Uniform projection designs under the stratified L_2-discrepancy

Yaping Wang

School of Statistics, East China Normal University

Joint work with Sixu Liu (Beijing Institute of Mathematical Sciences and Applications)

mODa14, June 18, 2026

Space-filling designs

- A good design for computer or physical experiments should be space-filling such that the points are arranged uniformly within the experimental region in some manner.
- Besides, a design is often chosen before we know which factors will be active.
- The response is commonly dominated by main effects and low-order interactions (often two-factor interactions).
- Therefore a good design, at least for the screening stage of the experiments, should have desirable low-dimensional projection properties.
- In projections, we would like to avoid replicates, clusters and large gaps.

Space-filling designs

- An (n, m, s) design has n runs and m factors, each taking levels in $\mathbb{Z}_s = \{0, 1, \ldots, s-1\}$.
- A U-type design is balanced: every level appears equally often in each column.
- Common choices include
  - Latin hypercube designs;
  - maximin, minimax designs;
  - uniform designs which minimizes a discrepancy;
  - Orthogonal-array based designs;
  - Maximum projection designs (Joseph et al., 2015);
  - Uniform projection designs (Sun et al., 2019);
  - Strong OA of various strengths (He and Tang, 2013; He et al., 2018), etc.
- For model-free computer experiments, "space-filling" is a robustness principle before a response model is selected.

From uniform design to uniform projection design

- Uniform designs often use an L_2-type discrepancy, for example the centered L_2-discrepancy CD (Hickernell, 1998a,b).
- However, full-dimensional discrepancy may have a dimensionality effect (Zhou, Fang and Ning, 2013):
  - good global uniformity can coexist with poor low-dimensional projections;
  - in screening stages, these low-dimensional projections are exactly what we inspect.
- Sun, Wang and Xu (2019) introduced the uniform projection criterion (based on the CD)

$$\Phi(D) = \frac{2}{m(m-1)} \sum_{|\mathbf{u}|=2} \mathrm{CD}(D_{\mathbf{u}}).$$

- A UPD minimizes the average two-dimensional projection discrepancy.

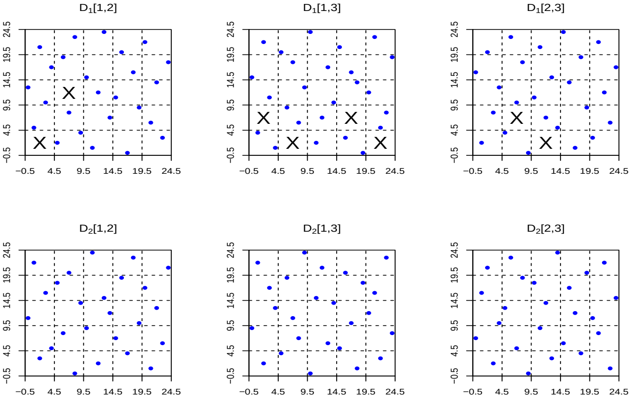
An example of uniform projection design (CD-based)

Table: Two 25 $\times$ 3 Latin hypercubes

D_1			D_2		
18	16	14	2	3	2
19	9	0	4	5	13
11	1	2	0	11	9
16	20	3	3	16	17
20	22	12	1	22	22
14	7	10	8	0	7
4	17	1	6	8	19
12	12	7	9	14	24
10	15	24	5	18	4
22	14	5	7	20	11
2	21	22	12	2	21
15	11	21	10	9	0
1	5	4	14	12	14
3	10	11	13	15	6
23	3	8	11	24	15
0	13	15	17	4	10
8	23	6	15	7	5
7	8	18	19	10	18
9	4	13	16	19	20
6	19	9	18	23	1
24	18	19	21	1	16
21	6	23	23	6	23
13	24	17	22	13	3
17	0	16	20	17	12
5	2	20	24	21	8

An example of uniform projection design (CD-based)

Figure: Bivariate projections of D_1 and D_2; 'X' marks an empty grid cell (Fig. 1 from Sun et al., 2019).



