Numer. Math. 54, 271–293 (1988)

Numerische Mathematik

On the Numerical Solution of the Equation $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2}-\left(\frac{\partial^2 z}{\partial x\,\partial y}\right)^2=f$ and Its Discretizations, I

V.I. Oliker* and L.D. Prussner*

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

Summary. The equation indicated in the title is the simplest representative of the class of nonlinear equations of Monge-Ampère type. Equations with such nonlinearities arise in dynamic meteorology, geometric optics, elasticity and differential geometry. In some special cases heuristic procedures for numerical solution are available, but in order for them to be successful a good initial guess is required. For a bounded convex domain, nonnegative f and Dirichlet data we consider a special discretization of the equation based on its geometric interpretation. For the discrete version of the problem we propose an iterative method that produces a monotonically convergent sequence. No special information about an initial guess is required, and to initiate the iterates a routine step is made. The method is self-correcting and is structurally suitable for a parallel computer. The computer program modules and several examples are presented in two appendices.

Subject Classification: AMS (MOS): 65N05 and 65G10; CR: G1.8.

In Euclidean space R^2 with Cartesian coordinates x, y consider a bounded domain Ω, a nonnegative function f: $\Omega \to [0, \infty)$, and a continuous function ϕ: $\partial\Omega \to R$. The purpose of this work is to investigate a numerical method for solving the nonlinear Dirichlet problem

$$M(z) \equiv \frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2}-\left(\frac{\partial^2 z}{\partial x\,\partial y}\right)^2=f \quad \text{in} \quad \Omega, \tag{0.1}$$

$$z|_\Gamma=\phi, \qquad \Gamma=\partial\Omega. \tag{0.2}$$

The Eq. (0.1) is perhaps the most simple representative of the class of nonlinear equations of Monge-Ampère type (see [8], p. 324). Such equations have been studied by many authors ([1, 4, 6, 7, 11, 12, 14 and 15]) mainly in connec-

* The work of both authors was supported by AFOSR Grants # 83-0319 and 84-0285

tion with problems of existence and uniqueness of surfaces with prescribed metric or curvature functions. However, they have also other important applications. In particular, the leading term in the "balance equation" in dynamic meteorology has the form (0.1) [9, 10]. In a more complicated setting the operator M appears in the von Karman system of equations in elasticity [19] and also in geometrical optics [22].

It has been established that the problem (0.1), (0.2) admits a unique smooth solution under appropriate smoothness assumptions on the data; see Caffarelli-Nirenberg-Spruck [6] and further references there. Existence of generalized (non-smooth) solutions has been established by Bakelman [4] and Aleksandrov [1]. On the other hand we could not find in the literature any convergent numerical procedures for solving equations of Monge-Ampère type. A heuristic approach was suggested by meteorologists in [3, 18], F.G. Shuman; On solving the balance equations. Private communication, April 1982. It is based on a finite difference approximation and linearization and requires a good initial guess. If the latter is not available this method will in general produce a divergent sequence of iterations. An example of such a situation is given here at the end of Appendix B.

Our solution of (0.1), (0.2) is based on a classical approach of Minkowski and Aleksandrov (cf. [1, 15]). For the convenience of the reader we briefly outline it here. More details are given in Sects. 1 and 2.

Let $z \in C^2(\Omega)$ and satisfy the Eq. (0.1) with $f>0$ and continuous. Then the function z is either convex or concave since $M(z)$ is the determinant of the matrix of $d^2 z$. For definiteness, we discuss below only concave functions. Introduce a Euclidean plane P with Cartesian coordinates p, q and consider the map $\gamma(x, y) \rightarrow \left(p=\frac{\partial z}{\partial x}, q=\frac{\partial z}{\partial y}\right)$. Since z is smooth and concave, γ is a diffeomorphism and the operator M is its Jacobian. Then for any Borel subset $\omega \subset \Omega$ we obtain with the use of the equation (0.1)

$$\mu(\omega) \equiv \int_{\omega} f\, dx\, dy = \int_{\omega} M(z)\, dx\, dy = \int_{\gamma(\omega)} dp\, dq \equiv \nu_z(\omega).$$

The function $\mu(\omega)$ is a completely additive measure on Borel subsets $\omega \subset \Omega$, and $\nu_z(\omega)$ is the area of the image $\gamma(\omega)$. Thus, the solution z must satisfy the equation in measures

$$\mu(\omega) = \nu_z(\omega), \qquad \omega \subset \Omega. \tag{0.3}$$

Since the measure $\nu_z(\omega)$ can be defined for nonsmooth concave functions, this identity allows us to widen the class of admissible solutions to include arbitrary concave functions. This is the class in which the numerical solution of the problem is constructed.

In order to define ν_z for a nonsmooth concave z we consider again the map γ. By definition $\gamma(\bar{x}, \bar{y}) = (p, q) \in P$, where p, q are the angular coefficients of the tangent plane

$$z - z(\bar{x}, \bar{y}) = p(x-\bar{x}) + q(y-\bar{y})$$

to the graph S_z at the point $(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}))$. If z is an arbitrary concave function then at each point of its graph there exists at least *one supporting plane, that is, a plane that leaves S_z in one of the halfspaces determined by it* [5]. Now instead of the map γ we consider the map $\tilde{\gamma}$ which associates with each $(\bar{x}, \bar{y}) \in \Omega$ the set of vectors $(p, q) \in P$ of angular coefficients of all supporting planes to S_z at $(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}))$. For a Borel set $\omega \subset \Omega$, $\tilde{\gamma}(\omega)$ is measurable [5] and we put $v_z(\omega) = \text{area of } \tilde{\gamma}(\omega)$. In this setting the right-hand side of (0.3) is defined for any concave function on Ω, and such a function is called a *generalized* solution of (0.1), (0.2) if it satisfies (0.3) and (0.2).

Next, the measure μ on the left of (0.3) is approximated by measures concentrated at a finite number of points inside Ω. For a point-concentrated measure μ and appropriately discretized boundary condition (0.2) the problem (0.3) is finite dimensional and its solution can be found among concave piecewise linear functions. This solution is unique. Under some restrictions this process generates a sequence of concave piecewise linear functions converging (in the $C^{0,1}$-norm) to the generalized solution of (0.3), (0.2).

The discretization step results in a nonlinear algebraic system to be solved in the class of concave functions and to which the known numerical methods do not apply. It is interesting to note that the operator of the discrete version of (0.3) is an M-function as defined by Rheinboldt [16]. However, the surjectivity property required for global convergence of iterative methods in [16, 17] is not available in our case, and these general results are not applicable.

The main objective of this paper is to present a convergent numerical method for solving the discrete variant of (0.3), (0.2). On Fig. 0 we illustrate it in the simplest case when the measure μ is concentrated at one point and the boundary data is homogeneous. Our construction is motivated by a remarkable geometric idea due to Pogorelov [14, 15].

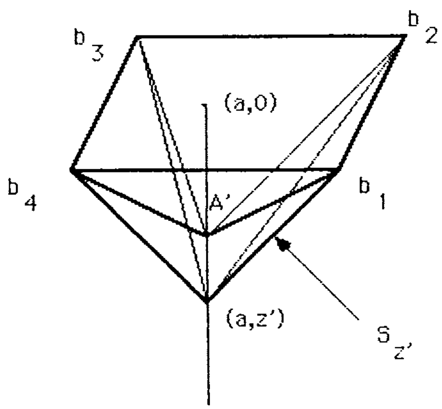

Fig. 0

Suppose $\bar{\Omega}$ is a rectangle as on Fig. 0, $a \in \Omega$, and for $\omega \subset \Omega$ $\mu(\omega) = \mu(a) > 0$ if $a \in \Omega$ and $\mu(\omega) = 0$ if $a \notin \omega$.

We want to solve (0.3) for a concave piecewise linear function z vanishing on $\partial\Omega$. Put $A = (a, 0)$. As the point A slides down to a position A' we construct convex hulls of the points b_1, b_2, b_3, b_4 and A'. For each of these convex hulls, its boundary with the top Ω removed defines a piecewise linear concave function z' for which $v_{z'}(\omega)=v_{z'}(a)$ if $a\in\omega$ and $v_{z'}(\omega)=0$ if $a\notin\omega$.

