Technical University of Denmark

DTU

Singularities in minimax optimization of networks

Madsen, Kaj; Schjær-Jacobsen, Hans

DTU Library
Technical Information Center of Denmark

Singularities in Minimax Optimization of Networks

KAJ MADSEN AND HANS SCHJÆR-JACOBSEN

Abstract—A theoretical treatment of singularities in nonlinear minimax optimization problems, which allows for a classification in *regular* and *singular* problems, is presented. A theorem for determining a singularity that is present in a given problem is formulated. A group of problems often used in the literature to test nonlinear minimax algorithms, i.e., minimax design of multisection quarter-wave transformers, is shown to exhibit singularities and the reason for this is pointed out.

Based on the theoretical results presented an algorithm for nonlinear minimax optimization is developed. The new algorithm maintains the quadratic convergence property of a recent algorithm by Madsen *et al.* when applied to *regular* problems and it is demonstrated to significantly improve the final convergence on *singular* problems.

I. INTRODUCTION

A NONLINEAR MINIMAX optimization algorithm with quadratic final convergence has recently been presented and applied to circuit optimization problems [1], [2]. However, as briefly discussed in [2], ill-conditioning may cause slow convergence of the algorithm even if certain parameters governing the optimization strategy were tentatively preadjusted.

The purpose of this paper is to present a theoretical treatment of singularities giving rise to ill-conditioned minimax problems and to develop a new method to solve singular problems. Since it is not always possible to determine beforehand whether a particular problem is regular or singular it is desirable to construct an automatic procedure for detection of singularities inherent in a given problem. Hence, the new algorithm is able to maintain the ultimate quadratic convergence when applied to regular problems and to reestablish fast final convergence in cases where singularities have been detected during optimization. The efficiency of the new approach is established by comparisons with some algorithms reported in the literature. But first we turn to a discussion of general aspects of singularities.

II. SINGULARITIES IN MINIMAX APPROXIMATION

We formulate the minimax problem as that of minimizing

$$F(\underline{x}) = \max_j |f_j(\underline{x})| \tag{1}$$

where

$$f_j(\underline{x}) = f_j(x_1, \cdots, x_n), \qquad j = 1, \cdots, m \quad m > n \tag{2}$$

is a set of residuals which are nonlinear functions of the design parameters $(x_1, \cdots, x_n)$. The following smoothness condition is assumed:

$$f_j(\underline{x} + \underline{h}) = f_j(\underline{x}) + \underline{h}^T \underline{f}_j'(\underline{x}) + o(\underline{h}) \tag{3}$$

where $\|o(\underline{h})\| / \|\underline{h}\| \to 0$ for $\underline{h} \to 0$. The solution of the minimax problem is denoted by $\underline{x}^*$, and the number of residuals for which the maximum (1) is attained at $\underline{x} = \underline{x}^*$ is r. (In order to simplify the notation we let the corresponding r functions be $f_1, \cdots, f_r$). This means that

$$|f_i(\underline{x}^*)| = F(\underline{x}^*), \qquad i = 1, \cdots, r$$
$$|f_i(\underline{x}^*)| < F(\underline{x}^*), \qquad i = (r+1), \cdots, m. \tag{4}$$

Definition 1

The minimax problem is *singular* with respect to the solution $\underline{x}^*$ if the matrix

$$\underline{\underline{D}} = \{\partial f_i / \partial x_j(\underline{x}^*)\} \quad \begin{matrix} i = 1, \cdots, r \\ \\ j = 1, \cdots, n \end{matrix} \tag{5}$$

has rank less than n. Otherwise the problem is *regular*.

In order to characterize singular problems the following definition is useful.

Definition 2

If the limit

$$F_{\underline{v}}'(\underline{x}) \equiv \lim_{t \to 0} \frac{F(\underline{x} + t\underline{v}) - F(\underline{x})}{t}, \qquad \underline{v} \neq \underline{0} \tag{6}$$

exists, we say that F is differentiable at $\underline{x}$ along the line through $\underline{x}$ and $\underline{x} + \underline{v}$.

