Derivative free optimization under constraints

Hoël Langouët and Delphine Sinoquet, IFP, hoel.langouet@ifp.fr

Optimization takes place in many IFP applications: estimation of parameters of numerical models from experimental data (earth sciences, combustion in engines), design optimization (networks of oil pipelines), optimizing settings of experimental devices (engine calibration, catalysis). These optimization problems consist of minimizing a functional that is complex (nonlinearities, noise), expensive to estimate (solution of a numerical model based on differential systems or experimental measurements), and for which derivatives are often not available. Moreover, often nonlinear constraints are introduced and sometimes several objectives are simultaneously optimized in order to find the best compromise.

We are interested in minimizing, under constraints, a function, which is expensive to evaluate, and for which the derivatives are not available [2]:

$$\left\{\begin{array}{cl} \min\limits_{x\in\mathbb{R}^n} f(x), & \\ b_l \leqslant x \leqslant b_u, & b_l, b_u \in \mathbb{R}^n, \\ Ax \leqslant b, & A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, \\ c_i(x) \geqslant 0, & i = 1, ..., l. \end{array}\right.$$

In practice, these problems are often solved by non linear optimization methods (SQP [11] for example) with derivatives approximated by finite differences. Although these methods are particularly efficient for the determination of active constraints, the number of function evaluations is usually too high for industrial problems with expensive simulators. Furthermore, the choice of the step of the finite differences, crucial for the convergence of this method, is generally cumbersome, because it depends on the accuracy of f which is difficult to estimate in practice.

Direct methods optimize a function without derivative calculation. Among them, Genetic Algorithms [3] and Pattern Search methods [5] are insensitive to inaccuracies in the calculation of f but require many evaluations. To overcome this difficulty, a cheap surrogate model of the objective function is typically used in the optimization process to limit the number of evaluations of the expensive function. These surrogate models are usually global models of f on the domain under study, constructed from a limited number of evaluations of f chosen according to a relevant criterion related to the estimation of the prediction error of these models and expected improvement on the minimization of the function [4, 13]. These models can be kriging models [4, 10, 13], Radial Basis Functions (RBF), splines ... But they are limited to problems of small size (10-20 param.).

Then, another class of methods based on local surrogate models was proposed by [1, 8, 9, 12]: these methods are inspired by SQP methods with trust region globalization [1]. A quadratic model is constructed at each iteration in a neighborhood of the current point, the size of this neighborhood is updated according to the comparison of the reduction predicted by the model and the effective reduction calculated by evaluating f. In particular, Powell [8, 9] proposed an efficient method, without constraints for problems of medium size (hundreds of parameters).

```
Algorithm SQA (Sequential Quadratic Approximation)

(1) Initialization : Select $m$ initial interpolation points. $x_{opt}$ is the initial point for which
    $f$ is minimal among the $m$ points. Determine the first quadratic model $Q \approx f$.
    $\rho = \Delta = \rho_{beg}$

(2) Solve the problem for a precision $\rho$
    (a)  (i) Minimize the model : $\min_d Q(x_{opt} + d)$
             s.t. $\begin{cases} \|d\| \leqslant \Delta, & \\ b_l \leqslant x_{opt} + d \leqslant b_u, & b_l, b_u \in \mathbb{R}^n \\ A \cdot (x_{opt} + d) \geqslant b, & A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m \\ c_i(x) \geqslant 0, & i = 1, ..., l \end{cases}$
        (ii) If $\|d\| < \frac{1}{2}\rho$ : $\rightarrow$ (2)(b) because it is necessary to ensure the validity of
             the model before doing small steps.
       (iii) Calculate $f(x_{opt} + d)$ and update the trust region radius $\Delta$ from the pre-
             dictivity of the quadratic model $R = \frac{f(x_{opt}) - f(x_{opt}+d)}{Q(x_{opt}) - Q(x_{opt}+d)}$.
             Update interpolation points : $x_{opt} = x_{opt} + d$ if $f(x_{opt} + d) < f(x_{opt})$ and
             update the model $Q$ in order to interpolate the model in $x_{opt} + d$.
        (iv) If $R > 0.1 \rightarrow$ (2)(a)(i)
             Otherwise continue.
    (b) Test the validity of the model $Q$. The model is considered valid if all the current
        interpolation points $x_i$, $i = 1, ..., m$ are relatively close to the optimal point $x_{opt}$,
        i.e. if the Euclidean distance $\|x_i - x_{opt}\| < 2\Delta$ for all $x_i$, $i = 1, ..., m$.
         (i) If the model is not valid :
             Improve the quality of the model $Q$ : $\max_{d \in \mathcal{R}^n} |l_t(x_{opt} + d)|$
             s.t. $\begin{cases} \|d\| \leqslant \overline{\Delta}, & \\ b_l \leqslant x_{opt} + d \leqslant b_u, & b_l, b_u \in \mathbb{R}^n, \\ A \cdot (x_{opt} + d) \geqslant b, & A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m, \\ c_i(x) \geqslant 0, & i = 1, ..., l. \end{cases}$
             Calculate $f(x_{opt} + d)$, the worst point of the interpolation set is replaced
             by a new point closer to the current minimum.
             The accuracy of $Q$ after its update should be better $\rightarrow$ (2)(a)(i).
        (ii) If the model is valid : If $\|d\| > \rho \rightarrow$ (2)(a)(i)
             Otherwise continue
    (c) If $\rho > \rho_{end}$ : reduction of $\rho$ and $\Delta \rightarrow$ (2)(a)
        Otherwise End of the algorithm.
```

