EUROPEAN CONSORTIUM FOR MATHEMATICS IN INDUSTRY

# Identifying and Optimizing Parameter Domains

Chupei Lin
Yan Aklog
Simon Arnold
**Prof. Dr. Thomas Goetz**

37th ECMI Modelling Week
23.–29.06.2024 Catania, Italy

**Given**

- A dataset by ZF containing 2001 datapoints (N) with 11 parameters each.

**Information about the data**

- Data was generated by Finite Element Simulations.
- Values of Parameters are standardized.
- Meaning of the parameters is unknown.
- Data is labeled into class 1 (feasible) and class 0 (non-feasible).

**Task**

- Find a parameter domain such that the volume of the domain is maximized and that at least 95% of the points inside the domain are feasible. The domain should be as simple as possible.

## Data

- Data points are randomly distributed within a unit hypercube.

## Idea

- Use statistical learning tools to get a first impression.

| | Support-Vector-Machine | K-Nearest-Neighbors | Random Forest |
|---|---|---|---|
| Acc. | 95.25% | 84.36% | 100.00% |
| Prec. | 96.0% | 97.3% | 100.00% |
| Character. | Hyperplane | Decision Boundary | Majority Vote |

$\Rightarrow$ Datapoints are separable in the $11^{th}$ dimension.

$\Rightarrow$ Monte Carlo Methods can be used to measure the volume of the domain in high-dimensional space.

# Definition of Our Optimization Problem

EUROPEAN CONSORTIUM FOR MATHEMATICS IN INDUSTRY

## Definition

The optimization problem can be expressed in a functional form that is

$$\max_{\beta} |\Omega(\beta)| \tag{1}$$

$$\text{s.t.} \quad \frac{n_{f,\Omega}(\beta)}{n_{\Omega}(\beta)} \geq 0.95 \tag{2}$$

where $\Omega$ is the domain, $n_{\Omega}$ the number of datapoints in $\Omega$ and $n_{f,\Omega}$ the number of feasible datapoints in $\Omega$.

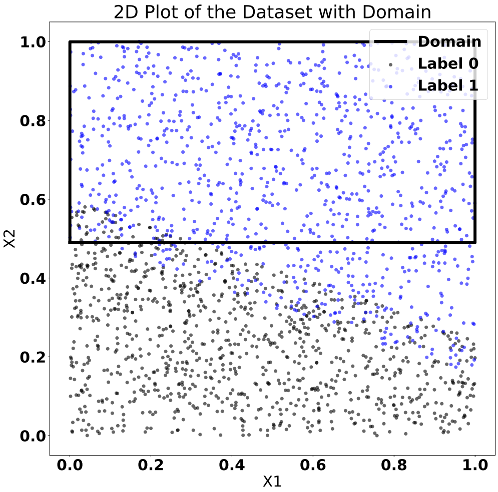
# Choice of Optimization Method

| Sequential Least Squares Programming (SLSQP) | Differential Evolution (DE) |
|---|---|
| Gradient-based Optimization | Population-based Stochastic Search |
| Find local optima | Find global optima |
| High efficiency for small problems providing accurate solutions | Robust to discontinuous objective functions |
| Struggle with problems where gradients are not available | Slower compared to gradient-based method |

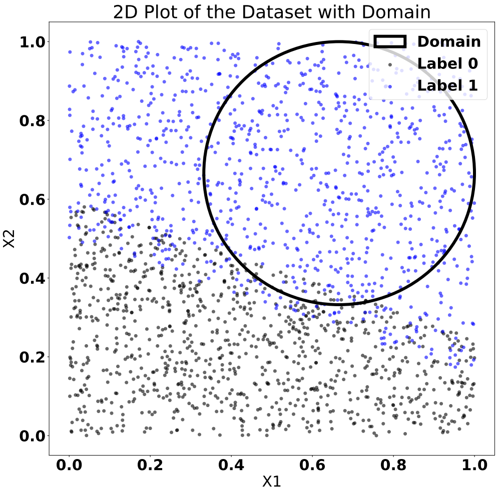
## Our Optimization Problem

- The objective function (1) may exhibit discontinuities.
- The constraint function (2) exhibits discontinuities.
- In the $11^{th}$-dimensional space, the count of variables is substantial.

