Semi-Deterministic Recursive Optimization Methods for Multichannel Optical Filters

Benjamin Ivorra[1], Bijan Mohammadi[1], Laurent Dumas[2] and Olivier Durand[3]

[1] I3M, University of Montpellier 2, 34095 Montpellier, France
ivorra@math.univ-montp2.fr
[2] Jacques-Louis Lions Laboratory, University of Paris 6, 75252 Paris, France
[3] Alcatel Research and Innovation, 91460 Marcoussis, France

Summary. In this paper, we reformulate global optimization problems in terms of boundary value problems. This allows us to introduce a new class of optimization algorithms. Indeed, many optimization methods, including non-deterministic ones, can be seen as discretizations of initial value problems for differential equations or systems of differential equations. Two algorithms included in this new class are applied and compared with a genetic algorithm for the design of multichannel optical filters.

1 Introduction

Global minimization (or maximization) problems are of great practical importance in many applications. For this reason, Genetic Algorithms (**GA**) have received a tremendous interest in recent years [1, 2, 5]. However, the main difficulties with these algorithms remain their computational time and their slow convergence.

Many minimization algorithms can be viewed as discrete forms of Cauchy problems for an ordinary differential equation (**ODE**) or a system of ODEs in the space of control parameters. We will see that if one introduces an extra information on the infimum, these algorithms can be formulated as Boundary Value Problems (**BVP**) for the same equations [3, 4]. A motivating idea is therefore to apply algorithms solving BVPs to global optimization. It is in particular shown that GAs be interpreted as a discrete form of BVPs for a set of coupled ODEs. Therefore, the BVP analysis has also been applied to them to improve their performances leading to the construction of a new algorithm called **HGSA**. All the algorithms issued from our BVP analysis, presented in section 2, are compared in section 3 to a classical GA for the design of a multichannel optical filter.

2 Global optimization methods

In this section, we consider a function $J : \Omega_{ad} \to \mathbb{R}$ to be minimized, where the optimization parameter x belongs to a compact admissible set $\Omega_{ad} \subset R^N$.

A unified dynamical system formulation is given for some stochastic and deterministic optimization algorithms. In particular, even if GAs are issued from evolutionary considerations, it is possible to associate to them a set of stochastic coupled ODEs (see subsection 2.2). A new class of global minimization methods is thus constructed, based on the solution of associated BVPs.

2.1 Semi-deterministic recursive optimization methods

We make here the following assumptions on the functional: $J \in C^2(\Omega_{ad}, \mathbb{R})$ and is coercive [3]. In this case, many deterministic minimization algorithms which perform the minimization of J can be seen as discretizations of the following dynamical system [3, 4]:

$$\begin{cases} M(\zeta)\dfrac{dx(\zeta)}{d\zeta} = -d(x(\zeta)) \\ x(0) = x_0 \end{cases} \tag{1}$$

where ζ is a fictitious parameter, M is a local metric transformation, d a direction in Ω_{ad} and $x_0 \in \Omega_{ad}$ is the initial condition.

For example if $d = \nabla J$ is the gradient of the functional J and $M = Id$, we recover the classical steepest descent method, while with $d = \nabla J$ and $M = \nabla^2 J$ the Hessian of J, we recover the Newton method.

A global optimization of J with system (1), called here *core optimization method*, is possible if the following boundary value problem has a solution:

$$\begin{cases} M(\zeta)\dfrac{dx(\zeta)}{d\zeta} = -d(x(\zeta)) \\ x(0) = x_0 \\ J(x(Z_{x_0})) = J_m \quad \text{with a finite } \ Z_{x_0} \in \mathbb{R} \end{cases} \tag{2}$$

where J_m denotes the minimum of J in Ω_{ad}. In practice, when J_m is unknown, we set J_m to a lower value (for example $J_m = 0$ for an inverse problem) and look for the best solution for a given complexity and computational effort.

The BVP (2) is over-determined as it includes two conditions and only one derivative. The over determination can be removed for instance by considering $x_0 = v$ in (1) as a new variable to be found by the minimization of the new functional:

$$h(v) = J(x_v(Z_v)) - J_m$$

where $x_v(Z_v)$ is the solution of (1) found at $\zeta = Z_v$ starting from v.

