Hybrid Evolution Strategy-Downhill Simplex Algorithm for Inverse Light Scattering Problems

Demetrio Macías, Gustavo Olague, and Eugenio R. Méndez

Centro de Investigación Científica y de Educación Superior de Ensenada,
División de Física Aplicada,
Apdo. Postal 2732, 22800 Ensenada, B. C., México.
{dmacias, olague, emendez}@cicese.mx

Abstract. The rough surface inverse scattering problem is approached with a combination of evolutionary strategies and the simplex method. The surface, assumed one-dimensional and perfectly conducting, is represented using spline curves. Starting from rigorously calculated far-field angle-resolved scattered intensity data, we search for the optimum profile using the evolutionary strategies $(\mu/\rho^{+},\lambda)$. After a fixed number of iterations, the best surface is finally recovered with the downhill simplex method. Aspects of the convergence and lack of uniqueness of the solution are discussed.

1 Introduction

The interaction of optical waves with rough surfaces is a familiar phenomenon. Most of what we see is light that has been scattered by objects (surface and/or volume scattering), and we are always trying to infer information about these objects from the scattered light. Not surprisingly, the scattering of electromagnetic waves from rough surfaces finds applications in many areas of science and technology. Among others, we can mention remote sensing, microscopy, surface metrology, and the characterization of optical surfaces.

In the context of the present paper, the direct scattering problem consists of determining the angular distribution of the scattered intensity from a given surface and conditions of illumination. Although to date there is no general solution to this problem, the subject has been studied extensively [1,2]. The inverse scattering problem deals with the reconstruction of the profile that gave rise to the angular distribution of the scattered intensity. The problem is more complex but, from a practical standpoint, much more important. It is also poorly understood.

Most of the previous works on inverse scattering make use of amplitude data [3,4,5]. Since optical detectors are not phase sensitive, the need to have amplitude, rather than intensity data, constitutes an important drawback that all these methods share. Another important limitation is that these methods are based on approximate models for the interaction between the incident light and the surface, and fail in situation when multiple scattering is important.

S. Cagnoni et al. (Eds.): EvoWorkshops 2003, LNCS 2611, pp. 399–409, 2003.

In a previous paper [6], we have studied the inverse scattering problem considering it as a problem of constrained optimization. We focused our attention on the case of surfaces that constitute realizations of a pre-specified Gaussian random process; that is, the random profiles belong to a well-defined statistical class. This assumption enabled us to use the so-called spectral method [7,8] for generating and representing the surfaces numerically. One of the drawbacks of using that kind of representation, is that its nature precludes the use of recombination and other optimization techniques, such as the downhill simplex method [9]. The other problem is that, as it stands, it cannot deal with deterministic surfaces; the profiles must be random, and we need to know the statistical class they belong to.

In the present paper, we continue along the lines of our previous work, removing some of the limitations. For this, we adopt a representation of the surfaces based on B-spline curves. This, not only permits the treatment of more general surfaces, but also enables us to use a wider range of optimization strategies. As in our previous paper, and unlike other works on the subject, the method described here does not rely on approximate expressions for the field-surface interaction.

The organization of this paper is as follows. In Sect. 2 we introduce the notation, define the physical situation considered, and formulate the inverse scattering problem as an optimization problem. We describe in some detail the mathematical representation of the surfaces. Section 3 is devoted to the description of the algorithms studied. The main results are presented in Sect. 4, together with a discussion and, finally, in Sect. 5 we present our main conclusions.

2 Direct and Inverse Scattering

We consider the scattering of light from a one-dimensional, perfectly conducting rough surface defined by the equation $x_3 = \zeta(x_1)$. The region $x_3 > \zeta(x_1)$ is vacuum, the region $\zeta(x_1) > x_3$ is a perfect conductor, and the plane of incidence is the x_1x_3-plane. With reference to Fig. 1, the surface is illuminated from the vacuum side by a monochromatic s-polarized plane wave making an angle θ_0 with the x_3 axis.

