AI505

Optimization

# Sampling Plans

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

# Outline

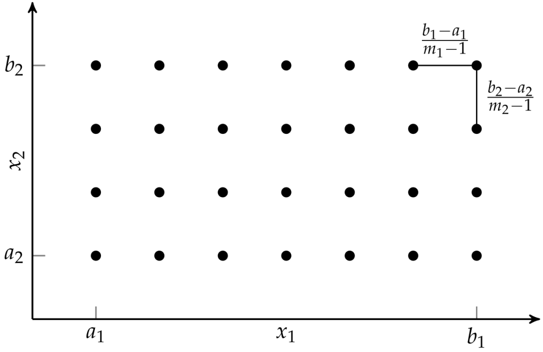
# Sampling Plans

- In all nonlinear non convex optimization, to generate good initial design points
- With computationally costly functions, to create an initial set of design points from where to build a **surrogate models** to optimize in place of the original function
- In hyperparameter tuning

# Full Factorial Design

- Factors and levels, terms from the field of Experimental Design in Statistics
- Uniform and evenly spaced samples across domain
- Simple, easy to implement, and covers domain
- Optimization over the points known as **grid search**
- Sample count grows exponentially with dimension: $n^m$
- Can be coarse and miss local features

$a_i \leq x_i \leq b_i$ for each component $i$.
grid with $m_i$ samples in the $i$th dimension



# Random Sampling

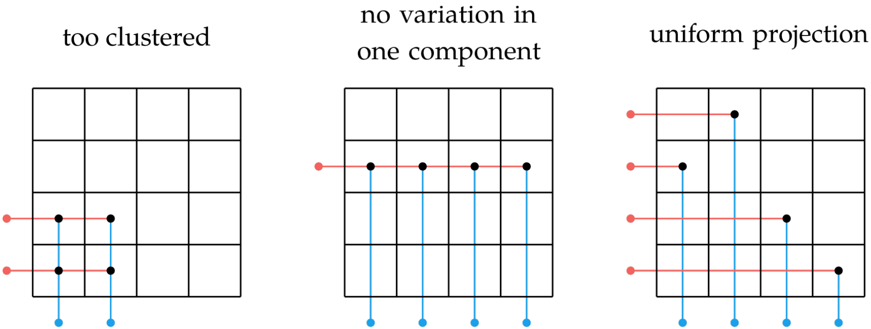
- Uses pseudorandom number generator to define samples according to our desired distribution
- If variable bounds are known, a common choice is independent **uniform distributions** across domains of possible variable values
  $[a_1, b_1] \times \ldots \times [a_n, b_n]$
- Ideally, if enough points are sampled and the right distribution is chosen, the design space will be covered

# Uniform Projection Plans

- A **uniform projection plan** is a sampling plan over a discrete grid where the distribution over each dimension is uniform.

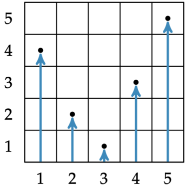
Example

In 2D, $m \times m$ sampling grid (as in full factorial), but, instead of taking all $m^2$ samples, we want to sample only $m$ positions.



# Uniform Projection Plans

Example (Random $m$-permutations)




$p = 4\ 2\ 1\ 3\ 5$

Example (Latin square)

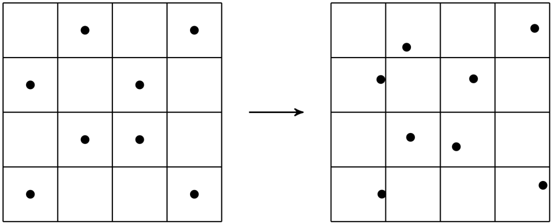
**Latin squares** are $m \times m$ grids where each row contains each integer 1 through $m$ and each column contains each integer 1 through $m$.

