SOME RESULTS ON EXISTENCE AND UNIQUENESS OF SOLUTION OF NONLINEAR NETWORKS

by

T. Fujisawa and E. S. Kuh

Memorandum No. ERL-M288

5 August 1970

Electronics Research Laboratory

College of Engineering
University of California, Berkeley
94720

This paper was presented at the 1970 IEEE International Symposium on Circuit Theory, Atlanta, Georgia, December 14-16, 1970. Research was sponsored by the National Science Foundation Grants, GK-10656X/GF-357.

T. Fujisawa is with the Dept. of Control Engineering, Osaka University, Toyonaka, Osaka, Japan.

E. S. Kuh is with the Dept. of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, California 94720

ABSTRACT

This paper deals with nonlinear networks which can be characterized by the equation $\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{y}$, where $\underset{\sim}{f}(\cdot)$ maps the real Euclidean n-space R^n into itself and is assumed to be continuously differentiable. $\underset{\sim}{x}$ is a point in R^n and represents a set of chosen network variables, and $\underset{\sim}{y}$ is an arbitrary point in R^n and represents the input to the network. The authors first derive sufficient conditions for the existence of a unique solution of the equation for all $\underset{\sim}{y} \in R^n$ in terms of the Jacobian matrix $\partial\underset{\sim}{f}/\partial\underset{\sim}{x}$. It is shown that if a set of cofactors of the Jacobian matrix satisfies a "ratio condition", the network has a unique solution. The class of matrices under consideration is a generalization of the class P recently introduced by Fiedler and Pták, and it includes the familiar uniformly positive-definite matrix as a very special case.

The authors next consider the solution of the equation based on the method of steepest descent. It is shown that the method always converges, and furthermore, if $\underset{\sim}{f}$ is continuously twice differentiable the rate of convergence is of a geometric progression.

I. INTRODUCTION

It has been known that the analysis of general nonlinear RLC networks depends on the analysis of three one-element-kind subnetworks [1-5]. Furthermore, in analyzing the one-element-kind networks, it is important to know the conditions under which the network has a unique solution. Various useful sufficient conditions for existence and uniqueness of solution have been found by many workers in terms of the element characteristics [1], [4-10]. However, by and large, these conditions are rather restrictive, and furthermore, they essentially imply that all nonlinear elements must be locally passive.

In this paper, nonlinear resistive networks of the most general form will be considered, which include passive as well as active elements, nonlinear resistors as well as coupled elements. It is well known that most nonlinear resistive networks can be characterized by the equation

$$\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{y} \tag{1}$$

where $\underset{\sim}{f}(\cdot)$ maps the real Euclidean n-space R^n into itself. $\underset{\sim}{x}$ is a point in R^n and it represents a set of chosen network variables, and $\underset{\sim}{y}$ is an arbitrary point in R^n, which represents an arbitrary set of input to the network. The necessary and sufficient conditions for the existence of a unique solution of (1) for all $\underset{\sim}{y} \in R^n$ has been given by Palais [11, 12]. More specifically, Palais' theorem states that the necessary and sufficient conditions for the mapping $\underset{\sim}{f}:R^n \to R^n$ to be a C^1-diffeomorphism of R^n onto itself are: (i) $\underset{\sim}{f}(x)$ is continuously differentiable, (ii) det $\underset{\sim}{J} \neq 0$ for all $\underset{\sim}{x}$ in R^n, where $\underset{\sim}{J} = \partial\underset{\sim}{f}/\partial\underset{\sim}{x}$ is the Jacobian matrix, and (iii) $\lim \|\underset{\sim}{f}(\underset{\sim}{x})\| = \infty$ as $\|\underset{\sim}{x}\| \to \infty$, where $\|\cdot\|$ is the Euclidean norm. Note that C^1-diffeomorphism implies that $\underset{\sim}{f}^{-1}$ is also continuously differentiable which is of crucial importance in this paper.

Two special cases of eq. (1) have been considered recently. The first one is a sufficient condition due to Ohtsuki and Watanabe [4], which states that if the Jacobian matrix $\underset{\sim}{J}$ is uniformly positive definite, there exists a unique solution for all $\underset{\sim}{y} \in R^n$. The second deals with a subclass of the equation in (1), which is of the form

$$\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{F}(\underset{\sim}{x}) + \underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{y} \tag{2}$$

where $\underset{\sim}{F} \in \mathcal{F}^n$ and it represents a "diagonal nonlinear mapping" of R^n onto itself. For i = 1, 2, $\cdots$, n, the i-th component of $F(\cdot)$ is a strictly monotone increasing function of x_i, which maps R^1 onto itself. Sandberg and Wilson have shown [9] that the necessary and sufficient condition on the n by n constant matrix $\underset{\sim}{A}$ for eq. (2) to have a unique solution for all $\underset{\sim}{F}(\cdot) \in \mathcal{F}^n$ and all $\underset{\sim}{y} \in R^n$ is that $\underset{\sim}{A}$ be of class P_o.

Class P_o and class P matrices were introduced by Fiedler and Ptâk [13], when they considered generalizations of positive definiteness and monotonicity. A constant square matrix is said to be of class P_o if its determinant and all principal minors are non-negative. Similarly, a constant square matrix is said to be of class P if its determinant and all principal minors are positive.

In this paper the present authors derive two theorems which are sufficient conditions for $\underset{\sim}{f}$ to be a C^1-diffeomorphism. In this case, the inverse mapping $\underset{\sim}{f}^{-1}$ is continuously differentiable, and hence the dependence of solution $\underset{\sim}{x}$ on input $\underset{\sim}{y}$ is smooth. The conditions are stated in terms of the Jacobian matrix $\underset{\sim}{J}(\underset{\sim}{x}) = \partial \underset{\sim}{f}/\partial \underset{\sim}{x}$. The first is a direct generalization of the matrix of class P, and the second represents a further generalization of the first. Various remarks and examples are given to compare the results with existing ones, and to illustrate further implications of the theorems.

The numerical method of obtaining the solution of eq. (1) is next considered. Gersho [14] and Sandberg [10] applied the method of steepest decent to obtain the global minimum of a scalar function, which yields the solution of eq. (1), and they have shown that the method always converges under the assumption that the conditions of Palais' theorem are satisfied. In this paper, under a fairly general additional condition that $\underset{\sim}{f}$ is continuously twice differentiable it is shown that the rate of convergence is of a geometric progression. Some computational difficulties underlying in this method are also discussed.