The scattering amplitude $R_s(q|k)$ can be written in the form [8]

$$R_s(q|k) = \frac{-i}{2\alpha_0(k)} \int_{-\infty}^{\infty} dx_1 \left[\exp\{-i\alpha_0(q)\zeta(x_1)\}F(x_1|\omega)\right] \exp\{-iqx_1\}, \qquad (1)$$

where

$$F(x_1|\omega) = \left(-\zeta'(x_1)\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_3}\right) E_2(x_1, x_3)\Big|_{x_3=\zeta(x_1)} \qquad (2)$$

represents the source function, $E_2(x_1, x_3)$ is the only nonzero component of the electric field, $\alpha_0(k) = \sqrt{\omega^2/c^2 - k^2}$, and $\alpha_0(q) = \sqrt{\omega^2/c^2 - q^2}$. The angles of incidence θ_0 and scattering θ_s are related to the components of the incident

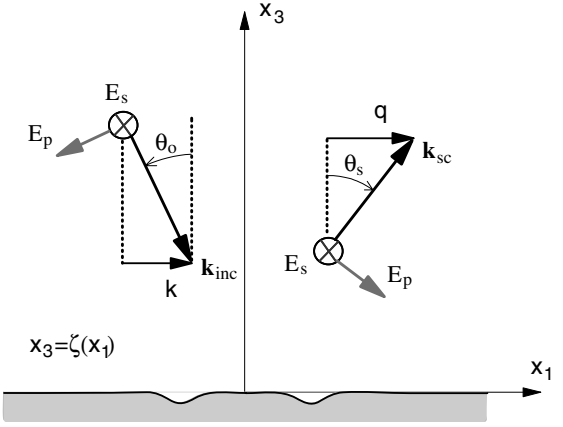

Fig. 1. Geometry of the scattering problem considered.

and scattered wavevectors that are parallel to the mean surface through the expressions

$$k = \frac{\omega}{c} \sin\theta_0 \,, \qquad\qquad q = \frac{\omega}{c} \sin\theta_s \,, \tag{3}$$

where c is the speed of light and ω is the frequency of the optical field. The far-field intensity $I_s(q|k)$ is defined as the squared modulus of the scattering amplitude $R_s(q|k)$. The angular distribution of the intensity is then represented by $I(q|k)$ as a function of q.

Due to the lack of general analytical solutions for the direct scattering problem, and the evident complexity of the inverse scattering problem, we approach it as an optimization problem. The goal is to retrieve the unknown surface profile function from scattered intensity data. The angle-resolved scattered intensity used as input data for our algorithms is obtained by rigorous numerical techniques [8]. For this, the source function $F(x_1|\omega)$ is determined by solving numerically an integral equation [8], and the scattering amplitude $R_s(q|k)$ is found through Eq. (1).

As we have mentioned, although the far-field scattered intensity depends on the surface profile function in a complicated way, for one-dimensional surfaces the direct problem can be solved numerically [8]. The closeness of a proposed profile, $z_c(x_1)$, to the original one can be estimated through the difference between the measured angular distribution of intensity $I^{(m)}(q|k)$, and the angular distribution of intensity $I^{(c)}(q|k)$, obtained by solving the direct scattering problem with the trial profile $z_c(x_1)$. The goal then would be to find a surface for which the condition $I^{(c)}(q|k) = I^{(m)}(q|k)$ is satisfied. When this happens, and if the solution to the problem is unique, the original profile has been retrieved.

We, thus, define our fitness (objective) function as:

$$f(\zeta(x_1)) = \sum_{i=1}^{N_{\text{ang}}} \int \left| I_s^{(m)}(q|k_i) - I_s^{(c)}(q|k_i) \right| dq, \tag{4}$$

where N_{ang} represents the number of angles of incidence considered, and the $k_i's$ are related to those angles through Eq. (3).

The inverse scattering problem is then reduced to the problem of minimizing $f(\zeta(x_1))$.

2.1 Representation of the Object Variables

From the preceding discussion it is natural to take the surface heights, or a quantity related to them, as the object variables. It is convenient to choose a representation scheme based on Shoenberg's variation diminishing approximation [10,11]. Assuming that the function $\zeta(x_1)$ consists of planar open curve defined within the interval $[-L/2, L/2]$, it can be represented parametrically in terms of spline curves.