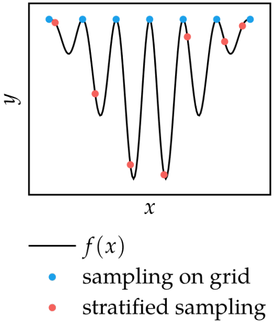
**Latin-hypercubes** are a generalization to any number of dimensions (note that the points remain $m$) N rooks on a chess board without threatening each other

| | | | |
|---|---|---|---|
| 4 | 1 | 3 | 2 |
| 1 | 4 | 2 | 3 |
| 3 | 2 | 1 | 4 |
| 2 | 3 | 4 | 1 |

# Stratified Sampling

- Each point is sampled uniformly at random within each grid cell instead of the center
- Cells decided by Full Factorial or Uniform Projection Plans
- Can capture details that regularly-spaced samples might miss

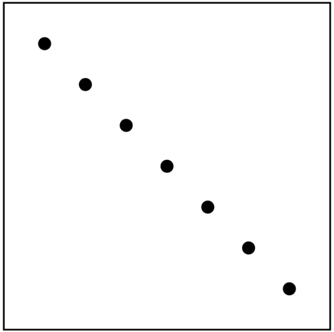




# Space Filling Metrics

- A sampling plan may cover a search space fully, but still leave large areas unexplored

Example (Uniform Projection Plan)



- **space-filling metrics** quantify this aspect measuring the degree to which a sampling plan $X \subseteq \mathcal{X}$ fills the design space

# Space-Filling Metrics: Discrepancy

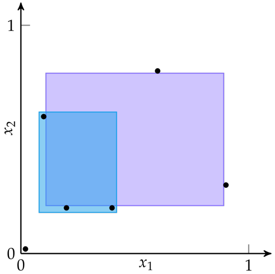
- **Discrepancy**: measure of ability of the sampling plan $X$ to fill a hyper-rectangular design space
- It is given by hyper-rectangular subset $\mathcal{H}$ with the maximum difference between the fraction of samples in $\mathcal{H}$ and the volume of $\mathcal{H}$'s.

$$d(X) = \underset{\mathcal{H}}{\text{supremum}} \left| \frac{\#(X \cap \mathcal{H})}{\#X} - \lambda(\mathcal{H}) \right|$$

$\lambda(\mathcal{H})$ is the $n$-dimensional volume of $\mathcal{H}$, ie, the product of the side lengths of $\mathcal{H}$

We wish to have a plan $X$ with low discrepancy

Often very difficult to compute directly




$d$ for the purple rectangle is $>$ than $d$ for the blue rectangle

# Space-Filling Metrics: Pairwise Distances

- Method of measuring relative space-filling performance of two $m$-point sampling plans
- Better spread-out plans will have larger pairwise distances:
    1. compute all pairwise distances between all points within each sampling plan
    2. sort the pairwise distances of each set in ascending order
    3. the plan with the first pairwise distance exceeding the other is considered more space-filling
- Suggests simple algorithm:
    1. produce a set of randomly distributed sampling plans,
    2. pick the one with greatest pairwise distances
- Possible also for uniform projection plans, by mutating them with swaps and simulated annealing.

# Space-Filling Metrics: Morris-Mitchell Criterion

- Alternative to previously suggested algorithm that simplifies the optimization problem

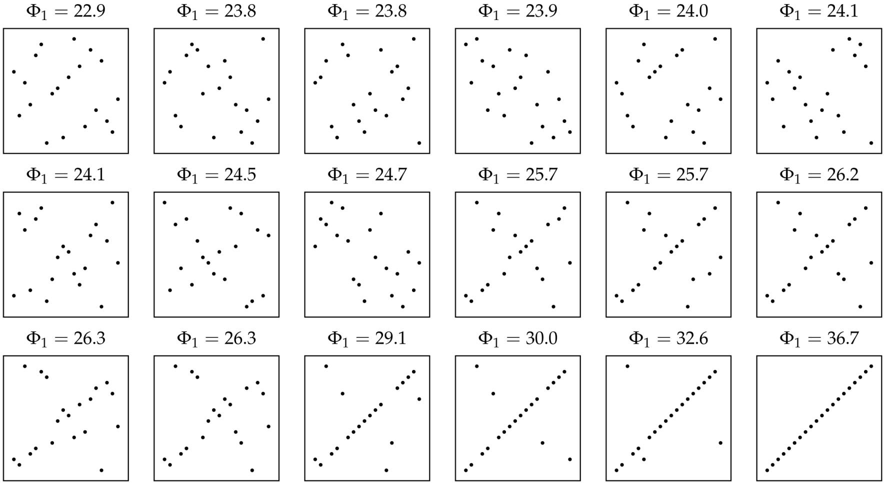
$$\underset{X}{\text{minimize}} \underset{q\in\{1,2,3,10,20,50,100\}}{\text{maximize}} \Phi_q(X)$$

$$\Phi_q(X) = \left(\sum_i d_i^{-q}\right)^{\frac{1}{q}}$$

where $d_i$ is the $i$th pairwise distance between points in $X$ and $q > 0$ is a tunable parameter. Larger values of $q$ give higher penalties to large distances.