The first approximation is $z'\equiv 0$. In this case $v_{z'}(a)=0$, but as A slides down to $-\infty$, the image $\tilde{\gamma}(a)$ extends over the entire plane P and correspondingly $v_{z'}(a)$ changes continuously and monotonically from 0 to ∞. On the figure A' indicates one such intermediate position. Obviously, for some position A' $\mu(a)=v_{z'}(a)$ and the corresponding z' is the solution. In the actual algorithm $v_{z'}(a)$ is represented by a quadratic in $z'(a)$ with coefficients depending on the coordinates of the vertices of Ω. The required position of A' is determined by explicitly solving for $z'(a)$ the equation $\mu(a)=v_{z'}(a)$.

When we deal with a measure $\mu=(\mu_1, \ldots, \mu_n)$ concentrated at $n(>1)$ points $a_1, \ldots, a_n$ and homogeneous boundary data one can also start the procedure with $z\equiv 0$. In this case we first triangulate Ω so that $a_1, \ldots, a_n$ and $b_1, \ldots, b_4$ are the only vertices of this triangulation. Then for each a_i a quadratic, expressing $v_{z'}(a_i)$ in terms of coordinates of adjacent vertices, is constructed. Using this quadratic and μ_i, an admissible downward displacement of $(a_i, 0)$ is determined. Then a convex hull of the resulting set of points and boundary vertices is constructed. The boundary of this convex hull with the top Ω removed is a graph of a piecewise linear concave function z' and this is the updated approximation. It is shown that for this approximation $v_{z'}(a_i)\leqq\mu_i$ for all i. The procedure is now repeated with z' as a start up.

There are several complications in this scheme and a substantial amount of work and a number of modifications are needed in order to turn it into a feasible numerical algorithm.

It should also be pointed out that though this procedure may somewhat resemble the finite difference and finite element schemes, it differs from them in principle. The standard classical schemes are not well suited for dealing with the nonlinearity in (0.1) and the concavity of solution. As a result they do not lead to a convergent sequence of discretizations.

We show that the process described above converges at least linearly. Clearly, it is self-correcting, and there is great flexibility in choosing the starting function. Another important feature of the method is that within each iteration step it uses only the local information pertaining to the nodes adjacent to the one for which the current solution is being improved. This allows a major saving in memory and computer time. It also makes it suitable for realization on a parallel computer. Our main results describing the method and proving its convergence are in Sect. 3.

The most time-consuming step in the proposed method is the construction of portions of convex hulls of points in R^3. We developed our own procedure, suitable for our purposes here, though some general algorithms are available [20]. A comparison shows that our algorithm is simpler than (but not as general as) the one in [20], and the running times for both have the same order of magnitude.

In section 4, we discuss local and global convergence, and, in particular, establish local convergence of the Newton method. In Appendix A, we describe the computer program modules which implement the method. Among these modules we have included a Newton step which achieves quadratic convergence once the current approximation is sufficiently close to the true solution. In

Appendix B we give numerical examples, including some with large gradients. The experience gained in testing several types of examples shows that the method is computationally feasible and is quite effective when used in combination with a Newton-type scheme.

In our subsequent papers, we intend to discuss two-sided approximations of the solution, extensions to more general classes of equations, the solution of the continuous problem, and the questions of convergence of its discretizations.

It is a pleasure to thank Professor P. Waltman for many stimulating discussions we have had with him while working on this paper.

1. Generalized Solutions

1.1. In the following, we consider the problem (0.1), (0.2) with $f \in C(\bar{\Omega})$ and $f \geqq 0$ in Ω. The condition $M(z) \geqq 0$ determines a subset in $C^2(\Omega)$. This subset consists of functions that are concave or convex over Ω, since $M(z)$ is the determinant of the matrix of the second differential of z. In the following we consider only concave functions. The case of convex functions reduces to this one by changing the sign of z.

Concavity can also be defined for nonsmooth functions. Namely, let Ω be a convex domain in R^2 and $z(u) \in C(\bar{\Omega})$, $u=(x, y)$. The function z is concave if for any $u_1, u_2 \in \bar{\Omega}$ and any $\lambda \in [0, 1]$,

$$z[\lambda u_1 + (1-\lambda) u_2] \leqq \lambda z(u_1) + (1-\lambda) z(u_2).$$

The class of continuous concave functions over $\bar{\Omega}$ we denote by W. It is a cone in $C(\bar{\Omega})$. If $z \in W \cap C^2(\Omega)$ then the second differential $d^2 z$ is a nonnegative quadratic form in Ω.

1.2. Consider a function $z \in W$ and its graph $S_z = \{(u, z(u)) \colon u \in \bar{\Omega}\}$. As it was explained in the introduction with each supporting plane at a point $(\bar{u}, z(\bar{u})) \in S_z$ with the equation

$$z - z(\bar{u}) = p(x-\bar{x}) + q(y-\bar{y}) \tag{1.1}$$

we associate the point (p, q) on the auxiliary plane P. Thus, the point $\bar{u}=(\bar{x}, \bar{y})$ is mapped into the point $v=(p, q) \in P$ and we have a map (generally multivalued) $\tilde{\gamma} \colon \Omega \to P$. By continuity, $\tilde{\gamma}$ is also defined on the boundary of S_z.

If ω is a Borel set of Ω, it is known [5] that $\omega' = \tilde{\gamma}(\omega)$ is also a Borel set on P and one may consider the measure

$$v_z(\omega) = \int_{\omega'} dp\, dq.$$

Thus, the following definition is appropriate [1, 4]. *A generalized solution of* (0.1), (0.2) *is a function $z \in W$ such that*

$$v_z(\omega) = \int_{\omega} f\, dx\, dy \qquad \textit{for any Borel subset } \omega \subset \Omega \tag{1.2}$$

and

(1.3) $$z|_{\partial\Omega}=\phi.$$

1.3. For a smooth $z\in W\cap C^2(\Omega)$, (1.2) implies (0.1). In this case f is the Radon-Nikodym derivative of v_z and, because of the continuity of f and d^2z, (0.1) holds everywhere in Ω. Since the main concern of this paper is the discrete version of (0.1), (0.2), we omit the details of this connection here.

1.4. The relation (1.2) is an equation in measures, and can be extended to equations where the right-hand side is a measure concentrated at a finite number of points. Then the general plan for solving (1.2), (1.3) is as follows. For a given function f, partition Ω into n disjoint subdomains Ω_i, pick a point $a_i\in\Omega_i$, and put

$$\mu_i=\int_{\Omega_i} f\,dx\,dy.$$

Then solve for $z_n\in W$ the discrete problem

$$v_{z_n}(\omega)=\sum_{a_i\in\omega}\mu_i,\qquad \text{for any } \omega\subset\Omega,$$

with some appropriate discretization of (1.3). The solution z_n is sought in the class of special concave piecewise linear functions which can be represented by a finite number of parameters (in Sect. 2 below an exact description of this class is given). Thus, the discrete problem becomes finite dimensional. Now let max diam $(\Omega_i)\to 0$. It is known that under certain conditions this process produces a sequence $z_n\in W$, converging to some $z\in W$ and the measures v_{z_n} converge weakly to v_z. Thus, z is the solution of (1.2), (1.3) and, consequently, a generalized solution of (0.1), (0.2). Since the purpose of this paper is to study only numerical solutions of the discretization of (1.2), (1.3) we refer the reader for further details to [1, 4, 15].

2. Formulation of the Discrete Analogue of (0.1), (0.2)

2.1. We now specialize the type of domain and boundary data for which the discrete problem will be formulated and solved.

In the rest of the paper, we consider only bounded convex domains on R^2 whose boundary is a closed polygon with a finite number of edges. Let Ω be such a domain and Γ its boundary. We denote the vertices of Γ by $b_1, \dots, b_N$, $(b_{N+i}=b_i)$. It is assumed that no b_i is an interior point of an edge.

