Automated Network Design with Optimal Tolerances

J. W. BANDLER, MEMBER, IEEE, AND P. C. LIU

Abstract—A new approach to network design to obtain optimal parameter values simultaneously with an optimal set of component tolerances is proposed. An automated scheme could start from an arbitrary initial acceptable or unacceptable design and under appropriate restrictions stop at an acceptable design which is optimum in the worst case sense for the obtained tolerances.

I. Introduction

It IS the purpose of this paper to present a new concept in the network design and tolerance selection problem. The concept of a "floating and expanding polytope" suggests that the two procedures of finding an acceptable nominal point and an optimal set of tolerances be replaced by one automated scheme. Using a suitable nonlinear programming technique, any arbitrary initial acceptable or unacceptable design may be used as a starting point. The scheme would stop at an acceptable design which is optimal in the worst case sense of obtained tolerances. The most suitable objective function to be minimized would seem to be one that best describes the cost of fabrication of the circuit, as suggested by some authors [1]-[6]. Several objective functions have been investigated and the results are discussed.

II. Theoretical Considerations

The Tolerance Region

A point $\phi \triangleq [\phi_1 \; \phi_2 \cdots \phi_k]^T$ is a vector of k elements and corresponds to the component values of the network. A nominal point $\phi^0 \triangleq [\phi_1^0 \; \phi_2^0 \cdots \phi_k^0]^T$ is a point associated with a set of nonnegative tolerances $\epsilon \triangleq [\epsilon_1 \; \epsilon_2 \cdots \epsilon_k]^T \geqslant \mathbf{0}$ such that the tolerance region R_t is given by

$$R_t \triangleq \{\phi \mid \phi_i^0 - \epsilon_i \leqslant \phi_i \leqslant \phi_i^0 + \epsilon_i, \quad i = 1, 2, \cdots, k\}. \quad (1)$$

Obviously, R_t is a polytope of k dimensions with sides of length $2\epsilon_i$, $i = 1, 2, \cdots, k$, and centered at ϕ^0. The polytope has 2^k vertices. Each vertex will be indexed from an index set $H \triangleq \{1, 2, \cdots, 2^k\}$ such that

$$\phi^1 \triangleq \begin{bmatrix} \phi_1^0 - \epsilon_1 \\ \phi_2^0 - \epsilon_2 \\ \vdots \\ \phi_k^0 - \epsilon_k \end{bmatrix}, \quad \phi^2 \triangleq \begin{bmatrix} \phi_1^0 + \epsilon_1 \\ \phi_2^0 - \epsilon_2 \\ \vdots \\ \phi_k^0 - \epsilon_k \end{bmatrix}, \quad \phi^3 \triangleq \begin{bmatrix} \phi_1^0 - \epsilon_1 \\ \phi_2^0 + \epsilon_2 \\ \vdots \\ \phi_k^0 - \epsilon_k \end{bmatrix}, \cdots, \quad \phi^{2^k} \triangleq \begin{bmatrix} \phi_1^0 + \epsilon_1 \\ \phi_2^0 + \epsilon_2 \\ \vdots \\ \phi_k^0 + \epsilon_k \end{bmatrix}. \quad (2)$$

A possible outcome of a circuit with a nominal design ϕ^0 and tolerance ϵ falls somewhere in or on the polytope. Depending on the location of ϕ^0 and the size of ϵ, a circuit with parameters ϕ may or may not be acceptable.

The Acceptable Region

The following discussion refers to the frequency-domain design of linear time-invariant circuits, but the results can be applied to the time domain as well. Let the set of frequency points under consideration be $\Omega = \{\omega_1, \omega_2, \cdots, \omega_u, \omega_{u+1}, \cdots, \omega_{u+l}\}$. Upper specifications $S_u(\omega_i)$, $i = 1, 2, \cdots, u$ are assigned to the first u frequency points and lower specifications $S_l(\omega_i)$, $i = u + 1, \cdots, u + l$ to the rest. Frequency points that have both upper and lower specifications may appear twice in the set. Let the response of the network at frequency ω_i be $F(\phi, \omega_i)$.

An acceptable region R_a is given by

$$R_a \triangleq \{\phi \mid S_u(\omega_i) - F(\phi, \omega_i) \geqslant 0, \quad i = 1, 2, \cdots, u$$
$$F(\phi, \omega_j) - S_l(\omega_j) \geqslant 0, \quad j = u + 1, \cdots, u + l\}. \quad (3)$$

Obviously, a design $\{\phi^0, \epsilon\}$ is an acceptable design only if $R_t \subseteq R_a$.

