

Robust optimization with simulated annealing

Dimitris Bertsimas · Omid Nohadani

Received: 17 March 2009 / Accepted: 11 November 2009
© Springer Science+Business Media, LLC. 2009

Abstract Complex systems can be optimized to improve the performance with respect to desired functionalities. An optimized solution, however, can become suboptimal or even infeasible, when errors in implementation or input data are encountered. We report on a robust simulated annealing algorithm that does not require any knowledge of the problems structure. This is necessary in many engineering applications where solutions are often not explicitly known and have to be obtained by numerical simulations. While this nonconvex and global optimization method improves the performance as well as the robustness, it also warrants for a global optimum which is robust against data and implementation uncertainties. We demonstrate it on a polynomial optimization problem and on a high-dimensional and complex nanophotonic engineering problem and show significant improvements in efficiency as well as in actual optimality.

Keywords Robust optimization · Simulated annealing · Global optimization · Nonconvex optimization

1 Introduction

Optimization has had a distinguished history in engineering and industrial design. Most approaches, however, assume that the input parameters are precisely known and that the implementation does not suffer any errors. To accommodate these sources of errors, sensitivity analysis techniques were developed which find solutions that are least sensitive among a larger set of optima. However, these methods do not provide designs that are intrinsically robust against errors.

D. Bertsimas · O. Nohadani (✉)
Sloan School of Management and the Operations Research Center, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
e-mail: nohadani@mit.edu

D. Bertsimas
e-mail: dbertsim@mit.edu

There has been evidence illustrating that if errors (in implementation or estimation of parameters) are not taken into account during the design process, the actual phenomenon can completely disappear. A prime example is optimizing the truss design for suspension bridges. The Tacoma Narrows bridge was the first of its kind to be optimized to divert the wind above and below the roadbed [1]. Only a few months after its opening in 1940, it collapsed due to moderate winds which caused twisting vibrational modes. In another example, Ben-Tal and Nemirovski demonstrated that only 5% errors can entirely destroy the radiation characteristics of an otherwise optimized phased locked and impedance matched array of antenna [2]. Therefore, taking errors into account during the optimization process is a first order effect.

These considerations have motivated the field of robust optimization. Recent works have been devoted to problems with convex objectives and constraints (e.g. linear) [3–5]. These works have shown that a convex optimization problem with parameter uncertainty can be transformed to another convex optimization problem. This transformation can be either exact or through a relaxation. However, the final problem can be more complex or can have a significantly larger number of constraints. Despite significant advancements, all these results are limited to convex problems.

Modern engineering design has objectives and constraints that are not explicitly given. Moreover, solutions are often obtained through numerical simulations. This means that no internal structure can be exploited. The challenge, thus, lies in robustly designing engineered and engineering systems which are described through simulations and are exposed to errors.

More recently, local search methods have been expanded to robust optimization of simulated-based problems [6–8]. They entail finding descent directions and iteratively taking steps along these directions to optimize the robustness, i.e., reducing the worst-case cost. A second order cone problem was solved to determine a direction that makes the largest angle to all worst-case neighbors that were found to deviate the most from the desired solution. The number of regarded worst-case scenarios and the dimensionality determined the computational efficiency of these algorithms [8]. These approaches, however, provided only local robust optima.

Here, we present a more generic method, that is independent of the problem structure and, thus, is not restricted to dimensionality or other topological features. In fact, it only requires a black-box that provides the function evaluation and the derivative for a given design. Moreover, we show that this technique can provide a global robust optimum at computationally reasonable costs. We showcase its performance on a 100-dimensional optimization problem in nanophotonic design, that has a highly nonconvex cost function as well as nonconvex search space.

2 Method

In general, to measure the deviations between the desired and current performance of a design \mathbf{x} , a cost function f is defined. For the purpose of generality, we do not impose any assumption of the structure on f , i.e., f can be nonconvex or simply given through a numerical simulation. The nominal optimization aims to minimize f . There are numerous techniques to achieve this goal for which we refer to the standard literature [9–11]. In real-world problems, the design variables \mathbf{x} are subject to errors $\Delta\mathbf{x}$. Therefore, possible designs are regarded to reside within an uncertainty set

$$\mathcal{U} = \{\Delta\mathbf{x} \in \mathbb{R}^n \mid \|\Delta\mathbf{x}\| \leq \Gamma\}, \quad (1)$$