II. THEOREMS ON THE EXISTENCE OF A UNIQUE SOLUTION

The two theorems of this section are the principal results of this paper.

Theorem 1. Let $\tilde{f}$ be a continuously differentiable mapping of R^n into itself, and let $\tilde{A}_k$ be the matrix consisting of the first k-rows and the first k-columns of the Jacobian matrix $\tilde{J} = [\partial f_i/\partial x_j]$ for $k = 1, \cdots, n$. Then, for any $\tilde{y} \in R^n$ there exists one and only one solution of eq. (1) if there exist positive constants $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\cdots$, $\varepsilon_n > 0$ such that

$$\left| \det \tilde{A}_1 \right| \geq \varepsilon_1, \quad \left| \frac{\det \tilde{A}_2}{\det \tilde{A}_1} \right| \geq \varepsilon_2, \cdots, \left| \frac{\det \tilde{A}_n}{\det \tilde{A}_{n-1}} \right| \geq \varepsilon_n \qquad \text{for all } \tilde{x} \in R^n \tag{3}$$

The above condition will be referred to as the "ratio condition" for convenience. Before presenting the proof of the theorem, two important facts are pointed out. First, the ratio condition implies that the following principal cofactors of the Jacobian matrix $\tilde{J}(\tilde{x})$ are non-zero for all $\tilde{x}$:

$$\det \tilde{A}_k(x) \neq 0 \qquad k = 1, 2, \cdots, n \tag{4}$$

For convenience, this will be referred to as the "cofactor condition". Note that if the cofactor condition holds, it does not necessarily follow that eq. (1) has a solution.

For example, let

$$y_1 = f_1(x_1, x_2) = x_1 \cosh x_2,$$

$$y_2 = f_2(x_1, x_2) = \int_0^{x_2} \frac{dz}{\cosh z}.$$

Then,

$$\underset{\sim}{J}(\underset{\sim}{x}) = \begin{bmatrix} \cosh x_2 & x_1 \sinh x_2 \\ 0 & \dfrac{1}{\cosh x_2} \end{bmatrix}$$

It is seen that $\det \underset{\sim}{A}_1 = \cosh x_2$ and $\det \underset{\sim}{A}_2 = 1$, which are nonzero for all $\underset{\sim}{x}$. However, $\det \underset{\sim}{A}_2 / \det \underset{\sim}{A}_1 = 1/\cosh x_2$, which does not satisfy the ratio condition. It is easy to check that the equation does not always have a solution because the range of f_2 is bounded. It will be seen in the proof below that if the cofactor condition is satisfied, the solution is unique if it exists. Thus, the ratio condition guarantees, in addition, that $\underset{\sim}{f}$ is an onto mapping.

The second fact is that $\det \underset{\sim}{J} = \det \underset{\sim}{A}_n \neq 0$ implies that for any $\underset{\sim}{y} \in R^n$ there is a neighborhood of $\underset{\sim}{y}$ where the inverse of $\underset{\sim}{f}$ can be defined, due to the well-known theorem on implicit functions in classical analysis [15]. Hence, if $\underset{\sim}{f}^{-1}$ is uniquely determined globally, this global inverse has to coincide with any of local inverses which are continuously differentiable [15]. Therefore, the fact that $\underset{\sim}{f}$ maps R^n onto itslef as a one-to-one correspondence implies that $\underset{\sim}{f}$ is a C^1-diffeomorphism. It also has to be remarked that the continuous differentiability of $\underset{\sim}{f}^{-1}$ implies the following: $\underset{\sim}{f}^{-1}$ satisfies a Lipschitz condition on any bounded region.

This local Lipschitzian condition will play an important role later in the section on solution method.

Other implications of the theorem will be examined after giving a proof.

<u>Proof</u>. It is easy to see that the statement of the theorem holds for $n = 1$, since the condition $|\det \underset{\sim}{A}_1| = |df_1/dx_1| \geq \varepsilon_1$ implies that f_1 is a strictly monotone increasing mapping of R^1 onto itself. Hence, the theorem is proven by induction.

Under the assumption that the statement is valid for $n = k-1$, the case of $n = k$ is considered. In this case,

$$y_i = f_i(x_1, \cdots, x_{k-1}, x_k) \qquad (i = 1, \cdots, k) \tag{5}$$

or in vector notation

$$\underset{\sim}{y} = \underset{\sim}{f}(\underset{\sim}{x}). \tag{6}$$

For the purpose of notational simplicity the following convention is introduced:

$$\underset{\sim}{x}_{-k} = [x_1, \cdots, x_{k-1}]^t \quad \underset{\sim}{y}_{-k} = [y_1, \cdots, y_{k-1}]^t,$$
$$\underset{\sim}{f}_{-k} = [f_1, \cdots, f_{k-1}]^t \tag{7}$$

and hence

$$\underset{\sim}{y}_{-k} = \underset{\sim}{f}_{-k}(\underset{\sim}{x}_{-k}, x_k) \tag{8}$$

where the superscript t denotes the transpose.

If the value of x_k is kept fixed, the mapping $\underset{\sim}{f}_{-k}$ of R^{k-1} into itself is continuously differentiable. Furthermore, it is clear that the mapping

$\underset{\sim}{f}_{-k}$ satisfies the ratio condition for $n = k-1$, and therefore $\underset{\sim}{f}_{-k}$ is a C^1-diffeomorphism of R^{k-1} onto itself. Thus, $\underset{\sim}{x}_{-k}$ can be represented as a function of $\underset{\sim}{y}_{-k}$ and x_k as follows:

$$\underset{\sim}{x}_{-k} = \underset{\sim}{f}_{-k}^{-1} (\underset{\sim}{y}_{-k}, x_k). \tag{9}$$