A new algorithm $A_1(v_1, v_2)$ with parameters v_1 and v_2 is then defined as:

`1-` $(v_1, v_2) \in \Omega_{ad} \times \Omega_{ad}$ `given,` $v_1 \neq v_2$

```
2-   Find v ∈ argmin_{w∈O(v2)} h(w) where O(v2) = IR v1v2 ∩ Ω_ad
3-   return the best v found during step 2
```

The line search minimization in A_1 might fail. For instance, a secant method degenerates on plateau and critical points. To avoid this problem, we add an external level to the algorithm A_1, keeping v_1 unchanged, and looking for v_2 by minimizing a new functional $w \mapsto h(A_1(v_1, w))$. This leads to the following two-level algorithm $A_2(v_1, v'_2)$:

1- $(v_1, v'_2) \in \Omega_{ad} \times \Omega_{ad}$ given, $v_1 \neq v'_2$

2- Find $v' \in argmin_{w \in \mathcal{O}(v'_2)} h(A_1(v_1, w))$ where $\mathcal{O}(v'_2) = \mathbb{R}\overrightarrow{v_1 v'_2} \cap \Omega_{ad}$

3- return the best v' found during step 2

The choice of the initial conditions in this algorithm is its only non-deterministic feature. The algorithm A_2 is thus called Semi-Determinist Algorithm (**SDA**). A mathematical background for this approach as well as a validation on academic test cases or on problems including solutions of nonlinear partial differential equations are available [3, 5, 7, 10].

Remarks:

- The construction can be pursued recursively considering

$$h^i(v_2^i) = \min_{v_2^i \in \Omega_{ad}} h^{i-1}(A_{i-1}(v_1, v_2^i))$$

  with $h^1(v) = h(v)$ and where i denotes the external level, justifying the name of recursive optimization methods.

- In practice, this algorithm succeeds if the trajectory passes close enough to the infimum (i.e. in $B_\varepsilon(x_m)$ where ε defines the chosen accuracy in the capture of the infimum). Hence, in the algorithm above, $x_w(Z_w)$ is replaced by the best solution found over $[0, Z_w]$.

2.2 Genetic algorithms

Genetic algorithms approximate the global minimum (or maximum) of any functional $J : \Omega_{ad} \to \mathbb{R}$, also called fitness function, through a stochastic process based on an analogy with the Darwinian evolution of species [1]: a first family, called 'population', $X^0 = \{x_l^0 \in \Omega_{ad}, \quad l = 1, ..., N_p\}$ of N_p possible solutions of the optimization problem, called 'individuals', is randomly generated in the search space Ω_{ad}. Starting from this population, we build recursively N_{gen} new populations $X^i = \{x_l^i \in \Omega_{ad}, l = 1, ..., N_p\}$ with $i = 1, .., N_{gen}$ through three stochastic steps, called selection, crossover and mutation. With these three basic evolution processes, it is generally observed that the best obtained individual is getting closer after each generation to the optimal solution of the problem [1]. An example of such stochastic processes is given below in order to show the analogy with the resolution of a discrete dynamical system.

We first rewrite X^i using the following (N_p, N)-real valued matrix form:

$$X^i = \begin{bmatrix} x_1^i \\ \vdots \\ x_{N_p}^i \end{bmatrix} \quad (3)$$

Selection: each individual x_l^i is ranked with respect to its fitness value $J(x_l^i)$ and N_p elements are then selected among the population to become 'parents'.

Introducing $\mathcal{S}^n(J(X^n))$ a binary (N_p, N_p)-matrix with, for each line i, a value 1 on the jth row when the jth individual has been selected and 0 elsewhere, we define

$$X^{n+1/3} = \mathcal{S}^n(J(X^n))X^n \quad (4)$$

Crossover: this process leads to a data exchange between two 'parents' and the apparition of two new individuals called 'childrens'.

Introduce $\mathcal{C}^n$ a real-valued (N_p, N_p)-matrix where for each couple of consecutive lines $(2i-1, 2i)$ $(1 \leq i \leq \frac{N_p}{2})$, the coefficients of the lth and kth rows are given by a 2×2 matrix of the form

$$\begin{bmatrix} \lambda_1 & 1-\lambda_1 \\ \lambda_2 & 1-\lambda_2 \end{bmatrix}$$

In this expression, $\lambda_1 = \lambda_2 = 1$ if no crossover is applied on the selected parents l and k or are randomly chosen in $[0, 1]$ in the other case (with a probability p_c). This step can be summarized as:

$$X^{n+2/3} = \mathcal{C}^n X^{n+1/3} \quad (5)$$

Mutation: this process leads to new parameters values for some individuals of the population. More precisely, each children is modified (or mutated) with a fixed probability p_m.

Introduce for instance a random perturbation matrix $\mathcal{E}^n$ with a i-th line equal to 0 if no mutation is applied to the ith children and a random value $\epsilon_i \in \mathbb{R}^N$ in the other case. This step can then take the following form:

$$X^{n+1} = X^{n+2/3} + \mathcal{E}^n \quad (6)$$

or more generally

$$X^{n+1} = f(X^{n+2/3}) \quad (7)$$

for a certain stochastic operator f in the space of (N_p, N)-real valued matrices.