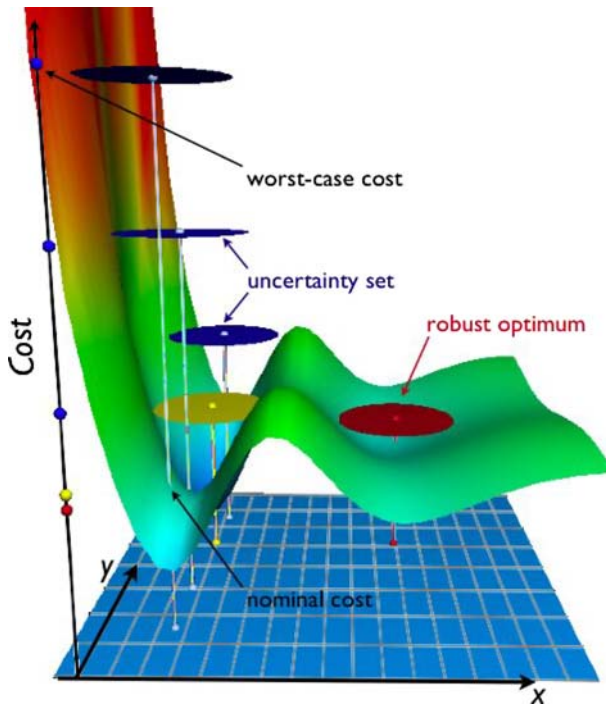


Fig. 1 2D-Sketch: Robustness is determined by worst-case costs within the uncertainty set (discs) of a design (point in x - y plane). Robust optimization seeks to place the lowest disc (red), even lower than the nominal optimum (yellow)

where Γ denotes the size of \mathcal{U} , determined by empirical error bars. While our approach applies to other norms $\|\Delta\mathbf{x}\|_p \leq \Gamma$ in (1) (p being a positive integer, including $p = \infty$), we present the case of $p = 2$. To find a design that is robust against all possible errors in \mathcal{U} , the most conservative approach is to find worst-case scenarios $g(\mathbf{x}) = \max_{\Delta\mathbf{x} \in \mathcal{U}} f(\mathbf{x} + \Delta\mathbf{x})$ and seek to minimize them by modifying the design appropriately. This *robust optimization problem* can be expressed through

$$\min_{\mathbf{x}} g(\mathbf{x}) \equiv \min_{\mathbf{x}} \max_{\Delta\mathbf{x} \in \mathcal{U}} f(\mathbf{x} + \Delta\mathbf{x}). \tag{2}$$

Since f is nonconvex and possibly not given in an analytically closed form, the inner maximization problem will be nonconvex and without a closed form as well. In our method, its solution is found using local searches within the uncertainty set of the design (see Fig. 1). These searches, such as gradient ascent sequences, identify a potentially discrete set $\mathcal{M}(\mathbf{x})$ of “bad neighbors” $\hat{\mathbf{x}}$ which have the highest costs in \mathcal{U} . Once $\mathcal{M}(\mathbf{x})$ is assembled for a given design \mathbf{x} , the robust optimization method seeks to iteratively update \mathbf{x} in order to exclude the elements of $\mathcal{M}(\mathbf{x})$ and consequently finds designs that have lower worst-case neighbors, thus improve robustness. In earlier works, the robust local method was used to find a local robust optimum [6, 8].

To find the global robust minimum, we introduce the method of robust simulated annealing (RSA). In a pioneering work, Kirkpatrick et al. introduced the method of simulated annealing for discrete optimization problems [12]. Relying on one of the early works in scientific computing by Metropolis et al. [13], they replaced the energy of an ensemble by the cost

function of a configuration and showed the connection between optimizing a multivariate function and the behavior of many-body systems with a large degree of freedom in thermal equilibrium. Motivated by these works, we rewrite $g(\mathbf{x})$ from Eq. 2 as follows

$$g(\mathbf{x}) = \lim_{\beta \rightarrow \infty} g_{\beta}(\mathbf{x}) = \lim_{\beta \rightarrow \infty} \left\{ \frac{1}{\beta} \log \int_{\hat{\mathbf{x}} \in \mathcal{M}(\mathbf{x})} e^{\beta f(\hat{\mathbf{x}})} d(\hat{\mathbf{x}}) \right\}. \tag{3}$$

We interpret the parameter β as the inverse temperature. We set β_0 to be the highest cost in the bad neighbors set $\mathcal{M}(\mathbf{x})$ (see Eq. 4 below). In the RSA algorithm, $g_{\beta}(\mathbf{x})$ is computed as a sum over the members of $\mathcal{M}(\mathbf{x})$ and the $\lim_{\beta \rightarrow \infty}$ in Eq. 3 is approximated by the largest available β . The RSA algorithm maximizes the robustness by iteratively minimizing the worst-case scenario of the current design \mathbf{x}_k as following:

RSA Algorithm:

Step 0. Initialization:

- (i) Set iterate $k = 0$, acceptance number $h = 0$, and annealing index $n = 0$.
- (ii) Let \mathbf{x}_0 be the initial decision vector that is arbitrarily chosen.
- (iii) Assemble $\mathcal{M}(\mathbf{x}_0)$ through a series of gradient ascent sequences within $\mathcal{U}(\mathbf{x}_0)$ [8].
- (iv) Compute the initial inverse temperature β_0 according to

$$\beta_0 = \frac{1}{\max_{\hat{\mathbf{x}} \in \mathcal{M}(\mathbf{x}_0)} f(\hat{\mathbf{x}})}. \tag{4}$$

- (v) Compute $g(\mathbf{x}_0) \equiv g_{\beta_0}(\mathbf{x}_0)$ according to Eq. 3.
- (vi) Determine the Boltzmann-weight $W(\mathbf{x}_k)$ through

$$W(\mathbf{x}_k) = e^{-\beta g(\mathbf{x}_k)}. \tag{5}$$

Step 1. Trial design:

- (i) From a Gaussian probability distribution, randomly generate $\mathbf{x}_{\text{trial}}$ such that $\|\mathbf{x}_{\text{trial}} - \mathbf{x}_k\|_2 \leq \Gamma$.
- (ii) Assemble $\mathcal{M}(\mathbf{x}_{\text{trial}})$ within $\mathcal{U}(\mathbf{x}_{\text{trial}})$ as above.
- (iii) Compute $g(\mathbf{x}_{\text{trial}}) \equiv g_{\beta}(\mathbf{x}_{\text{trial}})$ according to Eq. 3.
- (iv) Determine the Boltzmann-weight $W(\mathbf{x}_{\text{trial}})$ as in Eq. 5.

Step 2. Compute the acceptance probability:

$$P(\mathbf{x}_k \rightarrow \mathbf{x}_{\text{trial}}) = \min \left(1, \frac{W(\mathbf{x}_{\text{trial}})}{W(\mathbf{x}_k)} \right) \tag{6}$$

from the respective weights to permit an update, corresponding to the Metropolis update scheme [13].

Step 3. Generate a random number $R \in [0, 1]$:

- (i) if $P \geq R$: $\mathbf{x}_{\text{trial}}$ is accepted
 - (a) $k = k + 1$ and $\mathbf{x}_{k+1} = \mathbf{x}_{\text{trial}}$
 - (b) $h = h + 1$, and
 - (A) if $(h \bmod 100) \neq 0$, then $g(\mathbf{x}_{k+1}) = g(\mathbf{x}_{\text{trial}})$ and $W(\mathbf{x}_{k+1}) = W(\mathbf{x}_{\text{trial}})$.

(B) else $n = n + 1$ and update the inverse temperature via

$$\beta_n = \beta_0 e^{cn}, \tag{7}$$

with $c = 1 + \epsilon$. Recompute $g(\mathbf{x}_{k+1})$ and $W(\mathbf{x}_{k+1})$ with the updated β .

(c) Back to Step 1 and iterate.

(ii) if $P < R$: $\mathbf{x}_{\text{trial}}$ is rejected. Back to Step 1 and reiterate.

Note that in this algorithm, there are 2 iteration indices: n denotes the annealing index, which increases whenever the inverse temperature is updated, and k denotes the trial index within each annealing iterate in order to find a more robust trial decision $\mathbf{x}_{\text{trial}}$ at the same inverse temperature.

Theorem 1 *Suppose that an arbitrary cost function $f(\mathbf{x})$ with a bounded set of minimum points has a global robust optimum. Then the RSA Algorithm converges to the global optimum of the robust optimization problem (2), if all designs \mathbf{x} are accessible with equal probability as $n \rightarrow \infty$.*

Proof Geman and Geman have shown that a generic simulated annealing algorithm converges to a global optimum, if β is selected to be not faster than $\beta_n = \ln(n)/\beta_0$ and if all accessible states are equally probable for $n \rightarrow \infty$ [14].

The first condition is satisfied by the choice of the annealing schedule in Eq. 7. For the second condition, let $H(\mathbf{x})$ be the probability density of the randomly generated designs $\mathbf{x}_{\text{trial}}$. In order to show that any design \mathbf{x} in the D -dimensional configuration space can be sampled infinitely often during the annealing time, we have to show that the probability of not visiting \mathbf{x} during the annealing time vanishes. In other words, we have to show that

$$\prod_{n=0}^{\infty} (1 - H_n(\mathbf{x})) = 0, \tag{8}$$

which can be expressed as

$$\sum_{n=0}^{\infty} H_n(\mathbf{x}) = \infty. \tag{9}$$

Ingber et al. [15] have shown that the functional form of random generators for Gaussian-Markovian based systems, like the one we employed to generate $\mathbf{x}_{\text{trial}}$, is given by

$$H(\mathbf{x}) = \left(\frac{\beta}{2\pi}\right)^{\frac{D}{2}} \cdot e^{-\frac{\beta}{2}\Delta\mathbf{x}^2}, \tag{10}$$

where $1/\beta$ is a measure of the fluctuations (standard deviation) and $\Delta\mathbf{x}$ is difference between the current and the previous design. As the algorithm progresses, β is updated. Thus, the Eq. 10 can be expressed as

$$H_n(\mathbf{x}) = \left(\frac{\beta_0 e^{cn}}{2\pi}\right)^{\frac{D}{2}} \cdot e^{-\frac{\beta_0 e^{cn}}{2}\Delta\mathbf{x}^2} \geq e^{-\ln(n)}. \tag{11}$$

