



**HAL**  
open science

## Revisiting the paper helicopter project using an adaptive surrogate-based approach

Victor Picheny, Rodolphe Le Riche

► **To cite this version:**

Victor Picheny, Rodolphe Le Riche. Revisiting the paper helicopter project using an adaptive surrogate-based approach. [University works] INRA MIA-T; Ecole des Mines de Saint Etienne (EMSE). 2015. hal-01116601

**HAL Id: hal-01116601**

**<https://hal.science/hal-01116601>**

Submitted on 13 Feb 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Revisiting the paper helicopter project using an adaptive surrogate-based approach

Victor Picheny, Rodolphe Le Riche

February 13, 2015

## Abstract

The paper helicopter project has been given to graduate students of the French engineering school *Ecole des Mines de Saint Etienne*, as a practical work of two lectures: design of experiments and global optimization. The project consists of designing a very simplified paper helicopter for maximum flying time. Its main difficulty lies in the very high level of uncertainty, as flying times for the same design can vary very substantially due to experimental conditions or manufacturing imprecisions. The students were asked to combine design of experiments and surrogate modeling techniques (in order to deal with noise) with global optimization strategies. Despite its apparent simplicity, the project was found quite challenging and the strategies experimented by students were eventually close to what would require a real-case problem.

## 1 Problem description

The paper helicopter problem has been initially proposed by Box (1992) to introduce design of experiments and regression concepts. Since then, it has been re-used and modified many times (see for instance Antony and Jiju Antony, 2001; Annis, 2005; Viana et al, 2011), as it offers many pedagogical and practical qualities.

We present here a variant of the problem, as it has been given to graduate students of the French engineering school *Ecole des Mines de Saint Etienne*, as a practical work of two lectures: design of experiments and global optimization.

The helicopter is composed of two superimposed rectangles (a slightly simpler version than the original of Box, 1992); the upper rectangle is cut in two and folded to constitute the rotor and the lower rectangle constitutes the tail. The rotor wings make a right angle with the tail. For stability, two paper clips are attached at the bottom of the tail (Figure 1). The objective is to design the helicopter for maximum flying time.

The helicopter must be cut from an A4 sheet without gluing; hence, we have the following constraints:

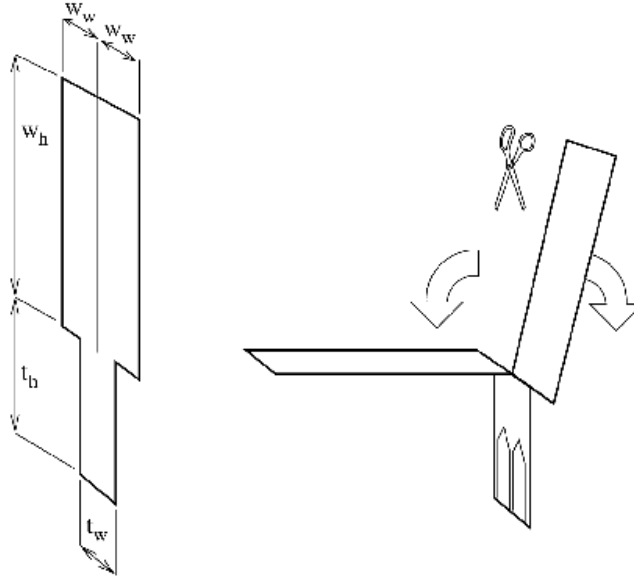
$$\begin{aligned}t_h + w_h &\leq 29.7 \text{ cm} \\ 2w_w &\leq 21 \text{ cm}\end{aligned}$$

The helicopter should hold two large paper clips ( $3 \times 1\text{cm}$ ) placed vertically next to each other, which gives a minimal size for the tail:

$$\begin{aligned}t_w &\geq 2 \times 1 \text{ cm} \\ t_h &\leq 3 \text{ cm}\end{aligned}$$

To ease in particular the design of experiments process, we decided to simplify those constraints to interval constraints on the design variables:

$$\begin{aligned}3.0 &\leq w_h \leq 13.0 \text{ cm} \\ 1.0 &\leq w_w \leq 7.0 \text{ cm} \\ 3.5 &\leq t_h \leq 10.0 \text{ cm} \\ 2.2 &\leq t_w \leq 6.0 \text{ cm}\end{aligned}$$



The flying time  $T$  is measured by throwing the helicopters from a fixed height, as large as possible to minimize timing error (typically, from two storey high), preferably in a wind free environment (a stairwell). This time is a random variable, due mainly to three factors: the experimental conditions (wind, throwing gesture, ...), the measuring error and the manufacturing uncertainties (in particular, the folding angle between the tail and the wings). Hence, we choose as an objective the maximization of the **expected flying time**  $\bar{T}$ , which is measured by the average time of five throws.

