Robust optimization in electromagnetic scattering problems

Dimitris Bertsimas, Omid Nohadani,[a] and Kwong Meng Teo
Operations Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 29 November 2006; accepted 26 January 2007; published online 13 April 2007)

In engineering design, the physical properties of a system can often only be described by numerical simulation. Optimization of such systems is usually accomplished heuristically without taking into account that there are implementation errors that lead to very suboptimal, and often, infeasible solutions. We present a robust optimization method for electromagnetic scattering problems with large degrees of freedom and report on results when this technique is applied to optimization of aperiodic dielectric structures. The spatial configuration of 50 dielectric scattering cylinders is optimized to match a desired target function such that the optimal arrangement is robust against placement and prototype errors. Our optimization method inherently improves the robustness of the optimized solution with respect to relevant errors and is suitable for real-world design of materials with unconventional electromagnetic functionalities, as relevant to nanophotonics. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2715540]

I. INTRODUCTION

The search for attractive and unconventional materials in controlling and manipulating electromagnetic field propagation has identified a plethora of unique characteristics in photonic crystals (PCs). Their peculiar functionalities are based on diffraction phenomena, which require periodic structures. While three-dimensional PC structures are still far from commercial manufacturing, two-dimensionally periodic PCs have already been introduced to integrated-device applications, e.g., through PC fibers.[1] However, technical difficulties such as ability to manufacture and disorder control pose restrictions on functionality and versatility. Upon breaking the spatial symmetry, additional degrees of freedom are revealed which allow for additional functionality and, possibly, for higher levels of control. Previous studies introduced the broken symmetry to PC structures of dielectric scatterers by diluting sites and optimizing the location of the missing scattering sites. Because of the underlying periodic structure, the additional degrees of freedom and, hence, their benefit have been very restricted.[2–4] More recently, unbiased optimization schemes were performed on the spatial distribution (aperiodic) of a large number of identical dielectric cylinders.[5,6] The resulting aperiodic structure, using an effective gradient-based optimization was reported to match a desired target function up to 95%.[6] While these works demonstrate the advantage of optimization, the robustness of the solutions still remains an open issue.

When implemented in the real world, however, the performance of many engineering designs often deviates from the predicted performance in the laboratory. A key source of this deviation lies in the presence of uncontrollable implementation errors. Traditionally, a sensitivity or postoptimality analysis was performed to study the impact of perturbations on specific designs. While such an approach can be used to compare designs, it does not intrinsically find one with lower sensitivities. Another class of robust design methods explored interactions between the uncertainties and the design variables by conducting a series of designed experiments.[7–9] This approach can fail for highly nonlinear systems with a large number of design variables.[10] Alternatively, the original objective function was replaced with a statistical measure consisting of expected values and standard deviations.[11] This method requires the knowledge of the probability distribution governing the errors, which usually cannot be easily obtained. Another approach suggested adding the first and, possibly, the second order approximation terms to the objective function.[12] Consequently, this is not suitable for highly nonlinear systems with sizeable perturbations. At the other end of the spectrum, the mathematics programming community has made much advances in the area of robust optimization over the past decade.[13,14] However, their results are confined to problems with more structures; for example, convex problems defined with linear, convex quadratic, conic-quadratic and semidefinite functions. Since our intention is not to review the rich literature in robust, convex optimization, we refer interested readers to Refs. 13 and 15.

In this article, we provide a robust optimization method for electromagnetic scattering problems with large degrees of freedom and report on results when this technique is applied to optimization of aperiodic dielectric structures. A key characteristic of our work is that it applies to nonconvex objective functions. Previous works did not take into account implementation errors that can lead to very suboptimal, and often, infeasible solutions. Our optimization method inherently improves the robustness of the optimized solution with respect to relevant errors and is suitable for real-world implementation. The objective is to mimic a desired power distribution along a target surface. The model is based on a two-dimensional Helmholtz equation for lossless dielectric scatterers. Therefore, this approach scales with frequency and allows to model nanophotonic design. Moreover, the ro-