Now, we can insert Eq. 11 into Eq. 9 and obtain

$$\sum_{n=0}^{\infty} H_n(\mathbf{x}) \geq \sum_{n=0}^{\infty} e^{-\ln(n)} = \sum_{n=0}^{\infty} \frac{1}{n} = \infty. \tag{12}$$

Therefore, as $n \rightarrow \infty$, all designs can be generated infinitely often. Moreover, the Metropolis weights are chosen such that the detailed balance condition

$$P(\mathbf{x} \rightarrow \mathbf{x}') \cdot W(\mathbf{x}) = P(\mathbf{x}' \rightarrow \mathbf{x}) \cdot W(\mathbf{x}'), \tag{13}$$

is satisfied which warrants for an equal probability that all states are accessible in the search space. Therefore, we can conclude that the local search based RSA algorithm converges to a global optimum as $n \rightarrow \infty$. \square

Note that the definition of the robustness in Eq. 3 contains β which varies throughout the annealing process. However, this does not affect the probability density $H(\mathbf{x})$ of the designs. Moreover, within each iteration of the RSA algorithm (and in fact over 100 acceptances), β is constant and acts only as a local renormalization factor. Even when β is updated in Step 3-ii-B, $g(\mathbf{x}_k)$ and $W(\mathbf{x}_k)$ are recomputed so that the current and the next trial design are compared at the same constant β ensuring a constant local probability density. Therefore, this enhancement to the weighting in $g(\mathbf{x})$ does not affect the convergence criteria of the RSA algorithm. Furthermore, the convergence behavior of RSA is similar to that of adaptive simulated annealing, since their cores are comparable. However, since the convergence rate is highly dependent on the underlying problem and our aim to provide the general foundation of RSA, we consider this discussion outside the scope of this work.

We demonstrate the performance of the RSA algorithm by applying it to two different optimization problems. The first application is a robust polynomial optimization problem that is intended to serve as a demonstration of the main features of RSA as well as its capability of providing a robust global optimum. This is a simple enough problem in which we can calculate the robust global optimum. The second application is on the robust optimization in engineering and is of direct relevance to nanophotonics and nanotechnology design, where small uncertainties may result in complete failure of an otherwise optimized design. This problem is high-dimensional, highly non-linear and its solution is not known explicitly, thus useful as a prototype for modern engineering design problems. This application is intended to demonstrate the performance of the RSA algorithm in a high-dimensional search space as well as its efficiency of providing the robust solution.

3 Application in polynomial optimization

For the sake of better demonstration of the RSA algorithm, we employ a two dimensional polynomial problem. We define a nonconvex polynomial function

$$\begin{aligned} f(x, y) = & 2x^6 - 12.2x^5 + 21.2x^4 + 6.2x - 6.4x^3 \\ & - 4.7x^2 + y^6 - 11y^5 + 43.3y^4 - 10y \\ & - 74.8y^3 + 56.9y^2 - 4.1xy - 0.1y^2x^2 \\ & + 0.4y^2x + 0.4x^2y. \end{aligned} \tag{14}$$

The implementation errors are bound as $\Delta = (\Delta x, \Delta y)$ by $\|\Delta\|_2 \leq 0.5$. Therefore, the robust optimization problem can be posed as

$$\min_{x,y} g(x, y) \equiv \min_{x,y} \max_{\|\Delta\|_2 \leq 0.5} f(x + \Delta x, y + \Delta y). \tag{15}$$

We need to stress that even though this problem is only two-dimensional, it is already a difficult problem. Henrion and Lasserre successfully introduced relaxation methods to solve

polynomial optimization problems [16, 17]. Applying the same technique to Problem (15), however, leads to polynomial semidefinite programs (SDP) which have the form of

$$\begin{aligned} \min_{x,y} & h(x, y) \\ \text{s.t.} & \mathbf{A}(x, y) \succeq \mathbf{0}. \end{aligned} \tag{16}$$

Note that the entries of the semidefinite constraint are assembled by multivariate polynomials A_{ij} . In order to solve this problem approximately, we need to convert it into a significantly larger SDP. More importantly, the size of the SDP grows with the size of the original problem, the maximum degree of the polynomials, and the number of variables. Because of this, it is evident that polynomial SDPs are not widely used in practice [18].

