Feasible Adjoint Sensitivity Technique for EM Design Exploiting Broyden's Update

Reza Safian, Natalia K. Nikolova, *Member, IEEE*, Mohamed H. Bakr, *Member, IEEE*, and John W. Bandler, *Fellow, IEEE*

McMaster University, Hamilton, Ontario, Canada L8S 4K1

Abstract — **The EM-FAST feasible adjoint sensitivity technique has been proposed for use with frequency domain electromagnetic solvers. It employs finite differences to approximate the derivatives of the system matrix with respect to the design parameters. Here, we propose to estimate and update these derivatives by the classical Broyden technique. This significantly accelerates the response sensitivity analysis when EM-FAST is used for gradient-based optimization. Regardless of the number of design parameters, the response sensitivity is obtained with computational overhead negligible in comparison with the system analysis. Our EM-Broyden technique is illustrated through the optimization of a Yagi-Uda antenna.**

I. Introduction

Design sensitivity information is crucial in gradient-based optimization. Its purpose is to estimate the gradient of the system's response with respect to the design parameters.

A feasible adjoint sensitivity technique (FAST) based on finite-difference approximation of the Jacobian of the system was first applied to nonlinear circuits [1]. Recently, a FAST has been proposed for applications with full-wave EM solvers (EM-FAST) [2],[3]. It applies finite differences to approximate the derivatives of the system matrix with respect to the design parameters and subsequently uses them to estimate the response sensitivities.

Here, we propose the use of the Broyden update [4], [5] for the estimation of the system matrix derivatives. Thus, the computational overhead due to the sensitivity calculations is reduced approximately by a factor of n, the number of design parameters, in comparison with the EM-FAST during a design optimization process.

We start with a brief overview of the EM-FAST and its computational requirements. Then, we describe Broyden's update of the gradients of the elements of the system matrix to accelerate the performance of the sensitivity analysis algorithm. We illustrate the implementation of the algorithm with the method of moments (MoM) in the optimization of the input impedance of a Yagi-Uda antenna.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants OGP0007239, OGP0227660-00 and STR234854-00, through the Micronet Network of Centres of Excellence and Bandler Corporation.

II. Adjoint Sensitivity Analysis

We focus on frequency-domain EM computational approaches. In the MoM, the system of equations arising from the discretization of a linear EM problem is

$$\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{I} = \boldsymbol{V} \tag{1}$$

where $\boldsymbol{x} = [x_1 \cdots x_n]^T$ is the vector of design parameters, $\boldsymbol{I} = [I_1 \cdots I_m]^T$ is the state variable vector (complex-valued current densities), and $\boldsymbol{V}$ is the corresponding global excitation vector. The design parameters are related to the geometry and/or the materials of the structure. $\boldsymbol{Z}$ is the system matrix whose complex coefficients depend on the design parameters. We now represent (1) as an equivalent real-valued system:

$$\begin{bmatrix} \Re \boldsymbol{Z} & -\Im \boldsymbol{Z} \\ \Im \boldsymbol{Z} & \Re \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \Re \boldsymbol{I} \\ \Im \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \Re \boldsymbol{V} \\ \Im \boldsymbol{V} \end{bmatrix}. \tag{2}$$

In (2), $\Re$ and $\Im$ denote the real and the imaginary parts, respectively, of a complex-valued quantity. Thus we can arrive at a sensitivity expression for complex linear systems using an algebraic approach for real-valued systems [6]. For brevity, we introduce the notation

$$\boldsymbol{Z}_r = \begin{bmatrix} \Re \boldsymbol{Z} & -\Im \boldsymbol{Z} \\ \Im \boldsymbol{Z} & \Re \boldsymbol{Z} \end{bmatrix}, \ \boldsymbol{I}_r = \begin{bmatrix} \Re \boldsymbol{I} \\ \Im \boldsymbol{I} \end{bmatrix}, \ \boldsymbol{V}_r = \begin{bmatrix} \Re \boldsymbol{V} \\ \Im \boldsymbol{V} \end{bmatrix}. \tag{3}$$

The size of $\boldsymbol{Z}_r$, $\boldsymbol{I}_r$ and $\boldsymbol{V}_r$ in the real system is twice the size of $\boldsymbol{Z}$, $\boldsymbol{I}$ and $\boldsymbol{V}$ in the complex system.