Since $\underline{x}^*$ is a minimum of F we have for $t > 0$

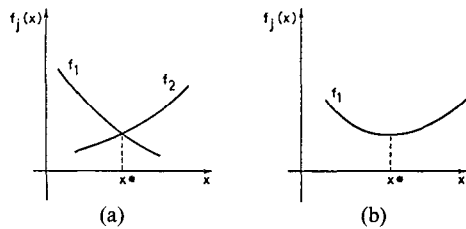$$\frac{F(\underline{x}^* + t\underline{v}) - F(\underline{x}^*)}{t} \geqslant 0 \tag{7}$$

Fig. 1. One-dimensional minimax problems. (a) Regular problem, $n=1, r=2$. (b) Singular problem, $n=1, r=1$.

for any direction $\underline{v}$, and therefore, if $F'_{\underline{v}}(x^*)$ exists it must be 0.

The proof of the following theorem is found in the Appendix.

Theorem 1

If the matrix $\underline{\underline{D}}$ has rank q then there exists a subspace $V \subseteq R^n$ of dimension $(n-q)$ such that

i) $\underline{v} \in V \Rightarrow F'_{\underline{v}}(x^*) = 0$;

ii) $\underline{w} \notin V \Rightarrow F'_{\underline{w}}(x^*)$ does not exist, and there is a positive number c such that

$$\left| \frac{F(\underline{x}^* + s\underline{w}) - F(\underline{x}^*)}{s} - \frac{F(\underline{x}^* + t\underline{w}) - F(\underline{x}^*)}{t} \right| \geqslant c \tag{8}$$

for all values $s < 0$ and $t > 0$ sufficiently close to 0.

Thus when the minimax problem is regular $V = \{\underline{0}\}$, and it therefore follows from (8) that the function $\phi_w(\xi) \equiv F(x^* + \xi w)$ has an edge at $\xi = 0$ for any $w \neq 0$. In the singular case there is at least one line along which the slope of F is 0, and if $q > 0$ there are other lines along which F is edged. Consequently, when the problem is singular, the solution is not so well determined numerically as in the regular case, and if $0 < q < n$ the minimum is situated in the bottom of an edged valley. Therefore many algorithms, especially descent methods, perform rather badly on singular problems.

The following theorem has been proved by Curtis and Powell [3], and a generalization is given by Bandler [4].

Theorem 2

The $r \times n$ matrix $\underline{\underline{D}}$ has rank less than r.

Thus r must be at least $(n+1)$ in order that the rank of $\underline{\underline{D}}$ is n.

Theorem 3

If $r < n+1$ then the minimax problem is singular.

This is a very simple sufficient criterion for testing if a problem is singular.

The intuitive interpretation of the theorem is straightforward. Consider Fig. 1 where the situation is sketched in one dimension. Here we have $n=1$ and $r=2$ in Fig. 1 (a) (a regular problem) but $r=1$ in Fig. 1(b) (a singular

problem). Clearly the minimum is much better defined numerically in the regular case than in the singular one.

III. A New Nonlinear Minimax Optimization Algorithm

The iterative algorithm described in [2] has quadratic final rate of convergence, when the problem to be solved is regular. However, when solving singular problems the algorithm may perform much slower. One important reason for this is that it is a descent method, i.e.,

$$F(\underline{x}_{k+1}) \leqslant F(\underline{x}_k) \tag{9}$$

is required. Consequently, when the current approximation $\underline{x}_k$ is in a steep valley, the step length is forced to be very small. Therefore in the present approach we occasionally allow for a step for which (9) does not hold. The change we make to the algorithm will not have any influence on the final rate of convergence in nonsingular cases.

As in [2] the increment $\underline{h}_k$ to add to $\underline{x}_k$ is found as a solution of the linear minimax problem

$$\max_j |f_j(\underline{x}_k) + \underline{f}'_j(\underline{x}_k)^T \underline{h}_k|$$

$$= \min_{\|\underline{h}\| < \lambda_k} \left\{ \max_j |f_j(\underline{x}_k) + \underline{f}'_j(\underline{x}_k)^T \underline{h}| \right\} \tag{10}$$

where $\|\underline{h}\| \equiv \max_i |h_i|$. The bound λ_k is adjusted automatically and will be small when a valley is reached. We define this to be the case when