Uniform criterion vs. uniform projection criterion

Uniform criterion

$$\mathrm{CD}(D)$$

- Full-dimensional uniformity.
- Useful when all factors are equally important.
- May hide poor two-factor projections.

Uniform projection criterion

$$\Phi(D)$$

- A projection-based substitute.
- Easier to compute and interpret.
- Closely related to distance criteria.
- Lower and upper bounds are easier to obtain.

Table: Comparison of D_1 and D_2

Criterion	D_1	D_2	Preferred
min L_1-dist	9	10	larger
min L_2-dist	29	38	larger
$\rho_{\mathrm{ave}}(D)$	0.015	**0.008**	smaller
$\mathrm{CD}(D)$	**0.00142**	0.00153	smaller

- The CD selects D_1 by full-dimensional uniformity.
- The projection criterion selects D_2, agreeing with larger minimum L_1- and L_2-distances and smaller average correlation.
- This motivates replacing a global criterion by a projection-robust one.

Uniform projection criterion under the stratified L_2-discrepancy

- Tian and Xu (2025) introduced the stratified L_2-discrepancy (SD) for domains partitioned into subregions.
- SD is RKHS-based and has a local count-balance interpretation.
- Stratification prevents concentration of design points in specific regions, especially in high-dimensional spaces.
- We keep the UPD idea, but replace the CD by the SD.
- Therefore SD-UPD asks for:

$$\text{simplification} + \text{good two-factor projections} + \text{good multi-resolution stratification.}$$

- The goal is a projection-robust criterion that is still easy to analyze.

Why use the stratified L_2-discrepancy?

- Classical L_2-type discrepancies give global distributional summaries, but high-dimensional product terms can obscure local stratification patterns.
- Strong orthogonal arrays (SOAs) and stratification pattern enumerators give a combinatorial multi-resolution view.
- The SD connects these two viewpoints:
  - local count balance in stratified regions;
  - flexible resolution weights;
  - RKHS and pairwise-kernel representation for arbitrary point sets.
- Our question:

Can the average two-dimensional SD be
simple to compute and easy to analyze as the $\Phi(D)$ for CD?

Notation

- (n, m, s^p): an n-run design with m factors, represented by an $n \times m$ matrix $D = (x_{ik})$, each taking s^p levels from
$$\mathbb{Z}_{s^p} = \{0, 1, \ldots, s^p - 1\}.$$
- $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, t)$: allow mixed levels.
- A general strong orthogonal array (Tian and Xu, 2022), denoted by $\mathrm{GSOA}(n, m, s^p, t)$, is an (n, m, s^p) design such that any g columns can be collapsed into
$$\mathrm{OA}(n, g, s^{u_1} \times \cdots \times s^{u_g}, g), \quad u_1 + \cdots + u_g = t, \quad u_i \leq p.$$
- Collapsing to s^u levels is done by keeping the leading base-s digits:
$$x \mapsto \lfloor x / s^{p-u} \rfloor.$$

The stratified L_2-discrepancy

- Denote $\mathbb{Z}_s = \{0, 1, \ldots, s-1\}$ and $\mathbb{Z}_{p+1} = \{0, 1, \ldots, p\}$.
- For any $\mathbf{u} = (u_1, \ldots, u_m) \in \mathbb{Z}_{p+1}^m$ (resolution vector) and $\mathbf{x} = (x_1, \ldots, x_m) \in [0, 1)^m$ (a point), the stratified region $R_{\mathbf{u}}(\mathbf{x})$ is defined by

$$R_{\mathbf{u}}(\mathbf{x}) = \bigotimes_{k=1}^{m} R_{u_k}(x_k),$$

$$R_{u_k}(x_k) = \left[\lfloor s^{u_k} x_k \rfloor s^{-u_k}, \lfloor s^{u_k} x_k \rfloor s^{-u_k} + s^{-u_k} \right).$$