The predefined scalar function $f(\boldsymbol{x}, \bar{\boldsymbol{I}}(\boldsymbol{x}))$ can represent a *response function* of the linear system or an *objective function* in an optimization problem. We assume that this function is differentiable. The objective is to determine the gradient of $f(\boldsymbol{x}, \bar{\boldsymbol{I}}(\boldsymbol{x}))$ at the current solution $\bar{\boldsymbol{I}}$ of (1) with respect to the design parameters $\boldsymbol{x}$:

0-7803-7695-1/03/$17.00 © 2003 IEEE

$$\nabla_x f, \text{ subject to } \boldsymbol{Z}(\boldsymbol{x})\boldsymbol{I} = \boldsymbol{V} \quad (4)$$

We define gradient operators as row operators, e.g.,

$$\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}. \quad (5)$$

When the gradient operator acts on a vector, e.g. $\boldsymbol{V}$, the result appears in the form

$$\nabla_x \boldsymbol{V} = \begin{bmatrix} \frac{\partial V_1}{\partial x_1} & \cdots & \frac{\partial V_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial V_m}{\partial x_1} & \cdots & \frac{\partial V_m}{\partial x_n} \end{bmatrix}. \quad (6)$$

Following straightforward matrix manipulations, the response sensitivity $\nabla_x f$ is expressed as [2], [3], [6]

$$\nabla_x f = \nabla_x^e f + \hat{\boldsymbol{I}}_r^T \left[\nabla_x \boldsymbol{V}_r - \nabla_x (\boldsymbol{Z}_r \overline{\boldsymbol{I}}_r) \right]. \quad (7)$$

Here, $\nabla_x^e f$ reflects only the explicit dependence of f on $\boldsymbol{x}$. The matrix $\nabla_x \boldsymbol{V}_r$ would typically be analytically available. In fact, the excitation is often insensitive to changes in the design parameters, i.e., $\nabla_x \boldsymbol{V}_r = \boldsymbol{0}$. In $\nabla_x (\boldsymbol{Z}_r \overline{\boldsymbol{I}}_r)$, $\overline{\boldsymbol{I}}_r$ is a constant vector representing the solution of (1) at the current design in the form given by (3). The vector

$$\hat{\boldsymbol{I}}_r^T = \nabla_{I_r} f \cdot \boldsymbol{Z}_r^{-1} \quad (8)$$

is the real-valued *adjoint variable vector*. Clearly, it is a solution to the real-valued system of equations

$$\boldsymbol{Z}_r^T \hat{\boldsymbol{I}}_r = [\nabla_{I_r} f]^T. \quad (9)$$

We can represent $\hat{\boldsymbol{I}}_r$ as

$$\hat{\boldsymbol{I}}_r = \begin{bmatrix} \Re \hat{\boldsymbol{I}} \\ \Im \hat{\boldsymbol{I}} \end{bmatrix}. \quad (10)$$

This defines the complex-valued adjoint variable vector $\hat{\boldsymbol{I}}$. The real-valued system (9) can now be expanded as

$$\begin{bmatrix} \Re \boldsymbol{Z}^T & \Im \boldsymbol{Z}^T \\ -\Im \boldsymbol{Z}^T & \Re \boldsymbol{Z}^T \end{bmatrix} \begin{bmatrix} \Re \hat{\boldsymbol{I}} \\ \Im \hat{\boldsymbol{I}} \end{bmatrix} = [\nabla_{\Re I} f \quad \nabla_{\Im I} f]^T \quad (11)$$

which is equivalent to the complex-valued system

$$\boldsymbol{Z}^H \hat{\boldsymbol{I}} = [\nabla_I f]^T. \quad (12)$$

Here, $\boldsymbol{Z}^H$ is the Hermitian transpose of $\boldsymbol{Z}$, and $\nabla_I f = \nabla_{\Re I} f + j \nabla_{\Im I} f$, $j = \sqrt{-1}$. Equation (12) defines the complex-valued *adjoint problem*. It is characterized by: (i) a matrix which is the Hermitian transpose of that of the original problem; and (ii) an excitation which is the gradient of the response function with respect to the state variables. Equations (7) and (12) represent the essence of the adjoint variable approach in the case of complex-valued linear system sensitivity analysis.

Notice that the solution to the adjoint problem (12) adds very little overhead to the analysis at the current design. The *LU* factored $\boldsymbol{Z}^H$ matrix is easily obtained from the *LU* factored $\boldsymbol{Z}$ matrix, thus avoiding a second system analysis. The overhead is due only to the forward-backward substitution in (12).