[a]Electronic mail: nohadani@mit.edu

 © 2007 American Institute of Physics

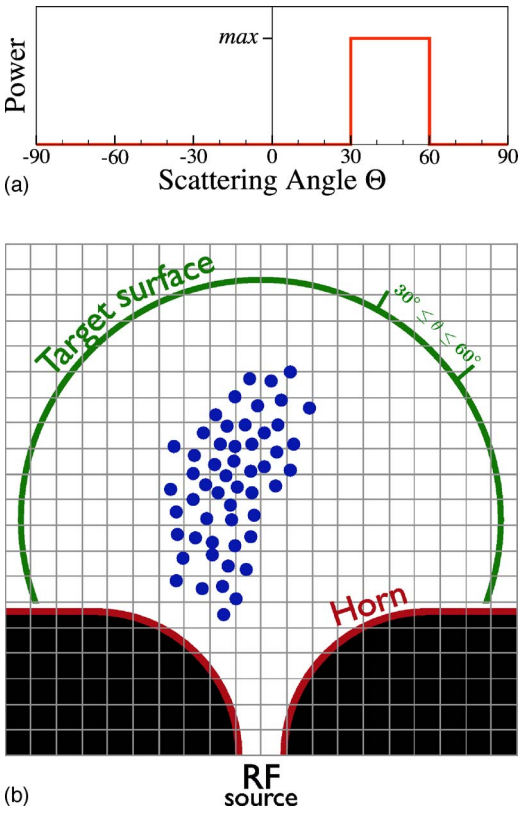

FIG. 1. (Color online) (a) The desired top-hat power distribution along the target surface. (b) Schematic setup: the radio frequency-source couples to the wave guide. Blue circles sketch the positions of scattering cylinders for a desired top-hat power profile.

bust optimization scheme requires only the function evaluation. Therefore, it is generic and can be applied to various problems arising in electromagnetics.

II. MODEL

To study the real-world aspect of robust optimization in design of dielectric structures, we adopted the model of Ref. 6 which was adapted to the laboratory experiment. In the following, we summarize the essentials of the physical model for the sake of completeness and refer for more details to Ref. 6. The incoming electromagnetic field couples in its lowest mode to the perfectly conducting metallic wave-guide (Dirichlet boundary conditions, therefore only the lowest transverse electric mode $TE_{1,0}$). Figure 1(b) sketches the horizontal setup. In the vertical direction, the domain is bound by two perfectly conducting plates, which are separated by less than 1/2 the wavelength, in order to warrant a two-dimensional wave propagation. Identical dielectric cylinders are placed in the domain between the plates. The sides of the domain are open in the forward direction. In order to account for a finite total energy and to warrant a realistic decay of the field at infinity, the open sides are modeled by perfectly matching layers.[16] The objective of the optimization is to determine the position of the cylinders such that the forward electromagnetic power matches the shape of a desired power distribution, as shown in Fig. 1(a).

For the power distribution, the electromagnetic field over the entire domain, including the scattering cylinders, is determined. As in the experimental measurements, the frequency is fixed to $f=37.5$ GHz.[6] Furthermore, the dielectric scatterers are nonmagnetic and lossless. Therefore, stationary solutions of the Maxwell equations are given through the two-dimensional Helmholtz equations, taking the boundary conditions into account. This means, that only the z component of the electric field E_z can propagate in the domain. The magnitude of E_z in the domain is given through the partial differential equation (PDE)

$$[\partial_x(\mu_{r_y}^{-1}\partial_x) + \partial_y(\mu_{r_x}^{-1}\partial_y)]E_z - \omega_0^2\mu_0\varepsilon_0\varepsilon_{r_z}E_z = 0, \tag{1}$$

with μ_r as the relative and μ_0 as the vacuum permeability. ε_r denotes the relative and ε_0 the vacuum permittivity. Equation (1) is numerically determined using an evenly meshed square-grid (x_i, y_i). The resulting finite-difference PDE approximates the field $E_{z,i,j}$ everywhere inside the domain including the dielectric scatterers. The imposed boundary conditions (Dirichlet condition for the metallic horn and perfectly matching layers) are satisfied. This linear equation system is solved by ordering the values of $E_{z,i,j}$ of the PDE into a column vector. Hence, the finite-difference PDE can be rewritten as

$$\mathbf{L}\cdot\mathbf{E}_z = \mathbf{b}, \tag{2}$$

where $\mathbf{L}$ denotes the finite-difference matrix, which is complex-valued and sparse. $\mathbf{E}_z$ describes the complex-valued electric field, that is to be computed, and $\mathbf{b}$ contains the boundary conditions. With this, the magnitude of the field at any point of the domain can be determined by solving the linear system of Eq. (2).

The power at any point on the target surface $(x(\theta), y(\theta))$ for an incident angle θ is computed through interpolation using the nearest four mesh points and their standard Gaussian weights $\mathbf{W}(\theta)$ with respect to $(x(\theta), y(\theta))$ as

$$s_{\mathrm{mod}}(\theta) = \frac{\mathbf{W}(\theta)}{2}\cdot\mathrm{diag}(\mathbf{E_z})\cdot\mathbf{E_z}. \tag{3}$$

In the numerical implementation, we utilized the UMFPACK-library to LU decompose $\mathbf{L}$ as well as to solve the linear system directly.[17] Furthermore, our implementation uses the Goto-BLAS library for basic vector and matrix operations.[18] By exploiting the sparsity of $\mathbf{L}$, we improved the efficiency of the algorithm significantly. In fact, the solution of a realistic forward problem ($\sim 70\,000\times 70\,000$ matrix), including 50 dielectric scatterers requires about 0.7 s on a commercially available Intel Xeon 3.4 GHz. Since the size of $\mathbf{L}$ determines the size of the problem, the computational efficiency of our implementation is independent of the number of scattering cylinders.