First, the interval $[-L/2, L/2]$ is divided into $(m-1)$ equally-spaced intervals of length $\Delta\tau = L/(m-1)$, where m represents the number of control points. The j^{th} control point, P_j, has coordinates (τ_j, α_j), with $\tau_j = -L/2 + (j-1)\Delta\tau$. The heights α_j of the control points will be chosen as the object variables.

The knots are points at which adjacent spline curves join. The number of knots, ν, is related to the number of control points and the order of the spline, k, by the relation $\nu = m+k$. Throughout this work the knot sequence is assumed uniform but, to force the ends of the curve coincide with the first and last control points, we take

$$t_1 = t_2 = \cdots = t_k = -L/2, \qquad t_{m+1} = t_{m+2} = \cdots = t_{m+k} = L/2. \tag{5}$$

The location of the other (internal) knots is computed through the expression

$$t_{k+j} = \frac{\tau_j + \cdots + \tau_{j+k-1}}{k-1} \qquad j = 1, \ldots, m. \tag{6}$$

The j^{th} B-spline function of order $k = 1$ is defined as

$$B_j^{(1)}(x_1) \triangleq \begin{cases} 1 & \text{for } t_j \leqslant x_1 < t_{j+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

For $k > 1$, the B-spline functions are evaluated recursively using the Cox-de Boor Algorithm [12,13],

$$B_j^{(k)}(x_1) \triangleq \omega_j^{(k)}(x_1) B_j^{(k-1)}(x_1) + [1 - \omega_{j+1}^{(k)}(x_1)] B_{j+1}^{(k-1)}(x_1) \tag{8}$$

where

$$\omega_j^{(k)}(x_1) \triangleq \begin{cases} \frac{x_1 - t_j}{t_{j+k-1} - t_j} & \text{for } t_j \neq t_{j+k-1}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The surface profile function can then be written as

$$\zeta(x_1) = \sum_{j=1}^{m} \alpha_j B_j^{(k)}(x_1). \tag{10}$$

An example of a one-dimensional random profile represented as a spline curve is shown in Fig. 2. The solid line is the random surface contained by the control polygon, shown with a dashed line, and the control points are represented by the diamonds. In this case, the order of the B-spline functions was $k = 4$. Note that the multiplicity of the knots at the beginning and end of the interval expressed by Eq.(5) implies that the profile $\zeta(x_1)$ is tangent to the control polygon at the first and last control points.

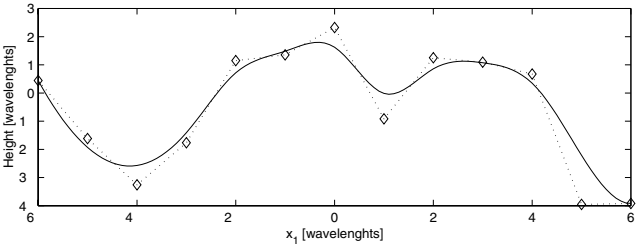

Fig. 2. Random profile represented as a spline curve (solid line). The control polygon is shown with a dashed line and its vertices (control points) with diamonds.

3 Description of the Inversion Algorithms

To retrieve the surface profile, we have studied some hybrid inversion algorithms, combining evolutionary strategies (ES)[14,15] and the simplex method of Nelder and Meade[9]. Using Schwefel's notation,[14] the evolutionary strategies employed are the elitist $(\mu/\rho+\lambda)$ and the non-elitist $(\mu/\rho, \lambda)$ strategies with recombination. Here μ is the number of parents in the initial population, λ is the number of offsprings generated by means of the genetic operators, and ρ is the number of members of the population that generate an intermediate population through recombination[15]. For the case $\rho = 1$, we have the canonical ES. Once the evolution strategies have reached the termination criterion, a local search algorithm (the simplex) is employed to improve the solution.

It is worth mentioning that we have previously used λ to denote the wavelength of the light, which is the usual notation in optical work. It is believed that due to the different context in which the two quantities are employed, the use of the same symbol to denote both should not lead to much confusion.