The nominal cost of $f(x, y)$ has multiple local minima and a global minimum at $(x^*, y^*) = (2.8, 4.0)$, where $f(x^*, y^*) = -20.8$. We determined the global minimum by using the Gloptipoly software as discussed in Reference [16] and verified using multiple gradient descents. We evaluated the worst-case cost function $g(x, y)$ by enumerating the nominal costs of neighbors using a fine discrete mesh. The Contour plot in Fig. 2 illustrates the worst-case cost surface and suggests that $g(x, y)$ has multiple local minima.

We applied the Robust Simulated Annealing algorithm to this polynomial problem using arbitrary initial designs (x, y) . Figure 2 shows one exemplary run and demonstrates the convergence towards to robust global minimum. The same global optimum is found when starting from other arbitrary initial points. The algorithm decreases the worst-case cost significantly while increasing the nominal cost slightly, which is the “price of robustness”. A much lower number of iterations is required when starting from a point that is not in the immediate proximity of a local (robust or nominal) minimum. At the termination point, the last configuration is surrounded by neighbors that have higher worst-case costs. In fact, detailed analysis of the distribution of bad neighbors at the termination showed that they lie on the boundary of the uncertainty set. Any additional trial configuration will lead to a vanishing update probability

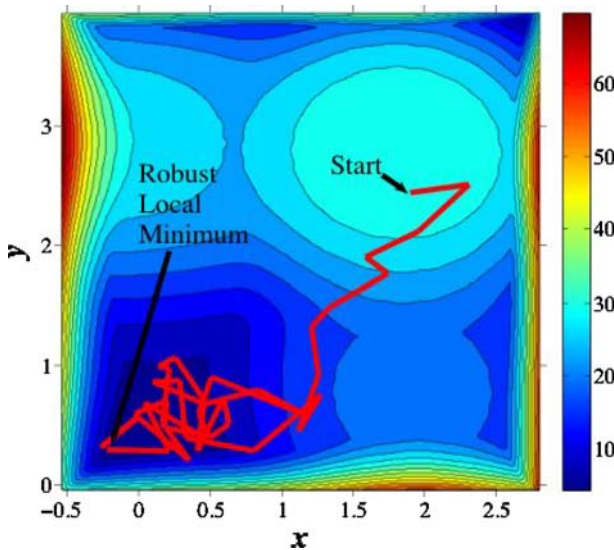


Fig. 2 2D-Sketch: Contour plot of worst-case cost function $g(x, y)$. The path of the progress of the Robust Simulated Annealing algorithm is shown with red lines as it starts at an arbitrary initial point and converges towards to robust global minimum

and, thus, will be rejected. The surface plot also confirms that the solution is indeed the true robust global minimum.

4 Application in nanophotonic optimization

In the second application, we demonstrate the efficiency and the performance of the RSA algorithm by applying it to a prototype electromagnetic problem, as it occurs in nanophotonic engineering. A plethora of unique characteristics in photonic crystals identified them as prime candidates for unconventional materials in controlling and manipulating electromagnetic field propagation [19]. Their peculiar functionalities are based on diffraction phenomena, which require periodic structures. Upon breaking the spatial symmetry, additional degrees of freedom are revealed which allow for additional functionality and higher levels of control. Geremia et al. broke this symmetry by diluting sites and optimizing the location of the missing scattering sites [20]. However, due to the periodicity requirement of these crystals, additional degrees of freedom and, hence, their benefit remained restricted. More recently, Seliger et al. performed unbiased gradient-based optimizations on the spatial distribution of a large number of dielectric cylinders [21]. The reported aperiodic structure matched a desired target function up to 95%.

When implemented in the real-world, however, the performance of many of these designs substantially deviates from the theoretical prediction. A key source of this deviation lies in the presence of uncontrollable implementation errors, such as incorrect positioning and erroneous shape of scatterers. Since nanophotonic structures exploit nonlinear features, they are highly susceptible to small perturbations. Therefore, a robustly optimized design warrants the desired performance and sustains errors.

We demonstrate this by applying the RSA method to an inverse design problem, that seeks to match the performance of 50 dielectric cylinders to a desired target function by varying the position of these scattering centers. In the following, we summarize the essentials of the physical model for the sake of completeness and refer for more details to References [6, 21]. The model is based on a two-dimensional Helmholtz equation for dielectric scatterers, which are lossless and non-magnetic. Therefore, this approach scales with frequency and allows to probe nanophotonic designs. The wave propagation is strictly two-dimensional, since the domain is bound by two metallic plates separated by less than half the incoming wave length. Figure 3 illustrates the setup of the domain along with the corresponding electric field strength. To conform with the laboratory experiment of Seliger et al. [21], we describe the electric field E_z through the partial differential equation (PDE)

