Explicit Size-Reduction of Circularly Polarized Antennas Through Constrained Optimization With Penalty Factor Adjustment

MARZIEH MAHROKH[ID][1] **AND SLAWOMIR KOZIEL**[ID][1,2], **(Senior Member, IEEE)**
[1]Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland
[2]Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland

Corresponding author: Marzieh Mahrokh (marziehm@ru.is)

ABSTRACT Modern communication systems of high data capacity incorporate circular polarization (CP) as the preferred antenna radiation field configuration. In many applications, integration of the system circuitry with antennas imposes size limitations on CP radiators, which makes their development process a challenging endeavor. This can be mitigated by means of simulation-driven design, specifically, constrained numerical optimization. Majority of the performance-related constraints are expensive to evaluate, i.e. require full-wave electromagnetic (EM) analysis of the system. Their practical handling can be realized using a penalty function approach, where the primary objective (antenna size reduction) is complemented by contributions proportional to properly quantified constraint violations. The coefficients determining the contribution of the penalty terms are normally set up using designer's experience, which is unlikely to render their optimum values in terms of the achievable miniaturization rates as well as constraint satisfaction. This paper proposes a procedure for automated penalty factor adjustment in the course of the optimization process. Our methodology seeks for the most suitable coefficient levels based on the detected constraint violations and feasibility status of the design. It is validated using two CP antenna structures. The results demonstrate a possibility of a precise constraint control as well as superior miniaturization rates as compared to the manual penalty term setup.

INDEX TERMS Circular polarization antennas, compact antennas, constrained optimization, penalty functions, simulation-driven design.

I. INTRODUCTION

With the growing demands for reliable high-capacity data transfer, an increasing attention has been given to incorporation of CP antennas into modern communication systems. The orthogonal radiation field configuration can assure the reliability of these systems due to attractive features, including a reduction of polarization mismatch and multipath losses [1], as well as mitigation of the Faraday's effect [2]. The continuing trend towards miniaturization enforces CP antennas to be compatible with space constraints in applications such as Aerospace and Synthetic Aperture Radar (SAR) [3], Global Positioning System (GPS) [4],

picosatellites [5], 5G communication systems [6], or wearable and on-body devices.

While preservation of high CP purity along with satisfying other electrical and field performance requirements is already challenging, ensuring compact size is an additional contribution to the design complexity. Several miniaturization techniques based on topological modifications of the antenna structure have been proposed, including the use of slots and fractals [7], defected ground structure [8], fractal metasurfaces and fractal resonators [9], or mushrooms and reactive impedance surfaces (RIS) [10]. These techniques have been successful in working out a compromise between the compact size and performance figures of CP antenna. Notwithstanding, the evolution of antenna topology into complex multi-parameter geometries hinders the process of finding an optimum design, especially with conventional,

manual or trial-and-error efforts. A workaround is numerical optimization, which, depending on the nature of the design problem and available resources, may resort to either global [11]–[13], or local search [14], [15]. Full-wave EM analysis is most often used as the computational model, the accuracy of which determines the reliability of the optimization process. Yet, EM models tend to be expensive to evaluate. Cost-efficient optimization methods have been developed to mitigate this problem, including space mapping [16], incorporation of adjoint sensitivities [17]–[19], data-driven surrogate-based methods [15], [20]–[22], and machine learning approaches [23], [24].

EM-driven miniaturization is the most efficient when size reduction is explicitly handled as the primary objective. On the other hand, the need for ensuring the appropriate levels of electrical performance figures necessitates constrained optimization, with the constraints being expensive to evaluate. A convenient way of constraint control is the penalty function approach [25]. Therein, properly quantified constraint violations appear as additional terms complementing the main objective. The efficacy of this approach relies on the proper adjustment of the penalty factors. Setting these too large leads to an extreme steepness of the objective function in the vicinity of the feasible region boundary. Having them too small results in excessive constraint violations. A possible workaround is adaptive adjustment of the acceptance threshold for the maximum in-band reflection level [26]. Other approaches include feasible space boundary exploration procedure [25], or alternating the size-reduction- and constraint-improvement-oriented search steps [26]. However, in all these cases, the performance of the optimization process depends on a proper manual selection of the penalty factors.