Substituting this relation into the k-th equation of (5), y_k is represented as a function of $\underset{\sim}{y}_{-k}$ and x_k. The dependence of y_k on x_k can be determined in the following way provided that the value of $\underset{\sim}{y}_{-k}$ is kept fixed [16]. The differentiation of (6) yields

$$d\underset{\sim}{y} = \underset{\sim}{J}\, d\underset{\sim}{x}$$

or equivalently

$$\begin{bmatrix} d\underset{\sim}{y}_{-k} \\ dy_k \end{bmatrix} = \underset{\sim}{J} \begin{bmatrix} d\underset{\sim}{x}_{-k} \\ dx_k \end{bmatrix}. \tag{10}$$

Since $d\underset{\sim}{y}_{-k} = 0$ and since $\det \underset{\sim}{A}_k \neq 0$, Cramer's rule can be used to obtain

$$\frac{dy_k}{dx_k} = \frac{\det \underset{\sim}{A}_k}{\det \underset{\sim}{A}_{k-1}} . \tag{11}$$

The cofactor condition (4) implies that if the value of $\underset{\sim}{y}_{-k}$ is kept fixed, y_k is a strictly monotone increasing or decreasing function of x_k for $-\infty < x_k < \infty$. The ratio condition (3) implies that the range of y_k covers the whole real line $-\infty < y_k < \infty$.

Let $\underset{\sim}{y}^* = [y_1^*, \cdots, y_{k-1}^*, y_k^*]^t$ be an arbitrary given point of R^k. For any value $x_k = s$ there is one and only one point $\underset{\sim}{x}(s) = [x_1(s), \cdots, x_{k-1}(s), s]^t$ such that

$$[y_1^*, \cdots, y_{k-1}^*, y_k]^t = \underset{\sim}{f}(\underset{\sim}{x}(s)). \tag{12}$$

It is now clear from the property of the dependence of y_k on $x_k = s$ discussed in the preceeding paragraph that there exists one and only one value of $s = s^*$ for which the following relation holds:

$$\underset{\sim}{y}^* = \underset{\sim}{f}(\underset{\sim}{x}(s^*)). \tag{13}$$

Thus, the point $\underset{\sim}{x}(s^*)$ is the unique solution to the equation $\underset{\sim}{y}^* = \underset{\sim}{f}(\underset{\sim}{x})$. This completes the proof of Theorem 1.

Remark 1. Ohtsuki and Watanabe have shown [4] that the uniform positive definiteness implies the same conclusion. It is not difficult to show that their statement is a special case of the Theorem 1. It has to be noticed first that the uniform positive definiteness guarantees a global Lipshitz condition for the inverse mapping $\underset{\sim}{f}^{-1}$. There easily follows from the derivation of the relation (11)

$$\|d\underset{\sim}{y}\| = \|dy_k\| = |\det \underset{\sim}{A}_k/\det \underset{\sim}{A}_{k-1}|\cdot|dx_k| \leq |\det \underset{\sim}{A}_k/\det \underset{\sim}{A}_{k-1}|\ \|d\underset{\sim}{x}\|.$$

Then, it is evident that a global Lipschitz condition of $\underset{\sim}{f}^{-1}$ requires the existence of the positive lower bound for the ratio of the two determinants. Therefore, the uniform positive definiteness satisfies the ratio condition.

The consideration of the case of uniform positive definiteness suggests to further pursue the question of whether or not a global Lipschitz condition follows from the ratio condition. So far no answer has been found.

Remark 2. The class P matrix of Fiedler and Pták is closely related to the cofactor condition (4). If the Jacobian matrix is of class P for all $\underset{\sim}{x}$, obviously the cofactor condition is satisfied. However, the cofactor condition allows det $\underset{\sim}{A}_k$ to be either positive or negative.

For example, let

$$y_1 = f_1(x_1, x_2) = x_1 + x_1^3$$

$$y_2 = f_2(x_1, x_2) = -x_2$$

The Jacobian matrix is not of class P, yet it satisfies the ratio condition, hence the cofactor condition. Therefore, the equation has a unique solution for all $\underset{\sim}{y}$. In addition, it has to be noted that a matrix with positive det $\underset{\sim}{A}_k$, $k = 1, 2, \cdots, n$, may not be of class P. For example, in

$$\underset{\sim}{A} = \begin{bmatrix} 1 & 5 & 0 \\ 0 & 1 & 5 \\ 1 & 5 & 1 \end{bmatrix}$$

one principal cofactor is negative. All these point out that the ratio condition is a broad generalization of the class P matrix so long as one deals with the question of uniqueness of solution of nonlinear equations. Sandberg and Wilson pointed out that class P matrix is useful in transistor circuit analysis [9, 10]. It is then clear that the class of circuits under consideration can include various forms of nonreciprocal and active nonlinear devices. The following example illustrates these points.

Example. A special case is considered, where a network consists of transistors, junction diodes, linear resistors and independent current sources. If the Ebers-Moll model is used to represent transistors, all nonlinear elements are voltage-controlled, although there exist nonlinear couplings between elements due to the existence of transistors. The cut-set analysis can be applied under the assumption that there exists a tree containing all nonlinear elements. Then, it turns out that eq. (1) is of the form

$$f_i(\underset{\sim}{x}) = \sum_{j=1}^{n} b_{ij} \, e^{\frac{x_j}{\lambda j}} + \sum_{j=1}^{n} c_{ij} x_j + d_i \tag{14}$$

for i = 1, 2, ..., n, where b_{ij}, c_{ij} and d_i are constants and λ_j are positive constants. In this case, the Jacobian matrix $\underset{\sim}{J}(\underset{\sim}{x})$ has the following form:

$$\underset{\sim}{J}(\underset{\sim}{x}) = [b_{ij} \, \zeta_j + c_{ij}] \tag{15}$$

where

$$\zeta_j = \frac{1}{\lambda_j} e^{\frac{x_j}{\lambda j}} \tag{16}$$

and ζ_j can assume any positive values for j = 1, 2, ⋯, n.

