Les Cahiers du GERAD ISSN: 0711–2440

Dynamic improvements of static surrogates in direct search optimization

C. Audet,
J. Côté-Massicotte

G–2019–10

January 2019

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2019
– Bibliothèque et Archives Canada, 2019

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2019
– Library and Archives Canada, 2019

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Dynamic improvements of static surrogates in direct search optimization

Charles Audet
Julien Côté-Massicotte

GERAD, Montréal (Québec), Canada, H3T 2A7

Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec) Canada, H3C 3A7

charles.audet@gerad.ca
julien.cote-massicotte@polymtl.ca

January 2019
Les Cahiers du GERAD
G–2019–10

Abstract: The present work is in a context of derivative-free optimization involving direct search algorithms guided by surrogate models of the original problem. These models are classified into two categories: static surrogates and dynamic models. This work introduces the quadratic hybrid model (HQM), that dynamically corrects information from a static surrogate. Instead of bringing an additive or multiplicative correction, the HQM generalizes these two types of corrections by considering the static model as an input variable of the quadratic model. Numerical tests are performed with the Mads algorithm on three multidisciplinary and one simulation-based engineering problems. The results show that the contribution of the HQM to the Mads algorithm is to solve problems at greater precision for the same computational budget.

Keywords: Surrogate-assisted optimization, Static surrogate, Quadratic model, Mesh Adaptive Direct Search

---

Acknowledgments: This work is supported by the NSERC grant RGPIN-2015-05311.

1 Introduction

The surrogate management framework [13] proposes ways to use a surrogate of an optimization problem as a tool to guide a direct search optimization algorithm. This framework has been used in many contexts [1, 21, 25, 35, 38]. A key feature of a surrogate is that it evaluates much faster than the true problem. Surrogates can be classified into two categories [7]. Static surrogates are usually simplifications of the true optimization problem, they include simplified physic models, or models in which internal tolerances are relaxed. Static surrogates are not required to be good approximations of the true problem in order to be useful to the optimization process. The second category are models, including quadratic, RBF and Gaussian models among others. These models can be updated when new information is available, and are designed to be used as approximation of the true function. These models are functions of the n optimization variables.

The Mesh Adaptive Direct Search (Mads) algorithm [5] is an example of a direct search method that exploits surrogates to solve an optimization problem. It has been shown that Mads is generally more efficient when assisted by a surrogate [4, 8, 9, 16, 37].

Research on surrogate-assisted direct search optimization usually involves the use of either the static surrogate or the dynamic models. Some authors have studied combination of both. For example, [24] propose additive and multiplicative ways of combining static surrogates with dynamic models. The objective of the present research is to propose an hybrid strategy to build a quadratic model whose input is not only the n optimization variables, but also a supplementary variable taking the value of the static surrogate. This yields flexible models that inherit from the global properties of the static surrogate and the local precision of quadratic model.

The paper is structured as follows. Section 2 brushes a picture of existing local and global tools in surrogate-assisted optimization. Section 3 proposes an hybrid strategy that builds a quadratic model in which one of its input is the static surrogate value. Numerical experiments are conducted in Section 4, and the proposed approach is compared with those that use either the static or dynamic model, and with the one that uses neither. Concluding remarks close the paper.

2 Surrogate-assisted optimization

This work target general optimization problems of the form :

$$\min_{x \in \Omega} f(x), \tag{1}$$

where $\Omega = \{x \in X : c_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n$ denotes the feasible region and where the functions $f, c_j : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ for $j \in J = \{1, 2, \ldots, m\}$ are costly to evaluate, only computable through a simulation and whose derivatives are not accessible. The set X is a subset of $\mathbb{R}^n$, typically of the form $X = \{x : \ell \leq x \leq u\}$ where ℓ and u are in $\overline{\mathbb{R}}^n$. The simulation cannot be launched at points outside of the set X. The contraints c_j, $j \in J$ can be violated, and the simulation output is valid. These constraints must be satisfied at the final solution produced by the optimization algorithm. Detailed descriptions of this class of problems as well as algorithms to solve them are found in the textbooks [7, 17]. One efficient way to solve problems from this class is through surrogate-assisted optimization.

2.1 Surrogates and models

Launching the simulation to evaluate the functions defining Problem (1) may be time-consumming. There are situations where each call to the simulation requires seconds [15] (valve train design), minutes [3] (spent potliner treatment), hours [10] (parameter tuning) or even days [26] (trailing-edge noise reduction). One way to deal with such problem is to use a simplified formulation called a *surrogate problem*.

Surrogates can be constructed by using simplified physics, or by relaxing internal tolerances within the blackbox simulation. These surrogates are often provided by the designer of the simulation. We refer to these as *static surrogates*. Surrogates can also be built and updated as the optimization process is deployed. Interpolation or regression methods can be applied to mimic the output of the simulation using quadratic [19, 16] or polynomial [30] approximations, DACE Kriging [12, 23, 36, 18], treed Gaussian processes [21], LOWESS models [37], radial basis functions [11, 31, 34, 41] or even ensembles of surrogates [9]. We refer to these as *dynamic models*.

The surrogate management framework [13] details how to exploit a surrogate problem to reduce the overall computational optimization time. It is summarized in Algorithm 1, where the surrogate functions are denoted by $\tilde{f}$ and $\tilde{c}$.