In the context of constrained optimization using genetic algorithm, the aforementioned problem has been mitigated by incorporating an adaptive, tune-free penalty function [27], non-stationary penalty function [28], or a self-adaptive penalty function [29]. The abovementioned methods have been demonstrated successful in identifying feasible solutions without any manual tuning of the penalty function.

This paper proposes a novel procedure for explicit size-reduction of antenna structures featuring an automated penalty factor adjustment throughout the optimization process. The adjustment process employs monitoring of the feasibility status of the current design and the constraint violation levels. Our methodology is validated using two CP antenna structures miniaturized under reflection and axial ratio constraints. Extensive benchmarking demonstrates superior size reduction rates along with a possibility of precise control over the design constraints as compared to the manual penalty term setup.

II. EXPLICIT SIZE-REDUCTION THROUGH CONSTRAINED OPTIMIZATION

This section recalls a formulation of EM-driven antenna size reduction problem, as well as outlines the standard trust-region-based algorithm employed as the main optimization engine.

A. PROBLEM FORMULATION

We denote by $\boldsymbol{R}(\boldsymbol{x})$ the response of the EM simulation antenna model, where $\boldsymbol{x}$ is a vector of the geometry parameters. The task is to minimize the antenna size $A(\boldsymbol{x})$, subject to performance-related constraints of the form

$$s_j(\boldsymbol{x}) \leq S_j, \quad j = 1, \ldots, k \quad (1)$$

The constraints are handled implicitly, using the penalty function approach. The objective function takes the form of

$$U_A(\boldsymbol{R}(\boldsymbol{x})) = A(\boldsymbol{x}) + \sum_{j=1}^{k} \beta_j c_j(\boldsymbol{x})^2 \quad (2)$$

The penalty function c_j quantifies relative violation of the jth constraint as $c_j(\boldsymbol{x}) = \max\{\zeta_j/S_j, 0\}$, with the absolute violation defined as

$$\zeta_j = s_j(\boldsymbol{x}) - S_j \quad (3)$$

The penalty coefficients β_j determine the contribution of the c_j-measured violation to (2).

The design problem is formulated as

$$\boldsymbol{x}^* = \arg \min_{\boldsymbol{x} \in X} U_A(\boldsymbol{R}(\boldsymbol{x})) \quad (4)$$

B. TRUST-REGION GRADIENT BASED ALGORITHM

Our core optimization procedure utilizes the standard trust-region gradient-based algorithm [30], therein, a series of candidate solutions to (4) are obtained as

$$\boldsymbol{x}^{(i+1)} = \arg \min_{\boldsymbol{x};\; ||\boldsymbol{x}-\boldsymbol{x}^{(i)}|| \leq \delta} U_A(\boldsymbol{L}^{(i)}(\boldsymbol{x})), \quad i = 0, 1, \ldots \quad (5)$$

where $\boldsymbol{L}^{(i)}(\boldsymbol{x})$ is a first-order Taylor approximation of $\boldsymbol{R}(\boldsymbol{x})$ at $\boldsymbol{x}^{(i)}$, constructed using the antenna response sensitivities estimated using finite differentiation. The vector $\boldsymbol{x}^{(i+1)}$ is accepted if $U_A(\boldsymbol{R}(\boldsymbol{x}^{(i+1)})) < U_A(\boldsymbol{R}(\boldsymbol{x}^{(i)}))$. The standard TR-based rules [32] are employed to adjust the search radius δ upon each iteration.