- Here, $R_{u_k}(x_k)$ is an interval of length s^{-u_k} that contains x_k, and its volume is

$$\mathrm{Vol}\{R_{\mathbf{u}}(\mathbf{x})\} = s^{-\sum_{j=1}^{m} u_j}.$$

Illustration of stratified regions

Resolution vectors shown:

$$\begin{array}{ccc} (0,1) & (1,0) & (1,1) \\ (1,2) & (2,1) & (2,2) \\ (2,3) & (3,2) & (3,3) \end{array}$$

Larger u_j gives a finer interval in coordinate j.

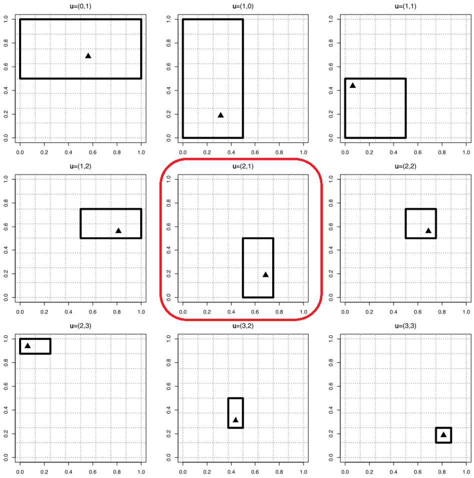
For the panel $\mathbf{u} = (2,1)$, take the displayed point as $\mathbf{x} = (0.7, 0.2)$. Then

$$\text{floor}(4 \cdot 0.7) = 2, \qquad \text{floor}(2 \cdot 0.2) = 0.$$

$$R_2(0.7) = \left[\frac{1}{2}, \frac{3}{4}\right), \qquad R_1(0.2) = \left[0, \frac{1}{2}\right).$$

Hence

$$R_{(2,1)}(\mathbf{x}) = \left[\frac{1}{2}, \frac{3}{4}\right) \times \left[0, \frac{1}{2}\right).$$



Tian and Xu (2025), Fig. 1

The stratified L_2-discrepancy

- For a finite point set $P \subset [0,1)^m$, the local discrepancy at resolution $\mathbf{u}$ is

$$\operatorname{disc}_{\mathbf{u}}^{R}(\mathbf{x}; P) = \operatorname{Vol}(R_{\mathbf{u}}(\mathbf{x})) - \frac{|P \cap R_{\mathbf{u}}(\mathbf{x})|}{n}.$$

- Its squared L_2-norm is

$$\left\|\operatorname{disc}_{\mathbf{u}}^{R}(P)\right\|^2 = \int_{[0,1]^m} \left|\operatorname{disc}_{\mathbf{u}}^{R}(\mathbf{x}; P)\right|^2 \mathrm{d}\mathbf{x}.$$

- If $r(\mathbf{u}) = \sum_k u_k$ and $\mathcal{P}_{\mathbf{u}}$ is the induced partition into $s^{r(\mathbf{u})}$ equal boxes, then equivalently

$$\left\|\operatorname{disc}_{\mathbf{u}}^{R}(P)\right\|^2 = s^{-r(\mathbf{u})} \sum_{B \in \mathcal{P}_{\mathbf{u}}} \left(s^{-r(\mathbf{u})} - \frac{N_B}{n} \right)^2.$$

- Thus the local term is a count-imbalance over one stratification.