A Theorem

It is impossible to test all the points in R_t to see whether they are in the acceptable region R_a. In order to make the problem tractable, a number of simplifying assumptions could be made to obtain a solution to the problem with reasonable computational effort. Obviously, if R_a is convex and if all the vertices of R_t are interior or boundary points of R_a, then $R_t \subseteq R_a$. It can be shown that the assumption of convexity is unnecessarily restrictive.

Theorem [1]: If the vertices of R_t are in R_a, then $R_t \subseteq R_a$ if, for all $j = 1, 2, \cdots, k$, the assumption that

$$\phi^a, \phi^{b(j)} = \phi^a + \alpha u_j \in R_a \quad (4)$$

where α is a scalar and

Manuscript received November 20, 1972; revised June 19, 1973, and July 19, 1973. This work was supported by the National Research Council of Canada under Grants A7239 and C154. This paper is based on a paper presented at the 1973 International Symposium on Circuit Theory, Toronto, Ont., Canada, April 9-11, 1973.

The authors are with the Group on Simulation, Optimization and Control and the Department of Electrical Engineering, McMaster University, Hamilton, Ont., Canada.

$$u_1 \triangleq \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad u_2 \triangleq \begin{bmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \cdots, \quad u_k \triangleq \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

implies that

$$\phi = \phi^a + \lambda(\phi^{b(j)} - \phi^a) \in R_a \tag{5}$$

for all λ satisfying

$$0 \leqslant \lambda \leqslant 1. \tag{6}$$

Under such assumptions, only the vertices of the polytope need be tested to ensure that $R_t \subseteq R_a$. It is easy to verify that the theorem holds for $k = 1$ and 2. The proof of the theorem follows by mathematical induction. A complete proof is presented by Bandler [1].

Other constraints such as parameter constraints can be considered. These constraints define a feasible region R_f. Then it is required that $R_t \subseteq (R_a \cap R_f) = R_c$.

The Nonlinear Programming Problem

A function $C_1(\phi^0, \epsilon)$ to be minimized may be

$$C_1 = \sum_{i=1}^{k} \frac{c_i \phi_i^0}{\epsilon_i} \tag{7}$$

where c_i is a weighting factor. See, for example, Pinel and Roberts [4].

Other possibilities are [1]

$$C_2 = \sum_{i=1}^{k} \frac{c_i}{\epsilon_i} \tag{8}$$

and

$$C_3 = \sum_{i=1}^{k} c_i \log_e \frac{\phi_i^0}{\epsilon_i}. \tag{9}$$

In (9) we would be minimizing the ratio of the volume of the polytope defined by the space diagonal ϕ^0 and the volume of the polytope defined by ϵ if the $c_i = 1$.

Let

$$g_{ij}(\phi^i, \omega_j) \triangleq \begin{cases} S_u(\omega_j) - F(\phi^i, \omega_j), & \text{for } 1 \leqslant j \leqslant u \\ F(\phi^i, \omega_j) - S_l(\omega_j), & \text{for } u+1 \leqslant j \leqslant u+l \end{cases} \tag{10}$$

for $i \in H$. That is, at each vertex ϕ^i, there are $l + u$ frequency constraints. There are 2^k vertices for a polytope of k dimensions. A total of $2^k(l + u)$ constraints have to be considered. Other constraints can be added.[1]

A suitable method for solving the nonlinear programming problem is to define [7]

$$B(\phi^0, \epsilon, r) = C(\phi^0, \epsilon) + \sum_{j=1}^{u+l} \sum_{i=1}^{2^k} \frac{r}{g_{ij}(\phi^i, \omega_j)} \tag{11}$$

and minimize B with respect to ϕ^0 and ϵ for appropriately decreasing values of r. Another more recent and efficient method of handling constrained minimization is by the least pth optimization [8], [9] of

$$V(\phi^0, \epsilon, \alpha) = \max_{i,j} \left[C(\phi^0, \epsilon), C(\phi^0, \epsilon) - \alpha_{ij} g_{ij}(\phi^i, \omega_j)\right], \quad \alpha_{ij} > 0. \tag{12}$$

For sufficiently large constant values α_{ij}, the unconstrained minimization of V with respect to ϕ^0 and ϵ yields exactly the constrained minimum of C. This nonlinear programming technique makes it possible to have any initial starting point, acceptable or otherwise, as shown by Bandler and Charalambous [8], [9].