III. AUTOMATED PENALTY FACTOR ADJUSTMENT

Formulation (2) facilitates handling of performance-related constraints yet the efficacy of this approach relies upon appropriate adjustment of the penalty coefficients (cf. Section I). This paper proposes a procedure for automated penalty factor adjustment, which eliminates the need for the trial-and-error, or experience-based penalty term setup, and leads to a more precise control of constraint violations as well as improved size reduction rates, as demonstrated in Section IV. This section discusses the underlying concept, the adjustment procedure, and the overall optimization algorithm.

A. AUTOMATED PENALTY FACTOR ADJUSTMENT

The setup of the penalty terms is instrumental in achieving top performance of the optimization process. Having the penalty coefficients too small results in excessive constraint

violations, whereas keeping them too large leads to numerical challenges in exploring the feasible region boundary.

Our technique aims at automating the penalty term setup using a set of rules, derived from current constraint violations along with a notion of sufficient improvement in successive iterations (i.e., between $x^{(i)}$, and $x^{(i+1)}$):

- If $x^{(i+1)}$ is feasible w.r.t. the jth constraint, reduce β_j;
- If $x^{(i+1)}$ is infeasible but there is sufficient improvement of the jth constraint violation w.r.t that of $x^{(i)}$, keep β_j intact;
- If $x^{(i+1)}$ is infeasible and there is either no improvement or insufficient improvement of the jth constraint, increase β_j;

Sufficient constraint violation improvement for the jth constraint, is defined as $\Delta_j = M\zeta_j$, where M is the improvement factor elaborated on below. In the following descriptions, ζ_j^{i+1} is the jth constraint violation at $x^{(i+1)}$ calculated as in (3), whereas Δ_j^{i+1} is defined for the last two consecutive vectors $x^{(i+1)}$ and $x^{(i)}$. Similarly, β_j^i stands for β_j at iteration i.

Rigorous formulation of the adjustment rules is provided in Step 4 of the pseudocode in Fig. 1. Therein, m_{incr} and m_{decr} are the modification factors for penalty coefficient modifications (here, set to $m_{incr} = 5$ and $m_{decr} = 1.25$). Keeping the penalty coefficient unchanged upon detecting sufficient constraint violation improvement allows for improving stability of the optimization process, i.e., avoiding over-multiplication of β_j, which would otherwise bounce back and forth throughout the optimization process.

1. Set $i = 0$; provide $x^{(0)}$;
2. Solve $x^{(i+1)} = \arg \min\limits_{x; \|x - x^{(i)}\| \leq \delta} U_A(L^{(i)}(x))$;
3. Compute gain ratio ρ_A and update TR size
$$\rho_A = \frac{U_A(R(x^{(i+1)})) - U_A(R(x^{(i)}))}{U_A(L^{(i)}(x^{(i+1)})) - U_A(L^{(i)}(x^{(i)}))} ;$$
   if $\rho_A < 0$,
      Go to 3;
   else
      $\Delta_j^{i+1} = M(s_j(x^{(i)}) - S_j)$;
   end
4. if $\zeta_j^i - \zeta_j^{i+1} > \Delta_j^{i+1}$ && $\zeta_j^{i+1} > 0$
      $\beta_j^{i+2} = \beta_j^{i+1} m_{incr}$; $i = i + 1$;
   elseif $\zeta_j^{i+1} \leq 0$
      $\beta_j^{i+1} = \beta_j^i / m_{decr}$; $i = i + 1$;
   else
      $\beta_j^{i+1} = \beta_j^i$; $i = i + 1$;
   end
5. if $\|x^{(i+1)} - x^{(i)}\| \leq \delta_x$ OR TR $\leq \delta_{TR}$
      END;
   else
      Go to 2;
   end
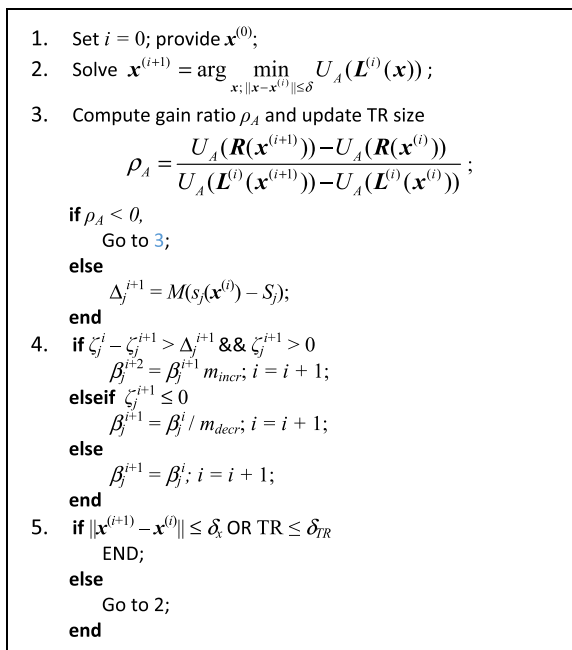
FIGURE 1. Operation of the proposed antenna size reduction algorithm with automated penalty factor adjustment.