For the notational simplicity, the following conventions are used:

$$\underset{\sim}{B} = [b_{ij}], \qquad \underset{\sim}{C} = [c_{ij}] \tag{17}$$

and $\underset{\sim}{B}_k$ and $\underset{\sim}{C}_k$ denote the matrices consisting of the first k-rows and first k-columns of $\underset{\sim}{B}$ and $\underset{\sim}{C}$, respectively. The notation $\underset{\sim}{C}_k\,(i_1, i_2, \cdots, i_j)$ for $1 \leq i_1 < i_2 < \cdots < i_j \leq k$ denotes the matrix consisting of the first k-rows and first k-columns of a matrix which is obtained from $\underset{\sim}{C}$ by replacing its i_1-th, i_2-th, ⋯, i_j-th columns by the corresponding i_1-th, i_2-th, ⋯, i_j-th columns of $\underset{\sim}{B}$, respectively. The following formula is easily derived:

$$\det \underset{\sim}{J}_k = \det \underset{\sim}{C}_k + \sum_{i \leq k} \zeta_i \det \underset{\sim}{C}_k(i) + \cdots$$

$$+ \sum_{i_1 < i_2 < \cdots < i_j \leq k} \zeta_{i_1} \zeta_{i_2} \cdots \zeta_{i_j} \det \underset{\sim}{C}_k(i_1, i_2, \cdots, i_j)$$

$$+ \cdots + \zeta_1 \zeta_2 \cdots \zeta_k \det \underset{\sim}{C}_k\,(1, 2, \cdots, k). \tag{18}$$

The necessary and sufficient conditions for $J(x)$ in (15) to satisfy the ratio condition (3) without taking absolute values are the following: (i) $\det C_p > 0$ for $p = 1, 2, \cdots, n$, (ii) $\det C_p\ (i_1, \ldots, i_j) \geq 0$ for $p = 1, 2, \cdots, n$, and for any $i_1 < \cdots < i_j \leq p$, and (iii) $\det C_p\ (i_1, \cdots, i_j) > 0$ implies $\det C_{p+1}\ (i_1, \cdots, i_j) > 0$ for $p = 1, 2, \cdots, k-1$ and for $i_1 < \cdots < i_j \leq p$.

Note that the ratio condition is now represented in terms of the signs of constant matrices. Therefore, it is easy to check whether or not the ratio condition is satisfied. For example, if B is a nonnegative diagonal matrix and C is of class P, it can be seen that the conditions stated above are automatically satisfied. In this case, $\det C_p$ $(p = 1,\ldots,n)$ are principal minors of $\det C$ and are all positive, and hence condition (i) is satisfied. On the other hand,

$$\det C_p(i_1, \cdots, i_j) = b_{i_1 i_1}\, b_{i_2 i_2} \cdots b_{i_j i_j} \cdot D$$

where D is one of the principal minors of det C and is positive. This, together with the fact that $b_{ii} \geq 0$ $(i = 1, \cdots, n)$, imply that conditions (ii) and (iii) are satisfied.

In the case of $n = 1$, conditions (i) and (ii) are equivalent to the relations $b_{11} \geq 0$ and $c_{11} > 0$. Hence, the necessity and sufficiency follow immediately because $\zeta_1 > 0$.

To prove the necessity and sufficiency of conditions (i), (ii) and (iii), it is assumed that the statement holds for $n = k-1$. Then, the ratio condition $\det J_k \geq \varepsilon_1 \cdots \varepsilon_k > 0$ requires that the right-hand side of (18) is not less than $\varepsilon_1 \cdots \varepsilon_k > 0$ for any positive values of $\zeta_1, \ldots, \zeta_k$. Therefore, there follow $\det C_k > 0$ and $\det C_k\ (i_1, \cdots, i_j) \geq 0$ for any

$i_1 < \cdots < i_j \le k$. Furthermore, substituting $\zeta_k = 0$ into (18), and comparing it with

$$\det \underset{\sim}{J}_{k-1} = \det \underset{\sim}{C}_{k-1} + \sum_{i \le k-1} \zeta_i \det \underset{\sim}{C}_{k-1}(i) + \cdots$$

$$+ \sum_{i_1 < \cdots < i_j \le k-1} \zeta_{i_1} \zeta_{i_2} \cdots \zeta_{i_j} \det \underset{\sim}{C}_{k-1}(i_1, i_2, \ldots, i_j)$$

$$+ \cdots + \zeta_1 \zeta_2 \cdots \zeta_{k-1} \det \underset{\sim}{C}_{k-1} (1, 2, \ldots, k-1)$$

it is easily seen that condition (iii) is necessary to guarantee the existence of a positive lower bound for $\det \underset{\sim}{J}_k / \det \underset{\sim}{J}_{k-1}$.

The sufficiency is also easily seen by noting that the value of $\det \underset{\sim}{J}_k$ is not less than its value for $\zeta_k = 0$ due to condition (ii). In this case, the ratio of $\det \underset{\sim}{J}_k \big|_{\zeta_k = 0}$ to $\det \underset{\sim}{J}_{k-1}$ is a ratio of one polynomial in $\zeta_1, \cdots, \zeta_{k-1}$ with nonnegative coefficients, $\det \underset{\sim}{C}_k + \cdots + \sum d^{(1)}_{i_1 \cdots i_j} \zeta_{i_1} \cdots \zeta_{i_j} + \cdots$, to another polynomial of the same kind, $\det \underset{\sim}{C}_{k-1} + \cdots + \sum d^{(2)}_{i_1 \cdots i_j} \zeta_{i_1} \cdots \zeta_{i_j} + \cdots$. It is easy to see from conditions (i) and (iii) that there exists a positive constant $\varepsilon_k > 0$ for which

$$\det \underset{\sim}{C}_k \big/ \det \underset{\sim}{C}_{k-1} \ge \varepsilon_k$$

and

$$d^{(1)}_{i_1 \cdots i_j} \big/ d^{(2)}_{i_1 \cdots i_j} \ge \varepsilon_k$$

for all $i_1, \cdots, i_j$. Therefore, the ratio is not less than $\varepsilon_k > 0$ regardless of the values of $\zeta_1, \cdots, \zeta_{k-1}, \zeta_k$. This completes the proof of the statement concerning this example.