The function ϕ: $\Gamma\to R$ is assumed to be continuous and piecewise linear so that over each edge $[b_i, b_{i+1}]$ it is a linear function. Then the curve $C=\{(u, \phi(u)), u\in\Gamma\}$ is a piecewise linear curve in R^3 and its vertices project (one-to-one) into $b_1, \dots, b_N$.

2.2. Let $\Omega\subset R^2$, ϕ: $\Gamma\to R$ be as in 2.1 and $a_1, \dots, a_n$ arbitrary (but distinct) points inside Ω. Let $z\in W$ and suppose that its graph S_z is a *polyhedral surface,* that is, it can be subdivided into a finite number of convex planar regions (faces). Assume further that

(i) $z|_\Gamma = \phi$,

(ii) the only possible vertices of S_z where there is a strictly supporting plane (intersecting S_z only in the vertex) and which project into interior points of Ω lie on vertical rays at the points $a_1, \ldots, a_n$.

The collection of such functions in W we denote by W_n.

2.3. Let $\mu_1, \ldots, \mu_n$ be given non-negative numbers. The following problem is posed.

(2.1) *Find a function $z \in W_n$ such that $v_z(a_i) = \mu_i$, $i = 1, \ldots, n$.*

It is clear that if such a z is found then $v_z(\omega) = \sum_{a_i \in \omega} \mu_i$ for any Borel set $\omega \subseteq \Omega$. Therefore this problem is just a special case of (1.2), (1.3), if the μ_i are related to a given f as explained in 1.4.

In order to see that (2.1) is indeed a problem of inverting a certain map between domains in n-dimensional Euclidean spaces, note first that any function $z \in W_n$ is entirely defined by its values $z_i = z(a_i)$. This is a well known property of convex polyhedral surfaces. In fact, any function from W_n can be obtained by the following construction. Let, as before, C denote the (relative) boundary of S_z for some arbitrary $z \in W_n$. Consider the portion of the vertical cylinder T whose generating curve is Γ, which is infinite in the direction $z > 0$, and whose boundary curve is C. Then S_z is the lower portion of the boundary of the convex hull H of T and all the points (a_i, z_i), that is, $S_z = (\partial H - T) \cup C$, and it is uniquely defined by the points (a_i, z_i), $i = 1, \ldots, n$, $(b_j, z(b_j))$, $j = 1, \ldots, N$ (see, for example, [2], Ch. 1).

Now to any function $z \in W_n$, we can put in correspondence a vector $\tilde{z} \in R^n$ with components $(z_1, \ldots, z_n)$, where $z_i = z(a_i)$. This correspondence will be a bijection only on the image of W_n in R^n.

In the next subsection we show that $v_z(a_i)$ can be effectively computed in terms of $z_1, \ldots, z_n$. Therefore, (2.1) is the discrete realization of (1.2), (1.3).

2.4. Let S_z be the graph of $z \in W_n$ and $A_i = (u_i, z(u_i))$, $i = 1, \ldots, m = n + N$, the interior and boundary vertices of S_z. Fix a vertex A_i corresponding to an interior point a_i and consider its image $\omega' = \tilde{\gamma}(a_i)$ on the plane P. It is easy to see that ω' can be only one of the following types according to whether A_i is:

(1) an interior point of a face of S_z; then ω' is a point on P;

(2) an interior point of an edge of S_z which is incident to faces contained in distinct planes; then ω' is a closed line segment on P;

(3) a "true" vertex of S_z, that is, a vertex at which there exists a strictly supporting plane to S_z; then ω' is a closed bounded convex figure on P with a piecewise linear polygonal boundary.

Let $V_i = \{A_k, k = 1, \ldots, i-1, i+1, \ldots, m\}$, and $Y_i = \{A_{ij}, j = 1, \ldots, n_i$, the set of vertices from V_i on S_z adjacent to $A_i\}$. The latter means the following.

For (1) we take $A_{i1}, \ldots, A_{in_i}$ to be the vertices in V_i of the largest flat convex polygon Q containing A_i; that is, Q is the face containing A_i.

For (2) we let $Y_i = \{A_{i1}, \ldots, A_{in_i^1}, A_{in_i^1+1}, \ldots, A_{in_i^2}\}$, where $A_{i1}, \ldots, A_{in_i^1}$ are the vertices in V_i of Q_1 and $A_{in_i^1}, A_{in_i^1+1}, \ldots, A_{in_i^2}, A_{i1}$ are the vertices in V_i

of Q_2 where the segment joining A_{i1} and A_{in_i} is the edge E containing A_i and Q_1 and Q_2 are the faces on S_z whose intersection consists only of E.

For (3) A_{ij} will be in Y_i if and only if there is a "true" edge E of S_z joining A_i and A_{ij}. By "true" we mean that E is incident to two faces of S_z which are not contained in the same plane.

It is assumed throughout that the elements of Y_i are enumerated in such a way that correspondingly $\{u_{ij}\}, j=1, \ldots, n_i$, are in counterclockwise order.

The plane α passing through the points A_i, A_{ij} and A_{ij+1} contains a face of S_z and its angular coefficients p_{ij}, q_{ij} in representation (1.1) give a vertex on the boundary (possibly degenerating into a point, of ω'. These coefficients are components of the normal $(p_{ij}, q_{ij}, -1)$ to α and can be found from the coordinates of A_i, A_{ij} and A_{ij+1}.

We introduce notation:

$$A_i=(x_i, y_i, z_i=z(x_i, y_i)), \qquad A_{ij}=(x_{ij}, y_{ij}, z_{ij}=z(x_{ij}, y_{ij})),$$
$$\Delta^i_{km}(\xi, \eta)=(\xi_i-\xi_{ik})(\eta_i-\eta_{im})-(\xi_i-\xi_{im})(\eta_i-\eta_{ik}),$$

where (ξ_i, η_i), (ξ_{ik}, η_{ik}) and (ξ_{im}, η_{im}) are ordered pairs of real numbers.

Since $v_z(a_i)=\int\limits_{\omega'=\gamma(a_i)} dp\,dq \equiv$ area of ω', we find it as

$$\begin{aligned}(2.2)\qquad v_z(a_i)&=\frac{1}{2}\sum_{j=1}^{n_i} p_{ij}q_{ij+1}-p_{ij+1}q_{ij}\\ &=\frac{1}{2}\sum_{j=1}^{n_i}\frac{1}{\Delta^i_{jj+1}(x,y)\,\Delta^i_{j+1\,j+2}(x,y)}\cdot[\Delta^i_{jj+1}(x,z)\,\Delta^i_{j+1\,j+2}(y,z)\\ &\quad-\Delta^i_{jj+1}(y,z)\,\Delta^i_{j+1\,j+2}(x,z)],\end{aligned}$$

where $n_i+1=1$, so that a full cycle is made around the vertex A_i.

2.4.1. *Remark.* If A_i is an interior point of a face on S_z, then it is quite obvious, but also follows from (2.2), that $v_z(a_i)=0$. The same is true if A_i is an interior point of an edge on S_z.

2.5. *With the class* W_n *we associate a map* $g\colon W_n \to R^n$, *where* $g(z)=(g_1(z), \ldots, g_n(z))$, $g_i(z)=v_z(a_i)$. *This is the discrete version of the opertor* M *on* W_n. The problem (2.1) becomes now equivalent to the problem of finding a $z^* \in W_n$ for which $g(z^*)=\mu$, where $\mu=(\mu_1, \ldots, \mu_n)$ is a vector with nonnegative components prescribed in advance.

3. Solution of the System $g(z)=\mu$

The procedure for solving the system

$$(3.1)\qquad g(z)=\mu,\ z\in W_n,$$

is iterative. Before we describe it, we need to establish some properties of the map g.