The aforementioned improvement factor M is selected as follows. Let us assume that the vector $x^{(i)}$ is infeasible, and a sufficient constraint violation improvement is observed for n consecutive iterations, from i to $i+n$. As $M < 1$, this results in
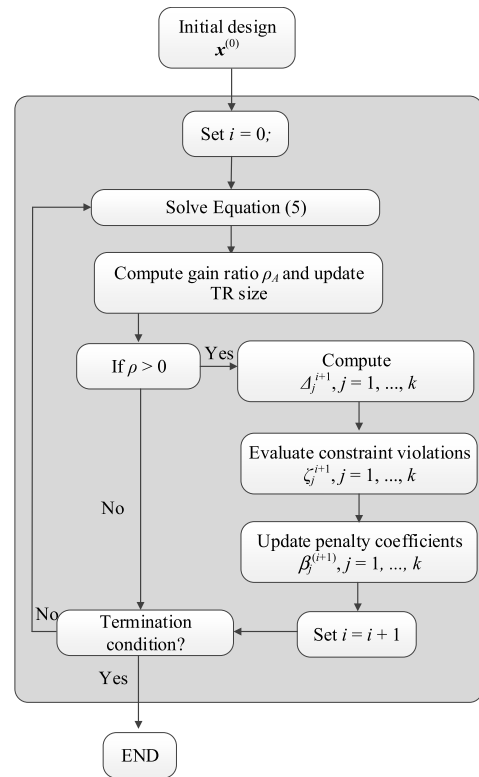
FIGURE 2. Operating flow of the proposed antenna size reduction algorithm with automated penalty factor adjustment.

a geometrical decrease of constraint violation, for which the upper bound at the iteration $i + n$ can be calculated as $\zeta_j^{i+n} \leq M\zeta_j^{i+n-1} \leq M\zeta_j^{i+n-2} \leq \cdots \leq M\zeta_j^i$. The improvement rate becomes faster as M gets smaller. On the other hand, the fulfillment of the sufficient improvement becomes more demanding. A value of $M = 0.5$ is chosen as a compromise.

B. OPTIMIZATION FRAMEWORK

The operation flow of the complete optimization algorithm has been shown in Fig. 1. The control parameters δ_x and δ_{TR} are the termination thresholds. Step 1 of the algorithm initializes the optimization procedure. Step 2 produces the candidate design by minimizing $U_A(L(x^{(i)}))$. Step 3 calculates the gain ratio, used to decide about the acceptance of rejection of $x^{(i+1)}$. Subsequently, constraint violation improvements are computed, which are used to update the penalty coefficients in Step 4.

IV. DEMONSTRATION CASE STUDIES

This section provides numerical validation of the automated adjustment procedure introduced in Section III. The verification case studies include two CP antennas optimized for minimum size with the constraints imposed on their axial ratio and reflection responses. The obtained results are compared to those produced with the fixed penalty coefficient setups ranging from the very relaxed to tight conditions regarding constraint satisfaction. The benchmark structures, experimental

setup, numerical results and their detailed discussion, are provided in Sections IV. *A* through IV. *C*, respectively.