Remark 3. As pointed out, the ratio condition is needed to guarantee that the range of y_k covers the whole real line, hence $\underset{\sim}{f}$ is an onto mapping. Based on this, one can state the following corollary of Theorem 1, which is of the Sandberg-Wilson's type [9]: Let $\underset{\sim}{f}$ be of the form

$$\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{F}(\underset{\sim}{x}) + \underset{\sim}{A}(\underset{\sim}{x}) \tag{19}$$

where $\underset{\sim}{F}(\cdot) \varepsilon \mathcal{F}^n$, and in addition each component of $\underset{\sim}{F}$ is continuously differentiable and admits a positive slope everywhere. Then, for any $\underset{\sim}{y} \varepsilon R^n$ there exists one and only one solution of eq. (19) if $\underset{\sim}{A}(\underset{\sim}{x})$ is continuously differentiable and the Jacobian $\partial\underset{\sim}{A}/\partial\underset{\sim}{x}$ belongs to the class P_0 for all $\underset{\sim}{x}$.

It is clear that the Jacobian matrix $\partial\underset{\sim}{f}/\partial\underset{\sim}{x}$ of eq. (15) never vanishes, and hence the cofactor condition is satisfied. The nature of $\underset{\sim}{F}(\cdot)$ guarantees that $\underset{\sim}{f}(\cdot)$ is an onto mapping, and this replaces the ratio condition. Sandberg and Wilson considered the case where $\underset{\sim}{A}(\underset{\sim}{x})$ is a linear function as in eq. (2) (Theorem 6 of [9]).

Remark 4. The following corollary is also immediate: Let $\underset{\sim}{f}$ be of the form (19), where the i-th component of $\underset{\sim}{F}(\cdot)$ is a continuously differentiable function of x_i, which maps R^1 onto itself. Then, for any $\underset{\sim}{y} \varepsilon R^n$ there exists one and only one solution of eq. (1) if $\underset{\sim}{A}(\underset{\sim}{x})$ is continuously differentiable and the Jacobian matrix $\partial\underset{\sim}{A}/\partial\underset{\sim}{x}$ belongs to the class P for all $\underset{\sim}{x}$.

In these two corollaries, det $\underset{\sim}{J}$ never vanishes, and hence C^1-diffeomorphism is guaranteed.

Theorem 1 can be further generalized. By applying interchange of rows and of columns to the Jacobian matrix, the following theorem can be easily proven.

<u>Theorem 2</u>. Let $\underset{\sim}{f}$ be a continuously differentiable mapping of R^n into itself. Then, for any $\underset{\sim}{y} \in R^n$ there exists one and only one solution of eq. (1) if there exist two permutations $(i_1, i_2, \cdots, i_n)$ and $(j_1, j_2 \cdots, j_n)$ of $(1, 2, \cdots, n)$, and there exist positive constants $\varepsilon_1 > 0$ $\varepsilon_2 > 0, \cdots, \varepsilon_n > 0$ for which

$$\left| \det \underset{\sim}{J} \begin{pmatrix} i_1 \\ j_1 \end{pmatrix} \right| \geq \varepsilon_1, \quad \left| \frac{\det \underset{\sim}{J} \begin{pmatrix} i_1, & i_2 \\ j_1, & j_2 \end{pmatrix}}{\det \underset{\sim}{J} \begin{pmatrix} i_1 \\ j_1 \end{pmatrix}} \right| \geq \varepsilon_2, \cdots$$

$$\left| \frac{\det \underset{\sim}{J} \begin{pmatrix} i_1, & i_2, & \cdots, & i_n \\ j_1, & j_2, & \cdots, & j_n \end{pmatrix}}{\det \underset{\sim}{J} \begin{pmatrix} i_1, & i_2, & \cdots, & i_{n-1} \\ j_1, & j_2, & \cdots, & j_{n-1} \end{pmatrix}} \right| \geq \varepsilon_n \quad \text{for all } \underset{\sim}{x} \in R^n \tag{20}$$

where $\underset{\sim}{J} \begin{pmatrix} i_j, & \cdots, & i_k \\ j_1, & \cdots, & j_k \end{pmatrix}$ is composed of the i_1-th, $\cdots$, i_k-th rows and j_1-th, $\cdots$, j_k-th columns of the Jacobian matrix $\underset{\sim}{J}$, in these orders.

This theorem allows the use of a set of non-principal cofactors in the ratio condition, thus enlarges applicability.

<u>Remark 5</u>. Consider the special case of eq. (1):

$$\underset{\sim}{f}(\underset{\sim}{x}) = \underset{\sim}{A}\underset{\sim}{x} + \underset{\sim}{w} = \underset{\sim}{y} \tag{21}$$

where $\underset{\sim}{A}$ is a constant matrix and $\underset{\sim}{w}$ is a constant vector. The ratio condition in (20) of Theorem 2 is equivalent to the condition $\det \underset{\sim}{A} \neq 0$. This can be seen as follows: For the constant matrix case, $\det \underset{\sim}{J} = \det \underset{\sim}{A} \neq 0$ implies that there exists at least one cofactor, $\det \underset{\sim}{J} \begin{pmatrix} i_1, & \cdots, i_{n-1} \\ j_1, & \cdots, j_{n-1} \end{pmatrix}$

which is nonzero, for otherwise det $\underset{\sim}{J}$ would be zero. Thus, the last condition in eq. (20),

$$\left| \det \underset{\sim}{J} \begin{pmatrix} i_1, & \cdots, & i_n \\ j_1, & \cdots, & j_n \end{pmatrix} \Big/ \det \underset{\sim}{J} \begin{pmatrix} i_1, & \cdots, & i_{n-1} \\ j_1, & \cdots, & j_{n-1} \end{pmatrix} \right| \geq \varepsilon_n > 0$$

is automatically satisfied. The same argument can be used to the second to the last condition in eq. (20), etc. Therefore, the statement in Theorem 2 represents the necessary and sufficient condition for the existence of a unique solution of eq. (21) for all $\underset{\sim}{y} \in R^n$.