Algorithm 1: Surrogate management framework in Mads (adapted from [7, 13])

Given the true functions $f, c_j : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $j \in J$, and their surrogates $\tilde{f}, \tilde{c}_j : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $j \in J$,

1. Search step (optional)
   - use the surrogate problem to generate a list $\mathcal{L}$ of trial points
   - evaluate the true problem functions at points in $\mathcal{L}$ in an opportunistic way
   - proceed to 3 if a new incumbent solution is generated, otherwise go to 2
2. Poll step
   - generate a set of poll points P, and order them using the surrogate problem
   - evaluate the true problem functions at points in P in an opportunistic way
3. Update optimization algorithm
   - update all algorithmic parameters and check stopping criteria
   - terminate or go to 4
4. Calibration of the surrogate (for dynamic surrogates)
   - recalibrate the surrogate functions using the new function values obtained in Steps 1 and 2

There are other types of surrogates. Polyak and Wetter [32] propose a framework to automatically select the precision of an adaptative surrogate. Chen and Kelley [14] propose a framework in which the true optimization cannot be evaluated, and only Monte-Carlo simulations estimations are available. Previous studies inclue additive ($g(x) \approx \tilde{g}(x) + A(x)$, $A : \mathbb{R}^n \mapsto \mathbb{R}$) and multiplicative ($g(x) \approx M(x)\tilde{g}(x)$, $M : \mathbb{R}^n \mapsto \mathbb{R}$) [2, 20, 24] corrections to the function $\tilde{g}$. We propose to dynamically correct a static surrogate. More precisely, we extend a quadratic model so that in addition to taking the n problem variables as input, it also takes the static surrogate value as input. The next subsection presents quadratic models.

2.2 Quadratic models

Let $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ be a generic function, that we wish to approximate. In what follows, g can be any function f or c_j from Problem (1). The contents of this section is covered into depth in [7, 16, 17, 19]. The general form of a quadratic model of the function g is $Q(x) = \alpha_0 + \alpha^\top x + \frac{1}{2}x^\top Hx$ where $\alpha_0 \in \mathbb{R}$, $\alpha \in \mathbb{R}^n$, $H = H^\top \in \mathbb{R}^{n\times n}$. This model depends on a total of $q + 1 = (n + 1)(n + 2)/2$ parameters because H is symmetric. Using a basis of the space of degree 2 polynomials, the quadratic function may be compactly rewritten as $Q(x) = \alpha^\top \rho(x)$, where

$$\rho(x)^\top = (\rho_0(x), \rho_1(x), \ldots, \rho_q(x)) = \left(1, x_1, x_2, \ldots, x_n, \tfrac{x_1^2}{2}, \tfrac{x_2^2}{2}, \ldots, \tfrac{x_n^2}{2},\ x_1x_2,\ x_1x_3, \ldots,\ x_{n-1}x_n\right)$$

and where $\alpha \in \mathbb{R}^{q+1}$ is redefined. The next definitions ensure existence and unicity of a quadratic model.

Definition 1 (Poised for quadratic regression) *The set of points points* $\mathbb{Y} = \{y^0, y^1, \ldots,\ y^p\} \subset \mathbb{R}^n$ *with* $p \geq q$, $q + 1 = (n + 1)(n + 2)/2$ *is* poised for quadratic regression *if the matrix*

$$M(\rho, \mathbb{Y}) = \begin{bmatrix} \rho_0(y^0) & \rho_1(y^0) & \ldots & \rho_q(y^0) \\ \rho_0(y^1) & \rho_1(y^1) & \ldots & \rho_q(y^1) \\ \vdots & \vdots & \vdots & \vdots \\ \rho_0(y^p) & \rho_1(y^p) & \ldots & \rho_q(y^p) \end{bmatrix} \in \mathbb{R}^{(p+1)\times(q+1)}$$

has rank equal to $q + 1$.

The minimum Frobenius norm model [19, 33] is an alternative when there are fewer than $q+1$ points available to construct the model.

Definition 2 (Poised for minimum Frobenius norm modelling) *Let* $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $\rho_L(x) = (1, x_1, x_2, \ldots, x_n)^\top$ *and* $\rho_Q(x) = (\frac{x_1^2}{2}, \frac{x_2^2}{2}, \ldots, \frac{x_n^2}{2},\ x_1x_2,\ x_1x_3, \ldots,\ x_{n-1}x_n)^\top$. *The set of points* $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$ *with* $n+1 < p+1 < (n+1)(n+2)/2$ *is* poised for minimum Frobenius norm modelling *if there is a unique optimal solution to:*

$$\begin{aligned} \min_{\alpha_L, \alpha_Q} \quad & \tfrac{1}{2}\|\alpha_Q\|^2 \\ s.t. \quad & g(y^i) = \rho_L(y^i)\alpha_L + \rho_Q(y^i)\alpha_Q \quad \text{for } i = 0, 1, \ldots, p\ , \end{aligned}$$

where $\alpha_L \in \mathbb{R}^{n+1}$, $\alpha_Q \in \mathbb{R}^N$ *and* $N = n(n+1)/2$.

To construct the quadratic function $Q(x)$, it suffices to identify $\alpha = [\alpha_L\ \alpha_Q]^\top$. If $n < p < q$ and $\mathbb{Y}$ is poised for minimum Frobenius norm modelling, then α is the unique solution to the problem in Definition 2. Otherwise, if $p \geq q$ et $\mathbb{Y}$ is poised for quadratic regression, then $\alpha \in \mathbb{R}^{q+1}$ is the solution that minimizes $\|M(\rho, \mathbb{Y})\alpha - g(\mathbb{Y})\|^2$ where $M(\rho, \mathbb{Y})$ is the matrix from Definition 1 and $g(\mathbb{Y}) = \big(g(y^0), g(y^1), \ldots, g(y^p)\big)^\top$.

3 A quadratic model involving a static surrogate

The objective of the present work is to derive a methodology exploiting the information contained in a static surrogate, and to improve it through the construction of a dynamic model.

Definition 3 (Hybrid Model) *Let* $\tilde{g} : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $\tilde{g} \neq 0$, *be a static surrogate of the function* $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$. *An* hybrid model $\hat{g} : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ *of* g *is a dynamic model of that takes into account the static model* $\tilde{g}$: $\hat{g}$ *depends on* $x = [x_1, x_2, \ldots, x_n]^\top$, *as well as on* $x_0 := \tilde{g}(x) \in \mathbb{R}$.