A. VERIFICATION EXAMPLES AND EXPERIMENTAL SETUP

Figure 3 shows the geometries of the two benchmark structures (Antenna I [31], and Antenna II [32]) employed for verification purposes. Antenna I is a stacked microstrip patch structure supposed to be optimized within the frequency band 5.36 GHz to 5.9 GHz, whereas, Antenna II is a circular patch structure with annular and rectangular slots, to be optimized within a frequency band from 8.1 GHz to 8.3 GHz. Table 1 provides the details of the substrate materials, designable variables, as well as the initial design vectors of both antennas. The computational models are simulated using the time domain solver of CST Microwave Studio. The initial designs have been obtained by an auxiliary optimization process so as to provide a reasonable margin for both the axial ratio and reflection coefficient constraints, thereby creating room for size reduction.
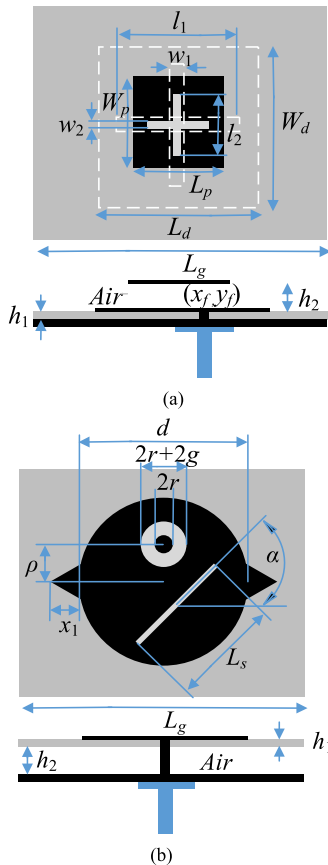
FIGURE 3. Geometries of the benchmark CP antennas: (a) Antenna I, (b) Antenna II.

The goal is to optimize the considered CP antennas for minimum size, defined as the substrate area $A(x)$. The constraints are imposed on the axial ratio $AR(x)$, and the reflection coefficient $|S_{11}(x)|$ of the antennas. In particular, we have $s_{AR}(x) \le 3$ and $s_{S11}(x) \le -10$, where $s_{AR}(x)$ and $s_{S11}(x)$ stand for the maximum value of $AR(x)$ and $|S_{11}(x)|$ respectively.

TABLE 1. Benchmark antenna structures.

	Antenna I [31]	Antenna II [32]
Substrate I	Arlon ($\varepsilon_r = 2.2$, $h = 1.575$ mm)	Arlon ($\varepsilon_r = 2.5$, $h = 3.8$ mm)
Substrate II	Air ($\varepsilon_r = 1$, $h = 3.8$ mm)	Air ($\varepsilon_r = 1.08$, $h = 2$ mm)
Design variables [mm]	$x = [x_f \, y_f \, l_1 \, l_2 \, W_p \, W_d \, L_p$ $L_d \, w_2 \, w_1 \, L_g]$	$x = [r \, g \, L_g \, d \, \rho \, L_s \, \alpha \, x_1]$
Initial design [mm]	$x = [4.16 \; 3.09 \; 8.26$ $12.08 \; 17.23 \; 12.93 \; 17.7$ $15.96 \; 1.15 \; 0.89 \; 26.04]$	$x = [1.58 \; 0.48 \; 21.7$ $12.46 \; 3.4 \; 9.4 \; 52.4$ $1.52]$

Correspondingly, the two penalty coefficients are defined as β_{AR} and β_{S11}. Note that descriptive subscripts AR and $S11$ are used here rather than numerical ones (cf. Section III) to allow for a better clarity.