Remark 6. The ratio conditions in Theorems 1 and 2 have a network interpretation. Consider an n-port resistive network with port current and voltage variables defined as follows: $y_1 = i_1$, $y_2 = i_2$, $\cdots$, $y_n = i_n$; $x_1 = v_1$, $x_2 = v_2$, $\cdots$, $x_n = v_n$. The ratio condition in Theorem 1 can then be interpreted in terms of n specific driving-point characteristics:

$$\det \underset{\sim}{A}_1 = \left. \frac{\Delta i_1}{\Delta v_1} \right|_{v_2, \cdots, v_n = \text{const.}}$$

$$\frac{\det \underset{\sim}{A}_2}{\det \underset{\sim}{A}_1} = \left. \frac{\Delta i_2}{\Delta v_2} \right|_{\substack{i_1 = \text{const.} \\ v_3, \cdots, v_n = \text{const.}}}$$

.................................,

$$\frac{\det \underset{\sim}{A}_n}{\det \underset{\sim}{A}_{n-1}} = \left. \frac{\Delta i_n}{\Delta v_n} \right|_{i_1, \cdots, i_{n-1} = \text{const.}}$$

Similarly the ratio condition in Theorem 2 can be interpreted in terms of n transfer and/or driving-point characteristics:

$$\det \underset{\sim}{J}\begin{pmatrix} i_1 \\ j_1 \end{pmatrix} = \frac{\Delta i_{i_1}}{\Delta v_{j_1}}\Bigg|_{v_{j_k} = \text{const.},\ k \neq 1}$$

$$\frac{\det \underset{\sim}{J}\begin{pmatrix} i_1, i_2 \\ j_1, j_2 \end{pmatrix}}{\det \underset{\sim}{J}\begin{pmatrix} i_1 \\ j_1 \end{pmatrix}} = \frac{\Delta i_{i_2}}{\Delta v_{j_2}}\Bigg|_{\substack{i_{i_1} = \text{const.} \\ v_{j_k} = \text{const.},\ k \neq 1,2}}$$

....

$$\frac{\det \underset{\sim}{J}\begin{pmatrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{pmatrix}}{\det \underset{\sim}{J}\begin{pmatrix} i_1, \cdots, i_{n-1} \\ j_1, \cdots, j_{n-1} \end{pmatrix}} = \frac{\Delta i_{i_n}}{\Delta v_{j_n}}\Bigg|_{i_{i_k} = \text{const.},\ k \neq n.}$$

III. SOLUTION METHOD

Gersho pointed out [14] that it is an old technique in numerical computation to convert the problem of solving the simultaneous nonlinear equation (1) to the one of minimizing the scalar function

$$g(\underset{\sim}{x}) = \| \underset{\sim}{f}(\underset{\sim}{x}) - \underset{\sim}{y} \|^2. \tag{22}$$

The applicability of the method of steepest descent to this minimization problem and its convergence behavior have been investigated by various authors [14, 15], [17-19]. It is assumed in this section that the function $\underset{\sim}{f}$ satisfies the conditions of Palais' theorem stated in the first section.

Then, $\|g(\underset{\sim}{x})\| \to \infty$ as $\|\underset{\sim}{x}\| \to \infty$ and hence the level sets

$$\{\underset{\sim}{x} : g(\underset{\sim}{x}) \leq c\} \tag{23}$$

are bounded for all c. The gradient of $g(\underset{\sim}{x})$ is given by

$$\nabla g(\underset{\sim}{x}) = 2J^t(\underset{\sim}{x}) \{\underset{\sim}{f}(\underset{\sim}{x}) - \underset{\sim}{y}\} . \tag{24}$$

It is clear from Palais' theorem and eq. (24) that the gradient vanishes if and only if $\underset{\sim}{x}$ is the solution of (1). All these imply that the application of the method of steepest descent yields a convergent algorithm [10, 14, 15].

In this section, the speed of convergence is considered under a fairly general additional condition that $\underset{\sim}{f}$ is continuously twice differentiable. There are available many forms for the method of steepest descent, and for definiteness, Curry's algorithm [17, 19] is selected as one of the representatives. This algorithm generates a sequence of points $\{\underset{\sim}{x}^{(k)}\}_{k = 1, 2, \ldots}$ starting with an arbitrarily chosen initial point $\underset{\sim}{x}^{(1)}$, of which one cycle of iteration is described as follows: (i) Calculate the gradient at $\underset{\sim}{x}^{(k)}$, $\nabla g(\underset{\sim}{x}^{(k)})$. If this vanishes, then $\underset{\sim}{x}^{(k)}$ is the solution to be found, and hence the iteration terminates here. If otherwise, go to the next step. (ii) Evaluate the behavior of a one-dimensional function

$$m_k(\lambda) = g(\underset{\sim}{x}^{(k)} - \lambda\nabla g(\underset{\sim}{x}^{(k)})) \tag{25}$$

for $\lambda \geq 0$. Find the first stationary point at $\lambda = \lambda_k$. Thus, $m_k'(\lambda) < 0$ for $0 \leq \lambda < \lambda_k$ and $m_k'(\lambda_k) = 0$. Set

$$\underset{\sim}{x}^{(k+1)} = \underset{\sim}{x}^{(k)} - \lambda_k \nabla g(\underset{\sim}{x}^{(k)}) \tag{26}$$

and go back to the step (i).

It is clear that the sequence $\{g(x^{(k)})\}$ is strictly monotone decreasing. It is shown later that the convergence of the value of $g(x^{(k)})$ to zero is of a geometric progression. That is, there exists a positive constant α_1 such that $0 < \alpha_1 < 1$ and

$$g(x^{(k+1)}) \leq \alpha_1 g(x^{(k)}) \tag{27}$$

for $k = 1,2, \cdots$. It has to be reminded here that the inverse f^{-1} satisfies a local Lipschitz condition. Let x^* be the solution to be found, $f(x^*) = y$. Since the set of points $f(x^*)$, $f(x^{(1)})$, $f(x^{(2)})$, $\cdots$ is contained in the bounded set $\{z : \|z-y\| \leq \|f(x^{(1)}) - y\|\}$, there exists a positive constant $H > 0$ such that

$$\begin{aligned} \|x^{(k)} - x^*\| &\leq H \; \|f(x^{(k)}) - f(x^*)\| = \\ &= H \; \|f(x^{(k)}) - y\| = H[g(x^{(k)})]^{1/2} \end{aligned} \tag{28}$$

From (27) and (28) there follows

$$\|x^{(k)} - x^*\| \leq \beta \cdot \alpha_2^{k-1} \qquad (k = 1, 2, \cdots) \tag{29}$$

where $\beta = H[g(x^{(1)})]^{1/2} > 0$, and $\alpha_2 = [\alpha_1]^{1/2}$ and hence $0 < \alpha_2 < 1$. This demonstrates that the speed of convergence of $\{x^{(k)}\}$ to the solution x^* is not less than some geometric progression.