3.1 The Evolution Strategies

The initial population is a set of μ random surfaces represented by means of Eq. (10) as

$$\zeta^{(i)}(x_1) = \sum_{j=1}^{m} \alpha_j^{(i)} B_j^{(k)}(x_1) \qquad \text{with } i = 1, 2, ..., \mu\,. \tag{11}$$

The initial ordinates $\alpha_j^{(i)}$ of the control points are generated using zero-mean normally distributed random numbers with standard deviation σ, and the fitness values of all the elements of the population are evaluated through Eq. (4).

When $\rho > 1$, recombination is applied[15], and the newly generated ordinate of the j^{th} control point is given by

$$\langle \alpha_j^{(i)} \rangle_\rho = \frac{1}{\rho} \sum_{i=1}^{\rho} \alpha_j^{(i)} \qquad \text{with } j = 1, 2, \ldots, m. \tag{12}$$

Mutation is introduced through random changes in the population [16]. In our implementation, it takes the form

$$\left(\langle \alpha_j^{(i)} \rangle_\rho \right)_{\text{mut}} = \langle \alpha_j^{(i)} \rangle_\rho + N(0, \sigma_{\text{mut}}), \tag{13}$$

where $N(0, \sigma_{\text{mut}})$ is a sequence of statistically independent zero-mean Gaussian random numbers with standard deviation σ_{mut}. No self-adaptation schemes were used in this work.

After mutation, the fitness values of all the elements of the population are evaluated through Eq. (4) and the selection scheme is applied.

In our implementation for the elitist strategies $(\mu+\lambda)$ and $(\mu/\rho+\lambda)$, we chose $\mu = \lambda$. From the union of the initial and the intermediate populations we select a secondary population that consists of the μ surfaces with the lowest associated values of $f(\zeta(x_1))$. This secondary set of surfaces constitutes the starting point for the next iteration of the algorithm. The process continues until the termination criterion is reached, which in our case was set by the maximum number of iterations g.

In the non-elitist strategies (μ, λ) and $(\mu/\rho, \lambda)$, we start with an initial population of μ random surfaces, and keep $\lambda = 10\mu$ during the entire process. Through the recombination and mutation operations, λ new surfaces are generated. This secondary population is evaluated and the best μ random surfaces, those with the lowest associated fitness values, are selected to be the initial population for the next iteration of the algorithm. As with the elitist strategies, the process continues until a termination criterion is reached.

3.2 Multidimensional Local Search Algorithm

At this point, the output of the evolutionary strategy is coupled to a multidimensional local search algorithm, represented by the simplex method [9]. The

element with the lowest fitness value is selected and included in the initial population of the simplex. Other elements are generated randomly, employing the representation given by Eq. (10). The dimensions of the simplex, i. e. its number of vertices, is determined by the number of parameters to be optimized. For the present application, this is the m heights α_j of the control points. Thus, the simplex will have $m+1$ vertices, each one of them associated with a surface and a fitness value $f(\zeta(x_1))$. Each element of the population may be visualized as different points in the search space.

The simplex conducts the search towards the optimum by changing its size through a series of defined operations [17]. The aim in each iteration of the algorithm is to find a new point for which the fitness value $f(\zeta(x))$ is lower than the worst fitness value of the previous iteration. The search process ends when a multiple contraction takes place and the simplex collapses to a point. When this happens, the simplex has converged to the optimum and all the members of the population are identical.

4 Results and Discussion

In principle, the data that serves as input to the algorithm should be obtained experimentally. However, in order to evaluate the performance of the proposed algorithms, in these studies we use data obtained through a rigorous numerical solution of the direct scattering problem [8].

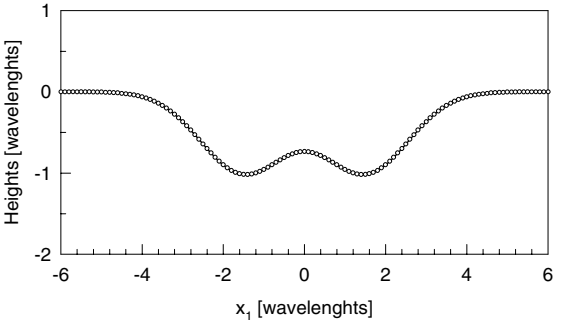

Fig. 3. The profile used in the generation of the scattering data.