III. Accelerated Optimization Using Broyden's Update

We write (7) explicitly as

$$\frac{\partial f}{\partial x_i} \doteq \frac{\partial^e f}{\partial x_i} + \hat{\boldsymbol{I}}_r^T \left[\frac{\partial \boldsymbol{V}_r}{\partial x_i} - \frac{\partial \boldsymbol{Z}_r}{\partial x_i} \overline{\boldsymbol{I}}_r \right], \; i = 1, 2, \ldots, n. \quad (13)$$

Here, the partial derivative $\partial / \partial x_i$ is applied to every element of the respective vector or matrix. The matrices $\partial \boldsymbol{Z} / \partial x_i$ $(i = 1, \ldots, n)$ in $\nabla_x (\boldsymbol{Z} \overline{\boldsymbol{I}})$ may be analytically available. If this is not the case, one can resort to the finite-difference approximation $\Delta \boldsymbol{Z} / \Delta x_i$ $(i = 1, \ldots, n)$. In both cases, the computation of the derivatives of $\boldsymbol{Z}$ is computationally expensive because it requires the equivalent of n additional $\boldsymbol{Z}$-matrix fills. This overhead becomes significant when $\boldsymbol{Z}$ is large.

The Broyden update refers to a rank-one formula proposed by Broyden [4],[5]

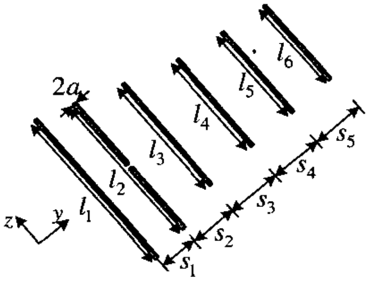
$$\boldsymbol{G}_{k+1} = \boldsymbol{G}_k + \frac{\boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{h}_k) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{G}_k \boldsymbol{h}_k}{\boldsymbol{h}_k^T \boldsymbol{h}_k} \boldsymbol{h}_k^T \quad (14)$$

where $\boldsymbol{G}_k$ is an approximation of the Jacobian $\nabla_x \boldsymbol{F}$ at $\boldsymbol{x}_k$, and $\boldsymbol{G}_{k+1}$ provides an updated Jacobian. $\boldsymbol{F}$ is the vector of functions under consideration, and $\boldsymbol{h}_k$ is an increment vector. The values of $\boldsymbol{F}$ at $\boldsymbol{x}_k$ and $(\boldsymbol{x}_k + \boldsymbol{h}_k)$ are assumed available. The updated approximation $\boldsymbol{G}_{k+1}$ satisfies the equation

$$\boldsymbol{F}(\boldsymbol{x}_k + \boldsymbol{h}_k) - \boldsymbol{F}(\boldsymbol{x}_k) = \boldsymbol{G}_{k+1} \boldsymbol{h}_k \quad (15)$$

In other words, $\boldsymbol{G}_{k+1}$ provides a perfect linear interpolation between the two points $\boldsymbol{x}_k$ and $(\boldsymbol{x}_k + \boldsymbol{h}_k)$.

In our problem, $\boldsymbol{F}$ is a vector that consists of all the elements of the $\boldsymbol{Z}_r$ matrix, and $\boldsymbol{G}_k$ is a matrix that consists of the derivatives of the elements of the $\boldsymbol{Z}_r$ matrix. To construct the vector $\boldsymbol{F}$, we stack all the columns of $\boldsymbol{Z}_r$ in a vector, therefore, when $\boldsymbol{Z}$ is an $(m \times m)$ matrix, $\boldsymbol{Z}_r$ is a $(2m \times 2m)$ matrix, and $\boldsymbol{F}$ is a vector with $4m^2$ elements. A row of the matrix $\boldsymbol{G}_k$ contains the derivatives of the respective element of the vector $\boldsymbol{F}(\boldsymbol{x}_k)$ with respect to all design parameters. Therefore, $\boldsymbol{G}_k$ is a $(4m^2 \times n)$



l_1/λ	l_2/λ	l_d/λ	s_1/λ	s_d/λ	a/λ
0.45	0.45	0.40	0.2	0.34	0.003
$l_3 = l_4 = l_5 = l_6 = l_d$; $s_2 = s_3 = s_4 = s_5 = s_d$					

Fig. 1. The geometry of the Yagi-Uda array.

matrix. The vector $\boldsymbol{h}_k$ has n elements corresponding to the increment in the design parameter space.