III. Examples

A Low-Pass Filter

A normalized 3-component LC low-pass ladder network, terminated with equal load and source resistances of 1 Ω, is considered. An insertion loss of 0.53 dB in the passband 0–1 rad/s and 26.0 dB in the stopband (band edge is 2.5 rad/s is realized by a minimax design without taking tolerances into account. The parameter values are $\phi_1^0 = L_1 = 1.6280$, $\phi_2^0 = C = 1.0897$, and $\phi_3^0 = L_2 = 1.6280$. The chosen set of frequency points is $\Omega = \{0.45, 0.50, 0.55, 1.0, 2.5\}$. $S_u = 1.5$ dB for the passband and $S_l = 25$ dB for the stopband are assigned. Two starting values $\phi_1^0 = 2$, $\phi_2^0 = 1$, $\phi_3^0 = 2$, and $\phi_1^0 = \phi_2^0 = \phi_3^0 = 1.5$ with 1-percent tolerances, have been studied. The first starting point is inside the acceptable region.

The sequential unconstrained minimization techniques (SUMT) method using C_1 of (7) and $c_i = 1$, $i = 1, 2, 3$, yields a solution of $\phi_1^0 = 1.9990$, $\phi_2^0 = 0.9058$, $\phi_3^0 = 1.9990$, and the corresponding tolerances are 9.89, 7.60, and 9.89 percent. Initially, $r = 1$. It is reduced by a factor of ten after each cycle of optimization. The adjoint network technique [10] and the Fletcher method [11] are used in the optimization process. A total of 185 function evaluations were performed to reduce C_1 from 300 to 33.38 for 6 complete cycles. One-hundred thirty-six function evaluations are needed to get the same results by the new nonlinear programming technique. The constants α_{ij}, $i = 1, \cdots, 8$, $j = 1, \cdots, 5$, are set uniformly to 100. p is increased from a starting value of 10–1000 for 2 cycles of optimization.

The SUMT method is not directly applicable with the second starting point which is outside the acceptable region. The same optimal point as before is reached with 105 function evaluations for 1 optimization by the new method. p is 1000 and α_{ij} is 100 for all i and j.

In contrast, if the nominal point is fixed, tolerances of 3.45, 3.18, and 3.45 percent are obtained for the three components.

A Bandpass Filter

The bandpass filter shown in Fig. 1 was studied by Butler [2], Karafin [3], and Pinel and Roberts [4]. An upper specification of 3 dB for the passband and a lower specification of

[1] Selecting, on physical or other grounds, constraints which are likely to be active at the solution to a nonlinear programming problem and discarding the rest can result in faster solution times, as is well known. Ultimately, all the constraints have to be satisfied.

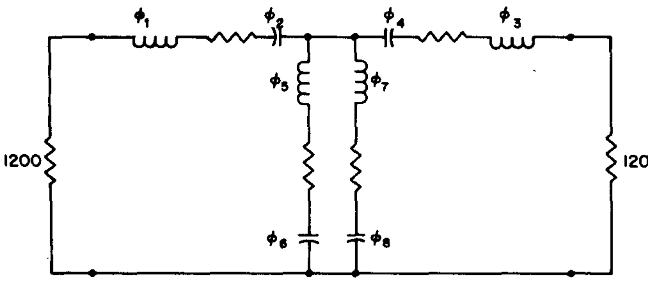

Fig. 1. Bandpass-filter example.

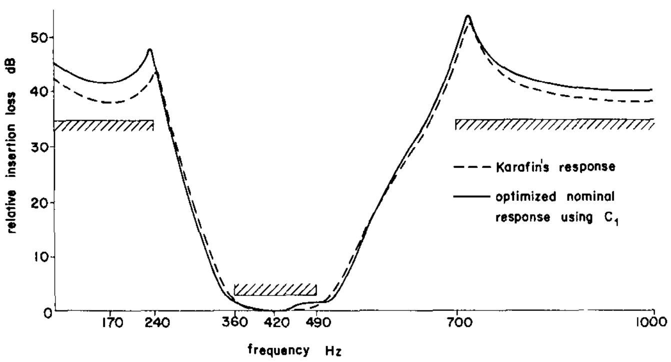

Fig. 2. Bandpass-filter response.