To verify this finite-difference technique for the power along the target surface (radius=60 mm from the domain center), we compared our simulations with experimental measurements from Ref. 6 for the same optimal arrangement of 50 dielectric scatterers ($\varepsilon_r=2.05$ and 3.175 ± 0.025 mm diameter). Figure 2(a) illustrates the good agreement between experimental and model data on a linear scale for an

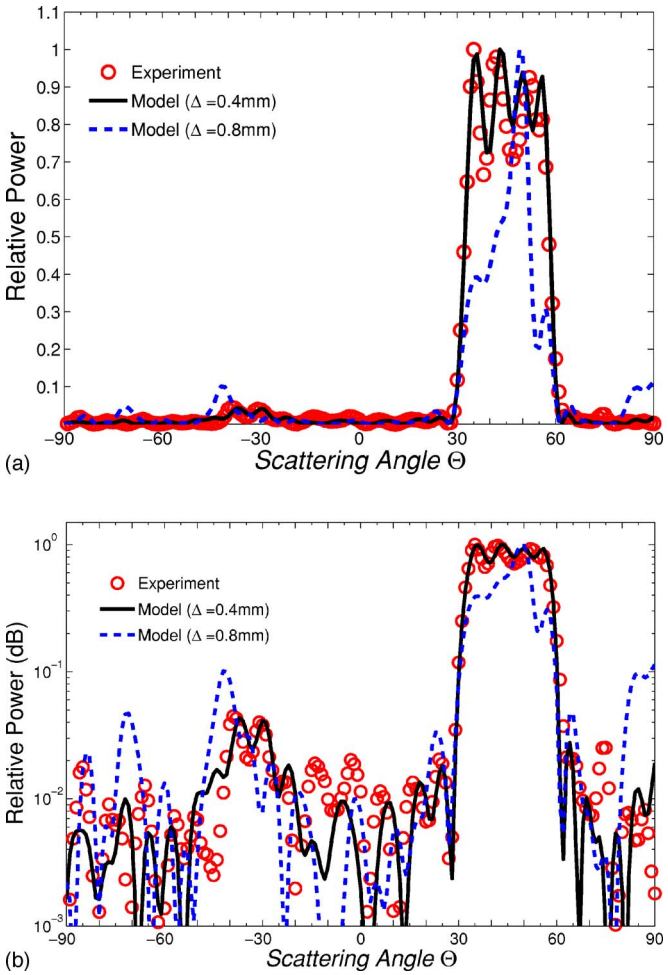

FIG. 2. (Color online) Comparison between experimental data (circles) (see Ref. 6) and simulations in (a) linear and (b) logarithmic scale. The solid lines are simulation results for smallest mesh size at $\Delta=0.4$ mm and the dashed lines for $\Delta=0.8$ mm.

objective top-hat function. The log scale in Fig. 2(b) emphasizes the relative intensities of the sidelobes to the peak. This comparison also shows that the simulation agrees well with experimental measurements only for a sufficiently small mesh-size of $\Delta=0.4\ \text{mm}\leq\lambda_0/20$. Other weak scattering models (e.g., Born approximation) fail to describe the true scattering behavior. We observed that higher excitation modes play a crucial role, which can be attributed to the small size of the scattering cylinders as well as the rapid jump in the dielectric constant which cannot be assumed continuos with respect to the wavelength.

III. ROBUST OPTIMIZATION PROBLEM

By varying the positions of 50 scattering cylinders a top-hat power profile over the target surface, as shown in Fig. 1(a), is sought. The desired objective function is denoted by s_{obj}. A cylinder configuration is given by a vector $\mathbf{p}\in\mathbb{R}^{100}$. The actual power profile along the target surface s_{mod} is computed using Eq. (3). 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

$$J(\mathbf{p})=\sum_{k=1}^{m}|s_{\text{mod}}(\theta_k)-s_{\text{obj}}(\theta_k)|^2. \tag{4}$$

Therefore, the optimization problem is to minimize the area between s_{obj} and s_{mod}. This *nominal optimization problem* is given through

$$\min_{\mathbf{p}\in\mathcal{P}} J(\mathbf{p}). \tag{5}$$

The minimization is with respect to the configuration vector $\mathbf{p}$ from a feasible set $\mathcal{P}$. Note that $J(\mathbf{p})$ is not convex in $\mathbf{p}$, and depends on $\mathbf{p}$ only over the linear system $\mathbf{L}(\mathbf{p})\cdot\mathbf{E}_z(\mathbf{p})=\mathbf{b}$.