Instead of simply applying an additive or multiplicative correction, the hybrid model treats the static model as an input. Thus, no information or approximation of the derivatives of the static model is necessary for the construction of the hybrid model, which is well suited in a derivative-free optimization context. A hybrid model can only be built if the static model $\tilde{g}$ is non-trivial. Indeed, in the trivial case $\tilde{g} = 0$, the static model would not bring any new information on g and would be detrimental to the construction of the hybrid model.

3.1 A Hybrid quadratic model

We introduce the *Hybrid Quadratic Model* (HQM) that takes advantage of the information contained in a static surrogate and of the local properties of a quadratic model. In order to achieve this, we revisit the principle of the quadratic model presented in section 2.1, by introducing a variable labelled x_0 representing the value of the static model. The HQM is constructed on a space of dimension $n+1$, where n is the number of variables in Problem (1). Formally, we consider the static surrogate $\tilde{g} : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ of the function $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ and

$$\phi(x)^\top = (\phi_0(x), \phi_1(x), \ldots, \phi_t(x)) = \left(1, x_0, x_1, \ldots, x_n, \tfrac{x_0^2}{2}, \tfrac{x_1^2}{2}, \ldots, \tfrac{x_n^2}{2},\ x_0x_1,\ x_0x_2, \ldots,\ x_{n-1}x_n\right)$$

where $x_0 = \tilde{g}(x)$ and $t+1 = (n+2)(n+3)/2$. The HQM of $g(x)$ is

$$\hat{g}(x) = \beta^\top \phi(x), \tag{2}$$

for some $\beta \in \mathbb{R}^{t+1}$. This construction requires a set of interpolation points $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$, $p \in \mathbb{N}$. The model is obtained by solving the least-square minimization problem

$$\min_{\beta \in \mathbb{R}^{t+1}} \sum_{y \in \mathbb{Y}} (g(y) - \hat{g}(y))^2. \tag{3}$$

The following definitions are needed to characterize the HQM, to ensure the existence and uniqueness of a HQM. They are similar to those from Section 2.1.

Definition 4 (Poised for hybrid quadratic regression) *The set of points* $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$ *with* $p \geq t$, $t + 1 = (n+2)(n+3)/2$ *and* $g(\mathbb{Y}) = \big(g(y^0), g(y^1), \ldots, g(y^p)\big)^\top$ *with* $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ *is said to be* poised for hybrid quadratic regression *if the matrix*

$$M(\phi, \mathbb{Y}) = \begin{bmatrix} \phi_0(y^0) & \phi_1(y^0) & \ldots & \phi_t(y^0) \\ \phi_0(y^1) & \phi_1(y^1) & \ldots & \phi_t(y^1) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_0(y^p) & \phi_1(y^p) & \ldots & \phi_t(y^p) \end{bmatrix} \in \mathbb{R}^{(p+1)\times(t+1)}$$

is of rank $t+1$ *and if* $g(\mathbb{Y}) \in \mathbb{R}^{p+1}$.

As with the quadratic model, we can use the Frobenius norm for cases with less than $t+1$ points.

Definition 5 (Poised for minimum Frobenius norm hybrid modelling) *Let* $\tilde{g} : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ *be a static model of* $g : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $\phi_L(x) = (1, x_0, x_1, \ldots, x_n)^\top$ *and* $\phi_Q(x) = (\frac{x_0^2}{2}, \frac{x_1^2}{2}, \ldots, \frac{x_n^2}{2},\ x_0x_1,\ x_0x_2, \ldots, x_{n-1}x_n)^\top$, *where* $x_0 = \tilde{g}(x)$. *The set of points* $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$, $n+2 < p+1 < (n+2)(n+3)/2$ *is* poised for minimum Frobenius norm hybrid modelling *if there is a unique optimal solution to*

$$\begin{aligned} \min_{\beta_Q \in \mathbb{R}^N} \quad & \tfrac{1}{2}\left\|\beta_Q\right\|^2 \\ \text{s.t.} \quad & g(y^i) = \phi_L(y^i)\beta_L + \phi_Q(y^i)\beta_Q \quad \text{for } i = 0, 1, \ldots, p, \end{aligned} \tag{4}$$

where $\beta_L \in \mathbb{R}^{n+2}$ *and* $N = (n+1)(n+2)/2$.

We next analyze into more detail the HQM specificities. Consider a set of $p+1$ interpolation points $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\}$. The analysis leading to the computation of $\beta = [\beta_L\ \beta_Q]^\top$ is partitioned into five cases.

Case I : $p = t$ **and the set** $\mathbb{Y}$ **is poised for hybrid quadratic regression.** Problem (3) reduces to solving

$$M(\phi, \mathbb{Y})\beta = g(\mathbb{Y}) = (g(y^0), g(y^1), \ldots, g(y^p)). \tag{5}$$

The system has a unique solution, because the matrix $M(\phi, \mathbb{Y})$ nonsingular.

Case II : $p > t$ **and the set** $\mathbb{Y}$ **is poised for hybrid quadratic regression.** The system (5) is over-determined, but the solution β is obtained using the pseudo-inverse matrix

$$\beta = [M(\phi, \mathbb{Y})^\top M(\phi, \mathbb{Y})]^{-1} M(\phi, \mathbb{Y})^\top g(\mathbb{Y}).$$