The proposed procedure is compared with the standard trust-region-based algorithm executed with fixed values of both of penalty coefficients i.e. $\beta_{AR} = 10^y$, $y = 1, 2, 3, 4$, and $\beta_{S11} = 10^z$, $z = 2, 3, 4, 5$. The values set for the termination thresholds are $\delta_x = \delta_{TR} = 10^{-3}$.

B. RESULTS

Figures 4 and 5 show the axial ratio and reflection coefficient responses along with the evolution of their corresponding penalty factors throughout the optimization process for Antennas I and II, respectively. Table 2 provides the optimization results of both the fixed and the automated adjustment setups for the two antennas. The data includes antenna size along with constraint violations of the axial ratio and the reflection coefficient, denoted by ζ_{AR} and ζ_{S11}, respectively. The final optimized design vectors are $x = [2.96 \; 3.16 \; 8.74 \; 14.10 \; 16.34 \; 13.32 \; 16.15 \; 15.80 \; 1.02 \; 1.00 \; 24.28]$ (mm) and $x = [1.77 \; 0.66 \; 19.32 \; 12.20 \; 2.97 \; 9.28 \; 52.56 \; 1.41]$ (mm) for Antenna I and Antenna II, respectively. Further discussion of the results can be found Section IV. *C*.

C. DISCUSSION

The analysis of the results reported in Table 2 allows for drawing several conclusions regarding the importance of the automated penalty factor adjustment, as well as the performance superiority of the proposed automated procedure over the conventional approach.

The major observations are as follows:

- The optimum values of the penalty coefficients are problem dependent, therefore, finding the appropriate setup beforehand is a matter of a guess work. Clearly, this affects the performance of the optimization process and increases its computational cost (e.g., if the initially adopted setup turns out to be sub-optimal).
- Using a penalty factor higher than the optimum value results in degradation of the performance in terms of the
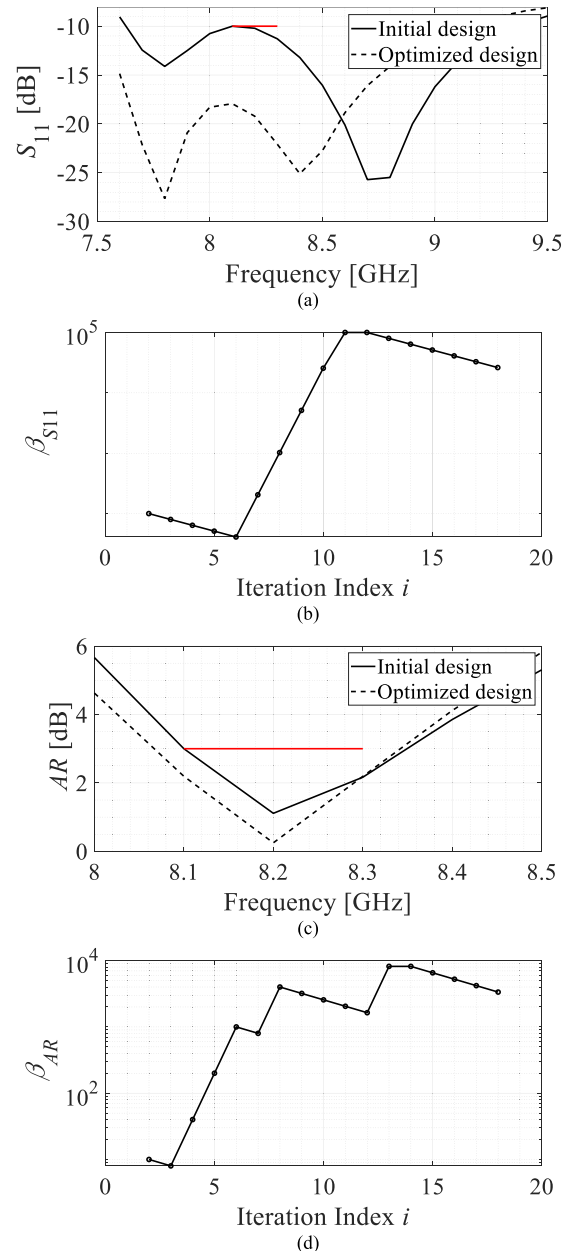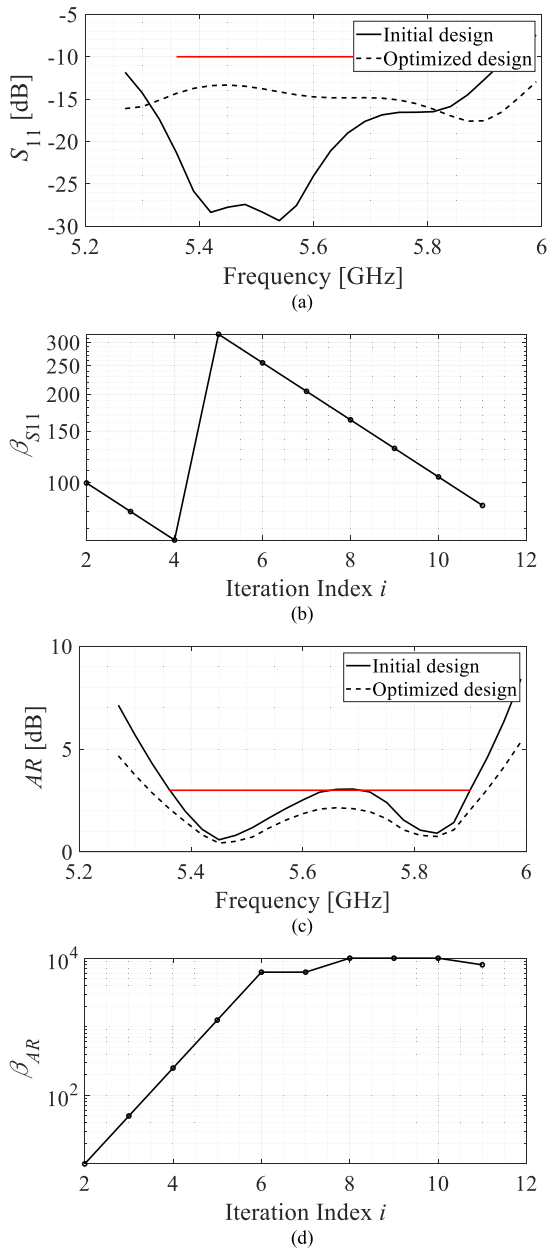
FIGURE 4. Antenna I: optimization results using the proposed automated penalty factor adjustment procedure: (a) reflection coefficient at the initial (- - -) and the optimized design (—); (b) axial ratio at the initial (- - -) and the optimized design (—); (c) evolution of the reflection coefficient penalty factor throughout the iterations of the optimization process; (d) evolution of the axial ratio penalty factor throughout the iterations of the optimization process.

FIGURE 5. Antenna II: optimization results using the proposed automated penalty factor adjustment procedure: (a) reflection coefficient at the initial (- - -) and the optimized design (—); (b) axial ratio at the initial (- -) and the optimized design (—); (c) evolution of the reflection coefficient penalty factor throughout the iterations of the optimization process; (d) evolution of the axial ratio penalty factor throughout the iterations of the optimization process.

achievable size reduction rates, whereas too low values, leads to significant constraint violation.

- The (fixed) penalty coefficient setup, which is optimum from the point of view of constraint violation, is still inferior in terms of achievable antenna size as compared to the proposed adaptive approach.
- Automated penalty factor adjustment allows to improve the quality of the final design by identifying the optimum values of the penalty coefficients for every iteration

throughout the optimization process. The history of the iteration-wise penalty factor adjustment for the reflection constraint, ζ_{S11}, of Antenna I is illustrated in Fig. 4(b). It starts with the minimum value of β_{S11} and continues with a decreasing trend for two subsequent iterations, i.e., as long as there is no constraint violation. The sudden increase of β_{S11} that can be observed between the fourth and the fifth iteration, is representative of constraint violation of the design obtained at the

TABLE 2. Optimization results for antennas I and II.