The optimization problem is then expressed as follow:

$$\begin{aligned}
 & \max_{t_w, t_h, w_w, w_h} && \bar{T} \\
 & s.t. && 3.0 \leq w_h \leq 13.0 \quad cm \\
 & && 1.0 \leq w_w \leq 7.0 \quad cm \\
 & && 3.5 \leq t_h \leq 10.0 \quad cm \\
 & && 2.2 \leq t_w \leq 6.0 \quad cm
 \end{aligned}$$

## 2 Methodology

### 2.1 General scheme

The overall objective of the project is the global optimization of the paper helicopter based on design of experiments and metamodeling techniques. As the class is divided in groups, each group will experiment a different strategy, all the results being eventually compared. However, all the strategies use the following general scheme:

1. construction of an initial design of experiments (DOE) (50 designs);
2. construction of a metamodel based on the initial set of observations;
3. improvement of the DOE and the metamodel (20 designs);
4. optimization of the helicopter based on the metamodel (5-10 designs).

A “cheaper” variant (timewise) is to consider only three repetitions, 40 initial experiments and 10 improving experiments.

Two types of metamodels are proposed: kriging and polynomial response surface (PRS). The DOE improvement and the final optimization strategy depends on the metamodel.

Steps 1, 2 and 4 are standard in surrogate-based optimization, while step 3 is more original. It consists of adding a set of observations that are chosen based on the initial information, in particular in order to improve the quality of prediction of the metamodel. At that stage, the design space can also be reduced if some regions are clearly identified as non-optimal. It would be possible to suppress this step and transfer the experimental effort (20 observations) to either step 1 or step 4; however, this step is motivated by several reasons:

- it is more efficient than step 1 only, as part of the observations are model-oriented and within a smaller region;
- step 4 requires to perform experiments one-by-one, which is very time consuming while they are done all at once in step 3;
- it is also much more robust than step 4 when the number of observations is too small to fit an accurate model.

## 2.2 Construction of the initial design of experiments and metamodel

The students are asked to implement space-filling methods: centroidal Voronoi tessellations (CVT) and latin hyper-cubes designs (LHD) (optimized with respect to the maximin criterion), and generate DOEs of 50 points (in four dimensions). The initial DOE for the project is the best generated according to several criteria (discrepancy, maximin, etc.).

An R script that creates a `LateX` file from a matrix is used to generate the drawings of the helicopters, which are then printed, cut and folded. The initial set of measurements is performed (250 throws), and the 50 observations are taken as the empirical means over the five repeated throws ( $\hat{T}_i$ ,  $1 \leq i \leq 50$ ). Note that keeping the 250 measurements is also a possibility. To account for the heteroskedasticity of the data, it is also possible to compute the estimated variance of  $\hat{T}_i$ , which is the empirical variance of the five measurements divided by five. This information is not critical when classical linear regression is used, but appears to be very useful for the kriging model.

The initial metamodels are fit using the R functions `lm()` (library `stats`) for PRS and `km()` (library `DiceKriging`) for kriging. On each case, the model is adjusted as best as possible to the data, by choosing the polynomial terms based on ANOVA and cross-validation for PRS, and by choosing the trend and covariance functions for kriging.

## 2.3 Improvement of the DOE and the metamodel

For this problem, 50 observations is quite insufficient to build an accurate metamodel. However, it may be sufficient to narrow some of the bounds and remove some of the non-optimal regions, which can be seen as a pre-optimization step.