Case III : $n+1 < p < t$ **and the set** $\mathbb{Y}$ **is poised for Frobenius** HQM**.** The system (5) is under-determined. It is however possible to find the unique solution β using the Frobenius quadratic model to the HQM by minimizing the norm of the quadratic coefficients of Equation (2). Writing $\beta = [\beta_L\ \beta_Q]^\top$ with $\beta_L \in \mathbb{R}^{n+2}$, $\beta_Q \in \mathbb{R}^N$ and $N = (n+1)(n+2)/2$ allows us to write Equation (2) as $\hat{g}(x) = \phi_L(x)\beta_L + \phi_Q(x)\beta_Q$, where $\phi_L(x) = (1, x_0, x_1, x_2, \ldots, x_n)^\top$, $\phi_Q(x) = (\frac{x_0^2}{2}, \frac{x_1^2}{2}, \frac{x_2^2}{2}, \ldots, \frac{x_n^2}{2},\ x_0x_1,\ x_0x_2, \ldots,\ x_{n-1}x_n)^\top$ and $x_0 = \tilde{g}(x)$. The vector β_Q is found by solving

$$\begin{aligned} \min_{\beta_Q \in \mathbb{R}^N} \quad & \tfrac{1}{2}\left\|\beta_Q\right\|^2 \\ \text{s.t.} \quad & g(\mathbb{Y}) = M(\phi_L, \mathbb{Y})\beta_L + M(\phi_Q, \mathbb{Y})\beta_Q, \end{aligned} \tag{6}$$

where $M(\phi_L, \mathbb{Y})$ and $M(\phi_Q, \mathbb{Y})$ are matrices of from Definition (4), but relative to the bases ϕ_L and ϕ_Q, respectively. The hypothesis on $\mathbb{Y}$ ensures a unique solution. Let $\mathcal{L}(\beta, \lambda) = \frac{1}{2}\left\|\beta_Q\right\|^2 - \lambda^\top (M(\phi_L, \mathbb{Y})\beta_L + M(\phi_Q, \mathbb{Y})\beta_Q - g(\mathbb{Y}))$ be the Lagrangean function of Problem (6), where $\lambda = [\lambda_0, \lambda_1, \ldots, \lambda_p]^\top$. The minimizer is found by solving the system of three equations $g(\mathbb{Y}) = M(\phi_L, \mathbb{Y})\beta_L + M(\phi_Q, \mathbb{Y})\beta_Q$, $0 = \lambda^\top M(\phi_L, \mathbb{Y})$ and $\beta_Q = \lambda^\top M(\phi_Q, \mathbb{Y})$ that are compactly written as

$$F(\phi, \mathbb{Y}) \begin{bmatrix} \lambda \\ \beta_L \end{bmatrix} = \begin{bmatrix} M(\phi_Q, \mathbb{Y}) M(\phi_Q, \mathbb{Y})^\top & M(\phi_L, \mathbb{Y}) \\ M(\phi_L, \mathbb{Y})^\top & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \beta_L \end{bmatrix} = \begin{bmatrix} g(\mathbb{Y}) \\ 0 \end{bmatrix}, \tag{7}$$

where $F(\phi, \mathbb{Y}) \in \mathbb{R}^{(p+n+3)\times(p+n+3)}$. The next proposition summarizes how to validate that a set is poised for minimum Frobenius norm hybrid modelling.

Proposition 1 *A set of points $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$, $n+2 < p+1 < (n+2)(n+3)/2$, is poised for minimum Frobenius norm hybrid modelling if and only if the system (7) has a unique solution, i.e., $F(\phi, \mathbb{Y})$ is nonsingular.*

Proof. Let $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\} \subset \mathbb{R}^n$, $n+2 < p+1 < (n+2)(n+3)/2$ be poised for Frobenius HQM. The proof follows from the equivalent statements:

$$\mathbb{Y} \text{ is poised for Frobenius HQM} \iff (4) \text{ has a unique solution} \iff (7) \text{ has a unique solution.}$$

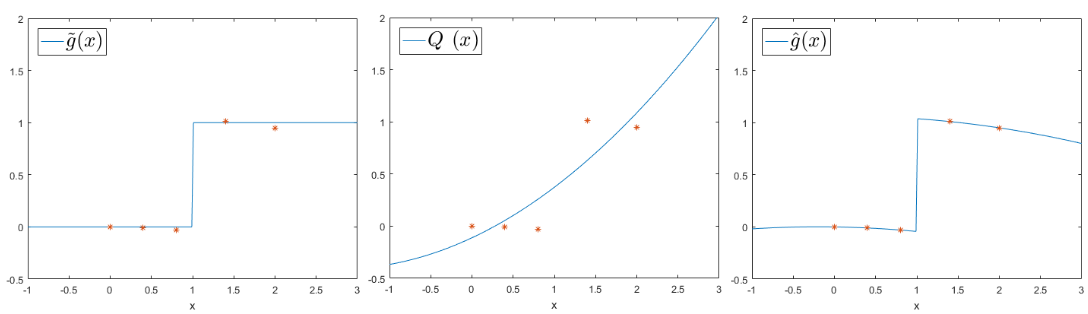
□

It follows that λ and β_L are obtained by solving Equation (7), and the remaining term is $\beta_Q = \lambda^\top M(\phi_Q, \mathbb{Y})$.

Case IV : $p > n+1$ but the set $\mathbb{Y}$ is not poised for any Frobenius HQM. There are infinitely many solutions to (5) or to (7). We select the one with least norm.

Case V : $p \le n+1$. Not enough information is collected to build a quadratic model.

We conclude this subsection with an example illustrating the motivation for building hybrid models. Let $\tilde{g}$ be a static surrogate that captures the discontinuity of a function $g : \mathbb{R} \mapsto \mathbb{R}$, defined as the step function $\tilde{g}(x) = 0$ if $x < 1$ anf $\tilde{g}(x) = 1$ otherwise. Suppose now that five function values are known for g, as illustrated by the dots in Figure 1. The plot on the left shows the static model, the central one shows the Frobenius norm model from Definition 2 and the plot on the right shows the hybrid HQM. This last model is the only one that captures both the discontinuity from the static model and the curvature from the quadratic model.