# Space-Filling Metrics: Morris-Mitchell Criterion

Uniform projection plans sorted from best to worst according to $\Phi_1$



# Space-Filling Subsets

- Often, the set of possible sample points is constrained to be a subset of available choices

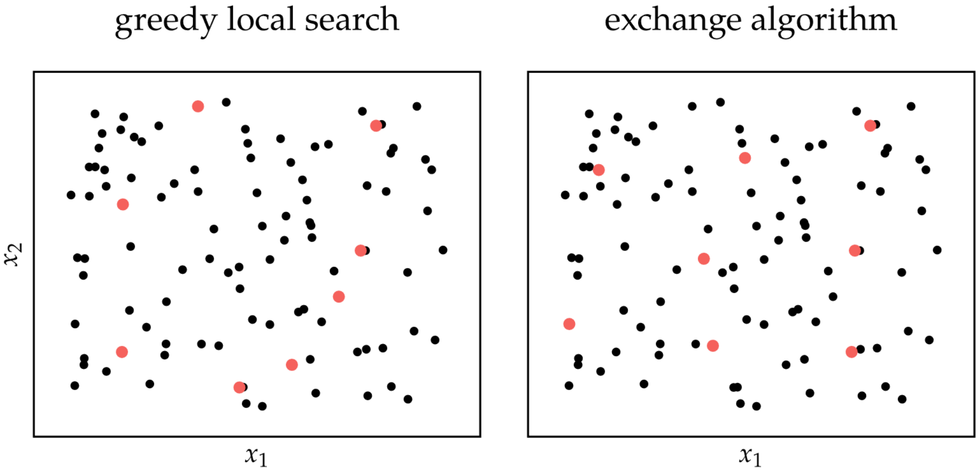
- A space-filling metric for a subset $S$ within a finite set $X$ is the maximum distance between a point in $X$ and the closest point in $S$, using a norm to measure distance

$$d_{\max}(X, S) = \underset{\boldsymbol{x} \in X}{\text{maximize}} \; \underset{\boldsymbol{s} \in S}{\text{minimize}} \; \|\boldsymbol{s} - \boldsymbol{x}\|_q$$

- A space-filling subset minimizes this metric

- Often computationally intractable, but heuristics like (repeated) greedy construction and exchange-search often produce acceptable results

# Space-Filling Subsets

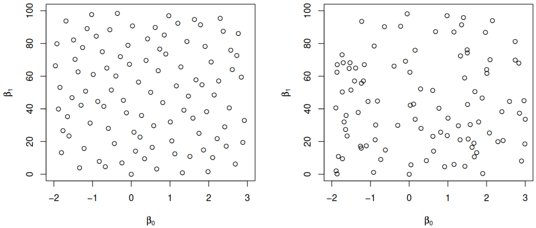


# Quasi-Random Sequences

- Also called **low-discrepancy sequences**, **quasi-random sequences** are deterministic sequences that systematically fill a space such that their integral over the space converges as fast as possible

- Used for fast convergence in Monte Carlo integration, which approximates an integral by sampling points in a domain

- Quasi-random sequences are typically constructed for the unit $n$-dimensional hypercube, $[0,1]^n$. Any multidimensional function with bounds on each variable can be transformed into such a hypercube.



# Quasi-Random Sequences

- Additive Recurrence: Recursively adds irrational numbers
- Halton Sequence: sequence of fractions generated with coprime numbers
- Sobol Sequence: recursive XOR operation with carefully chosen numbers

# Quasi-Random Sequences: Additive Recurrence

- Recursively adds irrational numbers

$$x_{k+1} = x_k + c \pmod 1$$

  $c$ irrational

$$c = 1 - \varphi = \frac{\sqrt{5} - 1}{2} \approx 0.618034$$

  $\varphi$ is golden ratio

- We can construct a space-filling set over $n$ dimensions using an additive recurrence sequence for each coordinate, each with its own value of $c$.