Taking into account constraints in these methods remains a difficulty, penalty methods are often inefficient in practice. In this paper, an extension of Powell's method for constraints is proposed. We assume here that derivatives of constraints are available.

This extension consists of :

1. choosing the interpolation points in the domain defined by the constraints,

2. minimizing the quadratic model under constraints in the trust region thanks to a SQP method.

Numerical results for smooth test functions constructed from the benchmark CUTEr [6] are presented. Bound constraints on parameters are introduced. The results obtained with the SQA method are compared to four different optimizations methods :

1. SQPAL, Sequential Quadratic method with an estimation of the gradient of the Lagrangian function by finite differences (step of 10^{-6}) and a BFGS approximation of the Hessian matrix [11],

2. EGO, which relies on a sequential method that constructs a global kriging model : the points to be evaluated are chosen according the expected improvement criterion [10],

3. NMSMAX, a Nelder-Mead simplex method [5],

4. CMAES, a genetic algorithm : Covariance Matrix Adaptation Evolutionary Strategy [3].

Data profiles and performance profiles [6] are used to compare the optimization methods. They are presented for two accuracies defined by : $\tau = \frac{f(x^*)-f_L}{f(x_0)-f_L}$. $f(x^*)$ is f evaluated on the current optimum obtained by the method, $f(x_0)$ is the initial value and f_L is either the optimum if known or the best solution found by the optimization methods. It measures the improvement towards this value f_L relatively to the initial point.

Data profiles in Figure 1 show that SQA solved the largest percentage of problems for all function evaluation budgets and for all required accuracy τ.