It needs to be emphasized that $\mathcal{P}$ is not a convex set. Instead, $\mathcal{P}$ is a 100-dimensional hypercube containing a large number of nonempty infeasible subsets, which represent nonphysical configurations with overlapping cylinders. Defining these infeasible subsets explicitly through introducing constraints in the optimization problem (5) is not practical due to the large number of constraints required. We took the alternative approach of avoiding configurations with overlapping cylinders.

To consider possible implementation errors $\Delta\mathbf{p}$, the *robust optimization problem* is defined as

$$\min_{\mathbf{p}\in\mathcal{P}}\max_{\Delta\mathbf{p}\in\mathcal{U}} J(\mathbf{p}+\Delta\mathbf{p}). \tag{6}$$

The uncertainty set $\mathcal{U}$ contains all the implementation errors, against which we want to protect the design. Therefore, the robust optimization problem minimizes the worst case cost under implementation error. These errors can arise due to misplacement of the scattering cylinders in laboratory experiments or actual manufacturing of the design.

We adopted a two-step strategy. In the initial step, a good configuration to the nominal optimization problem in Eq. (5) is found. This configuration is used as an initial solution to the second step, since, with all factors being equal, a configuration with a low *nominal cost* $J(\mathbf{p})$ will have a low *worst case cost* $\max_{\Delta\mathbf{p}\in\mathcal{U}} J(\mathbf{p}+\Delta\mathbf{p})$. In the second step, we iteratively update the configuration under evaluation with a more robust configuration through a local shift until terminating conditions are satisfied. This robust optimization algorithm does not assume any problem intrinsic structures. We discuss the nominal problem before we continue with the robust optimization problem.

A. Nominal optimization problem

To solve the nominal problem (5), we conducted a large number of random searches. Because of the large dimensionality, a coarse-grained random search did not deliver a significant and sufficiently fast improvement in $\max_{\mathbf{p}\in\mathcal{P}} J(\mathbf{p})$. Therefore, we developed two alternative algorithms that efficiently returned a good solution to the nominal optimization problem, as required in step 1 of the robust optimization method.

1. Gradient free stochastic algorithm

This algorithm is adapted from the simultaneous perturbation stochastic approximation (SPSA) algorithm.[19] It relies

only on function evaluations $J(\mathbf{p})$. Under general conditions, SPSA converges to a local minimum more efficiently, in expectation, than a gradient descent approach using finite-difference gradient estimates.[19]

In iteration k, the algorithm seeks a better configuration along a direction $\mathbf{d}_k$ emanating from the configuration vector $\mathbf{p}_k$. The direction $\mathbf{d}_k$ is a random vector generated from a symmetric Bernoulli distribution where $P(d_k^i = \pm 1) = \frac{1}{2}$, $\forall i$, and is deemed acceptable only when $\mathbf{p}_k \pm c_k \mathbf{d}_k$ is feasible. Here, d_k^i is the ith coordinate of $\mathbf{d}_k$ and c_k is a small positive scalar decreasing with k. Next, $\delta_k = J(\mathbf{p}_k + c_k \mathbf{d}_k) - J(\mathbf{p}_k - c_k \mathbf{d}_k)$ is evaluated. Consequently, $\mathbf{p}_{k+1}$ is set to $\mathbf{p}_k - \alpha_k \delta_k \mathbf{d}_k$, where α_k is another small positive scalar decreasing with k. Note that δ_k approximates the directional gradient $\mathbf{d}_k \cdot \nabla_{\mathbf{p}} J(\mathbf{p} = \mathbf{p}_k)$, if c_k is small enough.