The stratified L_2-discrepancy

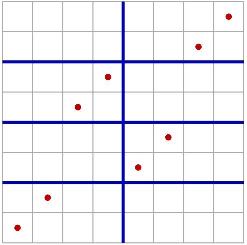
- Combine all resolutions with weights:

$$\mathrm{SD}(P) = \left(\sum_{\mathbf{u} \in \mathbb{Z}_{p+1}^m} \omega(\mathbf{u}) \left\| \mathrm{disc}_{\mathbf{u}}^R(P) \right\|^2 \right)^{1/2},$$

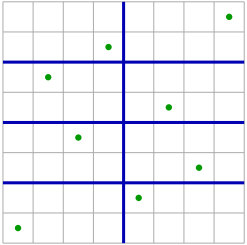
  where the product-type weight function $\omega(\mathbf{u}) = \prod_{i=1}^m \omega(u_i) > 0$.
- A common exponential weighting scheme suggested by Tian and Xu (2025) is

$$\omega(\mathbf{u}) = \prod_{i=1}^{m} \omega(u_i) = y^{\sum_{j=1}^m u_j}, \qquad y \in (0, 1].$$

An example from Tian and Xu (2025)



P_1: $\mathrm{SD} = 0.2416$



P_2: $\mathrm{SD} = 0.1389$

- Two 8-point designs in $[0,1)^2$, computed with $s = 2$, $p = 3$ and $\omega(\mathbf{u}) = 1$.
- The blue grid shows one local stratification, $\mathbf{u} = (1,2)$: two vertical strips and four horizontal strips.
- In P_1, four subregions have two points and four have no points; in P_2, this stratification is balanced.
- Aggregating all 16 stratifications gives $\mathrm{SD}(P_1) = 0.2416$ and $\mathrm{SD}(P_2) = 0.1389$.

The stratified L_2-discrepancy: kernel formula

- Like other L_2-type discrepancies (Hickernell, 1998a, b), the stratified L_2-discrepancy has an RKHS pairwise-kernel representation.
- Define
$$\delta_i(x_k, y_k) = \delta(\lfloor s^i x_k \rfloor, \lfloor s^i y_k \rfloor), \qquad i = 0, \ldots, p,$$
with δ being the Kronecker delta.
- Take
$$f^R(x, y) = \sum_{i=0}^{p} \frac{\omega(i)}{s^i} \delta_i(x, y), \qquad K^R(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^{m} f^R(x_k, y_k).$$
- We obtain the formula in Theorem 1 of Tian and Xu (2025):
$$\mathrm{SD}^2(P) \;=\; -\left(\sum_{i=0}^{p} \frac{\omega(i)}{s^{2i}}\right)^m \frac{1}{n^2} \sum_{a,b=1}^{n} \prod_{k=1}^{m} \left(\sum_{i=0}^{p} \frac{\omega(i)}{s^i} \delta_i(z_{ak}, z_{bk})\right).$$

Uniform projection criterion under the SD

- For an (n, m, s^p) design $D = (x_{ik})$, rescale the design points onto the unit hypercube $[0, 1)^m$ by
$$x_{ik} \mapsto z_{ik} = \frac{2x_{ik} + 1}{2s^p}.$$
- The uniform projection criterion under the SD is defined as the average of the squared SD over all two-dimensional projections
$$\Phi_{\mathrm{SD}}(D) = \frac{2}{m(m-1)} \sum_{|u|=2} \mathrm{SD}^2(D_u).$$
- A design D is called a uniform projection design under the SD if it minimizes Φ_{SD}.

Uniform projection criterion under the SD

Theorem

For a U-type (n, m, s^p) design $D = (x_{ik})$, let $z_{ik} = (2x_{ik} + 1)/(2s^p)$. Define

$$d_{ab} = \sum_{k=1}^{m} \sum_{i=0}^{p} \frac{\omega(i)}{s^i} \{1 - \delta_i(z_{ak}, z_{bk})\}.$$

Then

$$\Phi_{\mathrm{SD}}(D) = \frac{1}{n^2 m(m-1)} \sum_{a=1}^{n} \sum_{b=1}^{n} d_{ab}^2 + C_{\mathrm{SD}}(m, s, p),$$

where $C_{\mathrm{SD}}(m, s, p)$ is independent of the arrangement of D.

- Thus the design-dependent part is $G_D = \sum_{a,b} d_{ab}^2$.
- This gives an $O(n^2 m)$ evaluation formula.