3.1. It is clear that g is continuous as a function of z in the class W_n.

3.2. Lemma. *Let z and $z' \in W_n$ and $\tilde{z}$ and $\tilde{z}'$ be the corresponding vectors in R^n (see subsection 2.3). Suppose $g(z) - g(z') \geqq 0$. Then $\tilde{z}' - \tilde{z} \geqq 0$ and, therefore, in view of the arguments in 2.3, $z(u) \leqq z'(u)$ for $u \in \bar{\Omega}$. If $g(z) = g(z')$, then $z(u) \equiv z'(u)$.*

3.2.1. *Proof.* Suppose, on the contrary, that for some a_i, $z'_i = z'(a_i) < z_i = z(a_i)$. Since z and z' coincide on the boundary Γ, there exists a closed subdomain $\bar{\Omega}^0 \subset \bar{\Omega}$ (possibly $\bar{\Omega}$ itself) such that $z(u) = z'(u)$ on $\partial \bar{\Omega}^0$ and $z'(u) < z(u)$ for all $u \in \Omega^0$. Let α be a supporting plane to S_z at some point $u \in \Omega^0$. Then it cuts from the graph $S_{z'}$ a compact portion and, therefore, there exists a plane parallel to α supporting to $S_{z'}$. Hence, $v_{z'}(\Omega^0) \geqq v_z(\Omega^0)$. On the other hand, it is clear that there exist planes supporting to the portion of $S_{z'}$ over Ω^0 which are not supporting to the corresponding portion of S_z. Thus, $v_{z'}(\Omega^0) > v_z(\Omega^0)$. But that contradicts the hypothesis that $g(z) \geqq g(z')$. The uniqueness part is now obvious. The lemma is proved.

3.3. Lemma. *There exists a unique $z \in W_n$ such that $g_i(z) = 0$, $i = 1, \ldots, n$.*

3.3.1. *Proof.* Put $B_i = (b_i, \phi(b_i))$ where b_i are vertices on Γ, and consider the convex hull H of the set $\{B_1, \ldots, B_N\}$. Denote by C the polygonal curve $(u, \varphi(u))$, $u \in \Gamma$, and let T be the vertical polyhedral cylinder with generating curve Γ. Clearly $C \subset T$. We claim that $H \cap T = C$. Suppose, on the contrary, that there exists a point $B \in H \cap T$ and $B \notin C$. Let α be a vertical plane passing through B and supporting to T. Obviously α is a supporting plane to T and H. Furthermore, since there exists a neighborhood of B on T free of points of C, the plane α can be tilted towards the interior of T with B remaining in α and in this tilted position, $\alpha \cap C = \emptyset$. If we now move α parallel to itself towards the halfspace containing C, we will obtain a convex set which is smaller than H and contains C. This contradicts the fact that H is a convex hull.

The curve C divides the boundary ∂H of H into two portions. The lower portion (relative to the positive direction of the z-axis in R^3) is a graph of some function $z \in W$ and $z|_\Gamma = \phi$. Let us show that, in fact, $z \in W_n$. It suffices to show that S_z has no strictly supporting planes at points projecting into Ω. But obviously, any such strictly supporting plane would be also strictly supporting to H and then moving it towards the halfspace containing C, we again get a contradiction with the fact that H is a convex hull. Thus, every edge on S_z has its endpoints on the boundary curve C. Consequently, all $g_i(z) = 0$. Uniqueness of z follows from Lemma 3.2.

3.4. Theorem. *Let $z \in W_n$, $g_i(z) \leqq \mu_i$, $i = 1, \ldots, n$, and at least one of the inequalities be strict. Then there is a finite algorithm for defining a unique $z' \in W_n$ for which*

$$z' \leqq z, \; g_i(z') \leqq \mu_i, \text{ and } \sum_{i=1}^{n} g_i(z) < \sum_{i=1}^{n} g_i(z').$$

For convenience of exposition the **proof** is divided into several parts, using the notation of 2.4.

3.4.1. Among all vertices $\{A_i = (u_i, z(u_i))\}_{i=1}^m$ on S_z fix one that corresponds to an interior point a_i with $g_i(z) < \mu_i$. For $\xi \leqq z_i$, we let:

1. $A_i(\xi) = (u_i, \xi)$,

2. $S_i(\xi)$ be the lower portion (in the z-direction) of the boundary of the convex hull of $A_i(\xi)$ and the points of Y_i,

3. $Y_i(\xi)$ be the set of vertices on $S_i(\xi)$ adjacent to $A_i(\xi)$ (note $Y_i(\xi)$ may be different for different values of ξ),

4. $(P_{ij}(\xi), -1)=(p_{ij}(\xi), q_{ij}(\xi), -1)$ be the normal to the plane passing through $A_i(\xi)$, A_{ij} and A_{ij+1}.

As usual, we denote by $\partial S_i(\xi)$ the relative boundary of $S_i(\xi)$.

Observe that $S_i(z_i)$ is a portion of the graph S_z and from properties of convex hulls it follows that $\partial S_i(\xi)=\partial S_i(z_i)$ for any $\xi\in(-\infty, z_i]$.

3.4.2. We will need to use the following properties:

(a) The interval $(-\infty, z_i]$ can be partitioned into a finite number of intervals $\Lambda_s=(\alpha_s, \beta_s]$, $s=1, \ldots, M$, such that $Y_i(\xi)$ remains fixed for all $\xi\in\Lambda_s$. This follows from the fact that if $\xi_2<\xi_1$, then $Y_i(\xi_2)\subset Y_i(\xi_1)$. It is clear that if ξ is pulled sufficiently far down, then $Y_i(\xi)$ will consist only of vertices lying on $\partial S_i(z_i)$.

(b) Consider the intervals $\Lambda_s=(\alpha_s, \beta_s]$ and $\Lambda_{s+1}=(\alpha_{s+1}, \beta_{s+1}]$. If $A_{ij}\in Y_i(\xi)$ for $\xi\in\Lambda_s$ and $A_{ij}\notin Y_i(\xi)$ for $\xi\in\Lambda_{s+1}$ then A_{ij} is an interior point of a face on $S_i(\alpha_s)$, which appeared on $S_i(\alpha_s)$ when two faces on $S_i(\xi)$, $\xi>\alpha_s$, adjacent along an edge, assumed a coplanar position as $\xi\to\alpha_s$. Consequently, for the normals to these faces we have

$$\lim_{\xi\to\alpha_s^+}(P_{ij-1}(\xi)-P_{ij}(\xi))=0. \tag{3.2}$$

(c) Denote by $Y_i^-(\beta_s)(Y_i^+(\alpha_{s+1})\equiv Y_i^-(\beta_{s+1}))$ the adjacency set for $\xi\in(\alpha_s, \beta_s]$ (for $\xi\in(\alpha_{s+1}, \beta_{s+1}]$). The condition (3.2) allows us to determine the maximal increment $\beta_s-\alpha_s$, in ξ for which $Y_i(\xi)$ remains fixed. Indeed, let l_{ij} be the vector joining $A_i(\xi)$ with A_{ij} on $S_i(\xi)$, $\xi\in(\alpha_s, \beta_s]$. Then $P_{ij}(\xi)=(\Delta_{ij+1}(x, y))^{-1}\, l_{ij}\times l_{ij+1}$ (the notation Δ being as in 2.4 and "$\times$" being the vector cross product), and, since $l_{ij}=(x_{ij}-x_i, y_{ij}-y_i, z_{ij}-\xi)$, it is a linear function in ξ. Solving (3.2) for ξ and denoting the solution by ξ_i, we put $\alpha_s=\max \xi_j$, $j=1, \ldots, n_i(\beta_s)$. (Here $n_i(\xi)=|Y_i(\xi)|$.)

(d) On each interval Λ_s we use $Y_i^-(\beta_s)$ to form the coefficients in (2.2) and compute

$$v_\xi(a_i)=g_i'(z_1, \ldots, z_{i-1}, \xi, z_{i+1}, \ldots, z_n)=\frac{1}{2}\sum_{j=1}^{n_i(\xi)} p_{ij}(\xi)\, q_{ij+1}(\xi)-p_{ij+1}(\xi)\, q_{ij}(\xi),$$

where the prime indicates that only a subset of all vertices on S_z is used. Note that g_i' is a quadratic polynomial in ξ.