2. Modified gradient descent algorithm

We computed the gradient of the cost-functional $\nabla_{\mathbf{p}} J$ using the adjoint method. In general, since the linear operator $\mathbf{L}$ maps the entire vector space $\mathcal{C}^n \leftrightarrow \mathcal{C}^n$, the components of the cost-functional gradient can be determined from the Eqs. (3) and (4) through the adjoint equation as

$$\frac{\partial J}{\partial p_i} = \left\langle \mathbf{g} \middle| \frac{\partial E}{\partial p_i} \right\rangle \text{ with } \mathbf{g} = \frac{\partial J}{\partial s_{\text{mod}}} \frac{\partial s_{\text{mod}}}{\partial E}$$
$$= -\left\langle \mathbf{h} \middle| \frac{\partial \mathbf{L}}{\partial p_i} E \right\rangle \text{ with } \mathbf{L}^* \cdot \mathbf{h} = \mathbf{g}. \tag{7}$$

Note, that $\mathbf{L}^*$ is the adjoint operator to $\mathbf{L}$, which was regularized in the implementation to warrant that J is differentiable. Therefore, in order to compute the gradient, we have to solve the adjoint linear system $\mathbf{L}^* \cdot \mathbf{h} = \mathbf{g}$. This equation has the same structure and uses the same linear operator as the linear system for the function evaluation in Eq. (2). Consequently, we exploited the structure of the problem and utilize the LU decomposition of $\mathbf{L}$ for both the function and the gradient evaluation at practically no additional computational cost.

For this optimization problem, standard gradient descent steps quickly led to infeasible configurations and terminated at solutions with high cost. Nevertheless, gradient information is pertinent. To make use of it, we modified the standard algorithm to avoid configurations with overlapping cylinders. These modifications are: (1) if a gradient step leads to an infeasible configuration, the step size is repeatedly halved until a threshold, or (2) otherwise, apply the gradient step only to those cylinders that would not overlap. If the threshold is consistently breached, the algorithm approximates a coordinate descent algorithm which has similar convergence properties to standard gradient descent.[20]

B. Nominal optimization results

The starting configuration for the optimization is obviously significant for the performance. Due to the high-dimensional and nonconvex response surface, a global optimum can only be found through large-scale random searches, which is computationally exhaustive and, thus, beyond the scope of this work. Randomly generated initial configurations often lead to overlapping cylinders. These infeasible arrangements can only be overcome by human intervention, which we intended to omit. The performance of a large number of regular PC-like structures with and without random perturbation was simulated to obtain the best starting configuration. The inset of Fig. 3 illustrates this initial arrangement of the dielectric scattering cylinders, as it appears to be an intuitively good structure as well.

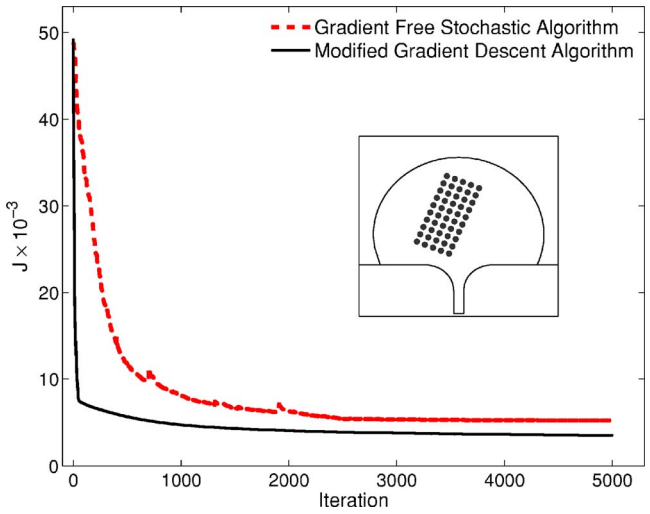

FIG. 3. (Color online) Performance comparison between the gradient free stochastic algorithm and the modified gradient descent algorithm on the nominal problem. Results show that modified gradient descent is more efficient and converges to a better solution.

We applied the gradient free stochastic algorithm and the modified gradient descent algorithm to this initial configuration. Figure 3 shows that the modified gradient descent algorithm reduces the objective function more efficiently. The gradient-free algorithm took ~2500 iterations to converge to an objective value of 0.0052. In contrast, the modified gradient algorithm required ~750 iterations to obtain configurations with a cost lower than 0.0052; it eventually converged to a configuration with a cost of 0.0032.

It is not surprising that the modified gradient descent algorithm outperforms the gradient free stochastic algorithm. Note, that at each iteration step, the gradient free algorithm uses two function evaluations and twice the time as compared to the modified gradient descent algorithm. The gradient free algorithm does not decrease the objective value monotonically, because, at any step, c^k and α^k may be too large. Adopting a strategy employing smaller scalars can alleviate the spikes but increase the overall time required to converge. Nevertheless, it is worthwhile to note the viability of using a gradient free optimization approach, since an efficient cost-functional gradient for such high-dimensional problems is not always available.

When the iteration count is high, both algorithms improve the objective value monotonically, albeit very slowly because infeasible configurations are encountered more often. Once the improvement rate went below a certain threshold, we terminated the search and used the final nominal configuration as the initial configuration for the robust optimization method.