Uniform projection criterion under the SD

projection average $\dfrac{2}{m(m-1)}\sum_{|u|=2}\mathrm{SD}^2(D_u)$ → single distance sum $G_D = \sum_{a,b} d_{ab}^2$ → sharp bounds equidistance condition

- The projection criterion is simplified to one row-pairwise hierarchical distance summary.
- The lower and upper bounds are easy to derive and evaluate once the criterion is written through G_D.
- The lower-bound equality condition has a clear interpretation: all row pairs should be equally separated across the stratification hierarchy.
- The same distance links Φ_{SD} to NRT (Niederreiter-Rosenbloom-Tsfasman) distance (Bierbrauer et al., 2002), stratification pattern enumerators and GSOA/SOA constructions.

Lower and upper bounds

- View d_{ab} as a weighted hierarchical discrete distance:

$$d_{ab} = \sum_{k=1}^{m} \sum_{i=0}^{p} \frac{\omega(i)}{s^i} \{1 - \delta_i(z_{ak}, z_{bk})\}.$$

- It suffices to consider lower and upper bounds for

$$G_D = \sum_{a=1}^{n} \sum_{b=1}^{n} d_{ab}^2,$$

since

$$\Phi_{\mathrm{SD}}(D) = \frac{G_D}{n^2 m(m-1)} + C_{\mathrm{SD}}(m, s, p).$$

Lower and upper bounds

Theorem

For a U-type (n, m, s^p) design D, let

$$A_0 = \sum_{i=0}^{p} \frac{\omega(i)}{s^i}, \qquad A_1 = \sum_{i=0}^{p} \frac{\omega(i)}{s^{2i}}, \qquad A_0^{(\ell)} = \sum_{i=0}^{\ell} \frac{\omega(i)}{s^i}.$$

Then

$$G_D^{\mathrm{LB}} \leq G_D \leq G_D^{\mathrm{UB}},$$

where

$$G_D^{\mathrm{LB}} = \frac{n^3 m^2}{n-1}(A_0 - A_1)^2,$$

$$G_D^{\mathrm{UB}} = n^2 m^2 \sum_{\ell=0}^{p-1} \frac{s-1}{s^{\ell+1}} \left(A_0 - A_0^{(\ell)}\right)^2.$$

Lower-bound equality condition

- The lower bound G_D^{LB} is achieved if and only if

$$d_{ab} = \bar{d} \quad \text{for all } 1 \le a < b \le n,$$

where

$$\bar{d} = \frac{nm}{n-1}(A_0 - A_1) = \sum_{i=0}^{p} \omega(i) \frac{(s^i - 1)nm}{s^{2i}(n-1)}.$$

- Thus the optimality condition is an equidistance condition under the weighted hierarchical distance.

When design is optimal under Φ_{SD}

- A sufficient condition: the lower bounds are attained if, for every row pair and every resolution i,
$$\sum_{k=1}^{m} \delta_i(z_{ak}, z_{bk}) = \frac{nm}{s^i(n-1)} - \frac{m}{n-1}, \quad 0 \leq i \leq p.$$
- Furthermore, when the weights are allowed to take arbitrary values, this lower bound is achieved if and only if the condition above holds.
- This condition fixes, for every pair of runs, how many coordinates stay in the same subinterval at each stratification resolution.
- This is a projection-averaged, multi-resolution analogue of equidistance.

