Bayesian Sequential Design For Multi-Objective Process Optimisation

Matteo Borrotti [1] & Antonio Pievatolo [2]

[1] *CNR-IMATI, Via Bassini 15, 20133 Milan (Italy), matteo.borrotti@mi.imati.cnr.it*
[2] *CNR-IMATI, Via Bassini 15, 20133 Milan (Italy), antonio.pievatolo@mi.imati.cnr.it*

Résumé.
La Complexité se pose dans différents domaines d'applications. Le nombre croissant de variables et les réponses du système utilisé pour décrire un problème expérimental limitent l'applicabilité des approches classiques du plan d'expériences (Design of Experiment - DOE) et du plan d'expérience séquentiel (Sequential Experimental Design - SED). Dans cette situation, plus d'efforts doivent être mis dans lélaboration d'approches méthodologiques pour les problèmes expérimentaux complexes avec réponses multiples. Ici, nous allons développer une nouvelle technique de planification d'expériences basée sur l'incorporation de la notion d'optimalité de Pareto dans le cadre bayésien du plan d'expériences séquentiel. L'un des aspects essentiels de l'approche impliquera la méthode de sélection des prochains points du plan d'expériences basés sur les informations actuelles et les réponses du système choisis. La nouvelle approche séquentielle a été testée sur une étude de cas simulé.

Mots-clés. Points de mesure optimaux, Optimum de Pareto , Utilité prédictive

Abstract. Complexity arises in different fields of applications. The increasing number of variables and system responses used to describe an experimental problem limits the applicability of classical approaches from Design of Experiments (DOE) and Sequential Experimental Design (SED). In this situation, more effort should be put into developing methodological approaches for complex multi response experimental problems. In this work, we will develop a novel design technique based on the incorporation of the Pareto optimality concept into the Bayesian sequential design framework. One of the crucial aspects of the approach will involve the selection method of the next design points based on current information and the chosen system responses. The novel sequential approach has been tested on a simulated case study.

Keywords. Optimal design points, Pareto optimality, Predictive utility

1 Introduction

Current research and development at both academic and industrial level is tackling the design and characterisation of sophisticated systems, where the presence of a high number of factors and variables limits the complete experimental screening towards actual optimisation. A possible example is the separation process for recycling waste electrical and electronic equipments (WEEE) [1], which is characterised by high number of process operating conditions (i.e. variables) and, often, more than one outcomes (i.e. responses). In this context, there is the urgent need of novel multi-objective approaches to find optimal solution (or a set of optimal solutions), which compromises between two or more conflicting system responses.

In this work, we propose a Bayesian sequential DOE approach to multi-objective designs. The novelty resides in the combination of the Bayesian paradigm and the unprocessed multiple responses.

Our approach has been tested in a simulated example related to the optimisation of electrostatic separation processes in recycling [2], showing promising initial results.

2 The Multi-objective Bayesian Sequential DOE Approach

2.1 The Multi-objective Optimisation Scheme

In this first attempt to develop a multi-objective Bayesian sequential DOE approach, we consider a general problem described by k experimental variables, $\mathbf{x} = (x_1, \ldots, x_k)$, and d objective functions, $\phi = (\phi_1, \ldots, \phi_d)$, each of which is used to calculate the respective system response forming the vector $\mathbf{y} = (y_1, \ldots, y_d)$. The ultimate aim is to simultaneously maximise all the d objective functions, which is not possible in general, so that a weaker optimality is sought instead: this is the set of all Pareto optimal design points, whose image in the multi-objective space is called the Pareto front [3]. (A design point $\mathbf{x}$ is called Pareto optimal if no other $\mathbf{x}^*$ exists such that $\phi_i(\mathbf{x}^*) \geq \phi_i(\mathbf{x})$ for all i). In order not to not put emphasis on one special objective function, ϕ_i, while doing multi-objective optimization, it is necessary to normalise the system response values. In our example the system responses lie in the interval $[0, 1]$ so no further action is required.

To begin the optimisation scheme, we randomly select a first initial design matrix X composed by n design points and evaluate them. We then create the matrix K composed by the initial design matrix X and the related matrix of system responses Y. At this step, matrix K has n rows and $k+d$ columns. At this point, we calculate d Bayesian one-dimensional utility functions (one for each system response) for the whole search space, based on the marginal predictive distribution of the response conditional on K and we identify the Pareto front based on the d Bayesian one-dimensional utilities. Among the points in the Pareto front, we select the next design point as the one with the minimun Euclidean distance from the *utopia* point [3] and add it to the matrix K along with the related vector of system responses. For example, a utopia point can be the ideal point formed by maximising the response within the Pareto front coordinate-wise. The process is repeated until a certain stopping criterion is satisfied.