If $\xi_1>\xi_2$ then by construction the point $A_i(\xi_1)\in S_i(\xi_1)$ lies above the point $A_i(\xi_2)\in S_i(\xi_2)$. Therefore, for any plane supporting to $S_i(\xi_1)$ at $A_i(\xi_1)$, there exists a parallel plane supporting to $S_i(\xi_2)$ at $A_i(\xi_2)$. It is also clear that at $A_i(\xi_2)$ there exist supporting planes for which there are no parallel planes supporting to $S_i(\xi_1)$ at $A_i(\xi_1)$. Hence, the set of normals to planes supporting to $S_i(\xi_2)$ at $A_i(\xi_2)$ contains as a proper subset the set of normals to planes supporting to $S_i(\xi_1)$ at $A_i(\xi_1)$. Therefore, $g_i'(\xi_1)<g_i'(\xi_2)$, that is, g_i' is strictly monotone. It is also evident that the further down the vertex $A_i(\xi)$ is pulled, the more 'vertical' become the supporting planes to $S_i(\xi)$ which contain the faces adjacent to $A_i(\xi)$. Then, from this and the definition of g_i' it follows that $g_i'(\xi)\to\infty$ as $\xi\to-\infty$.

(e) The function g_i' is obviously continuous on each Λ_s. We note that the sets $\{\lim_{\xi \to \beta_s^+} P_{ij}(\xi),\ j=1, \ldots, n_i(\xi)\}$ and $\{\lim_{\xi' \to \beta_s^-} P_{ij}(\xi'),\ j=1, \ldots, n_i(\xi')\}$ are equal. Hence, since $g_i'(\beta_s)$ measures the area on the p,q-plane of the figure bounded by the convex polygon with vertices $\{P_{ij}(\beta_s)\colon j=1, \ldots, n_i(\beta_s)\}$, we may conclude that g_i' is continuous on $(-\infty, z_i]$.

(f) We want to solve the equation

$$g_i'(z_i, \ldots, z_{i-1}, \xi, z_{i+1}, \ldots, z_n) = \mu_i, \tag{3.3}$$

and proceed as follows. Since $g_i(z) < \mu_i$, we consider the graph $S_i(z_i)$, the point $A_i(z_i) \in S_i(z_i)$ and the adjacency set $Y_i(z_i)$. Set $\beta_1 = z_i$, and determine α_1 as described in (c) above. On the interval $(\alpha_1, \beta_1]$ the function g_i' is a quadratic polynomial in ξ and the Eq. (3.3) either admits one solution $\bar{\xi}_i$ in $(\alpha_1, \beta_1]$ or no solution. In the first case we are done. If the second case occurs, we take the adjacency set $Y_i^-(\alpha_1)$ (simply by deleting the appropriate vertex A_{ij}), recompute the coefficients of g_i' and repeat the above procedure. Since g_i' is continuous on $(-\infty, z_i]$ and monotonically increasing as $\xi \to -\infty$, there exists a unique $\bar{\xi}_i$ solving (3.3). As a result, finding $\bar{\xi}_i$ requires solution of only a finite number of quadratic equations.

3.4.3. In the same way we find solutions $\bar{\xi}_i$ for all Eq. (3.3) corresponding to the system $g(z) = \mu$ and for which $g_i(z) < \mu_i$. When for some k $g_k(z) = \mu_k$, we simply set $\bar{\xi}_k = z(u_k)$. The equations are processed consecutively starting with $i=1$. Thus, for each interior point $a_i \in \Omega$, we have the point $(a_i, \bar{\xi}_i)$. The boundary points $(b_j, z(b_j))$ remain fixed in the process of finding $\bar{\xi}_i$, $i=1, \ldots, n$. We take the lower portion of the boundary of the convex hull of the points $(a_i, \bar{\xi}_i)$, $i=1, \ldots, n$, and $(b_j, z(b_j))$, $j=1, \ldots, N$. The resulting graph we denote by $S_{z'}$ corresponding to some function z'. Clearly, $z'(u) \leqq z(u)$, $u \in \bar{\Omega}$, $z'(u) = z(u)$, $u \in \Gamma$, and the only possible vertices of $S_{z'}$ are among the points $(a_i, z'(a_i))$. Thus, $z' \in W_n$. Further, $z'(u) \not\equiv z(u)$ if at least for one i $g_i(z) < \mu_i$.

Let us show that $g_i(z') \leqq \mu_i$, $i=1, \ldots, n$. Consider a vertex $A_i = (a_i, z'(a_i))$ on $S_{z'}$. If $A_i \in S_i(\bar{\xi}_i)$ then any supporting plane to $S_{z'}$ at A_i will also be supporting to $S_i(\bar{\xi}_i)$. Thus

$$g_i(z') \leqq g_i'(z_1, \ldots, z_{i-1}, \bar{\xi}_i, z_{i+1}, \ldots, z_n) = \mu_i .$$

If $A_i \notin S_i(\bar{\xi}_i)$ then it follows from the minimality property of convex hulls that A_i is either an interior point of a face or an edge. In either case $g_i(z') = 0 \leqq \mu_i$.

Finally we note that for any two functions z and $z' \in W_n$ such that $z(u) \geqq z'(u)$, $u \in \bar{\Omega}$ and $z \not\equiv z'$ the property $\sum_{i=1}^{n} g_i(z) < \sum_{i=1}^{n} g_i(z')$ holds. Indeed, since on $\partial\Omega$ $z(u) = z'(u)$, for any plane supporting to S_z there exists a parallel plane supporting to $S_{z'}$. But, obviously, there are planes supporting to $S_{z'}$ for which there are no parallel planes supporting to S_z.

The theorem is now completely proved.

3.5. Theorem. *With the use of the algorithm in Theorem 3.4 one can construct a monotone sequence of functions $z_t \in W_n$, $t = 1, 2, \ldots$, $z_t(u) \geqq z_{t+1}(u)$, converging in $C(\bar{\Omega})$-norm to a unique function $z^* \in W_n$ such that $g(z^*) = \mu$.*

3.5.1. Proof. As the function z_0 one can take, for example, the function constructed in Lemma 3.3. But any function z from W_n for which $g_i(z) \leqq \mu_i$ can be also used. The elements of the sequence $\{z_t\}$ are generated as in 3.4.3 and, if for some t $g_i(z_t) = \mu_i$ $i = 1, \ldots, n$, the procedure stops; otherwise it continues. The result is a monotone (possibly infinite) sequence $z_t \in W_n$, $t = 0, 1, \ldots$, for which $g_i(z_t) \leqq \mu_i$, $i = 1, \ldots, n$. We show now that this sequence is convergent.

Put $\sum_{i=1}^{n} \mu_i = \chi$. We claim that there exists a constant $C > 0$, depending only on χ and boundary values of $z_t(u)(= z_0(u)$ for all such that $|z_t|_\infty = \max_{\bar{\Omega}} |z_t| \leqq C$ for all t.

First of all note that since elements of W_n are piecewise linear functions, $|z_t|_\infty = \max\{\max_i |z_t(a_i)|, \max_j |z_t(b_j)|\}$. If $\max_{\bar{\Omega}} |z_t| = \max_{\Gamma} |z_t|$ the estimate is obvious. Suppose $\max |z_t| = |z_t(a_i)|$ for some i. If $z_t(a_i) \to -\infty$, let z_t' be the function in W_n whose graph is the lower portion of the boundary of the convex hull of $\{B_1, \ldots, B_N, (a_i, z_t(a_i))\}$. As $\sum_{i=1}^{n} g_i(z_t')$ measures the area of the image of Ω under the map $\tilde{\gamma}$ mentioned in 1.2 and as the plane passing through B_j, B_{j+1} and $(a_i, z_t(a_i))$ approaches a vertical plane as t approaches ∞, we may conclude that $\sum_{i=1}^{n} g_i(z_t')$ approaches ∞. However, as $z_t' \geqq z_t$, we know that $\sum_{i=1}^{n} g_i(z_t') \leqq \sum_{i=1}^{n} g_i(z_t)$, which contradicts the assumption $\sum_{i=1}^{n} g_i(z_t) \leqq \chi$. A detailed proof of such a compactness property for a more general class of concave functions can be found, for example, in [14] or [1].