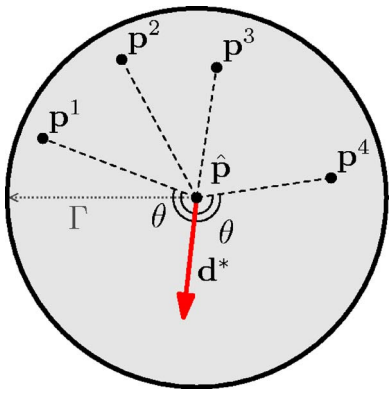

FIG. 4. (Color online) A two-dimensional illustration of the neighborhood $\{\mathbf{p}|\ \|\mathbf{p}-\hat{\mathbf{p}}\|_2 \leq \Gamma\}$. The solid arrow indicates an optimal direction $\mathbf{d}^*$ which makes the largest possible angle with the vectors $\mathbf{p}^i-\hat{\mathbf{p}}$ and points away from all bad neighbors $\mathbf{p}^i$.

C. Robust local search algorithm

In laboratory experiments, implementation errors $\Delta\mathbf{p}$ are encountered, when physically placing the cylinders. To include most of the errors, we define the uncertainty set $\mathcal{U}$ such that the probability $P(\Delta\mathbf{p}\in\mathcal{U})=99\%$. Consequently,

$$\mathcal{U}=\{\Delta\mathbf{p}|\ \|\Delta\mathbf{p}\|_2 \leq \Gamma\}, \tag{8}$$

where Δp_i is assumed to be independently and normally distributed with mean 0 and standard deviation 40 μm, as observed in experiments.[21] We chose Γ to be 550 μm.

Evaluating the worst cost under implementation errors involves solving an inner maximization problem

$$\max_{\Delta\mathbf{p}\in\mathcal{U}} J(\mathbf{p}+\Delta\mathbf{p}), \tag{9}$$

which does not have a closed form solution. Thus, we can only find an estimate of the worst case cost, $\tilde{J}_{\max}(\mathbf{p})$, through efficient local searches. These searches are conducted within the *neighborhood* $\mathcal{N}$ of a configuration $\hat{\mathbf{p}}$, defined as

$$\mathcal{N}=\{\mathbf{p}|\ \|\mathbf{p}-\hat{\mathbf{p}}\|_2 \leq \Gamma\}. \tag{10}$$

This set is illustrated in Fig. 4.

These searches form the initial part of the robust local search algorithm. The obtained worst case costs within $\mathcal{N}$ are used to find the next configuration with a local move, which aims to improve the worst case cost. The local move forms the second part of the robust local search algorithm. These two parts are repeated iteratively until the termination conditions are met. Next, we discuss these two parts in more detail.

1. Neighborhood exploration

The local search within a neighborhood $\mathcal{N}$ is conducted with several modified gradient ascents. To ensure that $\mathcal{N}$ is explored thoroughly, an additional boundary penalty is applied whenever an ascent step is near the boundary.

For this 100-dimensional problem, 101 gradient ascent sequences are carried out. The initial sequence starts from $\hat{\mathbf{p}}$ while the remaining 100 sequences start from $\hat{\mathbf{p}}+(\Gamma/3)\mathbf{e}_i$, if $[\partial J(\mathbf{p}=\hat{\mathbf{p}})]/\partial p_i \geq 0$ or from $\hat{\mathbf{p}}-(\Gamma/3)\mathbf{e}_i$, otherwise. p_i is a coordinate of $\mathbf{p}$ and $\mathbf{e}_i$ denotes the ith unit vector. A sequence is terminated when either a local maximum is obtained, a configuration outside the neighborhood is visited, or a time limit is exceeded.

Finally, the results of all function evaluations up to iteration k are stored in a set $\mathcal{H}_k$ and used to evaluate $\tilde{J}_{\max}(\mathbf{p}_k)$.

2. Robust local move

In the second part of the robust local search algorithm, we update the configuration $\mathbf{p}_k$ such, that the previously discovered "bad" neighbors are excluded from the updated neighborhood $\mathcal{N}_{k+1}$. We define the set of these bad neighbors as

$$\mathcal{M}_k=[\mathbf{p}|\mathbf{p}\in\mathcal{H}_k,\ \mathbf{p}\in\mathcal{N}_k,\ J(\mathbf{p})\geq\tilde{J}_{\max}(\mathbf{p}_k)-\sigma_k].$$

The cost factor σ_k governs the size of the set and may be changed within an iteration to ensure a feasible move.