$$|f_j(\underline{x}_k)| \geqslant F(\underline{x}_k)(1 - \delta) \tag{11}$$

holds for more than one value of j, δ being a small positive number. If this is true for $k = q$ the slope in the direction of steepest descent of F is estimated from the solution of the linear subproblem (10) at $\underline{x}_q$ with the bound $\lambda = \lambda_0$. This estimate is denoted α_q. Next five further iterations are taken, and the slope α_{q+5} is calculated. Now the distance d to the solution is estimated by linear extrapolation using the fact that the slope is zero at the solution $\underline{x} = \underline{x}^*$. Thus we get the estimate

$$d \cong d' \equiv l_{q+5} \| \underline{x}_{q+5} - \underline{x}_q \|$$

$$l_{q+5} = \max \left\{ 1, \min \left\{ 5, \frac{\alpha_{q+5}}{\alpha_q - \alpha_{q+5}} \right\} \right\} \tag{12}$$

where the heuristic bounds 1 and 5 are introduced in order to make the process stable. Now λ_{q+5} is chosen as d' with the exception that λ_k must not exceed a general upper bound Λ.

When this special step is taken the iteration is continued as before, but no further special step is allowed again until $F(\underline{x}_k)$ is less than $F(\underline{x}_{q+5})$. This ensures convergence.

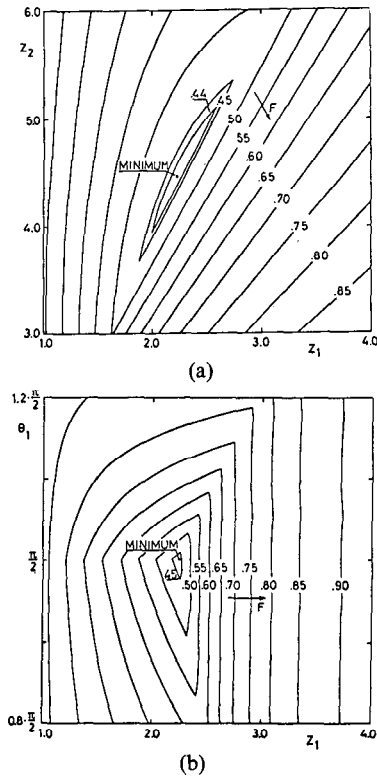It should be pointed out here that when a regular problem is solved the bound λ_k is inactive near the solu-

Fig. 2. Two-dimensional $(n=2)$ minimax problems arising from optimization of two-section transmission-line 10:1 transformers over 100-percent bandwidth. (a) Singular problem, $r=2$. Variation of Z_1 and Z_2, $\theta_1 = \theta_2 = \pi/2$. (b) Regular problem, $r=3$. Variation of Z_1 and θ_1, $Z_2 = 10/Z_1$, $\theta_2 = \theta_1$.

tion, and therefore the present algorithm maintains the quadratic final convergence of the algorithm presented in [2].

IV. ON SINGULARITIES IN A GROUP OF TEST PROBLEMS

A group of problems often used in the literature to test nonlinear minimax optimization algorithms consists in minimization of the maximum reflection coefficient in N-section transmission-line transformers. The (normalized) sample frequencies within the passband are chosen such that all frequencies at which the Chebyshev response attains its maxima are included. In other words, an *a priori* knowledge of the optimum minimax solution is used when selecting the sample frequencies. In the following we demonstrate that this may often create singularities if the optimization parameters are not carefully chosen.

Fig. 2(a), (b) reveals contour plots of the objective function

$$F(\underline{x}) = \max_j |\rho_j(\underline{x})|| \qquad (13)$$

where $\rho_j(x)$ is the complex reflection coefficient of a 10:1 transformer with 100-percent bandwidth sampled at 11 sample frequencies ψ_j. Vector x contains the optimization parameters. A discussion in the context of Section II is in order.

Consider first Fig. 2(a) which is the case of a two-section quarter-wave transformer where the characteristic impedances are varied, [5]. Clearly, the sufficient condition for a singularity presented in Theorem 3 is fulfilled and hence the problem is singular. It will be recalled from the frequency response of the optimum Chebyshev two-section transformer that the maximum reflection coefficient is attained at three frequencies, namely, the two band limits and the center frequency. The reason for having the singularity in Fig. 2(a) is that due to the symmetry of the frequency response, the two residuals at the band-edge sampling frequencies effectively create only one surface in the contour, thereby reducing r from 3 to 2.