- square roots of the primes are known to be irrational, and can thus be used to obtain different sequences for each coordinate:

$$c_1 = \sqrt{2}, c_2 = \sqrt{3}, c_3 = \sqrt{5}, c_4 = \sqrt{7}, c_5 = \sqrt{11}, \ldots$$

# Quasi-Random Sequences: Halton Sequence

- single-dimensional version, called van der Corput sequences, generates sequences where the unit interval is divided into powers of base $b$. For example, $b = 2$

$$X = \left\{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \ldots\right\}$$

  whereas $b = 5$

$$X = \left\{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{25}, \frac{6}{25}, \frac{11}{25}, \ldots\right\}$$

- Multi-dimensional space-filling sequences use one van der Corput sequence for each dimension, each with its own base $b$. The bases, however, must be **coprime** in order to be uncorrelated.

- Two integers are **coprime** if the only positive integer that divides them both is 1, eg, 8 and 9.

- Correlation can be avoided by the leaped Halton method, which takes every $p$th point, where $p$ is a prime different from all coordinate bases.

# Quasi-Random Sequences: Sobol Sequence

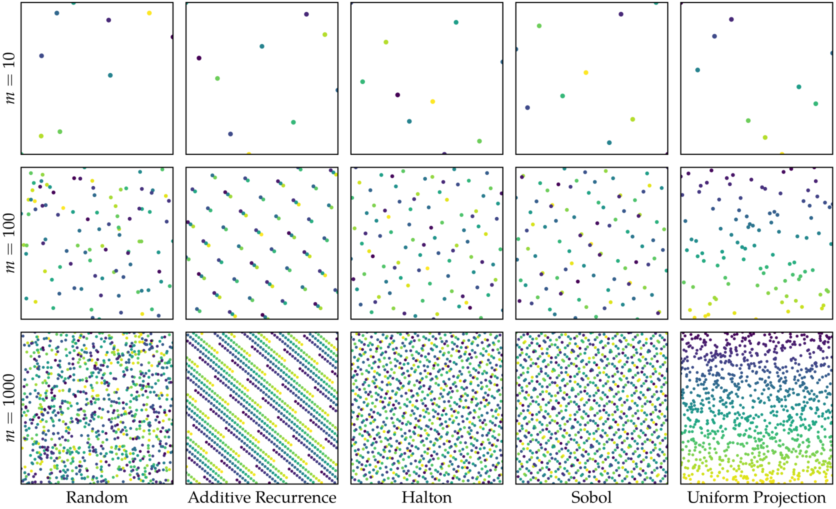
- Recursive XOR operation with carefully chosen numbers.
- XOR ($\oplus$) returns true if and only if both inputs are different
- For $n$-dimensional hypercube $I^n = [0,1]^n$, the $i$th point of the sequence $\boldsymbol{x}_i$ for dimension $j$ is calculated as:

  $$x_{i,j} = x_{i-1,j} \oplus v_{k,j}$$

  $v_{k,j}$ is the $j$th dimension of the $k$th direction number.
- **direction numbers** $v_{k,j} = (0.v_{k,j,1}v_{k,j,2}\ldots)_2$ where $v_{k,j,m}$ denotes the $m$th digit after the binary point.
- Tables of direction numbers with different properties have been proposed.
- Initialization: unit initialisation: $\ell$th left most bit set to one $v_{k,j,\ell} = 1$ for all $k$ and $j$ and all others to be zero

# Quasi-Random Sequences

space-filling sampling plans in two dimensions. Samples are colored according to the order in which they are sampled. The uniform projection plan was generated randomly and is not optimized.



# Summary

- Sampling plans are used to cover search spaces with a limited number of points
- Full factorial sampling, which involves sampling at the vertices of a uniformly discretized grid, requires a number of points exponential in the number of dimensions
- Uniform projection plans, which project uniformly over each dimension, can be efficiently generated and can be optimized to be space-filling
- Greedy construction and the exchange local search algorithm can be used to find a subset of points that maximally fill a space
- Quasi-random sequences are deterministic procedures by which space-filling sampling plans can be generated