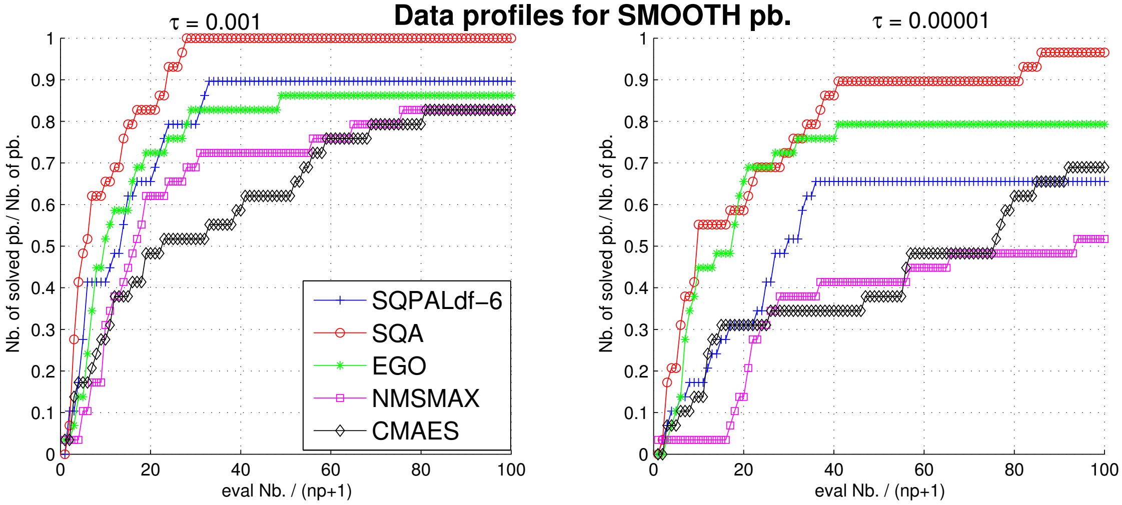

Figure 1: Data profiles for smooth problems of the benchmark proposed by [6] : the horizontal axis represents the number of evaluations divided by the number of parameters +1 which correspond to the "price" of computation of the gradient by finite differences ; the vertical axis represents the part of problems solved by the optimization method with the accuracy τ.

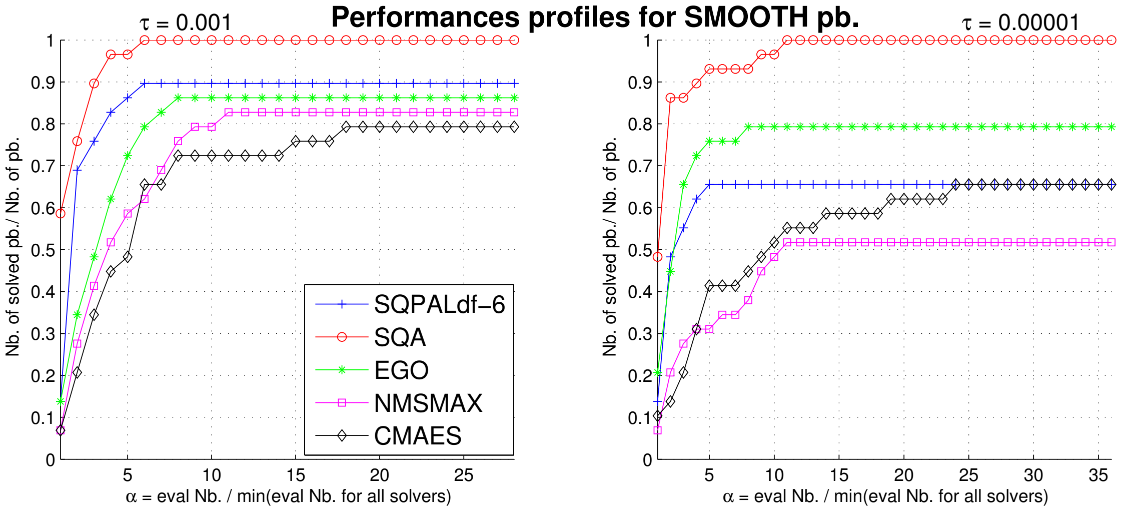

Figure 2: Performance profiles for smooth problems of the benchmark proposed by [6] : the horizontal axis represents the number of evaluations necessary to solve one problem divided by the minimal number of evaluations necessary to solve this problem among all the solvers.

Performance profiles of Figure 2 show that SQA is the fastest solver for a large percentage of problems, while EGO, SQPAL, NMSMAX and CMAES are the fastest for very small percentages of problems ($\alpha = 1$). In addition, the SQA method is the most robust method because it solves the largest percentage of problems for all levels of accuracy τ ($\alpha \to \infty$).