For the strategies explored, each element of the initial population consisted of a randomly rough surface. For the $(\mu/\rho+\lambda)$ strategies we chose $\mu = \lambda = 100$, whereas for the $(\mu/\rho, \lambda)$ strategies we set $\mu = 10$ and $\lambda = 100$. The number of elements to be selected from the initial population, depending on whether the recombination operation was performed or not in the search process, was respectively set to $\rho = 2$ or 1. The maximum number of iterations was $g = 300$, which also provided the termination criterion.

In the numerical experiments we considered the retrieval of the deterministic profile shown in Fig. 3. Since the time of computation required to find the optimum increases with the length of the surface, in order to keep the problem to a manageable size, we chose a surface of length $L = 12.1\lambda$. We used 13 control points, splines of order $k = 4$, and 17 knots. The heights α_j of the control points, were generated using normally distributed random numbers with zero mean and a standard deviation of $\sigma = 2\lambda$. For the scattering calculations we used a sampling of $\Delta x = \lambda/10$, so that we had $N = 121$ sampling points on the surface.

The data from which the profile is to be recovered were obtained by illuminating the surface in Fig. 3 from four different directions, defined by the angles of incidence $\theta_0 = -60°$, $\theta_0 = -30°$, $\theta_0 = 0°$, and $\theta_0 = 40°$. In Fig. 4, we show (solid curve) the scattering pattern produced by the surface in Fig. 3 for the case of normal incidence.

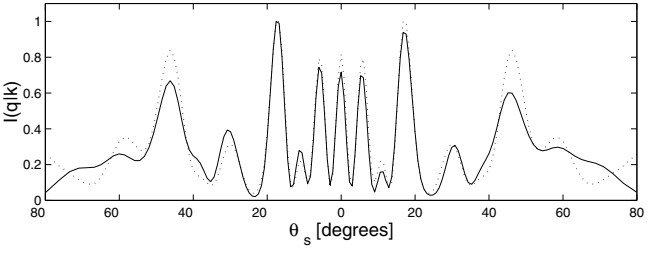

Fig. 4. Normalized scattered intensity produced by the surface depicted in Fig. 3 for the case of normal incidence. The dotted line shows the scattering pattern generated by the surface in Fig. 6(c) (solid curve) under the same conditions of illumination.

The described hybrid algorithms were tested for their relative success by searching for the solution starting from 30 different initial states. Each random initial state consisted of μ parents (surfaces). Not in all of the 30 attempts to recover the profile the algorithms converged to the correct surface. However, we found that a low value of $f(\zeta(x_1))$ corresponded, in most cases, to a profile that was close to the original one. So, the lowest value of $f(\zeta(x_1))$ was used as the criterion to decide whether the function profile had been reconstructed or not.

The convergence behaviour for one run of each of the studied algorithms is presented in Fig. 5. The plots corresponding to the $(\mu/\rho^+_{,}\lambda)$-ES are shown with solid lines to the left of $g = 300$, whereas a dotted line is used for the canonical strategies $(\mu^+_{,}\lambda)$. The fitness functions for iterations beyond $g = 300$, obtained by means of the simplex algorithm, are also shown in Fig. 5. The discontinuities in the plots (b) and (c) at $g = 300$ are due to the presence, in the initial populations of the simplex, of the best solutions found by the non-elitist strategy. It is worth pointing out that the highest final fitness values of Fig. 5 correspond to the canonical $(\mu^+_{,}\lambda)$-ES.