2.2 The Proposed Bayesian Utility Function

We will refer to $U^i(\mathbf{x}_{n+1}|X)$ as the i-*th* component of the multidimensional Bayesian utility function, where $i = 1, \ldots, d$ and $\mathbf{x}_{n+1}$ is the feasible design point in the design space that may be chosen for the next experimental step.

The use of $U^i(\mathbf{x}_{n+1}|X)$ has been inspired by the work of Verdinelli and Kadane (1992) [4].

$$\begin{aligned} U^i(\mathbf{x}_{n+1}|X) &= \int [\alpha\, y_{n+1,i} + \beta\, ln\, p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})] \times \\ &\quad \times p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1}) d\mathbf{y}_{n+1} d\theta, \end{aligned}$$

where $(y)_{1:(n+1)}$ collects $n+1$ d-dimensional system responses and θ parameterises the likelihood. Furthermore, α and β are non-negative weights, expressing the relative contribution that the experimenter is willing to attach to the two components of U^i. The number of system responses, d, is not higher than the the dimension of $\mathbf{y}$, which is the vector of response values. Generally, the posterior distribution of θ depends on all y.

The *i-th* utility function can be expanded as follows:

$$\begin{aligned} U^i(\mathbf{x}_{n+1}|X) &= \alpha \int\ y_{n+1,i}\ p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i} + \\ &\quad + \beta\ ln\ p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})]\ \times \\ &\quad \times\ p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i}\ d\theta \\ &= \alpha\ E[y_{n+1,i}|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1}]\ + \\ &\quad + \beta \int\ ln\ p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})\ \times \\ &\quad \times\ p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i}\ d\theta. \end{aligned}$$

The second term is related to the expected gain in Shannon information, obtained from adding the new design point to X, conditional on $\mathbf{y}_1 : n$. The simplest model for $\mathbf{y}_{1:n,i}$ is the following:

$$\mathbf{y}_{1:n,i} = X\ \theta_i\ +\ \epsilon_{1:n,i}, \tag{1}$$

where $\theta_i \sim \mathcal{N}(\theta_0, \sigma^2 R_0^{-1})$, σ^2 is known and $\epsilon_{1:n,i}$ is a vector of *iid* $\mathcal{N}(0, \sigma^2)$ random variables.

We know there are d system responses, each with its own U^i. However, without loss of generality, we can consider just one system response in the following steps. We then obtain:

$$y_{1:n} = X\ \theta\ +\ \epsilon_{1:n}. \tag{2}$$

Notice that the entries of Y are conditionally independent given θ.

We can now calculate the joint distribution of y_{n+1} and θ:

$$p(y_{n+1}, \theta|\mathbf{y}_{1:n}, X, \mathbf{x}_{n+1})\ =\ p(y_{n+1}|\theta, \mathbf{x}_{n+1})\ p(\theta|\mathbf{y}_{1:n}, X), \tag{3}$$

and

$$\theta|\mathbf{y}_{1:n}, X \sim \mathcal{N}\{(X^T X + R_0)^{-1} (X^T \mathbf{y}_{1:n} + R_0\theta_0), \sigma^2 (X^T X + R_0)^{-1}\}, \tag{4}$$

which is the posterior distribution after observing $\mathbf{y}_{1:n}$.

From Eq. 2 and Eq. 4, we obtain the two terms of the utility function. The factor multiplying β is:

$$\begin{aligned}\int ln\, p(\theta|\mathbf{y}_{1:(n+1)}, X, \mathbf{x}_{n+1}) \qquad & p(y_{n+1}, \theta|\mathbf{y}_{1:n}, X, \mathbf{x}_{n+1})\, dy_{n+1}\, d\theta \\ = -\frac{k}{2}\, ln(2\pi) - \frac{k}{2} + \frac{1}{2}\, ln\, det\{\sigma^{-2}\, & (\mathbf{x}_{(n+1)}\mathbf{x}_{(n+1)}^T + R)\},\end{aligned}$$

where $R = (X^T X + R_0)$.