The problem of determining a good direction $\mathbf{d}$, which points away from bad neighbors, can be formulated as

$$\begin{aligned} \min_{\mathbf{d},\,\varepsilon}\quad & \varepsilon \\ \text{s.t.}\quad & \|\mathbf{d}\|_2 \leq 1 \\ & \left(\frac{\mathbf{p}-\mathbf{p}_k}{\|\mathbf{p}-\mathbf{p}_k\|}\right)\cdot\mathbf{d} \leq \varepsilon\ \forall\ \mathbf{p}\in\mathcal{M}_k \\ & \varepsilon \leq 0. \end{aligned} \tag{11}$$

Because the first constraint is a conic quadratic constraint and all others are linear, this problem is a second order cone problem (SOCP), which can be solved efficiently using both commercial and noncommercial solvers. The optimal solution of this SOCP delivered a direction $\mathbf{d}^*$ forming the maximum possible angle with all the vectors $\mathbf{p}-\mathbf{p}_k$, $\mathbf{p}\in\mathcal{M}_k$, as shown in Fig. 4. This angle is at least 90° due to the constraint $\varepsilon\leq 0$. However, if a good direction is not found, we reduce σ_k, reassemble $\mathcal{M}_k$, and solve the updated SOCP. The terminating condition is attained, when σ_k decreases below a threshold.

D. Computation results

As the initial step of the robust optimization method, the nominal optimization also decreases the worst case cost significantly. For the PC-like initial configuration (see inset of Fig. 3), a worst case cost of $\tilde{J}_{\max}=0.05413$ was estimated, whereas the final nominal configuration delivered $\tilde{J}_{\max}(\mathbf{p}_1)=0.00646$, as shown in Fig. 5. While the nominal optimization primarily aims to reduce the nominal cost and increases the robustness indirectly, only the robust local search algorithm directly minimizes the worst case cost and, thus, improves the robustness further.

In the robust local search, the worst case cost at the terminating iteration step 65, $\tilde{J}_{\max}(\mathbf{p}_{65})$, was estimated with 110 000 configurations in the neighborhood of $\mathbf{p}_{65}$. As the iteration counts increase, the knowledge about the neighborhood grows and the more robust configurations are discovered. Figure 5 shows the improvement after 65 iterations of the robust local search algorithm. Here, the nominal cost of the design remains practically constant, while the estimated

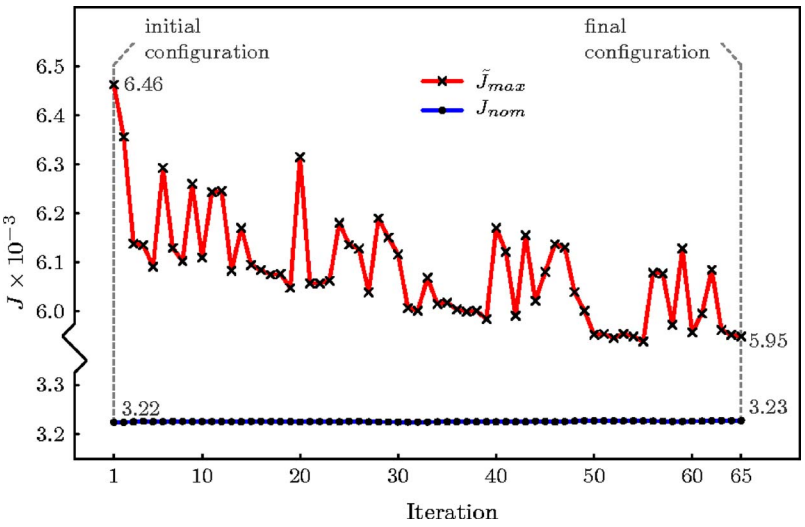

FIG. 5. (Color online) Performance of the robust local search algorithm. The worst case cost for the final configuration $\mathbf{p}_{65}$ is improved by 8% compared to $\mathbf{p}_1$, while the nominal cost remained constant. The overall improvement of the robustness is 90%.

worst case cost decreases significantly. Overall, we observe a 90% improvement in robustness of the final design, when compared to the initial PC-like structure.

Since we can only estimate the worst case cost by local searches, there is always a chance for late discoveries of worst implementation errors. Therefore, the decrease of the estimated worst case cost may not be monotonic.

IV. SENSITIVITY ANALYSIS

We have probed the neighborhood of $\mathbf{p}_1$ and $\mathbf{p}_{65}$ each with 10 000 normally distributed random perturbations. When the standard deviation of the perturbation is comparable to the assumed implementation errors, $\mathbf{p}_{65}$ is up to 2% less sensitive as $\mathbf{p}_1$.