Therefore, GAs can be seen as discrete dynamical systems, writing for instance in the presented case:

$$X^{n+1} = \mathcal{C}^n \mathcal{S}^n(J(X^n))X^n + \mathcal{E}^n \tag{8}$$

which is a particular discretization of a set of nonlinear first order ODEs of the type:

$$\dot{X}(t) = \Lambda_1(t, J(X(t)), p_c, p_m)X(t) + \Lambda_2(t, p_c, p_m) \tag{9}$$

where $\{p_c, p_m\}$ are fixed parameters and the construction of Λ_1 and Λ_2 has been described above. Finally, GAs can been interpreted as solving the following BVP:

$$\begin{cases} \dot{X}(t) = \Lambda_1(t, J(X(t)), p_c, p_m)X(t) + \Lambda_2(t, p_c, p_m) \\ X(0) = X^0 \\ \widehat{J}(X(T)) = J_m \end{cases} \tag{10}$$

where $\widehat{J}(X) = min\{J(x_i)/1 \leq i \leq N_p\}$ for any $X = {}^t(x_1, ..., x_{N_p})$

Engineers like GAs because these algorithms do not require sensitivity computation, perform global and multi-objective optimization and are easy to parallelize. However, their drawbacks remain their weak mathematical background, their computational complexity and their slow convergence. As a fine convergence is difficult to achieve with GA based algorithms, it is recommended when it is possible, to complete the GA iterations by a descent method. This is especially useful when the functional is flat around the infimum (see [2] for more complex coupling of GAs with descent methods).

2.3 Hybrid genetic/semi-deterministic algorithm

It is interesting to notice that once GA is seen as a dynamical system (9) for the population, it can be used as a core optimization method in the way presented in subsection 2.1. The aim here is to find a compromise between the robustness of GAs and the low-complexity features of SDAs.

In order to reduce the GA population size while keeping the efficiency of the method, we couple it with a SDA. The SDA provides information on the choice of the initial population X^0 whereas the GA performs global optimization starting from this population. We call this approach **HGSA** (Hybrid Genetic/Semi-deterministic Algorithm).

3 Application to multichannel optical filters design

Many important developments in optical fiber devices for telecommunications have been done in the recent years. Among them are Fiber Bragg Gratings

(**FBG**) which are an attractive alternative in applications such as multichannel multiplexing. FBGs are optical fibers with a modulated refractive index which reflects a part of the wavelength band, called reflected spectrum, and let pass the complementary band called transmitted spectrum [8].

The inverse problem considered here is the design of a given optical filter based on a FBG. More precisely, the objective is to construct a multichannel filter with a reflected spectrum that consists of $N_{peaks} = 16$ totally reflective identical channels spaced of $\Delta\lambda = 0.8$nm. The optimization space consists of all possible FBG refractive index modulation profiles for a given length, namely $L = 103.9$mm.

These refractive index modulation profiles are generated by spline interpolation through a number of $N = 9$ points equally distributed along the first half of the FBG and completed by parity with a maximum refractive index amplitude of $\overline{n}_{max} = 5 \times 10^{-4}$. Thus the search space is defined by $\Omega_{ad} = [-5 \times 10^{-4}, 5 \times 10^{-4}]^9$.

The functional to be minimized in Ω_{ad} is defined by:

$$J(x) = \sum_{i=1}^{N_c} (r(x, \lambda_i) - r_{target}(x, \lambda_i))^2 \qquad (11)$$

where:

- $r(x, .)$ is the reflected spectrum of the FBG with a refractive index modulation profile associated to $x \in \Omega_{ad}$. It is a function defined from the transmission band $[1.530, 1.545]$ (in microns) to $[0, 1]$ which is determined by solving a certain direct problem [9].
- $r_{target}(x, .)$ denotes the nearest perfect reflected spectrum to $r(x, .)$:

$$r_{target}(x, \lambda) = \begin{cases} 1 \; if \; \lambda \in \{\lambda_x, \lambda_x + \Delta\lambda, \ldots, \lambda_x + (N_{peaks} - 1)\triangle\lambda\} \\ 0 \text{ elsewhere} \end{cases}$$

for a certain λ_x in the transmission band.

Both functions $r(x)$ and $r_{target}(x)$ are evaluated on $N_c = 1200$ wavelengths equally distributed on the transmission band.

Results

The minimization of the cost function (11), has been tested with various algorithms presented in section 2, namely GA, HGSA and SDA algorithms.

The SDA method is applied with a core algorithm consisting of 10 iterations of a steepest descent method and a line search made of 5 iterations of a secant method for each level algorithm A_1 and A_2. The latter is initialized with random initial conditions v_1 and v_2'. As the minimal value of J is unknown, we set $J_m = 0$. Furthermore to reduce SDA computational time, gradient evaluations (representing 90% of this time) are done on a coarse mesh with $N_c = 300$ reducing the evaluation of a factor 4 (4s on a 3Ghz/512Mo Ram-Desktop computer) for a gradient variation of approximately 10%. Such method is called incomplete gradient approach [6].