The predictive mean multiplying α is:

$$\begin{aligned} E(y_{n+1}|\mathbf{y}_{1:n}) &= E(\theta^T x_{n+1} + \epsilon_{n+1} \mid \mathbf{y}_{1:n}) \\ &= E(\theta \mid \mathbf{y}_{1:n})^T\, x_{n+1} \\ &= [\{(X^T X) + R_0\}^{-1}\, (X^T \mathbf{y}_{1:n} + R_0\theta_0)]^T\, x_{n+1}. \end{aligned}$$

3 Simulated Case Study and Results

Our simulated case study is based on the work of Borrotti et al. (2015) [2]. The problem consists in separating metal (conductive) and nonmetal particles derived from WEEE by using the Corona electrostatic separation (CES) process. A typical industrial CES machine with fixed design parameters depends on the following controllable parameters (i.e., variables): (i) Electrostatic potential (x_1), or voltage, which ranges between -35.000 to -25.000 Volts; (ii) Drum speed (x_2), or simply speed, which ranges between 32 to 128 rpm; (iii) Feed rate (x_3), which ranges between 0.0028 to 0.028 kilograms per second (kg/s). As responses of the process we consider the recovery rate of conductive products ($R_{c,c}$) and the grade of conductive products ($G_{c,c}$, measuring purity).

From available data, we estimated two multiple regression models, ϕ_1 and ϕ_2, in order to simulate the responses $R_{c,c}$ and $G_{c,c}$ over the whole search space. Given a discretisation of the domain of the variables x_1, x_2 and x_3, we determined a search space of size 10^3 and deploying ϕ_1 and ϕ_2 we calculated the two system responses, from now on $\hat{y}_1$ and $\hat{y}_2$.

At this point, it was possible to calculate the real Pareto front of the whole search space obtaining a Pareto front of size 10. This result has been used to evaluate the performance of the multi-objective Bayesian sequential DOE approach.

Since we consider a randomly chosen initial design matrix, we computed $B = 50$ Monte-Carlo runs. The sample size n of the initial design was fixed to 20. A run was stopped after $T = 30$ iterations. Furthermore, we set R_0 as the identity matrix and σ^2 to 1. The values of α and β were varied in order to understand the contribution of the two terms in the Bayesian utility functions, $U^i(\mathbf{x}_{n+1}|X)$. In Table 1, the performance

(α, β)	$\overline{hd}$	$\overline{enp}$
$(0.00, 1.00)$	0.06 (0.03)	1.48 (0.79)
$(0.20, 0.80)$	0.06 (0.02)	1.70 (1.76)
$(0.50, 0.50)$	0.01 (0.01)	1.38 (0.87)
$(0.80, 0.20)$	0.01 (0.01)	2.38 (0.78)
$(0.90, 0.10)$	0.01 (0.01)	3.80 (1.03)
$(1.00, 0.00)$	0.01 (0.01)	8.40 (3.32)

Table 1: Average perfomance indicators with Monte-Carlo standard deviation in parentheses.

indicators, hypervolume distance (hd) and equal number of points (enp), are reported. The two indicators are the average values of hd and enp calculated over the 50 Monte-Carlo runs.

From Table 1, we notice that the larger the value of α the better the performance of multi-objective Bayesian sequential DOE approach. This behaviour indicates that the first term of $U^i(\mathbf{x}_{n+1}|X)$, which is the one devoted to maximising the expected system responses, is more important than the second term, which is related to the expected gain in Shannon information [5]. When $(\alpha, \beta) = (1.0, 0.0)$, we obtain an average equal number of points close to 8 out of 10 and an average hypervolume indicator close to 0. However, this behaviour can lead to an early convergence on locally optimal solutions with more complex search spaces. A compromise between α and β is probably then preferable.

The multi-objective Bayesian sequential DOE approach has shown promising results but it still needs further improvement. The Bayesian framework should be generalised to a non-linear setting. Furthermore, other sampling techniques for the initial design need to be considered and the multi-objective Bayesian sequential DOE approach should be compared to other methods in order to understand its real power.

References

[1] Grübler A. (2003), Technology and Global Change. Cambridge University Press, New York.

[2] Borrotti, M., Pievatolo, A., Critelli, I., Degiorgi, A., Colledani, M. (2016), A computer-aided methodology for the optimization of electrostatic separation processes in recycling. *Appl. Stoch. Model. Bus.*, 32, 133–148.

[3] Lu, L., Anderson-Cook, C.M., Robinson, T.J. (2011), Optimization of design experiments based on multiple criteria utilizing a Pareto frontier. *Technometrics*, 53, 353–365.

[4] Verdinelli, I., Kadane, J.B. (1992), Bayesian design for maximizing information and outcome. *J. Am.. Stat. Assoc.*, 87 , 510–515.

[5] Chaloner, K., Verdinelli, I. (1995), Bayesian experimental design: a review. *Stat. Sci.*, 10, 273–304.