$$(\partial_x(\mu_{r_y}^{-1}\partial_x) + \partial_y(\mu_{r_x}^{-1}\partial_y))E_z - \omega_0^2\mu_0\epsilon_0\epsilon_r E_z = 0, \tag{17}$$

where μ_r is the relative and μ_0 the vacuum permeability and ϵ_r denotes the relative and ϵ_0 the vacuum permittivity. Note, that in the experimental measurements, the frequency is fixed to $f = 37.5$ GHz and only stationary solutions of the Maxwell equations are sought, taking the boundary conditions into account. We compute the field values using the finite-difference representation of the PDE and solve a linear equation system. The power along the target surface for an incident angle θ is interpolated using the nearest four mesh points and their standard Gaussian weights $\mathbf{W}(\theta)$ as $s_{\text{mod}}(\theta) = \frac{\mathbf{W}(\theta)}{2} \cdot \mathbf{diag}(\mathbf{E}_z) \cdot \mathbf{E}_z$. The accuracy of this numerical model has been verified by actual laboratory experiments of identical configurations [6, 21].

The nominal optimization problem seeks to match the power profile along the detection surface to a top-hat target function s_{obj} , which has a constant maximum between $30^\circ \leq \theta \leq$

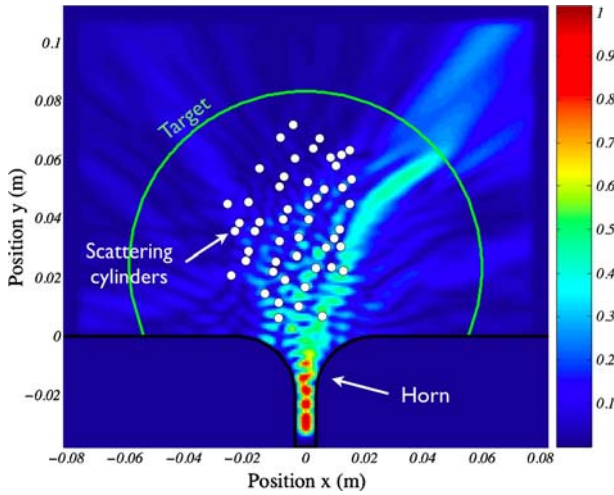


Fig. 3 The electric field of 50 dielectric scattering cylinders (*white circles*). The incoming radio frequency-source couples to the wave guide. For this configuration, the power is constant between $30^\circ \leq \theta \leq 60^\circ$ and amounts to 98% of the overall power

60° and is zero everywhere else. For any given discretized angle θ_k and configuration \mathbf{p} , a cost-functional J measures the deviation of s_{mod} from s_{obj} through

$$J(\mathbf{p}) = \sum_{k=1}^m |s_{\text{mod}}(\theta_k) - s_{\text{obj}}(\theta_k)|^2. \tag{18}$$

Note that $J(\mathbf{p})$ is not convex in \mathbf{p} , and depends on \mathbf{p} only through the PDE. Furthermore, the feasible set of all possible configurations is nonconvex, as cylinders should not overlap.

In laboratory experiments, implementation errors $\Delta \mathbf{p}$ are encountered, when physically placing the cylinders. To include most errors, we define the uncertainty set \mathcal{U} such that the probability $P(\Delta \mathbf{p} \in \mathcal{U})$ is 99%. Each component Δp_i is measured to be independently and normally distributed with a standard deviation $40\mu\text{m}$ [21]. Hence, we chose $\Gamma = 550\mu\text{m}$ for the size of the uncertainty set in Eq. 1.

Recently, a solution to this nanophotonic model was reported to be more robust by $\sim 8\%$ [6]. Since the underlying method relied on local searches, the reported robust optimum was a local optimum. However, it has been shown, that if the configurations are randomly generated through a Gaussian distribution, a Markovian system will converge to the global minimum, when a Boltzmann—or Cauchy cooling scheme is applied [13, 14]. In other words, the probability of the system not to generate a state which is the global minimum, vanishes. Therefore, applying the proposed RSA to this model warrants a global robust optimum.