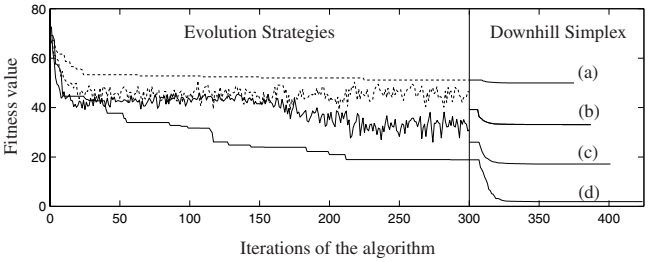

Fig. 5. Convergence behaviour for the best approximation found by the each one of the studied algorithms. The labels correspond, respectively to: (a) the $(\mu + \lambda)$-ES + downhill simplex, (b) the (μ, λ)-ES + downhill simplex, (c) the $(\mu/\rho, \lambda)$-ES + downhill simplex, and (d) the $(\mu/\rho + \lambda)$-ES + downhill simplex.

In Fig. 6 we present typical results obtained with the algorithms considered. The illumination conditions were the same as those used for the generation of the scattering information shown in Fig. 4. Figs. 6(a) and (b) correspond to the case $\rho = 1$, while Figs. 6 (c) and (d) to the case $\rho = 2$. To facilitate the visualization of the results the original profile is shown with circles.

When only the mutation operator was used in the search for the optimum ($\rho = 1$) we obtain the worst convergence behaviour. The curve shown in Fig. 5(a) corresponds to the reconstruction of Fig. 6(a). It is clear that the retrieved profile does not resemble the original one. The characteristic local search behaviour of the downhill simplex is illustrated by the fact that the improvement of the fitness function after the point $g - 300$ is small. A similar situation can be also seen in Fig. 6(b) for the (μ, λ)-ES.

For the case $\rho = 2$, we obtain a striking result that also illustrates the lack of uniqueness of the solution when intensity data are used. This is shown in Fig. 6(c). The output of the $(\mu/\rho, \lambda)$ evolution strategy is shown with the dotted line, while the solid line corresponds to the solution refined by the downhill simplex. The evolution of the corresponding fitness values is shown as curve (c) in Fig.(5). Although the reconstructions of Fig. 6(c) do not resemble the original profile, the scattering patterns of Fig. 4 demonstrate the fact that different profiles can generate the same, or similar, scattering information.

These results also illustrate a curious property of the scattering problem. If the reconstructions of Fig. 6(c) are reflected with respect to the x_1 and x_3 axis, the resulting profile resembles the sought one. It can be shown [18] that within the Kirchhoff approximation, and if polarization effects are not important, the far-field intensity is invariant under this kind of operation. Such situations lead to multiple solutions of the inverse scattering problem. Interestingly, polarization and multiple scattering effects reduce the number of possible solutions to the inverse problem.

The solution found with the $(\mu/\rho + \lambda)$-ES is depicted with a dotted line in Fig. 6(d). In this case the algorithm retrieved the profile quite well; there are only some subtle differences that are perhaps more noticeable near the ends of the surface. An improvement is obtained when the downhill simplex is used

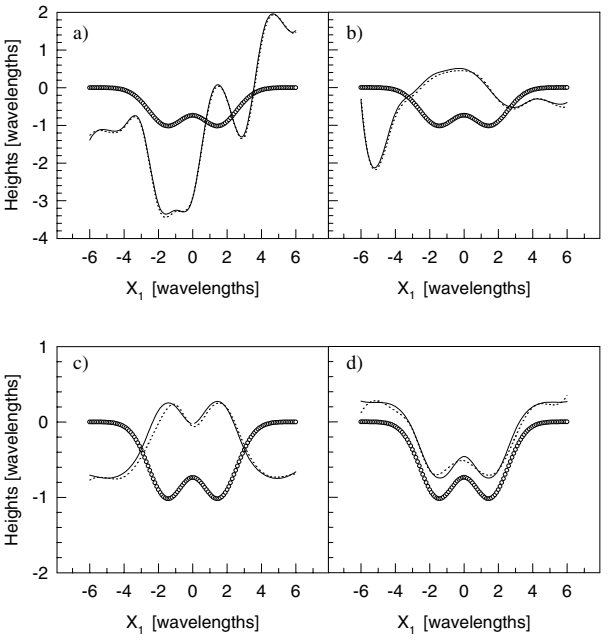

Fig. 6. Reconstruction of the surface profile using: (a) the $(\mu + \lambda)$-ES + downhill simplex, (b) the (μ, λ)-ES + downhill simplex, (c) the $(\mu/\rho, \lambda)$-ES + downhill simplex, (d) the $(\mu/\rho+\lambda)$-ES + downhill simplex. The original profile is plotted with circles and the reconstructions are depicted with a dotted curve for the two evolution strategies, and with a solid curve for the downhill simplex algorithm.

