

# Foundations of Optimal Experimental Design

## Design Matrix, Models, and Information Matrix

### The Design Matrix and Linear Model Representation

In regression-based DOE, each run is associated with a vector of regressor values derived from the chosen model. Collecting these vectors for all runs forms the design matrix (commonly denoted  $\mathbf{X}$ ). For example, if we plan to fit a second-order (quadratic) model in two factors  $X_1$  and  $X_2$ , the regressor for a run with factor levels  $(x_1, x_2)$  would be:

$$f(x_1, x_2) = [1, x_1, x_2, x_1x_2, x_1^2, x_2^2]$$

This includes an intercept, linear terms, interaction, and quadratics. The design matrix  $\mathbf{X}$  has one row per run and one column per model term. This matrix encapsulates how the experimental settings relate to model predictions.

Having a full-rank design matrix is essential: it ensures the model's parameters can be estimated uniquely. The number of runs should be at least equal to (and preferably larger than) the number of parameters in the model. In R, one can easily construct a design matrix using `model.matrix()`:

```
design_runs <- data.frame(x1 = c(-1,-1, 1, 1, 0, 0),  
x2 = c(-1, 1,-1, 1, 0, 0)) # example runs  
X <- model.matrix(~ x1 + x2 + I(x1^2) + I(x2^2) + x1:x2, data = design_runs)  
head(X)
```

	(Intercept)	x1	x2	I(x1^2)	I(x2^2)	x1:x2
1	1	-1	-1	1	1	1
2	1	-1	1	1	1	-1
3	1	1	-1	1	1	-1
4	1	1	1	1	1	1
5	1	0	0	0	0	0
6	1	0	0	0	0	0