# 2D (artificial) Problem

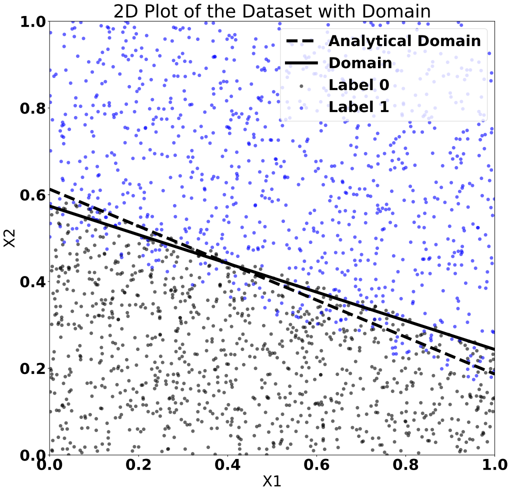



(a) $\max_\beta |\Omega| = (\beta_2 - \beta_1)(\beta_4 - \beta_3)$




(b) $\max_\beta |\Omega| = \pi\beta_r^2$

# 2D (artificial) Problem

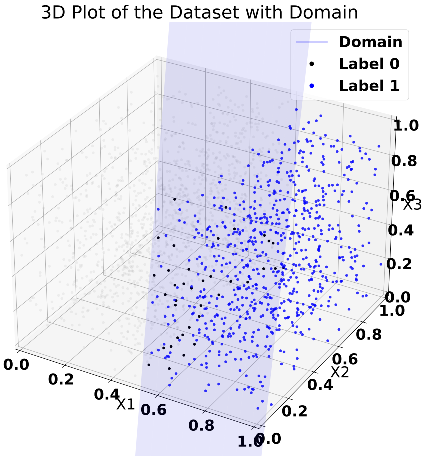



(c) $\max_\beta |\Omega| = ?$

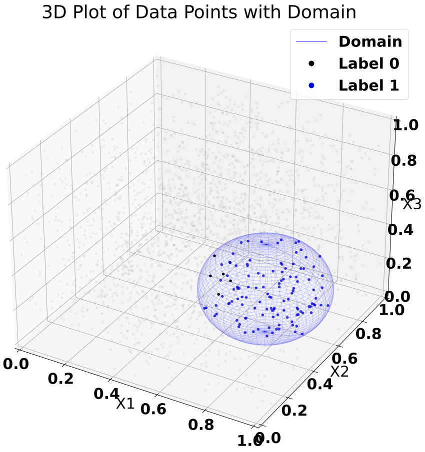
- Area of domain is difficult to calculate analytically for 2D $\rightarrow$ worse for 11D.
- **Idea:** Use Monte Carlo Integration to approximate area

$$|\Omega| \approx \frac{n_\Omega}{N}$$

| Shape | $\lvert\Omega\rvert$ | $n_\Omega$ | $n_{f,\Omega}$ |
|---|---|---|---|
| Rectangle | 0.51 | 810 | 770 |
| Circle | 0.35 | 569 | 541 |
| Line | 0.468 | 935 | 889 |
| Analytical | 0.6 | 925 | 879 |

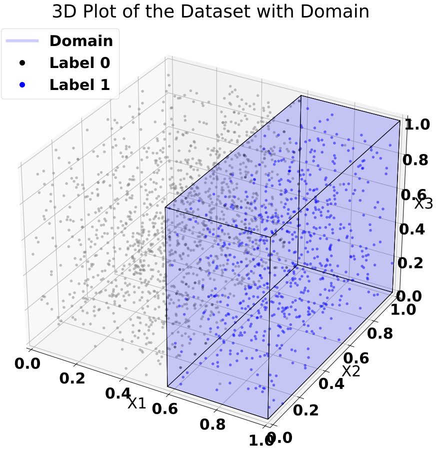



(d) $\max_\beta |\Omega| \approx \frac{n_\Omega}{N}$




(e) $\max_\beta |\Omega| = \frac{4}{3}\pi\beta_r^3$

# 3D (artificial) Problem




(f) $\max_{\beta} |\Omega| = (\beta_2 - \beta_1)(\beta_4 - \beta_3)(\beta_6 - \beta_5)$

- Employing cutting planes captures more feasible points.
- **Rotation:** In high-dimensional space, using a rotated cuboid as the objective function may yield better results, but it requires more variables to account for the rotation angles.

| Shape | $\|\Omega\|$ | $n_{\Omega}$ | $n_{f,\Omega}$ |
|---|---|---|---|
| Rectangular Cuboid | 0.42 | 725 | 689 |
| Sphere | 0.065 | 110 | 105 |
| Plane | 0.374 | 748 | 711 |

# 11D Problem (739 "Feasible" Points)

EUROPEAN CONSORTIUM FOR MATHEMATICS IN INDUSTRY

## Methods Comparison

| | HyperPlane | HyperRectangle | HyperSphere | SVM |
|---|---|---|---|---|
| Domain Volume | 36.33% | 13.08% | 0.03% | 34.98% |
| Feasibility Rate | 95.1% | 95.0% | 100.0% | 96.0% |
| # of Points In the Domain | 727 | 260 | 1 | 700 |

## High-Dimensional Space Characteristics

- The Volume of an $n$-dimensional hypersphere decreases exponentially as $n$ increases, assuming the radius remains constant.

  $\Rightarrow$ The Volume contained within the largest hypersphere inside an $11^{th}$ dimensional unit hypercube is only 0.092%.

# Summary and Outlook



- The Monte Carlo Method is an effective way to approximate the Volume of the domain in high-dimensional space for randomly distributed datasets.
- Differential Evolution can be used to find global optimizations for discontinuous and complex problems.
- Employing a hyperplane for a well-separated dataset can yield the optimal result in our problem.

## Outlook

- Using more hyperplanes may generate a larger domain while not violating the constraints.
- Other simple shapes could be employed.
- Allowing the sphere to extend beyond the hypercube may yield better results, but with a larger radius.


Any Questions?