**Criteria** The initial DOEs are improved by adding a set of 20 observations using model-oriented criteria, that is, the new observations are added to improve the quality of the initial model. For PRS, the *D-criterion* (determinant of the Fisher information matrix (see John and Draper (1975))) is used. For kriging, the *IMSE criterion* is used (see Sacks et al (1989)). On both cases, the criterion is of the form:

$$C(\mathbf{x}_{p+1}, \dots, \mathbf{x}_{p+n}) = \phi(\mathbf{x}_1, \dots, \mathbf{x}_{p+n}), \quad (1)$$

where  $\mathbf{x} = (t_w, t_h, w_w, w_h)$  denotes a design,  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is the current DOE,  $\mathbf{x}_{p+1}, \dots, \mathbf{x}_{p+n}$  is a set of candidate additional observations, and  $\phi$  is a scalar function defined by the model. Calculation details are given in 4.

Improving the DOE amounts to solve the following optimization problem:

$$\begin{aligned} \max \quad & C(\mathbf{x}_{p+1}, \dots, \mathbf{x}_{p+n}) \\ \text{s.t.} \quad & \mathbf{x}_{p+i} \in D \quad \forall 1 \leq i \leq n. \end{aligned} \quad (2)$$

**Optimization of the criteria** Solving this problem is challenging, as it can be of high dimension ( $4 \times p$ ), and the objective is strongly multimodal (it is for instance invariant by a permutation of any  $\mathbf{x}_{p+i}, \mathbf{x}_{p+j}$ ). Therefore, it constitutes an excellent application for global optimization algorithms.

The students have first implemented different strategies: Nelder-Mead, CMA-ES, DIRECT, etc. Then, they are tested on two analytical functions (a quadratic function, and Michalewicz’s function) before applying it to the DOE problem.

## 2.4 Global optimization of the helicopter

**Polynomial response surface** The final step consists of performing an experiment where the response surface is maximum. The metamodel can be updated with this observation, and the procedure repeated a couple of times if the maximum changes significantly. Finding the maximum of the surface requires solving a global optimization problem (in dimension 4). This is done using the algorithm previously implemented.

Note that here, the strategy assumes that the metamodel is sufficiently accurate to provide a good estimate of the maximizer of the objective function.

**Kriging** Contrarily to PRS, kriging is well-suited for sequential design, and the final step consists of several iterations of the Efficient Global Optimization (EGO) algorithm (see Jones et al (1998)). Similarly to the PRS case, this strategy amounts to find the maximizer of a function defined by the kriging mean and variance called the *Expected Improvement*, defined as:

$$EI_n(\mathbf{x}) = (m_n(\mathbf{x}) - y_{max})\Phi\left(\frac{m_n(\mathbf{x}) - y_{max}}{s_n(\mathbf{x})}\right) - s_n(\mathbf{x})\phi\left(\frac{m_n(\mathbf{x}) - y_{max}}{s_n(\mathbf{x})}\right), \quad (3)$$

where  $\Phi$  and  $\phi$  denote the Gaussian cumulative distribution function and probability density, respectively,  $m_n$  and  $s_n$  denote the kriging mean and variance and  $y_{max}$  is the maximum of the current set of observations.

As the observations are noisy, the original EGO strategy should not be applied, so a heuristic modification (called the *reinterpolation* procedure) is used instead (see Forrester et al (2006)).

On both cases, as doing experiments one at a time is particularly time-consuming, only a few designs are tested.

## 3 Discussion

**Scientific challenges and potential advanced topics** In our setting, a relatively large proportion of a space-filling design (say, 20 to 30%) experience unstable flights, which result in a lot shorter recorded times. Those instability are not systematic, and often happen two or three times out of five.

A first consequence is that noise is largely heteroskedastic, which requires a statistical treatment. It even seems reasonable to infer a noise model, either continuous (noise varying smoothly with some design variables) or discrete (by partitioning the space into stable/unstable regions).

Another interesting point is that optimal designs seem to be at the limit of stability (helicopter as big as possible but stable). An alternative approach would be to define a constrained optimization problem. However, the stochasticity of the constraint evaluation (stability) makes the problem particularly challenging.

Finally, it seems that a critical factor, hidden in our approach, is the angle between the wing and the body. The students were told to ply the paper to an approximate 90 degree angle, but this is quite difficult to control. As shown in Annis (2005), this angle has a strong influence on the helicopter velocity, and its uncertainty may explain a large proportion of the noise in the observations. Controlling this angle (by changing the helicopter design), or coupling the statistical approach with a physical model (Annis, 2005) might provide interesting solutions.

**Pedagogical aspects** Despite its apparent simplicity, this was found quite rich and challenging regarding both statistical and optimization aspects, which makes it an excellent tool for teaching.