Numerical results for noisy test functions (relative noise level of 10^{-2}) constructed from the benchmark CUTEr [6] are presented. The results obtained with the SQA method are compared with the same four methods. Data profiles and performance profiles [6] are also used to compare these five methods and presented for two accuracies.

Data profiles in Figure 3 show that the method SQA and the method EGO solve the largest percentage of problems for all function evaluation budgets.

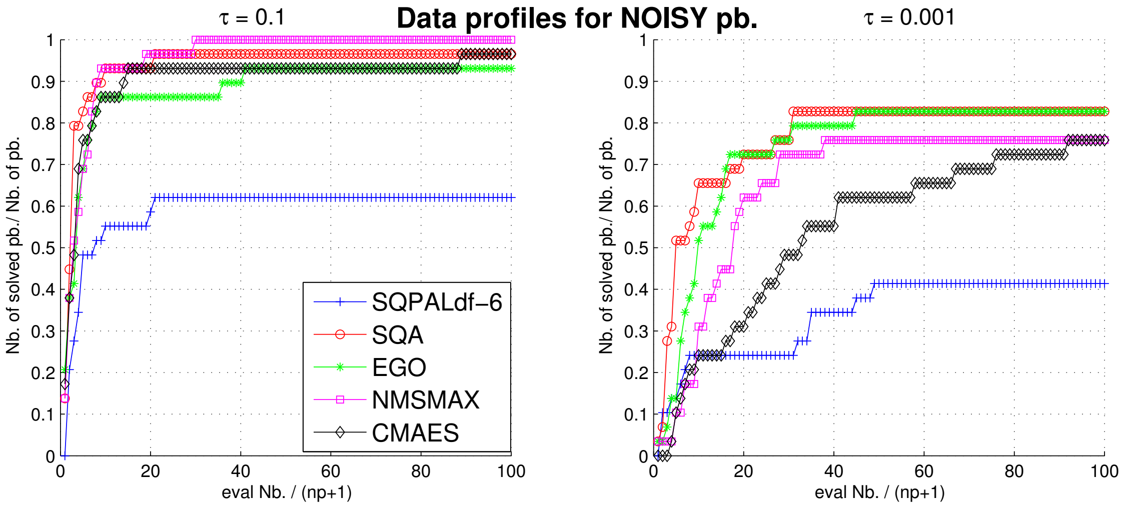

Figure 3: Data profiles for noisy problems of the benchmark proposed by [6]

Performance profiles of Figure 4 show that the method SQA is the fastest solver in at least 60% of the problems, while other methods EGO, SQPAL, NMSMAX and CMAES are the faster for less than 20% of the problems ($\alpha = 1$). In addition, the SQA method like EGO is the most robust because it solves the largest percentage of problems for accuracy $\tau = 10^{-3}$.