to refine the solution. This can be concluded from the behavior of the fitness function shown in Fig. 5(d). The solution found is shown with the solid line in Fig. 6(d). The vertical displacement of the reconstructed profiles of Fig. 6(d) is quite understandable, as the far-field intensity is insensitive to such shifts. It is also worth pointing out that the displacement is unimportant for practical profilometric applications.

5 Summary and Conclusions

A hybrid evolutionary approach to the inverse scattering problem has been successfully applied to the reconstruction of a one-dimensional, perfectly conducting rough surface. The use of B-splines provides an effective method for the representation of the object variables. In addition, its implementation and evaluation are straightforward. Furthermore, unlike the spectral method used in [6], the representation with B-splines allows the generation of deterministic surfaces.

For the numerical examples presented in this work, the recombination operator played a fundamental role in the convergence to the optimum. Also, the hybridization lead to an additional improvement of the solution found by the evolution strategy.

Acknowledgments. The authors are grateful to CONACyT (México) for financial support.

References

1. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, (Pergamon Press, London, 1963), p. 29.
2. J. A. Oglivy, *Theory of wave scattering from random rough surfaces*, (Institute of Physics Publishing, Bristol, 1991), p. 277.
3. Wombel, R. J., DeSanto, J. A.: Reconstruction of rough-surface profiles with the Kirchhoff approximation, J. Opt. Soc. Am. A **8**, 1892, (1991).
4. Quartel, J. C., Sheppard, C. J. R.: Surface reconstruction using an algorithm based on confocal imaging, J. Modern Optics **43**, 496, (1996).
5. Macías, D., Méndez, E. R., Ruiz-Cortés, V.: Inverse scattering with a wavefront matching algorithm, J. Opt. Soc. Am. A **19**, 2064, (2002).
6. Macías, D., Olague, G., Méndez, E. R.: Surface profile reconstruction from scattered intensity data using evolutionary strategies, in *Applications of Evolutionary Computing*, S. Cagnoni, *et. el*, eds., (Springer LNCS**2279**, Berlin), 233, (2002).
7. Thorsos, E. I.: The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum, J. Acoust. Soc. Amer. **83**, 78 (1988).
8. Maradudin, A. A., Michel, T., McGurn, A. R., Méndez, E. R.: Enhanced backscattering of light from a random grating, Ann. Phys. (N. Y.) **203**, 255 (1990).
9. Nelder, J., Mead, R.: A simplex method for function optimization *Computer Journal*, **7**, 308,(1965).
10. Boor, C. de: *A Practical Guide to Splines*, (Springer-Verlag, NY, 1978), p.154.
11. Boor, C. de: Spline Basics, In: Handbook of Computer Aided Geometric Design, G. E. Farin, J. Hoschek and M. Kim (eds.), Elsevier, pp. 141–164, (2002).
12. Cox, M. G.: The numerical evaluation of *B*-splines, *J. Inst. Math. Applics.*, **10**,134, (1972).
13. Boor, C. de: On calculating with *B*-splines, *J. Approx. Theory*,**6**, 50,(1972).
14. H. P. Schwefel, *Evolution and Optimum Seeking*, (John Wiley & Sons Inc., NY, 1995), p. 444.
15. H. G. Beyer, *The Theory of Evolution Strategies*, (Springer-Verlag, Berlin, 2001), p.380.
16. Th. Bäck, U. Hammel and H.-P. Schwefel: Evolutionary computation: Comments on the history and current state. IEEE Transactions on Evolutionary Computation, 1(1):3–17, 1997.
17. Press, W. H. (editor) *Numerical Recipes in Fortran*, (Cambridge University Press, Cambridge, 1992) p. 402.
18. Macías, D., Méndez, E. R.: Unpublished work (2001).