A small equidistance example

Let $s = 3,\ p = 2,\ n = 9,\ m = 4$. Example from Tian and Xu (2025, Supp Table 4):

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 3 & 4 & 5 \\ 2 & 6 & 8 & 7 \\ 3 & 7 & 1 & 4 \\ 4 & 1 & 5 & 6 \\ 5 & 4 & 6 & 2 \\ 6 & 5 & 2 & 8 \\ 7 & 8 & 3 & 1 \\ 8 & 2 & 7 & 3 \end{pmatrix}$$

$\mathrm{GSOA}(9, 4, 3^2, 2)$

Collapses by leading base-3 digits:

$$D^{(i)} = \left\lfloor D/3^{\,2-i} \right\rfloor, \qquad i = 0, 1, 2.$$

$$D^{(0)} = \mathbf{0},\, D^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 \end{pmatrix},\, D^{(2)} = D.$$

Then every two rows are equidistant at each resolution:

$$(H_0(a,b), H_1(a,b), H_2(a,b)) = (0, 3, 4), \qquad a < b,$$

where $H_i(a, b)$ is the Hamming distance between rows a and b in $D^{(i)}$.

A small equidistance example

- Hence, all row pairs have the same weighted hierarchical distance:

$$d_{ab} = \sum_{i=0}^{2} \frac{\omega(i)}{3^i} H_i(a,b) = 3\frac{\omega(1)}{3} + 4\frac{\omega(2)}{9} = \omega(1) + \frac{4}{9}\omega(2) = \bar{d}.$$

- This is a projection-averaged, multi-resolution analogue of equidistance.
- Since $N_i = 4 - H_i$, $(N_0, N_1, N_2) = (4, 1, 0)$ for every row pair.

$$d_{ab} \equiv \bar{d}, \qquad G_D = \sum_{a,b} d_{ab}^2 = n(n-1)\bar{d}^2.$$

- With constant weights, $\bar{d} = 13/9$:

$$G_D = 9 \cdot 8 \left(\frac{13}{9}\right)^2 = 150.222\ldots = G_D^{\mathrm{LB}}.$$

Optimal designs under Φ_{SD}

- Multiplication tables of Galois fields:

$$\mathrm{GF}(s^p) \quad \Longrightarrow \quad \mathrm{GSOA}(s^p, s^p - 1, s^q, t), \quad q \leq p.$$

- Generalized Hadamard matrices and Rao-Hamming type constructions broaden the parameter ranges, e.g.

$$\mathrm{GSOA}(s^2, s+1, s^2, 2).$$

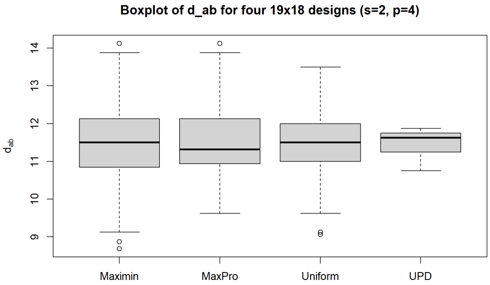
- Juxtaposition and collapsing preserve the equality condition.
- Some full-SD optimal GSOAs (Tian and Xu, 2025) are also optimal under Φ_{SD}.

A benchmark with four 19×18 LHDs

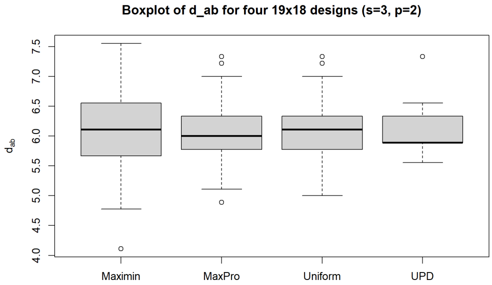
Design	CD	WD	MD	$\mathrm{SD}^{(2)}$	$\mathrm{SD}^{(3)}$	$\Phi_{\mathrm{SD}}^{(2)}$	$\Phi_{\mathrm{SD}}^{(3)}$
Maximin	1.2889	7.0488	25.2549	87.7170	6.0710	0.0227647	0.0072553
MaxPro	1.3090	**6.8823**	24.8515	87.6938	6.0468	0.0221945	0.0068359
Uniform	**1.2643**	6.9414	**24.8049**	87.6903	6.0496	0.0220757	0.0068860
UPD	1.2655	6.9352	24.8554	**87.6342**	**6.0365**	**0.0204921**	**0.0066817**

- Four LHDs: maximin, MaxPro, CD-uniform and CD-based UPD (Sun et al., 2019; Tian and Xu, 2025).
- Full-dimensional criteria select different winners.
- The stratified L_2 discrepancy criteria and the proposed projection criteria both favor the UPD.