Under such circumstances the monotone sequence of vectors $((z_t)_1, \ldots, (z_t)_n)$, $(z_t)_i = z_t(a_i)$ is bounded and converges to a vector $(z_1^*, \ldots, z_n^*)$. Taking the lower portion of the convex hull of the points (a_i, z_i^*), $i = 1, \ldots, n$, and $(b_j, z_0(b_j))$, $j = 1, \ldots, N$, we obtain a graph S_{z^*} of some function $z^* \in W_n$. Obviously, convergence of this sequence of vectors implies convergence in $C(\bar{\Omega})$-norm of z_t to z^*.

Let us show that for all i $g_i(z^*) = \mu_i$. Fixing i and z_t, we begin by showing that g_i' is Lipschitz over $X = \{\xi\colon z_i^* \leqq \xi \leqq (z_0)_i\}$ with a Lipschitz constant which is independent of the z_t for which g_i' is constructed.

For each A_s, $s = 1, \ldots, M$, let $U_i^s = \{u_{i1}^s, \ldots, u_{in_i(\beta_s)}^s\}$ be the projections on (x, y)-plane of elements of $Y_i^-(\beta_s)$. Next for $(\xi_1, \ldots, \xi_n) \in R^n$ let $(p_{ij}^s(\xi_1, \ldots, \xi_n)$, $q_{ij}^s(\xi_1, \ldots, \xi_n)$, $-1)$ be the normal to the plane passing through (u_{ij}^s, ξ_{ij}^s), $(u_{ij+1}^s, \xi_{ij+1}^s)$ and $A_i(\xi_i)$, where $\xi_{ij}^s = \xi_k$ if $u_{ij}^s = a_k$, and let $k_i^s(\xi_1, \ldots, \xi_n)$ be given by

$$\frac{1}{2} \sum_{j=1}^{n_i(\beta_s)} (p_{ij}^s(\xi_1, \ldots, \xi_n)\, q_{ij+1}^s(\xi_1, \ldots, \xi_n) - p_{ij+1}^s(\xi_1, \ldots, \xi_n)\, q_{ij}^s(\xi_1, \ldots, \xi_n)).$$

Clearly, k_i^s is quadratic in each of its arguments and $k_i^s((z_t)_1, \ldots, (z_t)_{i-1}, \xi, (z_t)_{i+1}, \ldots, (z_t)_n) = g_i'(\xi)$ for $\xi \in A_s$.

Since k_i^s is a polynomial in $\xi_1, \ldots, \xi_n$, k_i^s is Lipschitz over the bounded region X. Thus, in particular,

$$\begin{aligned}&|k_i^s((z_t)_1, \ldots, (z_t)_{i-1}, \xi_1, (z_t)_{i+1}, \ldots, (z_t)_n) - k_i^s((z_t)_1, \ldots, (z_t)_{i-1}, \xi_2, (z_t)_{i+1}, \ldots, (z_t)_n)|\\&\quad = |g_i'(\xi_1) - g_i'(\xi_2)| \leqq M_i^s |\xi_1 - \xi_2|\end{aligned}$$

for ξ_1 and ξ_2 in $\Lambda_s \cap X$ and for some $M_i^s > 0$. As M_i^s depends only on U_i^s and X and as there are only a finite number of choices for U_i^s, we may find $M_i > 0$, independent of z_t, so that $|g_i'(\xi_1) - g_i'(\xi_2)| \leqq M_i |\xi_1 - \xi_2|$ for ξ_1 and ξ_2 in $\Lambda_s \cap X$.

To show that g_i' is Lipschitz over X, we need only note that g_i' is continuous over X and $X = \bigcup_{s=1}^{M} \Lambda_s \cap X$.

Next we use this Lipschitz condition to bound $|\mu_i - g_i(z_t)|$. Let ξ_i be the solution to $g_i'(\xi) = \mu$. As $g_i'((z_t)_i) = g_i(z_t)_i$ and $(z_t)_i \geqq \xi_i \geqq (z_{t+1})_i$, we have

$$|\mu_i - g_i(z_t)| = |g_i'(\xi_i) - g_i'((z_t)_i)| \leqq M_i |\xi_i - (z_t)_i| \leqq M_i |(z_{t+1})_i - (z_t)_i|.$$

Finally as $z^* = \lim_{t\to\infty} z_t$ and, hence, $\lim_{t\to\infty} |(z_{t+1})_i - (z_t)_i| = 0$, we use the continuity of g_i over W_n to conclude that $|\mu_i - g_i(z^*)| = 0$. The theorem is proved.

3.6. Theorem. *Let $z_t \in W_n$, $t = 1, 2, \ldots$, be the sequence of iterates constructed as in Theorems 3.4 and 3.5 and converging to $z^* \in W_n$ where $g(z^*) = \mu$. Then*

$$|z_{t+1} - z^*|_\infty \leqq c_t |z_t - z^*|_\infty, \qquad t = 1, 2, \ldots$$

where constants $c_t < 1$ for all t.

3.6.1. Proof. The inequality is an obvious consequence of the fact that the iterates z_t constructed in the proof of the Theorem 3.4 are monotone in t.

3.7. Remark. The last estimate shows that the method is at least linearly convergent and in that sense is comparable to classical iterative procedures for linear systems. On the other hand finding sharp uniform bounds for c_t in terms of the right hand side μ of the equation remains an open and apparently difficult problem. A different type of estimate involving the right hand side μ is given in Theorem 4.6 in the next section.

4. Properties of the Derivative *Dg* and Local Convergence

4.1. Lemma (cf. [21]). Let $z \in W_n$. Then each function $g_i(z_1, \ldots, z_n)$, $z_k = z(a_k)$, is differentiable.

4.1.1. Proof. Consider the point $A_i = (a_i, z_i)$ on S_z and an adjacent point A_{ij}. Assume that the point A_i is displaced vertically by a distance δ, where $|\delta|$ is sufficiently small, and consider the graph $S_{z'}$ which is the lower portion of the boundary of the convex hull of the interior vertices $A_s = (a_s, z_s)$, $s = 1, \ldots, n$, $s \neq i$, $A_i' = (a_i, z_i + \delta)$, and boundary vertices $B_k = (b_k, z(b_k))$, $k = 1, \ldots, N$. Recall that $g_i(z_1, \ldots, z_n)$ is the area of the set $\tilde{\gamma}_z(a_i)$ on the plane P. We need to compare $g_i(z_1, \ldots, z_n)$ with $g_i(z_1, \ldots, z_{i-1}, z_i + \delta, z_{i+1}, \ldots, z_n)$.

To make the arguments more transparent, assume that P is a horizontal plane passing through the vertex A_i. (From the definitions of P and $\tilde{\gamma}$, this assumption can be always fulfilled.) Consider the set of normals $P_{ij}(\delta)=(p_{ij}(\delta), q_{ij}(\delta), -1)$, originating at A_i, to the supporting planes containing the edge $l_{ij}(\delta)$ joining A_i' with A_{ij}. These normals fill out on P a segment (possibly degenerating into a point) $Q_{ij}(\delta)$. The distance $d_{ij}(\delta)$ from the origin in P (which is assumed to coincide with A_i) to $Q_{ij}(\delta)$ can be found as follows. First of all note that the plane $R_{ij}(\delta)$ containing the $P_{ij}(\delta)$ is perpendicular to the edge $l_{ij}(\delta)$ and, consequently, the segment $Q_{ij}(\delta)$ is perpendicular for all δ to the projection of l_{ij} on the horizontal plane, that is, for all δ, $Q_{ij}(\delta)$ are parallel to each other. The distance $d_{ij}(\delta)=\tan(\alpha(\delta))$ where $\alpha(\delta)$ is the angle between $R_{ij}(\delta)$ and the downward vertical direction. On the other hand $\alpha(\delta)$ is equal to the angle between the edge $l_{ij}(\delta)$ and its projection v_{ij} on the plane $z=0$. Hence, $d_{ij}(\delta) =\frac{z_i+\delta-z_j}{|v_{ij}|}$, where $|v_{ij}|$ is the length of the projection of l_{ij}. Therefore, the change in the area $g_i(z_1, \ldots, z_n)$ due to the parallel displacement of the segment $Q_{ij}(0)$ is

$$\Delta_{ij}=-\operatorname{sgn}(\delta)\tfrac{1}{2}[(|Q_{ij}(\delta)|-|Q_{ij}(0)|)(d_{ij}(\delta)-d_{ij}(0))],$$

where $|Q_{ij}(\delta)|$ denotes the length of the segment. From this, it is clear that up to terms of higher order in $|\delta|$ $\Delta_{ij}=\frac{|Q_{ij}(0)|}{|v_{ij}|}\delta$. The total change in the area $g_i(z_1, \ldots, z_n)$ is $\sum_j \Delta_{ij}$ (up to terms of order higher than $|\delta|$). That means that g_i is a differentiable function in z_i.