It is evident, that a 100-dimensional random sampling is computationally challenging, e.g., when estimating $\tilde{J}_{\max}(\mathbf{p}_1)$, random sampling is far inferior to the multiple gradient ascent method: the best estimate attained by the former with 30 000 random samples is 96% of the estimate obtained with only 3000 multiple gradient ascent steps. Furthermore, a perturbative sensitivity analysis does not improve the worst case performance. There are no known practical approaches that improve sensitivities for a problem at such high dimensions. In contrast, our approach incorporates the widely used concept of probabilistic robustness through the variable size of the uncertainty set $\mathcal{U}$.

V. CONCLUSIONS

We have presented a robust optimization technique for electromagnetic scattering problems and applied it to the optimization of aperiodic dielectric structures. This generic method only assumes the capability of function evaluation. We have demonstrated that using a modified gradient descent will increase the efficiency of the robust algorithm significantly. However, if the function gradient is not accessible, a gradient-free stochastic algorithm can be utilized to obtain a robust solution. The application of our robust optimization method to improve the configuration of 50 dielectric cylinders showed that the final design configuration matches the shape of the desired function whose top-hat maximum is at $30° \leq \theta \leq 60°$. Since the problem is high-dimensional and highly non-convex, a global optimum can be estimated only through local searches. While the deviation from an optimal solution (perfect matching) is negligible, the robustness against implementation errors in laboratory experiments or manufacturing increased by 90%. Furthermore, laboratory measurements have verified our model.[6]

The generic aspect of the presented method allows it to be employed in various engineering problems in electromagnetics, in particular when function evaluation is provided. Moreover, the demonstrated approach for the dielectric scattering structure scales with frequency and can be applied to nanophotonic design to achieve unconventional and desired functionalities.

ACKNOWLEDGMENTS

The authors would like to thank D. Healy and A. Levi for encouragement and fruitful discussions. They also acknowledge C. Wang, R. Mancera, and P. Seliger for providing the initial structure of the numerical implementation. This work is supported by DARPA-N666001-05-1-6030.

[1] B. Temelkuran, S. D. Hart, G. Benoit, J. D. Joannopoulos, and Y. Fink, Nature (London) **420**, 650 (2002).
[2] J. M. Geremia, J. Williams, and H. Mabuchi, Phys. Rev. E **66**, 066606 (2002).
[3] L. Sanchis, A. Håkansson, D. López-Zanón, J. Bravo-Abad, and J. Sánchez-Dehesa, Appl. Phys. Lett. **84**, 4460 (2004).
[4] A. Håkansson, J. Sánchez-Dehesa, and L. Sanchis, IEEE J. Sel. Areas Commun. **23**, 1365 (2005).
[5] I. L. Gheorma, S. Haas, and A. F. J. Levi, J. Appl. Phys. **95**, 1420 (2004).
[6] P. Seliger, M. Mahvash, C. Wang, and A. F. J. Levi, J. Appl. Phys. **100**, 034310 (2006).
[7] G. Taguchi, Int. J. Prod. Res. **16**, 521 (1978).
[8] K. H. Lee, I. S. Eom, G. J. Park, and W. I. Lee, AIAA J. **34**, 1063 (1996).
[9] Y. Chen, R. Yu, W. Li, O. Nohadani, S. Haas, and A. F. J. Levi, J. Appl. Phys. **94**, 6065 (2003).
[10] W. Chen, J. Allen, K. L. Tsui, and F. Mistree, J. Mech. Des. **118**, 478 (1996).
[11] B. Ramakrishnan and S. S. Rao, Adv. Design Automat. **32**, 241 (1991).
[12] S. Sundaresan, K. Ishii, and D. R. Houser, Eng. Optimiz. **24**, 101 (1995).
[13] A. Ben-Tal and A. Nemirovski, Math. Program. **92**, 453 (2002).
[14] D. Bertsimas and M. Sim, Math. Program. **107**, 5 (2006).
[15] G. J. Park, T. H. Lee, K. H. Lee, and K. H. Hwang, AIAA J. **44**, 181 (2006).
[16] D. M. Kingsland, J. Gong, J. L. Volakis, and J. F. Lee, IEEE Trans. An-

tennas Propag. **44**, 975 (1996); J. P. Berenger, J. Comput. Phys. **127**, 363 (1996).
[17]T. A. Davis, ACM Trans. Math. Softw. **30**, 2 (2004).
[18]K. Goto, and R. A. van de Geijn, Univerity of Texas, Austin, Department of Computer Sciences, Tech. Report TR-2002–55, 2002.
[19]J. C. Spall, IEEE Trans. Autom. Control **37**, 332 (1992).
[20]Z. Q. Luo and P. Tseng, J. Optim. Theory Appl. **72**, 7 (1992).
[21]A. F. J. Levi (private communication).