An example of a regular minimax problem is subsequently shown in Fig. 2(b). The problem is generated using a two-section transformer, Z_1 and θ_1 being the optimization parameters with $Z_2 = 10/Z_1$ and $\theta_2 = \theta_1$. In this case we have $r=3$ corresponding to the situation that no directional derivative as defined by (6) exists at $\underline{x} = \underline{x}^*$.

V. NUMERICAL RESULTS

The proposed algorithm for solution of singular nonlinear minimax optimization problems has been applied to minimize the maximum reflection coefficient of 10:1 three- and four-section transformers with 100-percent bandwidth using different starting points for the impedances and lengths. According to the developments in Sections II and IV these optimization problems are singular and hence provide us with a good basis for evaluation of the algorithm performance. Moreover, our results can be compared with other algorithms applied to the same problems.

We define the residuals f_j by

$$f_j = \frac{1}{2} |\rho_j(\underline{x})|^2, \qquad j = 1, \cdots, 11 \qquad (14)$$

where the 11 sample frequencies are given by

$$\underline{\psi} = \{0.5, 0.6, 0.7, 0.77, 0.9, 1.0, 1.1, 1.23, 1.3, 1.4, 1.5\} \qquad (15)$$

in the three-section case and

$$\underline{\psi} = \{0.5, 0.6, 0.667, 0.8, 0.9, 1.0, 1.1, 1.2, 1.333, 1.4, 1.5\} \qquad (16)$$

in the four-section case.

The transformers are conveniently analyzed by scattering matrices and the derivatives $\partial f_j / \partial x_i$ are obtained by the adjoint network concept [6]. The basis of algorithm comparisons will be the number of network evaluations (including calculation of the gradients) required to bring the maximum reflection coefficient

$$\rho_{max} = \max_j |\rho_j| \qquad (17)$$

within 5 correct digits of the known optimum values.

A three-section quarter-wave transformer was first optimized with starting point $(Z_1, Z_2, Z_3) = (1.0, 3.16228, 10.0)$.

TABLE I
OPTIMIZATION TO 5 CORRECT DIGITS OF A THREE-SECTION 10:1
TRANSFORMER WITH 100-PERCENT BANDWIDTH, $\delta = 0.001$, $\Lambda = 1$

		1.0	1.5
Starting values	z_1	1.0	1.5
	θ_1	$\frac{\pi}{2}$	$0.8 \cdot \frac{\pi}{2}$
	z_2	3.16228	3.0
	θ_2	$\frac{\pi}{2}$	$1.2 \cdot \frac{\pi}{2}$
	z_3	10.0	6.0
	θ_3	$\frac{\pi}{2}$	$0.8 \cdot \frac{\pi}{2}$

		Number of network evaluations	
Madsen et al. [2]		707	253
Least p'th [7], average		155	95
Charalambous and Conn [8]		67	162
	λ_0		
	.1	48	29
	.2	68	18
	.3	53	44
Present	.4	82	36
algorithm	.5	46	38
	.6	42	32
	.7	94	32
	.8	67	19
	.9	89	44
	1.0	66	45
	Average	66	34

TABLE II
OPTIMIZATION TO 5 CORRECT DIGITS OF A FOUR-SECTION 10:1
TRANSFORMER WITH 100-PERCENT BANDWIDTH, $\delta = 0.001$, $\Lambda = 1$,
$\lambda_0 = 0.1$