(a) Static surrogate $\tilde{g}$ **(b) Quadratic model Q** **(c) Hybrid model $\hat{g}$**

Figure 1: A static, quadratic and hybrid model of a function $g(x)$

3.2 Algorithmic approach

Consider the situation where we are given the static surrogate of Problem (1):

$$\min_{x \in \tilde{\Omega}} \tilde{f}(x), \tag{8}$$

with $\tilde{\Omega} = \{x \in X : \tilde{c}_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n$ and $\tilde{f}, \tilde{c}_j : X \to \overline{\mathbb{R}}$ for $j \in J$. Problem (1) is referred to as the *real* problem and (8) as the *static surrogate*. Using the methodology from Section 3.1, we create $m+1$ local HQM models $\hat{f}^k \approx f, \hat{c}_1^k \approx c_1, \ \ldots, \ \hat{c}_{m-1}^k \approx c_{m-1}$ and $\hat{c}_m^k \approx c_m$ to be used in the poll step of Mads. More specifically, consider the poll set P^k at iteration k around the incumbent solution x^k. The static surrogate functions are evaluated at all points in P^k. The bounds $\max_i$ and $\min_i$, $i \in \{1, 2, \ldots, n\}$, as well as center μ_i and radius r_i are then computed for each $i \in \{1, 2, \ldots, n\}$:

$$\max_i = \max\{y_i : y \in P^k\}, \ \min_i = \min\{y_i : y \in P^k\}, \ \mu_i = \tfrac{1}{2}(\max_i + \min_i), \ r_i = \tfrac{1}{2}(\max_i - \min_i). \tag{9}$$

Let $\mathcal{B}_\infty(\mu; r) = \{x \in \mathbb{R}^n : |x_i - \mu_i| \leq r_i, i \in \{1, 2, \ldots, n\}\}$ be the smallest box containing P^k. We introduce a scalar $\rho \in (0, \infty)$ that may be provided by the user to delimit the region containing the trial points used to build the HQM. The interpolation points $\mathbb{Y} = \{y^0, y^1, \ldots, y^p\}$ will be selected from $\mathcal{B}_\infty(\mu; \rho r)$ as well as from the cache V^k (i.e. the set of points for which both the true and surrogate problem were evaluated by the start of iteration k).

The trial points in P^k are then ordered according to their HQM values $\hat{f}$ and $\hat{h}$ as detailled in [7, p 240]. The ordered poll points are then evaluated with an opportunist strategy. Iteration k is interrupted as soon as a new feasible incumbent solution is found. With the progressive barrier [6], the iteration is not interrupted if only a new infeasible incumbent solution is found. Algorithm 2 summarizes a Mads iteration with HQM.

Algorithm 2: Iteration k of Mads with an HQM

Given the true functions, their static surrogates $f, c_j, \tilde{f}, \tilde{c}_j : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $j \in J$, and starting point $x^0 \in X$

1. Standard search step
2. Poll step
   - 2.1. HQM construction
     - choose a poll set P^k and compute $\tilde{f}(y)$ and $\tilde{c}(y)$ for each $y \in P^k$
     - set μ and r using Equation (9) and let $\mathbb{Y} = \mathcal{B}_\infty(\mu; \rho r) \cap V^k$
     - if $|\mathbb{Y}| \geq n+2$ then construct the $m+1$ HQM $\hat{f}^k(x), \hat{c}_1^k(x), \ \ldots, \ \hat{c}_m^k(x)$
     - if $|\mathbb{Y}| < n+2$ then set $\hat{f}^k(x) \leftarrow \tilde{f}(x), \hat{c}_1^k(x) \leftarrow \tilde{c}_1(x), \ \ldots, \ \hat{c}_m^k(x) \leftarrow \tilde{c}_m(x)$
   - 2.2. Ordering
     - order the trial points in P^k according to the values $\hat{f}^k(x)$ and $\hat{h}^k(x)$
     - evaluate the functions of the true Problem (1) at points of P^k with an opportunist strategy
3. Standard parameter update step and termination

Figure 2 illustrates these steps on an example in $\mathbb{R}^2$. The dark circles represents the poll set P^k and the white circles . The central graph shows the center μ and the box $\mathcal{B}_\infty(\mu; r)$. The graph on the right shows the region $\mathcal{B}_\infty(\mu; \rho r)$ to construt $\mathbb{Y}$. The points y^4 and y^5 are not used in $\mathbb{Y}$.

4 Computational experiments

To test our approach, we need optimization problems that are accompanied by a static surrogate, and that are not contaminated by numerical noise, otherwise the quadratic models would be irrelevant. The experiments are conducted on three multidisciplinary optimization (MDO) problems from two families, and on a water treatment engineering problem, using version 3.8.0 of the NOMAD software package [22].

Four ordering strategies are compared for the poll step of Mads. The most basic one is called *success* and does not use any models at it orders the trial point by increasing angle with the last direction of success [5]. The second one is called *quad* and only uses quadratic models; it is the default in NOMAD

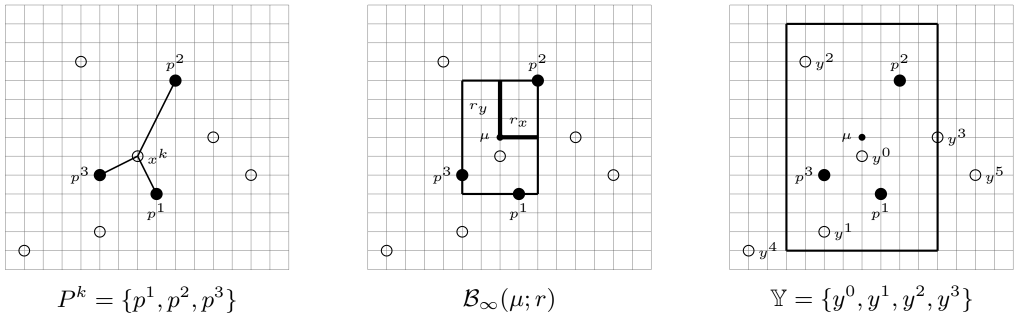


Figure 2: Construction of $\mathbb{Y}$ with ratio parameter $\rho = 2$

when no static surrogate is provided. The third one is called *static* and used exclusively the provided static surrogate. Finally, the last one is called *hqm* and correspond to the new strategy proposed in the present work. Results are presented summarized through data profiles [29] that show the proportion of problem solved within a fixed tolerance τ as a function of the number of calls to the true simulation. Three values of τ are tested.