To increase the efficiency of the search in this high-dimensional problem, we generate random states adapted to the annealing scheme. Starting from a uniform random number $r^i \in (0, 1)$ for the dimension i , we generate new random numbers

$$s^i = \text{sign}\left(r^i - \frac{1}{2}\right) \frac{1}{\beta} \left[(1 + \beta)^{|2r^i - 1|} - 1 \right], \tag{19}$$

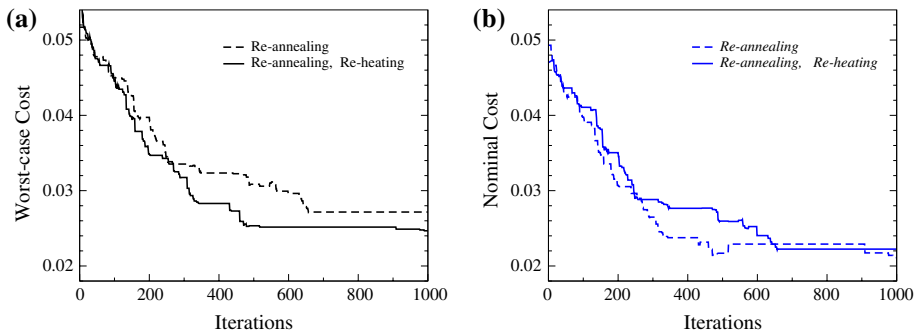


Fig. 4 Performance of different annealing schemes for RSA on the 100-dimensional nanophotonic problem, starting from a regular arrangement given by the underlying photonic crystal and using re-annealing at every 100 acceptances and re-heating at every 50 rejections: **a** worst-case cost performance and **b** nominal cost performance

reflecting the cooling schedule in Eq. 7. This adaptation creates random numbers $s^i \in (-1, 1)$. Consequently, we generate trial configurations \mathbf{p} according to

$$\mathbf{p}_{\text{trial}} = \mathbf{p} + \|\Delta\mathbf{p}\| \cdot \mathbf{s} \tag{20}$$

It has been shown that by successive application of this adaptive sampling of random states (for iteration number $k \rightarrow \infty$), the probability of not generating a state which corresponds to the global minimum vanishes [22]. To increase the efficiency of the algorithm, we advanced the cooling scheme by using a *re-annealing* technique, which resets the iteration index every 100 acceptances and increases β_0 by 5%. Additionally, we allowed a *re-heating* every 50 rejections, where the iteration index is reset and β_0 is decreased by 5%. The re-annealing pushes the system faster towards the minimum while avoiding quenching effects (local minima of small extend). The re-heating allows the search to escape local minima that have larger extends (bath-top shapes). Figure 4 illustrates the difference between these two techniques. While the re-annealing technique improves the nominal and the worst-case cost sufficiently fast, it gets stuck in an many small local minima for the robust cost, as shown in Fig. 4a. Using the additional re-heating, the algorithm improves both costs at comparable efficiency, however, it manages to escape local minima and, thus, achieves overall high-quality results faster. Note that the choice of the parameters is certainly correlated to the actual problem and, thus, cannot be generalized in this scope. Furthermore, since the algorithm minimizes only the worst-case cost, the performance of the nominal cost is merely a useful “side-product” of RSA.

To allow a comparison to previous methods, our starting configuration corresponds to the initial arrangement used in the robust local search (RLS), as reported in Ref. [6]. Since RSA conducts a global search, the initial configuration will neither affect the accessibility of the search space, nor alter the quality of the final solution. Figure 5 illustrates the performance of RSA for the nanophotonic problem, starting from a nominally optimized design, i.e., without taking robustness into account. The nominal and worst-case costs of the configuration at any given iteration step are plotted as the algorithm progresses. Starting from the initial configuration, within a few iterations, RSA improves the robustness significantly, i.e., the worst-case cost decreases. Since RSA updates the states according to the Boltzmann-weights, it rapidly scans large portions of the search space before a new updated state is accepted. As Fig. 5 shows, RSA outperforms previous robust optimization techniques. While both algorithms

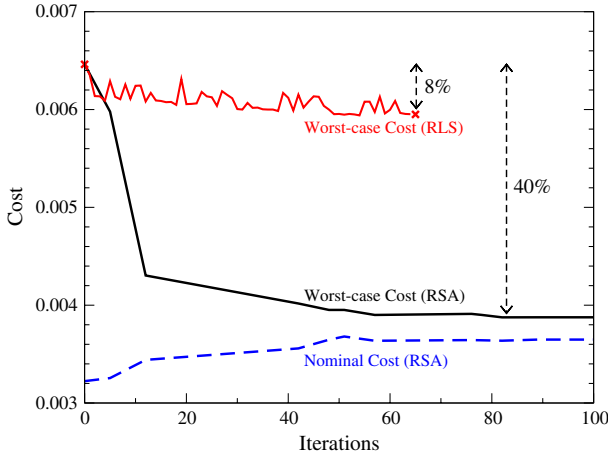


Fig. 5 Performance of the Robust Simulated Annealing on the 100-dimensional nanophotonic problem starting from a nominally optimized configuration. The nominal and worst-case costs of each configuration are plotted over the iteration steps. The cross displays the previously reported robust solution using the robust local search algorithm [6]. The nominal cost increases slightly as the “price” of improved robustness

require comparable times for the neighborhood search, RLS will require additional computation (approximately 20 min) to find a deterministic descent direction. Moreover, even though RSA is not deterministic and can yield different solutions for different runs and initial configurations, it will provide the lowest possible optimum for $k \rightarrow \infty$, thus, the global robust minimum. We terminated the simulations in Fig. 5 after the number of trials for the same acceptance step exceeded an empirical threshold.