HGSA and GA are applied with the following values: the population size is set to $N_p = 180$ for GA (respectively 10 for HGSA) and the generation number is set to $N_{gen} = 30$ for GA (resp. 10 for HGSA). The selection is a roulette wheel type [1, 2] proportional to the rank of the individual in the population. The crossover is barycentric in each coordinate with a probability $p_c = 0.45$. The mutation process is non-uniform with a probability $p_m = 0.15$ for GA (resp. 0.35 for HGSA). A one-elitism principle, that consists in keeping the current best individual in the next generation, has also been imposed. Finally, a steepest descent method is performed at the end of both algorithms.

The different optimization results are summarized in Table 1 whereas the corresponding reflected spectra obtained with the optimized profiles are presented on Figure 1. Although only SDA optimization produces 16 totally reflective peaks, GA and HGSA associated spectra are still industrially applicable due to the fact that in practice we only need 95%-reflective peaks [8]. The SDA method also gives the best result in terms of cost function minima and computational time. Note also that the HGSA technique over-perform a classical GA, providing an interesting alternative to SDA in cases where gradients cannot be evaluated.

Table 1. Optimization results

	SDA	HGSA	GA
minimal value of the cost function	3.0	4.3	5.8
Functional Evaluation Number	3000 (90% on coarse mesh)	2600	5500
Computational time	4h	11h	24h

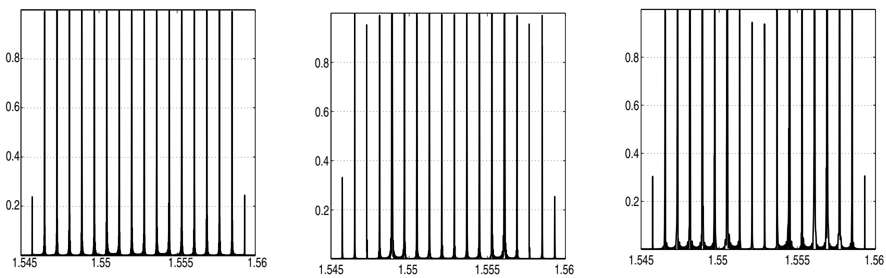

Fig. 1. Reflected spectra of optimized filters(reflexivity vs. wavelength (μm)) obtained with (**Left**) SDA, (**Center**) HGSA and (**Right**) GA.

4 Conclusions

A new class of semi-deterministic methods has been introduced. This approach allows us to improve deterministic and non-deterministic optimization algorithms. Both of them have been detailed and applied to the design of a multichannel optical filter for which the results obtained over-perform those obtained with a classical genetic algorithm.

It represents a new validation of theses methods on industrial problems involving multiple local minima after some previous others: temperature and pollution control in a bunsen flame [10], shape optimization of fast-microfluidic-mixer devices [5], shape optimization of under aerodynamic and acoustic constraints for internal and external flows [6].

References

1. Goldberg, D.: Genetic algorithms in search, optimization and machine learning. Addison Wesley (1989)
2. Dumas, L., Herbert, V., Muyl, F.: Hybrid method for aerodynamic shape optimization in automotive industry. Computers and Fluids. **33**(5), 849-858 (2004)
3. Mohammadi, B., Saiac, J-H.: Pratique de la simulation numérique. Dunod (2002)
4. Attouch, H., Cominetti, R.: A dynamical approach to convex minimization coupling approximation with the steepest descent method. Journal of Differential Equations. **128**(2), 519-540 (1996)
5. Ivorra, B., Mohammadi, B., Santiago, J-G., Hertzog, D-E.: Semi-deterministic and genetic algorithms for global optimization of microfluidic protein folding devices. International Journal of Numerical Methods in Engineering. **66**(2), 319-333 (2006)
6. Mohammadi, B., Pironneau, O.: Applied Shape Optimization for Fluids. Oxford University Press (2001)
7. Ivorra, B., Mohammadi, B., Redont, P., Dumas, L.: Semi-deterministic vs genetic algorithms for global optimization of multichannel Optical Filters. International Journal of Computational Science and Engineering. To be published. (2006)
8. Chow, J., Town, G., Eggleton, B., Ibsen, M., Sugden, K., Bennion, I.: Multiwavelength Generation in an Erbium-Doped Fiber Laser Using In-Fiber Com Filters. IEEE Photonic Technology Letter. **8**(1), 60-62 (1996)
9. Erdogan, T.: Fiber Grating Spectra. Journal of Lightwave Technology. **15**(8), 1277-1294 (1997)
10. Debiane, L., Ivorra, B., Mohammadi, B., Nicoud, F., Ern, A., Poinsot, T., Pitsch, H.: Temperature and pollution control in flames. Proceeding of the CTR Summer Program 2004. 367-375 (2004)