4.1 Multidisciplinary optimization problems

In MDO, disciplines are interconnected in such a way that the input of one discipline is the output of the others. Several iterations are necessary to stabilize input and output values. Two precision parameters are used here to define the true and the surrogate problem. A threshold $\epsilon > 0$ on the magnitude of the difference between consecutive iterations, and a bound $n_{\max}$ on the number of allowed iterations.

We first consider the Simple_MDO_2n_Variables [40] family of problems $\min_x a_1^*(x) + a_2^*(x)$ for $x \in X = [-100; 100]^{2n}$, with $2n$ variables where

$$a_1 = \frac{x_1^2 - 2x_1 + 1 + \sum_{i=2}^{n}(2x_i^2 - 2x_{i-1}x_i + 1)}{1 + 0.5a_2}, \quad a_2 = \left(\sum_{i=n+1}^{2n}(2x_i^2 - 2x_{i-1}x_i + 1)\right)\sqrt{1 + 0.5a_1}\ .$$

The stars appearing in a_1^* and a_2^* indicate that the values converged during the MDO process within the tolerance parameters: $\epsilon = 10^{-6}$, $n_{\max} = 10,000$ for the true problem, and $\epsilon = 1$, $n_{\max} = 5$ for the static surrogate. Two instances are tested, each from 100 randomly generated starting points with 10 random seeds for a total of $1,000$ runs per problem. Figure 3 show data profiles for 3 values of the tolerance parameter τ on the Simple_MDO_10_Variables (on top) and Simple_MDO_16_Variables (on bottom) problems.

Without any surprise, the success ordering strategy is outperformed by the others. For $\tau = 10^{-3}$, hqm and stat are faster than quad, but for a large number of evaluations quad solves more problems than stat. hqm and stat dominate quad when $\tau = 10^{-5}$. hqm dominates stat for that tolerance, but even more with $\tau = 10^{-7}$: hqm solves 25% more problems than stat on the problem with 10 variables and 55% more on the problem with 16 variables.

We next consider the SIMPLIFIED WING design optimization problem [39] involving aerodynamic, structural and performance disciplines. The problem has 10 bounded variables and 10 inequality constraints The parameters defining the true and static surrogate problems are $\epsilon = 10^{-12}$ and $n_{\max} = 100$ for the true problem, and $\epsilon = 10^{-1}$ and $n_{\max} = 5$ for the static surrogate. The data profiles of Figure 4 are generated from 100 starting points and 10 random seeds. Similar conclusions are drawn from the three graphs. The success ordering strategy is once again non-competitive. The three other are comparable, with a slight preference for the stat strategy. These results suggest that the correction brought by HQM is not always dominant.

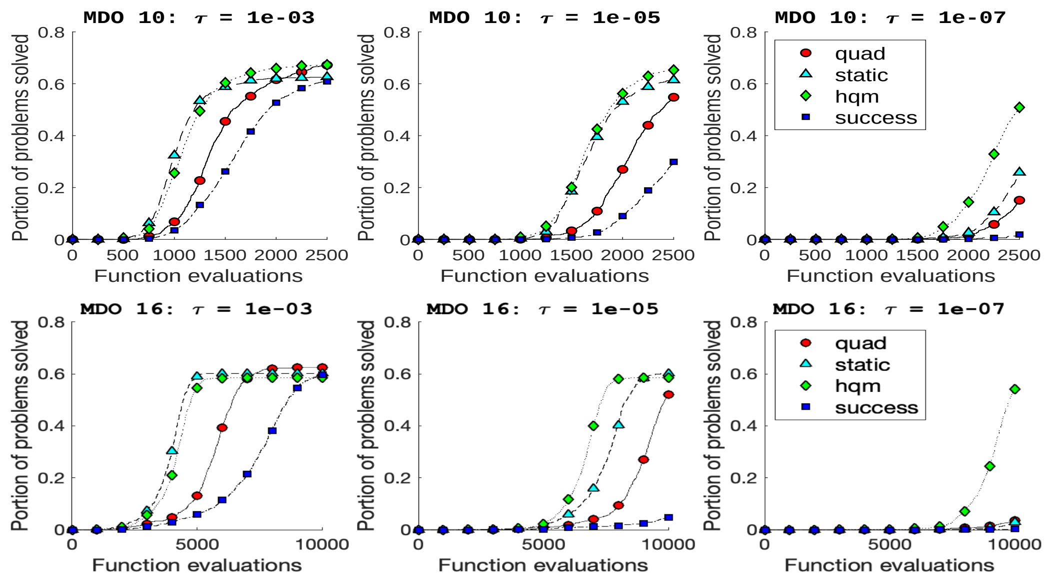


Figure 3: Data profiles on the Simple_MDO_10_Variables and Simple_MDO_16_Variables problems