Using the Broyden update leads to faster estimation of the consecutive $\partial Z/\partial x_i$ matrices in an iterative manner. In comparison with the EM-FAST, which uses the finite difference approximation for $\Delta Z/\Delta x_i$, this approach reduces the overhead significantly. Its computational requirements are much smaller compared to a matrix fill. The response and its gradient are obtained by a single system analysis regardless of the number of design parameters n. The Broyden-FAST approach does not require modifications of the EM analysis algorithms. During optimization, the derivatives of the **Z** matrix with respect to all design parameters are approximated using the **Z** matrix of the previous step, the **Z** matrix of the current step and the **Z** matrix gradient from the previous step. Thus, the **Z** matrix is filled only once per design iteration.

The approximate Jacobians generated by Broyden's update may be less accurate than those obtained by finite differences. Hence, the optimization may require more steps to reach a solution.

Broyden [4] has shown that for quadratic functions the update converges and reduces the overall computational effort. Although such properties cannot be proven for a general nonlinear problem, Generally, the Broyden update gives sufficient accuracy for mildly nonlinear functions. This property can be exploited when the behavior of the **Z** matrix coefficients with respect to geometrical perturbations is predictable in order to improve the accuracy of the matrix derivative estimation. Such an approach, however, is solver-dependent.

Broyden's update has been used in a number of applications such as gradient-based optimization where analytical sensitivities are not available [5], the aggressive space mapping technique [7], etc.

TABLE I
OPTIMIZATION OF THE INPUT IMPEDANCE OF THE YAGI-UDA ARRAY

K	l_{2n}	s_{1n}	s_{2n}	R_{in}	X_{in}	f
1	0.450	0.200	0.340	28.576	-9.977	0.6237
2	0.300	0.400	0.400	67.312	-20.849	0.2960
3	0.320	0.400	0.400	67.308	-20.072	0.2858
4	0.390	0.400	0.400	69.081	-14.573	0.2067
5	0.600	0.212	0.342	67.008	-2.373	0.0883
6	0.495	0.286	0.371	75.457	5.742	0.0856
7	0.455	0.285	0.367	78.479	2.108	0.0804
8	0.547	0.220	0.369	70.149	2.485	0.0518
9	0.544	0.241	0.340	70.879	-0.798	0.0310
10	0.521	0.261	0.353	75.229	0.116	0.0306
11	0.539	0.240	0.356	72.435	0.706	0.0124

IV. ILLUSTRATIVE EXAMPLE

The initial design of a six-element Yagi-Uda antenna is given in Fig. 1. All dimensions are normalized with respect to the free-space wavelength λ. We vary the normalized length $l_{2n} = l_2/\lambda$ and the normalized separations $s_{1n} = s_1/\lambda$ and $s_{2n} = s_1/\lambda$, i.e., $\boldsymbol{x} = [l_{2n}\ s_{1n}\ s_{2n}]^T$. We now proceed with the optimization of the Yagi-Uda antenna for an input impedance of $\bar{Z}_{in} = 73\ \Omega$. The objective function is defined as

$$f = \frac{|Z_{in} - \bar{Z}|}{\bar{Z}}. \tag{16}$$

The progress of the optimization is summarized in TABLE I. The gradient-based optimization routine of MATLAB® *fmincon* is used. An optimal solution is reached in ten iterations. At each design iteration, we update the $\partial \mathbf{Z}^{(k)}/\partial l_{2n}$, $\partial \mathbf{Z}^{(k)}/\partial s_{1n}$ and $\partial \mathbf{Z}^{(k)}/\partial s_{2n}$ matrices using (14). They are substituted in (7) to compute the response sensitivities.

The results are compared with the sensitivities generated by the original EM-FAST, which employs forward finite differences to approximate the system matrix derivatives. We emphasize that the optimization is driven by the gradient estimates provided by the Broyden-FAST. The EM-FAST sensitivities are calculated off-line only for validation. Both sensitivity curves are plotted in Figs. 2, 3 and 4. The Broyden update needs an initial value for the gradient at the start of the optimization. It is calculated by forward finite differences. Therefore, the sensitivity at the first step is the same for both EM-FAST and Broyden-FAST. In the next few steps, Broyden-FAST sensitivity values deviate from the EM-FAST values but, as the optimization progresses, they converge to the EM-FAST values. It is evident that our approach based on the

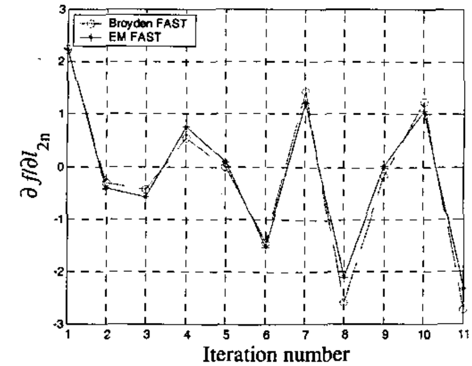


Fig. 2. Comparison of the sensitivity of the objective function with respect to the length of the driven element l_{2n} between Broyden-FAST and the original EM-FAST.