Starting values								Number of network evaluations	
z_1	θ_1	z_2	θ_2	z_3	θ_3	z_4	θ_4	Madsen et al. [2]	Present algorithm
1.0	$\frac{\pi}{2}$ 2.2361	$\frac{\pi}{2}$	4.4721	$\frac{\pi}{2}$	10.0	$\frac{\pi}{2}$		2300	262
1.6347	$\frac{\pi}{2}$ 3.1623	$\frac{\pi}{2}$	6.1173	$\frac{\pi}{2}$	10.0	$\frac{\pi}{2}$		2466	217
2.0	$0.8\frac{\pi}{2}$ 3.0	$1.2\frac{\pi}{2}$	4.0	$0.8\frac{\pi}{2}$	7.0	$1.2\frac{\pi}{2}$		61	31
1.0	$1.2\frac{\pi}{2}$ 2.0	$0.8\frac{\pi}{2}$	5.0	$1.2\frac{\pi}{2}$	8.0	$0.8\frac{\pi}{2}$		1097	165
3.0	$1.3\frac{\pi}{2}$ 4.0	$1.3\frac{\pi}{2}$	6.0	$1.3\frac{\pi}{2}$	9.0	$1.3\frac{\pi}{2}$		1680	320
0.5	$0.7\frac{\pi}{2}$ 1.0	$0.7\frac{\pi}{2}$	3.0	$0.7\frac{\pi}{2}$	7.0	$0.7\frac{\pi}{2}$		998	252

This optimization required 44 response evaluations whereas 99 evaluations were necessary with the algorithm [2].

Other results are shown in Table I for different values of the initial bound λ_0 on the variables. Also shown are results obtained by the least pth algorithm of Charalambous and Bandler [7], the algorithm of Madsen et al. [2], and the recent algorithm of Charalambous and Conn [8].

In Table II results are shown for the four-section transformer using the algorithm [2] and the present algorithm. Here $\lambda_0 = 0.1$ was chosen. The extremely low number of network evaluations used with one of the starting points is not typical for the type of problems considered but rather an indication of a luckily chosen starting point.

Clearly, the number of network evaluations required by our algorithm is much less than for the method in [2], thus emphasizing the relevance of the present approach.

VI. DISCUSSION AND CONCLUSION

Based on Theorem 3 a characterization of singular minimax problems as opposed to regular ones has been established providing a means for automatic detection of ill-conditioning on some types of problems during an optimization process. A relevant strategy for optimization of singular cases has been developed and implemented in an automatic computer algorithm. A discussion of the transmission-line impedance transformer viewed as a minimax optimization problem has explained why slow convergence has often been observed in the literature. It has been established that the new algorithm is significantly faster in obtaining the solution to singular minimax problems than was the algorithm [2].

As pointed out in [2] a singularity at the solution normally will not affect the initial rate of convergence, which is very rapid. The strategy adopted in the present paper when a singularity is detected is active only in the vicinity of a singularity. Hence, the present algorithm possesses sure over all convergence, quadratic final convergence with regular problems, and fast final convergence on singular problems.

The relative success of the present approach may suggest that similar improvements are obtainable with the recently developed nongradient algorithm [9]. Although the method in [9] has proved more efficient than other nongradient algorithms it still uses a relatively large number of residual calculations when singularities are present. This possibility still remains to be investigated.

APPENDIX

Proof of Theorem 1:

Without loss of generality we can suppose that

$$f_1(\underline{x}^*) = \cdots = f_r(\underline{x}^*) = F(\underline{x}^*). \tag{A1}$$

Since the rank of $\underline{\underline{D}}$ is q there exists a vector space $V \subseteq R^n$ (the null space of $\underline{\underline{D}}$) such that

$$\underline{\underline{D}}\,\underline{v} = \underline{0}, \qquad \underline{v} \in V$$

$$\underline{\underline{D}}\,\underline{w} \neq \underline{0}, \qquad \underline{w} \notin V.$$

This means that if $\underline{v} \in V$ and $1 \leqslant i \leqslant r$ then

$$f_i(\underline{x}^* + t\underline{v}) = f_i(\underline{x}^*) + t \underline{f}_i'(\underline{x}^*)^T \underline{v} + o(t)$$

$$= f_i(\underline{x}^*) + o(t) \tag{A2}$$

and therefore we have for sufficiently small values of t

$$F(\underline{x}^* + t\underline{v}) - F(\underline{x}^*) = \max_{i \leqslant r} \{ f_i(\underline{x}^*) + o(t) \}$$