Penalty Factor Setup $[\beta_{AR}, \beta_{S11}]$	Antenna I			Antenna II		
	Area A [mm^2]	Constraint violation ζ_{AR} [dB]	Constraint violation ζ_{S11} [dB]	Area A [mm^2]	Constraint violation ζ_{AR} [dB]	Constraint violation ζ_{S11} [dB]
$[10^1, 10^2]$	313.26	5.13	0.91	374.38	2.77	0
$[10^1, 10^3]$	313.26	5.13	0.91	323.37	2.24	0.2
$[10^1, 10^4]$	313.26	5.13	0.91	334.37	2.69	0.01
$[10^1, 10^5]$	313.26	5.13	0.91	340.24	2.15	0.01
$[10^2, 10^2]$	587.9	0.36	0	309.99	0.05	3.8
$[10^2, 10^3]$	587.9	0.36	0	361.73	0.33	0
$[10^2, 10^4]$	587.9	0.36	0	359.71	0.27	0
$[10^2, 10^5]$	587.9	0.36	0	356.18	0.20	0
$[10^3, 10^2]$	607.39	0.02	0	404.58	0.07	0
$[10^3, 10^3]$	607.39	0.02	0	421.75	0.05	0
$[10^3, 10^4]$	607.39	0.02	0	421.75	0.05	0
$[10^3, 10^5]$	607.39	0.02	0	421.75	0.05	0
$[10^4, 10^2]$	663.73	0.008	0	455.41	0.03	0
$[10^4, 10^3]$	663.73	0.008	0	455.41	0.03	0
$[10^4, 10^4]$	663.73	0.008	0	455.41	0.03	0
$[10^4, 10^5]$	663.73	0.008	0	455.41	0.03	0
Adaptive β (This work)	589.78	0.07	0	373.36	0	0

(Leftmost vertical label: Reference algorithms)

fourth iteration. The following decreasing trend up to the last iteration indicates a lack of constraint violation for the corresponding iterations. Similar trends can be observed for axial-ratio-related penalty factor, as well as for Antenna II. The current values of both β_{S11} and β_{AR} are set to either reduce constraint violation or to maintain the solution in the vicinity of the feasible region boundary.

In general, the described automated adjustment procedure, in turn, enables improved size reduction rates along with a better control over constraint violations.

V. CONCLUSION
This paper proposed a novel methodology for optimization-based antenna size reduction using local trust-region-based search routines. Our procedure can be incorporated into frameworks involving a penalty function approach for implicit handling of design constraints. Therein, the proper adjustment of the penalty factors strongly correlates with the efficacy as well as the reliability of the optimization process, both in terms of constraint satisfaction and the achievable size reduction rates, yet it is difficult to be identified beforehand. The proposed procedure virtually eliminates the need for

manual, or guess-work efforts by an automated penalty factor adjustment. The latter is conducted based sufficient constraint violation in successive iterations and consequently allowing for better size reduction rates while leading to a precise control over the design constraints as compared to the fixed penalty coefficient setup.

The proposed methodology has been validated using two CP antenna structures optimized for minimum size, with the constraints imposed on their axial ratio and reflection responses. Benchmarking against fixed penalty factor setups indicates superior performance of the automated adjustment in terms of the achievable size reduction rates and a precise control of the design constraints.

VI. ACKNOWLEDGMENT

REFERENCES
[1] Y.-H. Yang, B.-H. Sun, and J.-L. Guo, "A low-cost, single-layer, dual circularly polarized antenna for millimeter-wave applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 4, pp. 651–655, Apr. 2019.
[2] W. Wang, C. Chen, S. Wang, and W. Wu, "Circularly polarized patch antenna with filtering performance using polarization isolation and dispersive delay line," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, no. 8,