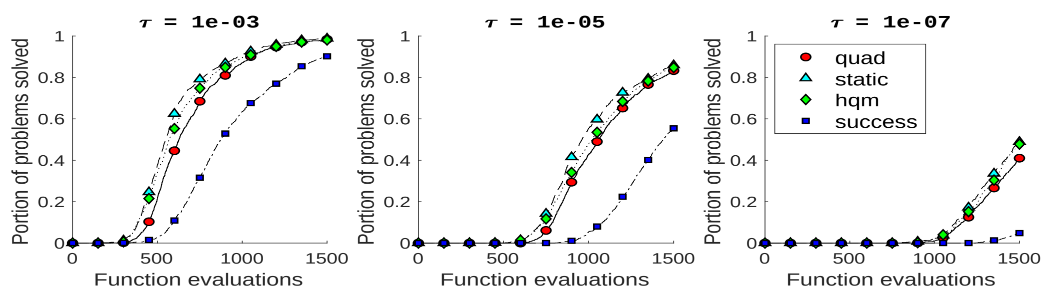


Figure 4: Data profiles on the Simplified Wing **problem**

4.2 The Lockwood problem

Lockwood [27, 28] is a pump-and-treat groundwater remediation problem from Montana Lockwood Solvent Groundwater Plume Site [27] with 6 bounded variables and 2 inequality constraints. A static surrogate is obtained by altering two parameters within the blackbox, with a fifth of the execution times compared to the true problem. The results illustrated in Figure 5 are obtained from 100 randomly generated starting points. The top the profiles are standard data profiles. In order to take into account the computational time of the static surrogate, the plots on the bottom are data profiles in whicht the horizontal axis is the elapse CPU time in seconds on an PC (Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz 8.00 GB RAM) under linux.

The data profiles separates the ordering strategies into two groups: the pair that uses the static surrogate, and the pair that does not. The former systematically dominates the latter for all three values of τ. The overall best strategy is HQM. Once again, the pair of strategies using the static surrogate are dominant, and the combined use of the static surrogate with the quadratic models give the best results.

5 Discussion

This work proposed a way to combine a static surrogate as input for a quadratic model of an optimization problem, to be used within the poll step of the Mads algorithm. Numerical comparisons were performed on three MDO problems, and on the Lockwood engineering test problem. The overall

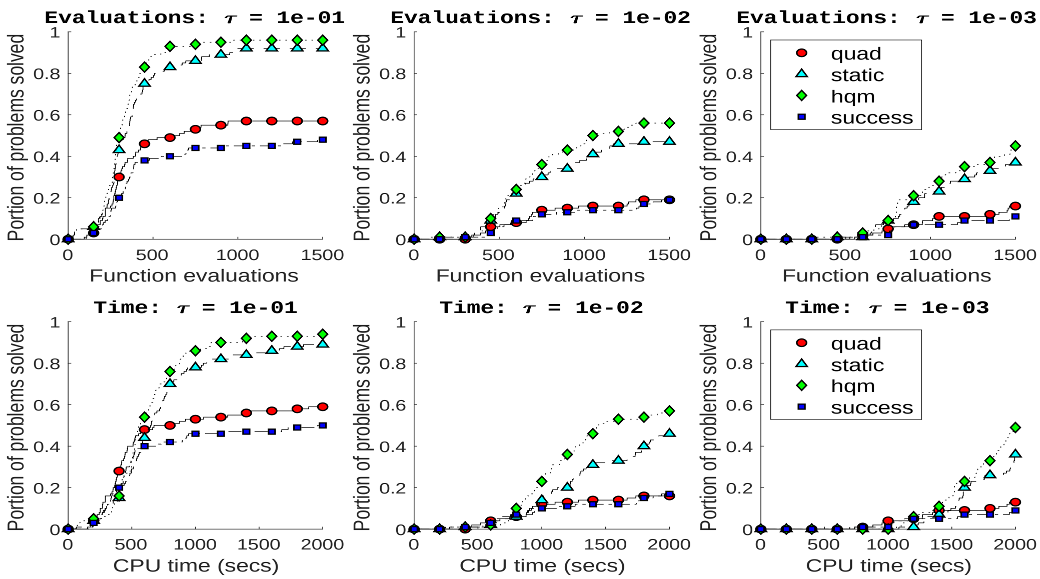


Figure 5: Data profiles on the Lockwood problem (vs number of evaluations on top, and time on bottom)

conclusion is that the HQM benefits from the global exploration aspect of the static surrogate and of the local performance of the quadratic model. The static surrogates is useful to identify a bassin containing a good solution, and the quadratic models helps in accelerating the convergence within that bassin.

Further research may combine a static surrogate with other types of dynamic models. In addition of quadratic models, some dynamic models of Talgorn et al. [37] may be useful to guide the global search step or the local poll step of Mads.

References

[1] M.A. Abramson, T.J. Asaki, J.E. Dennis, Jr., R. Magallanez, Jr., and M.J. Sottile. An Efficient Class of Direct Search Surrogate Methods for Solving Expensive Optimization Problems with CPU-time-related Functions. Structural Multidisciplinary Optimization, 45(1):53–64, 2012.

[2] N.M. Alexandrov, R.M. Lewis, C.R. Gumbert, L.L. Green, and P.A. Newman. Approximation and model management in aerodynamic optimization with variable-fidelity models. Journal of Aircraft, 38(6):1093–1101, 2001.

[3] C. Audet, V. Béchard, and J. Chaouki. Spent potliner treatment process optimization using a MADS algorithm. Optimization and Engineering, 9(2):143–160, 2008.

[4] C. Audet, V. Béchard, and S. Le Digabel. Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. Journal of Global Optimization, 41(2):299–318, 2008.

[5] C. Audet and J.E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. SIAM Journal on Optimization, 17(1):188–217, 2006.

[6] C. Audet and J.E. Dennis, Jr. A Progressive Barrier for Derivative-Free Nonlinear Programming. SIAM Journal on Optimization, 20(1):445–472, 2009.

[7] C. Audet and W. Hare. Derivative-Free and Blackbox Optimization. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2017.

[8] C. Audet, A. Ianni, S. Le Digabel, and C. Tribes. Reducing the Number of Function Evaluations in Mesh Adaptive Direct Search Algorithms. SIAM Journal on Optimization, 24(2):621–642, 2014.

[9] C. Audet, M. Kokkolaras, S. Le Digabel, and B. Talgorn. Order-based error for managing ensembles of surrogates in derivative-free optimization. Journal of Global Optimization, 70(3):645–675, 2018.