It has been well known [15] that the convergence of $\{g(x^{(k)})\}$ to zero and the one of $\{x^{(k)}\}$ to x^* satisfy the relations (27) and (29) if $g(x)$ is a quadratic function with a positive definite Hessian matrix. Goldstein has shown that there exists an index K_0 such that $\| x^{(k+1)} - x^*\|$

$\leq \alpha_2 \| \underset{\sim}{x}^{(k)} - \underset{\sim}{x}^* \|$ for $k \geq K_0$, where $0 < \alpha_2 < 1$, under the assumption that the Hessian matrix is positive definite at the solution $\underset{\sim}{x}^*$ [18]. The Hessian matrix of $g(\underset{\sim}{x})$ at $\underset{\sim}{x}^*$ is $2\underset{\sim}{J}^t(\underset{\sim}{x}^*)\ \underset{\sim}{J}(\underset{\sim}{x}^*)$ and hence is positive definite. Therefore, the fact that the speed of convergence in a neighborhood of $\underset{\sim}{x}^*$ is of a geometric progression follows from Goldstein's result. The importance of (27) is that this holds globally, but not in a neighborhood of $\underset{\sim}{x}^*$. Thus, the speed of a geometric progression is assured from the beginning of the iteration.

The proof of (27) is useful to make an underlying computational problem clear, and hence is included herein. Let S be a level set defined by

$$S = \{\underset{\sim}{x}:\ g(\underset{\sim}{x}) \leq g(\underset{\sim}{x}^{(1)})\} \tag{30}$$

The set S is compact and all the points of the sequence $\{\underset{\sim}{x}^{(k)}\}$ belongs to S. Since the continuous matrix $\underset{\sim}{J}(\underset{\sim}{x})$ is non-singular for all $\underset{\sim}{x} \in S$, it is easy to prove the following assertion: There exist two positive constants γ_1 and γ_2 such that $\gamma_1 > \gamma_2 > 0$ and

$$\gamma_1 \|\underset{\sim}{z}\| \geq \| \underset{\sim}{J}^t(\underset{\sim}{x})\ \underset{\sim}{z}\| \geq \gamma_2 \|\underset{\sim}{z}\| \tag{31}$$

for any $\underset{\sim}{x} \in S$ and for any vector $\underset{\sim}{z}$.

It is seen from (24) and (31) that $\underset{\sim}{x} \in S$ implies

$$\|\nabla g(\underset{\sim}{x})\| \leq 2\ \gamma_1 \|\underset{\sim}{f}(\underset{\sim}{x}^{(1)}) - \underset{\sim}{y}\| \tag{32}$$

the right-hand side of which is a constant denoted by γ_3 hereafter. Now let G be the set of points defined by

$$G = \{\underset{\sim}{x}:\ \|\underset{\sim}{z} - \underset{\sim}{Z}\| \leq \gamma_3 \text{ for some } \underset{\sim}{Z} \in S\} \tag{33}$$

then the set G is a compact set containing the set S. By making use of the fact that $g(\underset{\sim}{x})$ is continuously twice differentiable it is easy to prove that there exists a positive constant $K > 0$ for which the relation

$$\|\nabla g(\underset{\sim}{\xi}) - \nabla g(\underset{\sim}{\eta})\| \leq K\|\underset{\sim}{\xi} - \underset{\sim}{\eta}\| \tag{34}$$

holds for any $\underset{\sim}{\xi} \in G$ and any $\underset{\sim}{\eta} \in G$.

The following relation is readily derived from (25):

$$\begin{aligned} m_k'(\lambda) &= - \langle \nabla g(\underset{\sim}{x}^{(k)} - \lambda \nabla g(\underset{\sim}{x}^{(k)})), \nabla g(\underset{\sim}{x}^{(k)}) \rangle = \\ &- \| g(\underset{\sim}{x}^{(k)})\|^2 - \langle \nabla g(\underset{\sim}{x}^{(k)} - \nabla g(\underset{\sim}{x}^{(k)}) - \nabla g(\underset{\sim}{x}^{(k)}), \\ &\nabla g(\underset{\sim}{x}^{(k)})\rangle \end{aligned} \tag{35}$$

where the notation $\langle \cdot, \cdot \rangle$ denotes the usual scalar product. Then there follows from (24), (31) and (35)

$$\begin{aligned} m_k'(0) &= - \|\nabla g(\underset{\sim}{x}^{(k)})\|^2 = - 4 \|\underset{\sim}{J}^t(\underset{\sim}{x}^{(k)}) \{\underset{\sim}{f}(\underset{\sim}{x}^{(k)}) - \underset{\sim}{y}\}\|^2 \\ &\leq - 4 \gamma_2^2 \|\underset{\sim}{f}(\underset{\sim}{x}^{(k)}) - \underset{\sim}{y}\|^2 = - 4 \gamma_2^2 g(\underset{\sim}{x}^{(k)}). \end{aligned} \tag{36}$$

If $0 \leq \lambda \leq 1$, then $\|\lambda \nabla g(\underset{\sim}{x}^{(k)})\| \leq \gamma_3$ from (32), and hence

$$\underset{\sim}{x}^{(k)} \in S \subset G \text{ and } \underset{\sim}{x}^{(k)} - \lambda \nabla g(\underset{\sim}{x}^{(k)}) \in G. \tag{37}$$