It is easy to see that the same arguments apply if instead of A_i the vertex A_k, $k \neq i$, is displaced. The lemma is proved.

4.2. We denote by $Dg(z)$ the derivative map and put

$$Dg(z)\,h=\left\{\sum_{j=1}^{n} G_{ij}(z)\,h_j,\ i=1, \ldots, n,\ h\in R^n\right\}.$$

4.3. Lemma. *Let* $z\in W_n$, *and suppose*

$$g_i(z)>0, \qquad i=1, \ldots, n. \tag{4.1}$$

Then the system

$$\sum_{j=1}^{n} G_{ij}(z)\,h_j=0,\ i=1, \ldots, n \tag{4.2}$$

admits only a trivial solution.

4.3.1. Proof. Suppose the system (4.2) admits a solution $h\neq 0$. Lets s be the index for which $\max_i|h_i|=|h_s|$. We consider the case $h_s>0$. The case $h_s<0$ is treated similarly.

Put $\bar{h}_i=h_s-h_i$, $i=1, \ldots, n$, and let δ be a small positive number. The condition (4.1) means that none of the vertices $A_i=(a_i, z_i)$ on S_z belongs to the convex hull of the remaining vertices (including the ones on the boundary ∂S_z). Therefore, for sufficiently small δ all points $(a_i, z_i+\delta\bar{h}_i)$ will be vertices on the graph

$S_{z'}$ of some function $z' \in W_n$, and this function also satisfies (4.1). For the function z' we have: $z'(u) \geqq z(u)$, $u \in \bar{\Omega}$, $z'(a_s) = z(a_s)$, and, of course, $z'(u) = z(u)$ on $\partial\Omega$.

Consider the convex polyhedral angles Q and Q' with vertex A_s formed by the faces, correspondingly, of S_z and $S_{z'}$ adjacent to A_s. The angle Q is either equal to Q' or contains it. If $Q = Q'$ then S_z and $S_{z'}$ concide along the faces adjacent to A_s and, therefore, for vertices A_i adjacent to A_s we have $\bar{h}_i = 0$. We move then vertex A_s to an adjacent vertex, repeating this process until we come to a vertex where the corresponding angles Q and Q' do not coincide. If such a vertex does not exist, we end the process at a vertex on $\partial S_z = \partial S_{z'}$, and then all $\bar{h}_i = 0$.

Suppose that a vertex at which $Q \neq Q'$ is found. It may be assumed that this vertex is A_s. Since Q must contain Q', $g_s(z') > g_s(z)$. Then applying Lemma 4.1 and taking into account (4.2), we obtain

$$\begin{aligned} 0 < g_s(z') - g_s(z) &= \sum_{j=1}^{n} G_{sj}(z)\, \bar{h}_j + o(\delta) \\ &= \delta h_s \sum_{j=1}^{n} G_{sj}(z) + o(\delta). \end{aligned}$$

On the other hand consider the lower portion $S_{\bar{z}}$ of the boundary of the convex hull of the points $(a_i, z_i + \delta h_s)$, $i = 1, \ldots, n$, and the boundary vertices $(b_j, z(b_j))$, $j = 1, \ldots, N$. Again, for sufficiently small δ $S_{\bar{z}}$ is a graph of a function $\bar{z} \in W_n$ satisfying (4.1). It is easy to see by considering supporting planes to $S_{\bar{z}}$ and S_z at corresponding vertices that for any vertex A_i on S_z for which there is no edge joining A_i with ∂S_z, we have $g_i(\bar{z}) = g_i(z)$ and for all vertices that can be joined by an edge with ∂S_z, $g_i(\bar{z}) < g_i(z)$.

Then

$$g_s(\bar{z}) - g_s(z) = \delta h_s \sum_{j=1}^{n} G_{sj}(z) \leqq 0, \tag{4.3}$$

and we arrive at a contradiction. Therefore, $Q = Q'$, and, consequently, we have all $\bar{h}_i = 0$ and $h_i = h_s$.

Then (4.2) implies that

$$\sum_{j=1}^{n} G_{ij} = 0, \qquad i = 1, \ldots, n. \tag{4.4}$$

An analysis of the proof of Lemma 4.1 shows that under the condition (4.1) $\sum_{j=1}^{n} |G_{ij}| \neq 0$, $i = 1, \ldots, n$. Since in (4.3) the inequalities will be strict for vertices A_s adjacent to ∂S_z, we conclude from this and (4.4) that each $h_s = 0$. The lemma is proved.

4.4. Theorem. *For any $z \in W_n$ for which* (4.1) *is satisfied, the derivative map $Dg(z)$: $R^n \to R^n$ is an isomorphism. Consequently, if z^* is a solution of the system $g(z) = \mu$, where μ is a vector with positive components, then z^* is a point of attraction of the Newton method.*

4.4.1. This theorem follows from the invertibility of $Dg(z)$ for z sufficiently close to z^*, and Theorem 10.2.2 in [13].

4.4.2. Corollary 4.4 and the known results on convergence of Newton's iterations (see [13], Sects. 10.2.2 and 10.2.3) imply that when z^* is a point of attraction, the convergence is quadratic.

4.5. Our next result shows how the size of the right hand side in the system $g(z)=\mu$ affects global convergence. We will need the following

4.5.2. Lemma. *Let $z_1, z_2 \in W_n$, $z_1(u) \geqq z_2(u)$, $u \in \Omega$, and at least at some point in Ω, the inequality is strict. Then $\sum_{i=1}^{n} g_i(z_1) < \sum_{i=1}^{n} g_i(z_2)$.*

The lemma is proved using an argument similar to those in 3.4.2(d).

4.6. Theorem. *Let z^* be a solution from W_n of the system $g(z)=\mu$, where $\mu=(\mu_1, \ldots, \mu_n)$. Let $\{z_s\}$, $s=0, 1, \ldots$, be the approximation of z^* obtained in Theorem 3.5. Then*

$$|z_s - z^*|_\infty \leqq C \sqrt{\sum_{i=1}^{n} \mu_i} \sqrt{1-\gamma_s}, \qquad s=0, 1, \ldots,$$

where $0<\gamma_s \leqq 1$, $\gamma_s < \gamma_{s+1}$, $\gamma_s \xrightarrow[s\to\infty]{} 1$ and C is a constant depending only on the domain Ω.

4.6.1. Proof. Let z_0 be the initial approximation to z^*. If all $g_i(z_0)=0$, then by performing one step of the algorithm in Theorem 3.4 we will obtain a new approximation z'_0 such that $g_i(z'_0) \neq 0$ at least for one i. To simplify the notation we put $z_0 = z'_0$.