TABLE I
RESULTS FOR THE BANDPASS FILTER

	Karafin [3] Pinel and Roberts [4]	C_1	C_2	C_3
ϕ_1^o	1.824×10^{0}	3.0142×10^{0}	2.3206×10^{0}	2.7682×10^{0}
ϕ_2^o	7.870×10^{-8}	4.9750×10^{-8}	6.3694×10^{-8}	5.2611×10^{-8}
ϕ_3^o	1.824×10^{0}	2.9020×10^{0}	2.3206×10^{0}	2.7682×10^{0}
ϕ_4^o	7.870×10^{-8}	5.0729×10^{-8}	6.3694×10^{-8}	5.2611×10^{-8}
ϕ_5^o	4.272×10^{-1}	8.2836×10^{-1}	6.0517×10^{-1}	7.7895×10^{-1}
ϕ_6^o	9.880×10^{-7}	5.5531×10^{-7}	7.7708×10^{-7}	5.8726×10^{-7}
ϕ_7^o	1.437×10^{-1}	3.0319×10^{-1}	2.1677×10^{-1}	2.5438×10^{-1}
ϕ_8^o	3.400×10^{-7}	1.6377×10^{-7}	2.2630×10^{-7}	1.8981×10^{-7}
t_1	3 , 3.32	6.99	2.29	7.67
t_2	5 , 2.41	6.52	11.26	6.53
t_3	5 , 3.30	6.97	2.29	7.67
t_4	3 , 2.41	6.55	11.26	6.53
t_5	2 , 1.14	4.36	3.30	4.33
t_6	2 , 1.89	5.69	3.02	8.10
t_7	3 , 7.80	6.80	6.61	5.85
t_8	5 , 2.07	5.25	4.40	2.71
Cost	2.60 3.45	1.34	2.06	1.46

35 dB for the stopband relative to 0 dB at a central frequency at 420 Hz are assigned. See Fig. 2. $\Omega = \{360, 490, 170, 240, 700, 1000\}$ in which the first 2 frequencies are assigned to the upper specification and the last 4 to the lower specification. The frequency point of 420 Hz is not included as it is kept at zero. A constant Q is assumed for the four inductors and, therefore, the four corresponding resistances are dependent variables.

Nominal values used by Pinel and Roberts and a $\frac{1}{2}$-percent tolerance for each component are used as a starting point. Parameter values are scaled by normalizing with respect to the central frequency and the load resistance such that the inductors and capacitors will have the same order of magnitude to avoid ill-conditioning. Components ϕ_3 and ϕ_4 are assumed equal to ϕ_1 and ϕ_2, respectively, for the objective function C_2 and C_3. Only 2^6 vertices are taken. Initially, the same assumptions are made for the objective function C_1, but because of some violations a selection of the 2^8 vertices are subsequently taken.[2]

Using the SUMT method, initially, $r = 1$. r was reduced successively by a factor of ten. The adjoint network technique and the Fletcher method are again used in the optimization process. See Table I and Fig. 2 for some results. No more than 10 min on a CDC 6400 are needed to obtain the results for 2^6 vertices. Note that $c_i = 1$, $t_i \triangleq 100\epsilon_i/\phi_i^0$, and the cost is $\Sigma_{i=1}^8 1/t_i$. There are no violations observed for both the Monte Carlo and the worst case analyses at the specified test frequencies assuming 2^8 vertices. The relative insertion loss, however, becomes negative in some instances in the passband. The same assumptions were made as Pinel and Roberts [4] that the component distribution is uniformly concentrated within 5 percent of the extremes of the relative tolerances and 1000 simulations were made for the Monte Carlo analysis.

[2]The algorithm currently being used selects, for each vertex ϕ^j at a particular frequency, another vertex ϕ^i such that the signs of the components of $\phi^i - \phi^0$ are all opposite to the corresponding signs of the components of the gradient vector of the constraint evaluated at ϕ^j and that frequency. This usually leads to a substantially smaller number of constraints to be considered at each frequency during optimization. Periodic updating of the selected vertices and restarting of the optimization process is generally required.

IV. CONCLUSIONS

It has been shown that, by moving the nominal point, a set of larger tolerances can usually be obtained, and that an arbitrary initial design may be used to start the automated scheme. A drawback of this basic scheme is, of course, that a large number of constraints are used. Future work should, it is felt, be concentrated on methods of reducing them. Some preliminary ideas of reducing the number of constraints are currently being tested.[2] A complete solution to the problem is not claimed; however, it may be concluded that our approach is a promising one in network design subject to tolerance considerations.

ACKNOWLEDGMENT

The authors wish to thank Dr. B. J. Karafin and Dr. E. M. Butler of Bell Laboratories, Holmdel, N. J., and J. F. Pinel of Bell-Northern Research, Ottawa, Ont., Canada, for valuable discussions and for providing some unpublished results of their work. J. F. Pinel, in particular, pointed out errors in our preliminary results on the bandpass-filter problem [12]. The authors also wish to thank Dr. C. Charalambous, J. H. K. Chen, V. K. Jha, and Mrs. J. R. Popović for the help they provided.