Therefore from (34)

$$\|\nabla g(\underset{\sim}{x}^{(k)} - \lambda \nabla g(\underset{\sim}{x}^{(k)})) - \nabla g(\underset{\sim}{x}^{(k)})\| \leq K\lambda \| \nabla g(\underset{\sim}{x}^{(k)})\| \tag{38}$$

for $0 \leq \lambda \leq 1$. Combining this with (35) and applying Schwarz inequality, the following relation is obtained:

$$m_k'(\lambda) \leq - \|\nabla g(\underset{\sim}{x}^{(k)})\|^2 + K\lambda \|\nabla g(\underset{\sim}{x}^{(k)})\|^2 \tag{39}$$

for $0 \leq \lambda \leq 1$. Hence the relation

$$0 \leq \lambda \leq \lambda_o = \text{minimum } (1, 1/_{2K}) \tag{40}$$

implies

$$m_k'(\lambda) \leq -\frac{1}{2} \|\nabla g(\underset{\sim}{x}^{(k)})\|^2 < 0 \tag{41}$$

from which there follows

$$0 < \lambda_o < \lambda_k \tag{42}$$

Then, from (25), (36), (41) and (42)

$$\begin{aligned} g(\underset{\sim}{x}^{(k)}) - g(\underset{\sim}{x}^{(k+1)}) &= m_k(0) - m_k(\lambda_k) \geq \\ &\geq m_k(0) - m_k(\lambda_o) = -\int_0^{\lambda_o} m_k'(\lambda)\, d\lambda \geq \\ &\geq \frac{\lambda_o}{2} \|\nabla g(\underset{\sim}{x}^{(k)})\|^2 \geq 2\lambda_o\, \gamma_2^2\, g(\underset{\sim}{x}^{(k)}) \end{aligned} \tag{43}$$

where

$$0 < 2\lambda_o\, \gamma_2^2 < 1. \tag{44}$$

Therefore,

$$g(\underset{\sim}{x}^{(k+1)}) \leq (1 - 2\lambda_o \gamma_2^2)\, g(\underset{\sim}{x}^{(k)}) \tag{45}$$

for $k = 1, 2, \cdots$, and the constant α_1 in (27) is $1 - 2\lambda_o \gamma_2^2$. Thus, the relation (27) has been established.

Due to the desirable speed of convergence, the application of the method of steepest descent seems very attractive. However, difficulties are often encountered in the one-dimensional search problem of finding the first stationary point of $m_k(\lambda)$ [20]. In this respect, it might be preferable to fix the value of λ instead of finding λ_k in each cycle by a search procedure [18-20]. This is possible if a number $\tilde{\lambda}$ in place of the unknown λ_o can be found so that $0 < \tilde{\lambda} \leq \lambda_o$. If this is the case, the one-dimensional search problem can be eliminated by taking $\lambda_k = \tilde{\lambda}$ for $k = 1, 2, \cdots$, and furthermore the beautiful property of convergence of a geometric progression is preserved. The determination of $\tilde{\lambda}$, however, might cause another computational difficulty. The present authors will provide another iterative solution method based on piecewise-linear models of physical networks in a forthcoming paper [21].

REFERENCES

[1] C. A. Desoer and J. Katzenelson, "Nonlinear RLC Networks", Bell Sys. Tech. J., vol. 44, pp. 161-198; January, 1965.

[2] E. S. Kuh, "Representation of Nonlinear Networks", 1965 Proc. NEC, vol. 21, pp. 702-706; 1965.

[3] P. P. Varaiya and R. W. Liu, "Normal Form and Stability of a Class of Coupled Nonlinear Networks," IEEE Trans. on the Circuit Theory Group, vol. CT-13, pp. 413-418; December, 1966.

[4] T. Ohtsuki and H. Watanabe, "State-Variable Analysis of RLC Networks Containing Nonlinear Coupling Elements," IEEE Trans. on the Circuit Theory Group, vol. CT-16, pp. 26-38; February, 1969.

[5] T. E. Stern, "Theory of Nonlinear Networks and Systems," Addison-Wesley, Reading, Mass.; 1965.

[6] R. J. Duffin, "Nonlinear Networks (I, II, and IIb)," Bull. Amer. Math. Soc., vols. 52, 53, and 54, pp. 836-838, 963-971 and 119-127; 1946, 1947 and 1948.

[7] L. O. Chua and R. A. Rohrer, "On the Dynamic Equations of a Class of Nonlinear RLC Networks", IEEE Trans. on the Circuit Theory Group, vol. CT-12, pp. 475-489; December, 1965.

[8] A. N. Wilson, Jr., "On the Solutions of Equations for Nonlinear Resistive Networks," Bell Sys. Tech. J., vol. 47, pp. 1755-1773; October, 1968.

[9] I. W. Sandberg and A. N. Wilson, Jr., "Some Theorems on Properties of DC Equations of Nonlinear Networks," Bell Sys. Tech. J., vol. 48, pp. 1-34; January, 1969.

[10] I. W. Sandberg, "Theorems on the Analysis of Nonlinear Transistor Networks," Bell Sys. Tech. J., vol. 49, pp. 95-114; January, 1970.

[11] R. S. Palais, "Natural Operations on Differential Forms," Trans. Amer. Math. Soc., vol. 92, pp. 125-141; July, 1959.

[12] C.A. Holtzmann and R. W. Liu, "On the Dynamical Equations of Nonlinear Networks with n-Coupled Elements," 1965 Proc. of the 3rd Annual Allerton Conference on Circuit and System Theory, University of Illinois, pp. 533-545; 1965.

[13] M. Fiedler and V. Ptak, "Some Generalizations of Positive Definiteness and Monotoricity," Numerische Mathematik, vol. 9, pp. 163-172; 1966.

[14] A. Gersho, "Solving Nonlinear Network Equation using Optimization Techniques," Bell Sys. Tech. J., vol. 48, pp. 3135-3138; November, 1969.

[15] T. L. Saaty and J. Bram, "Nonlinear Mathematics," McGraw-Hill, New York, N. Y.; 1964.

[16] E. S. Kuh and I. Hajj, "Nonlinear Circuit Theory: Resistive Networks," to appear in Proc. IEEE.

[17] H. B. Curry, "The Method of Steepest Descent for Nonlinear Minimization Problems," Quart. Appl. Math., vol. 2, pp. 258-261; 1944.

[18] A. A. Goldstein, "Minimizing Functionals on Hilbert Space," Computing Methods in Optimization Problems, Academic Press, New York, N. Y., pp. 159-165; 1964.

[19] A. A. Goldstein, "On the Steepest Descent," J. SIAM Control, Ser. A, vol. 3, pp. 147-151; 1965.

[20] J. Kowalik and M. R. Osborne, "Methods of Unconstrained Optimization Problems," Elsevier, New York, N. Y.; 1968.

[21] T. Fujisawa and E. S. Kuh, "Piecewise-Linear Theory of Nonlinear Networks," to appear.