First, each observation has an actual cost (timely speaking), hence it really motivates the students to double-check their codes, designs of experiments, etc., before performing any actual experiment. Besides,

contrarily to computer experiments for instance, the study cannot be redone, so the students have to deal with the data they have acquired, which sometimes lead to interesting compromises.

Second, this problem share many characteristics with current challenging computer experiments and optimization problems: it can be seen as an expensive black-box, stochastic, multi-physics. Hence, while the primary purpose of the problem was to teach engineers statistics (Box, 1992; Antony and Jiju Antony, 2001), we feel that it can also serve as an introduction to advanced research topics.

## 4 Appendix

### 4.1 D criterion

Let the polynomial reponse surface be of the form:

$$y(\mathbf{x}) = \sum_{k=1}^p f_k(\mathbf{x}) + \varepsilon \quad (4)$$

For a design of experiments  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , the corresponding Fisher information matrix is given by:

$$\mathbf{M} = \mathbf{F}^T \mathbf{F}, \quad (5)$$

with  $F_{ij} = f_j(\mathbf{x}_i)$ . The D-criterion is then  $D = \det(\mathbf{F}^T \mathbf{F})$ .

In our context, part of the DOE is fixed (the observations already performed), and we would like to optimize the criterion with respect to a set of  $p$  new observations. The form of  $\mathbf{M}$  allows us to avoid computing the expensive determinant of the new  $(n+p) \times (n+p)$  matrix for every set of candidate new observations. Indeed, let  $\mathbf{F}_n$  be the part of  $\mathbf{F}$  corresponding to the current DOE and  $\mathbf{F}_p$  the part corresponding to the candidate points. We have:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_n \\ \mathbf{F}_p \end{bmatrix} \quad (6)$$

$$\det(\mathbf{F}^T \mathbf{F}) = \det(\mathbf{F}_n^T \mathbf{F}_n + \mathbf{F}_p^T \mathbf{F}_p) \quad (7)$$

$$= \det(\mathbf{F}_n^T \mathbf{F}_n) \det\left(\mathbb{I}_p + \mathbf{F}_p [\mathbf{F}_n^T \mathbf{F}_n]^{-1} \mathbf{F}_p^T\right) \quad (8)$$

### 4.2 IMSE criterion

The IMSE criterion is defined as the integral of the kriging prediction variance over the design space:

$$IMSE = \int_D s_n^2(\mathbf{x}) d\mathbf{x}, \quad (9)$$

where  $s_n^2(\cdot)$  denotes the prediction variance based on  $n$  observations. In our context,  $p$  observations are to be added to the design, so the criterion to optimize is:

$$IMSE(\mathbf{x}_{p+1}, \dots, \mathbf{x}_{p+n}) = \int_D s_{n+p}^2(\mathbf{x}) d\mathbf{x}. \quad (10)$$

The kriging prediction variance doe not depend on the value of the observations, hence  $s_{n+p}$  can be computed by updating the current kriging model with dummy observation values. Note that kriging update equations can be use for speed-up, but they are not considered here.

The integral over  $D$  must be done numerically, using Gauss quadrature or Quasi Monte-Carlo methods for instance. As the integral calculation is embedded in an optimization loop, it is important to always use the same integration points, which creates some bias but avoid having to treat numerical noise in the objective function.

## References

- Annis DH (2005) Rethinking the paper helicopter: Combining statistical and engineering knowledge. *The American Statistician* 59(4):320–326
- Antony J, Jiju Antony F (2001) Teaching the taguchi method to industrial engineers. *Work Study* 50(4):141–149
- Box GE (1992) Teaching engineers experimental design with a paper helicopter. *Quality Engineering* 4(3)
- Forrester A, Keane A, Bressloff N (2006) Design and Analysis of “Noisy” Computer Experiments. *AIAA journal* 44(10):2331
- John RS, Draper NR (1975) D-optimality for regression designs: a review. *Technometrics* 17(1):15–23
- Jones D, Schonlau M, Welch W (1998) Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 13(4):455–492
- Sacks J, Welch W, Mitchell T, Wynn H (1989) Design and analysis of computer experiments. *Statistical science* pp 409–423
- Viana FA, Haftka RT, Hamman R, Venter G (2011) Efficient global optimization with experimental data: revisiting the paper helicopter design. Gainesville: University of Florida