REFERENCES

[1] J. W. Bandler, "Optimization of design tolerances using nonlinear programming," in *Proc. 6th Princeton Conf. Information Sciences and Systems* (Princeton, N. J.), Mar. 1972, pp. 655–659; also in

J. Optimiz. Theory Appl., to be published.
[2] E. M. Butler, "Realistic design using large-change sensitivities and performance contours," *IEEE Trans. Circuit Theory (Special Issue on Computer-Aided Circuit Design)*, vol. CT-18, pp. 58–66, Jan. 1971.
[3] B. J. Karafin, "The optimum assignment of component tolerances for electrical networks," *Bell Syst. Tech. J.*, vol. 50, pp. 1225–1242, Apr. 1972.
[4] J. F. Pinel and K. A. Roberts, "Tolerance assignment in linear networks using nonlinear programming," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 475–479, Sept. 1972.
[5] A. K. Seth, "Electrical network tolerance optimization," Ph.D. dissertation, Univ. Waterloo, Waterloo, Ont., Canada, 1972.
[6] E. De Castro, G. Iuculano, and V. A. Monaco, "Component value spread and network function tolerances: An optimal design procedure," *Alta Freq.*, vol. 40, pp. 867–872, 1971.
[7] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* New York: Wiley, 1968.
[8] J. W. Bandler and C. Charalambous, "A new approach to nonlinear programming," in *Proc. 5th Hawaii Int. Conf. System Sciences* (Honolulu, Hawaii), Jan. 1972, pp. 127–129.
[9] ——, "Nonlinear programming using minimax techniques," *J. Optimiz. Theory Appl.*, vol. 13, 1974.
[10] J. W. Bandler and R. E. Seviora, "Current trends in network optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-18, pp. 1159–1170, Dec. 1970.
[11] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, pp. 317–322, Aug. 1970.
[12] J. W. Bandler and P. C. Liu, "Automated network design with optimal tolerances," in *Proc. 1973 Int. Symp. Circuit Theory* (Toronto, Ont., Canada), Apr. 1973, pp. 181–184.

An Optimal Pivoting Order for the Solution of Sparse Systems of Equations

M. NAKHLA, STUDENT MEMBER, IEEE, K. SINGHAL, MEMBER, IEEE, AND J. VLACH, SENIOR MEMBER, IEEE

***Abstract*—Analytic expressions for finding fill-in, the number of nonzero elements that change in value, and the number of "long operations" during each step of the *LU* decomposition are given.**

A new optimal pivot ordering algorithm is proposed which leads to a reduction of the overall fill-in and long operation count. Comparison is made with two other known algorithms.

I. Introduction

Several papers have been published on the solution of sparse systems of equations [1]. Results have indicated that considerable computational and storage savings as well as reduction of roundoff errors can be achieved in solving large sparse systems. It has also been shown that the order in which the variables are eliminated strongly affects the fill-in and the number of long operations (multiplications or divisions) required. Several algorithms [1]-[7] have been proposed for determining near optimal ordering.

In this paper, an alternative algorithm is proposed and compared with the algorithms published by Berry [2] and Hsieh-Ghausi [3]. It results in fewer operations and smaller fill-in.

Manuscript received March 30, 1973. This work was supported by the National Research Council under Grant A7398.
The authors are with the Department of Electrical Engineering, University of Waterloo, Waterloo, Ont., Canada.

II. Theorems

Two similar LU decompositions are used for an $n \times n$ matrix A, the first one being

$$\begin{aligned} l_{ij} &= a_{ij}^{(j-1)}, && i = j, \cdots, n \\ u_{ij} &= a_{ij}^{(i-1)}/l_{ii}, && j = i+1, \cdots, n \\ a_{ij}^{(r)} &= a_{ij}^{(r-1)} - l_{ir}\, u_{rj}, && i,j = r+1, \cdots, n \\ a_{ij}^{(0)} &= a_{ij} \end{aligned}$$

with the result that all the diagonal elements of U have the value one. The second algorithm is described by

$$\begin{aligned} u_{ij} &= a_{ij}^{(i-1)}, && j = i, \cdots, n \\ l_{ij} &= a_{ij}^{(j-1)}/u_{jj}, && i = j+1, \cdots, n \\ a_{ij}^{(r)} &= a_{ij}^{(r-1)} - l_{ir}\, u_{rj}, && i,j = r+1, \cdots, n \\ a_{ij}^{(0)} &= a_{ij} \end{aligned} \tag{2}$$

with the result that all the diagonal elements of L have the value 1.

Theorem 1: Assume that the $n \times n$ matrix A has completed the rth step of the LU decomposition described either by (1) or by (2). Let $A^{(r)}$ be the updated matrix.