[10] C. Audet and D. Orban. Finding optimal algorithmic parameters using derivative-free optimization. SIAM Journal on Optimization, 17(3):642–664, 2006.

[11] M. Björkman and K. Holmström. Global optimization of costly nonconvex functions using radial basis functions. Optimization and Engineering, 1:373–397, 2000.

[12] A.J. Booker. Well-conditioned Kriging models for optimization of computer simulations. Technical Report M&CT-TECH-00-002, Boeing Computer Services, Research and Technology, M/S 7L–68, Seattle, Washington 98124, 2000.

[13] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, V. Torczon, and M.W. Trosset. A Rigorous Framework for Optimization of Expensive Functions by Surrogates. Structural and Multidisciplinary Optimization, 17(1):1–13, 1999.

[14] X. Chen and C.T. Kelley. Optimization with hidden constraints and embedded Monte Carlo computations. Optimization and Engineering, 17(1):157–175, 2016.

[15] T.D. Choi, O.J. Eslinger, C.T. Kelley, J.W. David, and M. Etheridge. Optimization of automotive valve train components with implicit filtering. Optimization and Engineering, 1(1):9–27, 2000.

[16] A.R. Conn and S. Le Digabel. Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. Optimization Methods and Software, 28(1):139–158, 2013.

[17] A.R. Conn, K. Scheinberg, and L.N. Vicente. Introduction to Derivative-Free Optimization. MOS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[18] E.J. Cramer and J.M. Gablonsky. Effective parallel optimization of complex computer simulations. In Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference,, August 2004.

[19] A.L. Custódio, H. Rocha, and L.N. Vicente. Incorporating minimum Frobenius norm models in direct search. Computational Optimization and Applications, 46(2):265–278, 2010.

[20] M. Eldred, A. Giunta, and S. Collis. Second-order corrections for surrogate-based optimization with model hierarchies. 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004.

[21] R.B. Gramacy and S. Le Digabel. The mesh adaptive direct search algorithm with treed Gaussian process surrogates. Pacific Journal of Optimization, 11(3):419–447, 2015.

[22] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear Optimization with the MADS algorithm. ACM Transactions on Mathematical Software, 37(4):44:1–44:15, 2011.

[23] S. Lophaven, H. Nielsen, and J. Søondergaard. Dace: A matlab kriging toolbox version 2.0. Technical Report IMM-REP-2002-12, Informatics and Mathematical Modelling, Technical University of Denmark, 2002.

[24] X. Marduel, C. Tribes, and J.-Y. Trépanier. Variable-fidelity optimization: Efficiency and robustness. Optimization and Engineering, 7(4):479–500, 2006.

[25] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Optimal aeroacoustic shape design using the surrogate management framework. Optimization and Engineering, 5(2):235–262, 2004.

[26] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Trailing-edge noise reduction using derivative-free optimization and large-eddy simulation. Journal of Fluid Mechanics, 572:13–36, 2007.

[27] L.S. Matott, K. Leung, and J. Sim. Application of MATLAB and Python optimizers to two case studies involving groundwater flow and contaminant transport modeling. Computers & Geosciences, 37(11):1894–1899, 2011.

[28] L.S. Matott, A.J. Rabideau, and J.R. Craig. Pump-and-treat optimization using analytic element method flow models. Advances in Water Resources, 29(5):760–775, 2006.

[29] J.J. Moré and S.M. Wild. Benchmarking derivative-free optimization algorithms. SIAM Journal on Optimization, 20(1):172–191, 2009.

[30] R.H. Myers, D.C. Montgomery, and C.M. Anderson-Cook. Response surface methodology: process and product optimization using designed experiments. Wiley, Hoboken, N.J., 3 edition, 2009.

[31] R. Oeuvray and M. Bierlaire. BOOSTERS: A derivative-free algorithm based on radial basis functions. International Journal of Modelling and Simulation, 29(1):26–36, 2009.

[32] E. Polak and M. Wetter. Precision control for generalized pattern search algorithms with adaptive precision function evaluations. SIAM Journal on Optimization, 16(3):650–669, 2006.

[33] M.J.D. Powell. Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. Mathematical Programming, 100(1):183–215, 2004.

[34] R.G. Regis and C.A. Shoemaker. Constrained global optimization of expensive black box functions using radial basis functions. Journal of Global Optimization, 31:153–171, 2005.

[35] S. Sankaran, C. Audet, and A.L. Marsden. A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation. Journal of Computational Physics, 229(12):4664–4682, 2010.

[36] J. Søndergaard. Optimization Using Surrogate Models — by the Space Mapping Technique. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, 2003.

[37] B. Talgorn, C. Audet, S. Le Digabel, and M. Kokkolaras. Locally weighted regression models for surrogate-assisted design optimization. Optimization and Engineering, 19(1):213–238, 2018.

[38] B. Talgorn, S. Le Digabel, and M. Kokkolaras. Blackbox Optimization in Engineering Design: Adaptive Statistical Surrogates and Direct Search Algorithms. In N.D. Lagaros and M. Papadrakakis, editors, Engineering and Applied Sciences Optimization, volume 38 of Computational Methods in Applied Sciences, pages 359–383. Springer International Publishing, 2015.

[39] C. Tribes, J.-F. Dubé, and J.-Y. Trépanier. Decomposition of multidisciplinary optimization problems: formulations and application to a simplified wing design. Engineering Optimization, 37(8):775–796, 2005.

[40] C. Tribes and J.-Y. Trépanier. Coordination of multidisciplinary distributed analyses and optimizations. In Canadian Aeronautics and Space Institute AERO'09 Conference, 2009.

[41] S.M. Wild and C.A. Shoemaker. Global convergence of radial basis function trust region derivative-free algorithms. SIAM J. Optimization, 21(3):761–781, 2011.