Row-pairwise distances explain the ranking



$s = 2,\ p = 4$



$s = 3,\ p = 2$

- The UPD has a more concentrated distribution of d_{ab}.
- This matches the equality principle behind the lower bound.

Further connections

- The weighted hierarchical distance can be written through the NRT distance on $\mathbb{Z}_{s^p}$:

$$d_{ab} = \sum_{k=1}^{m} \{A_0 - A_0^{(p-\rho(x_{ak},x_{bk}))}\}.$$

Hence

$$\Phi_{\mathrm{SD}}(D) = \frac{1}{n^2 m(m-1)} \sum_{a=1}^{n} \sum_{b=1}^{n} \left[\sum_{k=1}^{m} \{A_0 - A_0^{(p-\rho(x_{ak},x_{bk}))}\} \right]^2 + C_{\mathrm{SD}}(m,s,p).$$

- Under a special exponential weighting scheme,

$$\Phi_{\mathrm{SD}}(D) = \frac{E_2(D;y) - 1}{(1-y)^2},$$

where $E_2(D;y)$ is a two-dimensional projection pattern enumerator.

- This links Φ_{SD} to space-filling patterns (Tian and Xu, 2022), stratification pattern enumerators (Tian and Xu, 2024) and SOAs of strength 2+ (He, Cheng and Tang, 2018).

Summary

- Projection uniformity matters when active factors are unknown.
- The stratified L_2-discrepancy adds multi-resolution control to projection uniformity.
- Φ_{SD} has a simple row-pairwise distance representation.
- Sharp bounds make optimality a hierarchical equidistance condition.
- Some existing GSOA, finite-field and generalized Hadamard constructions are automatically optimal.

Future work:

- Tighter bounds for small factor-to-run ratio designs.
- More algebraic constructions under Φ_{SD}.

Selected references

- Bierbrauer, J., Edel, Y. and Schmid, W. C. (2002). Coding-theoretic constructions for (t, m, s)-nets and ordered orthogonal arrays. *J. Combin. Des.* **10**, 403–418.
- He, Y. and Tang, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100**, 254–260.
- He, Y., Cheng, C.-S. and Tang, B. (2018). Strong orthogonal arrays of strength two plus. *Ann. Statist.* **46**, 457–468.
- Hickernell, F. J. (1998a). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67**, 299–322.
- Hickernell, F. J. (1998b). Lattice rules: how well do they measure up? *Lecture Notes in Statist.* **138**, 109–166.
- Joseph, V. R., Gul, E. and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika* **102**, 371–380.
- Liu, S. and Wang, Y. (2026). Uniform projection designs under the stratified L_2-discrepancy. arXiv:2605.19900.
- Sun, F., Wang, Y. and Xu, H. (2019). Uniform projection designs. *Ann. Statist.* **47**, 641–661.
- Tian, Y. and Xu, H. (2022). A minimum aberration-type criterion for selecting space-filling designs. *Biometrika* **109**, 489–501.
- Tian, Y. and Xu, H. (2024). Stratification pattern enumerator and its applications. *J. R. Stat. Soc. Ser. B* **86**, 364–385.
- Tian, Y. and Xu, H. (2025). A stratified L_2-discrepancy with application to space-filling designs. *J. R. Stat. Soc. Ser. B* **88**, 464–490.
- Zhou, Y. D., Fang, K.-T. and Ning, J. (2013). Mixture discrepancy for quasi-random point sets. *J. Complexity* **29**, 283–301.

Thank You!

School of Statistics

East China Normal University

Image: Shanghai skyline from the Bund, Wikimedia.