Through comparisons with existing criteria and methods, we demonstrate that the proposed criteria exhibit excellent performance and validate the robustness of the proposed batch sequential designs.
- Future work will extend this framework to high-dimensional and constrained design spaces.

Acknowledgement

Thank you for your attention!

Yongdao Zhou

Nankai University, China