$$- \max_{i \leqslant r} \{ f_i(\underline{x}^*) \} = o(t) \tag{A3}$$

which implies that $F_{\underline{v}}'(x^*) = 0$. Now suppose that $\underline{w} \notin V$. Then there exists $j \leqslant r$ such that $\underline{w}^T f_j'(x^*) \neq 0$, which means that if $\underline{u} = \{\text{sign}\,\underline{w}^T f_j'(x^*)\} \cdot \underline{w}$ then $\underline{u}^T f_j'(x^*) > 0$. Therefore we obtain the following:

$$F(\underline{x}^* + t\underline{u}) - F(\underline{x}^*) = F(\underline{x}^* + t\underline{u}) - f_j(\underline{x}^*)$$

$$\geqslant f_j(\underline{x}^* + t\underline{u}) - f_j(\underline{x}^*) = t\underline{u}^T \underline{f}_j'(\underline{x}^*) + o(t) \tag{A4}$$

and consequently, for $t > 0$ sufficiently small,

$$\frac{F(\underline{x}^* + t\underline{u}) - F(\underline{x}^*)}{t} > \frac{1}{2} \underline{u}^T \underline{f_j'}(\underline{x}^*) > 0. \quad (A5)$$

Since

$$\frac{F(\underline{x}^* + t\underline{u}) - F(\underline{x}^*)}{t} \leqslant 0, \quad \text{for } t < 0 \quad (A6)$$

the limit $F_u'(\underline{x}^*)$ cannot exist, which means that neither $F_{-w}'(\underline{x}^*)$ nor $F_w'(\underline{x}^*)$ exists. Since (8) is a consequence of (A5) and (A6) the theorem is proved.

REFERENCES

[1] K. Madsen, "An algorithm for minimax solution of overdetermined systems of non-linear equations," *J. Inst. Maths. Applics.*, vol. 16, pp. 321–328, 1975.

[2] K. Madsen, H. Schjær-Jacobsen, and J. Voldby, "Automated minimax design of networks," *IEEE Trans. Circuits and Systems*, vol. CAS-22, pp. 791–796, Oct. 1975.

[3] A. R. Curtis and M. J. D. Powell, "Necessary conditions for a minimax approximation," *Comp. J.*, vol. 8, pp. 358–361, Jan. 1966.

[4] J. W. Bandler, "Conditions for a minimax optimum," *IEEE Trans. Circuit Theory* (Correspondence), vol. CT-18, pp. 476–479, July 1971.

[5] J. W. Bandler and P. A. Macdonald, "Cascaded noncommensurate transmission-line networks as optimization problems," *IEEE Trans. Circuit Theory* (Correspondence), vol. CT-16, pp. 391–394, Aug. 1969.

[6] J. W. Bandler and R. E. Seviora, "Wave sensitivities of networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 138–147, Feb. 1972.

[7] C. Charalambous and J. W. Bandler, "New algorithms for network optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-21, pp. 815–818, Dec. 1973.

[8] C. Charalambous and A. R. Conn, "Optimization of microwave networks," *IEEE Trans. Mocrowave Theory Tech.*, vol. MTT-23, pp. 834–838, Oct. 1975.

[9] K. Madsen, O. Nielsen, H. Schjær-Jacobsen, and L. Thrane, "Efficient minimax design of networks without using derivatives," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 803–809, Oct. 1975.

✦

Kaj Madsen was born in Denmark in 1943. He received the cand. scient. degree in mathematics from the University of Aarhus, Aarhus, Denmark, in 1968.

He has been a Lecturer in numerical analysis at the Technical University of Denmark, Lyngby, since 1968, apart from the academic year 1973–1974, when he was at AERE Harwell, Didcot, England. His fields of interest in teaching and research are optimization and interval analysis.

✦

Hans Schjær-Jacobsen was born in Løgumkloster, Denmark, on December 30, 1943. After being educated as an electronics technician, he received the M.Sc. degree in electrical engineering from the Technical University of Denmark, Lyngby, Denmark, in 1972. At present he is with the Electromagnetics Institute, Technical University of Denmark, where he is working toward the Ph.D. degree.

He has carried out research in the fields of analysis and computer-aided design of antennas, microwave reflection amplifiers, and mocrowave filters. From 1972 to 1973, he worked on a research contract with ESTEC concerning multiple beam satellite antennas. His current research interest is in the development and application of numerical optimization methods for designing antennas and circuits.