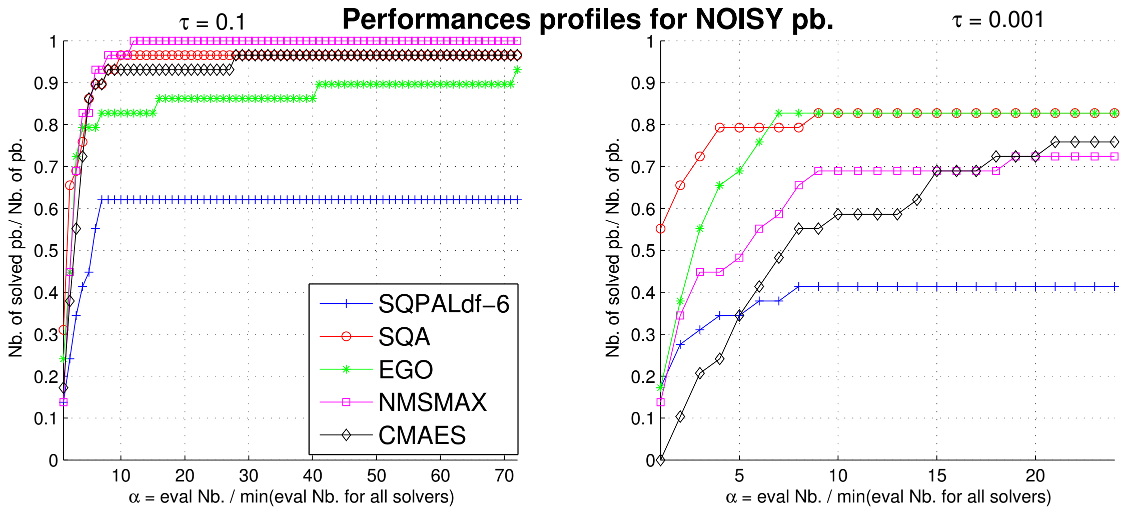

Figure 4: Performance profiles for noisy problems of the benchmark proposed by [6]

In conclusion, SQA is the fastest and more robust method for smooth problems of the presented benchmark and is equivalent to EGO for noisy problems. Improvements can be made by taking into account noise as proposed by [12]. Moreover SQA can also handle large size problems (hundreds of parameters) ; which is not the case of kriging methods (like EGO). Introduction of non linear constraints is being studied, a difficulty arises when derivatives of constraints are not available.

An application for production and seismic data matching in the field of oil reservoir characterization will also be presented.

References

[1] Conn A. R., Gould N. I. M. et Toint P. L., (2000) *Trust-Region Methods*, MPS-SIAM Series on Optimization.

[2] Conn A. R., Scheinberg K. et Vicente L. N., (2009) *Introduction to Derivative-Free Optimization*, MPS-SIAM Series on Optimization.

[3] Hansen, N. et Kern N., (2004) *Evaluating the CMA Evolution Strategy on Multimodal Test Functions*, In Eighth International Conference on Parallel Problem Solving from Nature PPSN VIII, Proceedings, pp. 282-291, Berlin: Springer.

[4] Jones D., Schonlau M. et Welch W., (1998) *Efficient global optimization of expensive black-box functions.*

[5] KOLDA T. G., LEWIS R; M. et TORCZON V., (2003) *Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods*, SIAM Review, Vol. 45, No. 3, pp. 385-482.

[6] MORÉ J. J. et WILD S. M., (2007) *Benchmarking Derivative-Free Optimization Algorithms*, Mathematics and Computer Science Division, Argonne National Laboratory, Preprint ANL/MCS-P1471-1207, http://www.mcs.anl.gov/ more/dfo/ceperf.pdf.

[7] OEUVRAY R., (2005) *Trust-Region Methods Based on Radial Basis Functions with application to Biomedical Imaging*, Thèse de doctorat, Ecole Polytechnique de Lausanne.

[8] POWELL M.J.D., (2004) *The NEWUOA software for unconstrained optimization without derivatives.*

[9] POWELL M.J.D., (2007) *Developments of NEWUOA for unconstrained minimization without derivatives.*

[10] SCHONLAU M., (1997) *Computer experiments and global optimization*, Thèse de doctorat, Université de Waterloo.

[11] SINOQUET D. et DELBOS F., (2008) *Adapted nonlinear optimization method for production data and 4D seismic inversion*, 11th ECMOR, Bergen, Norway, 8 - 11 September 2008.

[12] VANDEN BERGHEN F., (2004) *CONDOR : a constrained, non-linear, derivative-free parallel optimizer for continuous, high computing load, noisy objective functions*, Thèse de doctorat, Université de Bruxelles.

[13] VILLEMONTEIX J., (2008) *Optimisation de fonctions coûteuses, Modèles gaussiens pour une utilisation efficace du budget d'évaluations : théorie et pratique industrielle*, Thèse de doctorat, Université Paris-sud XI.

[14] WILD S. M., REGIS R. G. et SHOEMAKER C. A., (2008) *ORBIT: Optimization by Radial Basis Function Interpolation in Trust-Regions*, SIAM J. Scientific Computing, Vol. 30, No. 6, pp. 3197-3219.