Note that during the performance of RSA, only a few “surprises” are discovered, as opposed to RLS. This is due to the fact that in RSA steps are only taken, if no worst neighbors are residing around the new update, whereas in RLS the neighborhood search of the update configuration is conducted after the step has been taken.

These results in Fig. 5 illustrate that the presented method of robust simulated annealing increases the robustness of these prototype nanophotonic structures significantly. In fact, a design that is more robust against larger errors allows for alternative choices of manufacturing, i.e., costly high-precision implementation might become redundant, if the design is intrinsically robust against larger errors. In addition to the reduced production costs and higher manufacturing yield, these robust solutions are also protected against errors, whose sources were unknown at the time of the optimization. This feature extends the advantage of the robust design far beyond its conventional notion that was limited to protection against known uncertainties, thus, provides enhanced performance reliability of novel nano-structures.

5 Conclusion

We have introduced the generic algorithm of robust simulated annealing, which provides the global robust optimum of complex systems. Since it relies on the solver of the problem as a black-box, this method is generic and does not require any detailed knowledge of the structure of the problem, thus, is applicable to most engineering design problems which employ numerical simulations to compute relevant quantities. Furthermore, the method and

its efficiency are independent of the definition of the cost function or the dimensionality of the problem. We have applied the method to a robust polynomial optimization problem and shown that it finds efficiently the robust global optimum. Moreover, we have demonstrated its performance by applying it to an actual nanophotonic design problem and shown that the solution outperforms in terms of efficiency as well as absolute robustness results any previously available methods.

Acknowledgments We thank A.F.J. Levi for fruitful discussions.

References

1. Petroski, H.: Design Paradigms. Cambridge University Press, Cambridge (1994)
2. Ben-Tal, A., Nemirovski, A.: Robust optimization—methodology and applications. *Math. Progr.* **92**, 453 (2002)
3. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. *Math. Oper. Res.* **23**, 769 (1998)
4. Bertsimas, D., Sim, M.: Tractable approximations to Robust conic optimization. *Math. Progr.* **107**, 5 (2006)
5. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. *Math. Progr.* **98**, 49 (2003)
6. Bertsimas, D., Nohadani, O., Teo, K.: Robust optimization in electromagnetic scattering problems. *J. Appl. Phys.* **101**, 074507 (2007)
7. Levi, A.F.J., Haas, S.: Optimal Device Design, Chap. 6. Cambridge University Press, Cambridge (2010)
8. Bertsimas, D., Nohadani, O., Teo, K.: Robust optimization for unconstrained simulation-based problems. *Oper. Research* (to appear 2009). doi:[10.1287/opre.1090.0715](https://doi.org/10.1287/opre.1090.0715)
9. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific (1995)
10. Bertsimas, D., Tsitsiklis, J.N.: Introduction to Linear Optimization. Athena Scientific (1997)
11. Hartmann, A.K., Rieger, H.: New Optimization Algorithms in Physics. Wiley-VCH, New York (2004)
12. Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. *Science* **220**, 671 (1983)
13. Metropolis, N., Rosenbluth, A., Teller, M., Teller, E.: Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087 (1953)
14. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and bayesian resoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721 (1984)
15. Ingber, L., Rosen, B.: Genetic algorithms and very fast simulated reannealing: a comparison. *J. Math. Comput. Model.* **16**, 87 (1992)
16. Henrion, D., Lasserre, J.B.: GloptiPoly: global optimization over polynomials with matlab and SeDuMi. *ACM Trans. Math. Softw.* **29**, 165–194 (2003)
17. Lasserre, J.B.: A moment approach to analyze zeros of triangular polynomials sets. *Math. Program Ser. B* **107**, 275–293 (2006)
18. Kojima, M.: Sums of squares relaxations of polynomial semidefinite programs. Research report **B-397**, Tokyo Institute of Technology (2003)
19. Joannopoulos, J.D., Villeneuve, P.R., Fan, S.: Photonic crystals: putting a new twist on light. *Nature* **386**, 143 (1997)
20. Geremia, J.M., Williams, J., Mabuchi, H.: Inverse-problem approach to designing photonic crystals for cavity QED experiments. *Phys. Rev. E* **66**, 066606 (2002)
21. Seliger, P., Mahvash, M., Wang, C., Levi, A.: Optimization of aperiodic dielectric structures. *J. Appl. Phys.* **100**, 034310 (2006)
22. Ingber, L.: Very fast simulated re-annealing. *J. Math. Comput. Model.* **12**, 967 (1989)