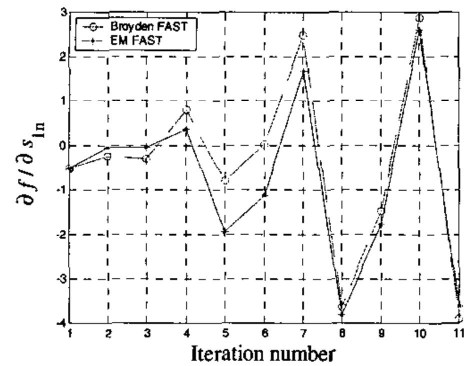


Fig. 3. Comparison of the sensitivity of the objective function with respect to the separation s_{1n} between Broyden-FAST and the original EM-FAST.

Broyden update produces sufficiently accurate sensitivity results for the purposes of gradient-based optimization.

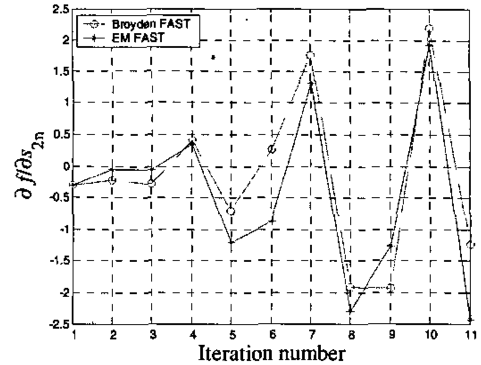


Fig. 4. Comparison of the sensitivity of the objective function with respect to the separation s_{2n} between Broyden-FAST and the original EM-FAST.

V. Conclusion

A new adjoint based method to design sensitivity analysis with MoM is proposed in which the Broyden formula accelerates the estimation and update of the derivatives of the system matrix. We show that the Broyden update is efficient, problem-independent and sufficiently accurate for the purpose of gradient-based optimization. It allows the computation of the system response and its gradient in the design parameter space through a single system analysis when it is integrated with the feasible adjoint sensitivity technique. The overhead associated with the gradient estimation is negligible in comparison with the computational requirements of one full-wave analysis. It is reduced even further than both the exact adjoint method and FAST, with respect to which it is validated.

References

[1] J.W. Bandler, Q.-J. Zhang, J. Song and R. M. Biernacki, "FAST gradient based yield optimization of nonlinear circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1701-1710, Nov. 1990.

[2] N.K. Georgieva, S. Glavic, M. Bakr and J. W. Bandler, "Feasible adjoint sensitivity technique for EM design optimization," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2751-2758, Dec. 2002.

[3] N.K. Georgieva, S. Glavic, M.H. Bakr and J. W. Bandler, "Feasible adjoint sensitivity technique for EM optimization," *IEEE MTT-S IMS Digest* 2002 (Seattle, Washington), pp. 971-974, June 2002.

[4] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Mathematics of Computation*, vol. 19, pp. 577-593, 1965.

[5] J.W. Bandler, S.H. Chen, S. Daijavad and K. Madsen, "Efficient optimization with integrated gradient approximations," *IEEE Trans. Microwave Theory Tech.*, vol. 36, pp. 444-455, Feb. 1988.

[6] E.J. Haug, K.K. Choi and V. Komkov, *Design Sensitivity Analysis of Structural Systems*. Orlando, Florida: Academic Press, Inc., 1986.

[7] J.W. Bandler, R.M. Biernacki, S.H. Chen, R.H. Hemmers and K. Madsen, "Electromagnetic optimization exploiting aggressive space mapping," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 2874-2882, Dec. 1995.

[8] F. Alessandri, M. Mongiardo and R. Sorrentino, "New efficient full wave optimization of microwave circuits by the adjoint network method," *IEEE Microwave and Guided Wave Letters*, vol. 3, pp. 414-416, Nov. 1993.

[9] J.W. Bandler, "Computer-aided circuit optimization," Chapter 6 in G.C. Temes and S.K. Mitra, Eds., *Modern Filter Theory and Design*. New York: John Wiley & Sons, 1973.

[10] C.A. Balanis, *Antenna Theory*, 2nd ed. New York: John Wiley & Sons, 1997.

[11] R. F. Harrington, *Field Computation by Moment Methods*. New York: IEEE Press, 1993.