From Theorem 3.5 and Lemma 4.5.1 it follows that there exist constants $\alpha, \beta_1, \ldots \beta_s, \ldots$ such that

$$\alpha \sum_{i=1}^{n} g_i(z^*) = \sum_{i=1}^{n} g_i(z_0), \qquad 0<\alpha \leqq 1,$$

where z_0 is the initial approximation, and

$$\beta_s \sum_{i=1}^{n} g_i(z_s) = \sum_{i=1}^{n} g_i(z_{s-1}), \qquad 0<\beta_s \leqq 1, \qquad s=1, 2, \ldots.$$

Put $\gamma_s = \alpha \Big/ \prod_{k=1}^{s} \beta_k$. It is clear from the constructions in 3.4 that $\gamma_s < \gamma_{s+1} < 1$, and $\gamma_s \to 1$. On the other hand, it follows from [21] (see also [12] for simplifications and generalizations) that for $z_s \in W_n$,

$$|z_s - z^*|_\infty \leqq C \sqrt{\sum_{i=1}^{n} g_i(z^*) - \sum_{i=1}^{n} g_i(z_s)},$$

and, consequently,

$$|z_s - z^*|_\infty \leqq C \sqrt{\sum_{i=1}^{n} \mu_i} \sqrt{1-\gamma_s},$$

where the constant C depends on geometric quantities related to the domain Ω. The theorem is proved.

Appendix A

Algorithms

We describe the algorithms used in solving (3.1). These correspond to the method presented in the proof of Theorem 3.4. A Newton step, based on Theorem 4.4, has been added in order to accelerate convergence once a solution close to the true solution has been found.

In these algorithms repeated mention will be made of four arrays. As it is easier on the computer not to distinguish between the vertices on the boundary and those in the interior, the first array is $B=\{b_i(=(u_i, z_i)): i=1, \ldots, J\}$, which are the points for which we wish to find a concave function $\bar{\psi}$: $\Omega \subset R^2 \to R$, $\Omega = \operatorname{conv}(u_1, \ldots, u_J)$, corresponding to $\mu_{M+1}, \ldots, \mu_J$. We further assume that the polygon in R^2 with vertices $u_1, \ldots, u_M$, $M<J$, is convex, contains $u_{M+1}, \ldots, u_J$ and that $u_1, \ldots, u_M$ are given in counterclockwise order.

As $\bar{\psi}$ is piecewise linear, the array PLANE $=\{\text{plane}_1, \ldots, \text{plane}_{\text{plane\_top}}\}$ of integer triples will completely characterize $\bar{\psi}$. For each k plane$_k[1]$, plane$_k[2]$, plane$_k[3]$ give the indices of the b_j which lie on the k-th face of the graph of $\bar{\psi}$.

For each $i=1, \ldots, J$, edge$_i$ begins a linked list with information about those vertices adjacent to b_i. In particular, if p is in the linked list with root edge$_i$ then $p(v2)$ gives the index of an adjacency of b_i and p(next) is the next node on the list.

The array adj$_1, \ldots,$ adj$_J$ contains the information in the previous array in a more convenient form. The number of adjacencies to b_i are given adj$_i$(n_adj) while adj$_i$(list$_1$), …, adj$_i$(list$_{\text{adj}_i(n_\text{adj})}$) are the indices of those adjacencies given in either clockwise or counterclockwise order.

At the end of the description of each algorithm we display the running time. NF stands for the number of faces in S where S is the graph of $\bar{\psi}$, and NA(b_i) gives the number of edges incident to b_i. The graph of S is triangulated so that all faces are recorded by giving three vertices.

Algorithm build_triangulation

Input: B.

Effect: The faces of S are recorded in PLANE as the triples $\{i, j, k\}$ such that b_i, b_j, b_k are the vertices of a face. The edges incident to b_i are recorded as a linked list with root edge$_i$. If the edge $b_i b_j$ is found, it is because the face with vertices b_i, b_j, b_k, for some k, has been found.

Running time: $O(NF^* J)$.

Algorithm find_adj

Input: B and EDGE $=\{\text{edge}_{M+1}, \ldots, \text{edge}_J\}$.

Effect: The array adj$_{M+1}, \ldots,$ adj$_J$ is formed where adj$_i$(n_adj) gives the number of adjacencies to b_i and adj$_i$(list$_0$), …, adj$_i$(list$_{\text{adj}_i(n_\text{adj})}$), adj$_i$(list$_0$)

$=\mathrm{adj}_i(\mathrm{list}_{\mathrm{adj}_i(n\ \mathrm{adj})})$, are the indices of those adjacencies given in either clockwise or counterclockwise order. Also if b_i does not lie on S, then $b_i(z)$ is changed so it does.

Running time: $O(J^* \max_i NA(b_i))$ if every b_i lies on S. If no $b_i (i>M)$ lies on S then the time is no worse than $O(J^* NF)$.

Algorithm newz

In newz() we find new_z such that $\mu_i(\text{new\_z})=\mu_i$, where

$$\mu_i(z)=\tfrac{1}{2}\left|\sum_{j=1}^{\mathrm{adj}_i(n\ \mathrm{adj})}[(DN_j\wedge DN_{j+1})\,z^2+(DN_j\wedge N_{j+1}+N_j\wedge DN_{j+1})\,z + N_j\wedge N_{j+1}]\cdot e_1\wedge e_2\right|.$$

Here

1. $p_j=b_{\mathrm{adj}_i(\mathrm{list}_j)}$,
2. $p=(b_i(x), b_i(y), z)$,
3. $\bar{N}_j=(p_j-p)\times(p_{j+1}-p)$,
4. $N_j=(\bar{N}_j(x), \bar{N}_j(y))/\bar{N}_j(z)|_{z=0}$,
5. $DN_j=\dfrac{\partial}{\partial z} N_j|_{z=0}$,
6. $e_1=(1,0)$ and $e_2=(0,1)$,
7. $(t\wedge u)\cdot(v\wedge w)=t\cdot v^* u\cdot w-t\cdot w^* u\cdot v$.

Input: index i corresponding to an interior element of B.

Output: new_z, the new z-coordinate of b_i. If each point in Y_i, the adjacency set of b_i, lies on S_i=lower portion of the boundary of the convex hull of Y_i and $(u_i, \bar{z}_i)$ where $\bar{z}_i$ solves $\mu_i(z)=\mu_i$, then new_z is $\bar{z}_i$. Otherwise, those elements of Y_i lying above S_i are removed from the adjacency set and $\mu_i(z)$ is redefind. After a finite number of these redefinitions a solution will be found.

Running time: $O(NA(b_i)^2$ is the worst case time. Usually, the time will be $O(NA(b_i))$.

Algorithm newton

Input: B and $\mu_{M+1}, \ldots, \mu_J$.

Effect: If the Newton scheme converges, the z-coordinates of $b_{M+1}, \ldots, b_J$ are changed so that B is the solution corresponding to the μ_i and the number of times necessary is returned. Otherwise, 0 is returned. The iterative step is $\mathrm{new}Z=\mathrm{old}Z+A^{-1}(MU-\mathrm{curr}\,MU)$, where $\mathrm{old}Z_i=b_i(z)$, $\mathrm{curr}\,MU_i=MU_i=0$ for $i\le M$, $MU_i=\mu_i$ and $\mathrm{curr}\,MU_i=\mu_i(\mathrm{old}Z_i)$ for $i>M$, and $A_{ij}=\delta_{ij}$ for i or $j\le M$, $A_{ii}=\dfrac{\partial}{\partial z_i}\mu_i(z)|_{z=b_i(z)}$ for $i>M$ and A_{ij} is computed by $\dfrac{\partial}{\partial z_j}\mu_i(b_i(z))$ for i and $j>M$. Thus, A_{ij} is 0 unless b_j is adjacent to b_i.

To satisfy the hypotheses of Theorem 4.4 it is necessary that all points in B lie on S. However, the curr MU_i corresponding to an interior point b_i lying on S might even so be zero, which would would make Theorem 4.4 inapplicable, and, thus, might mean that A is not invertible. Furthermore, even if A^{-1} exists initially, the initial approximation may not be close enough to the true solution to allow the Newton method to work. Thus, the fact that all points in B lie on S is clearly not sufficient for the success of newton(). Therefore, the newton() returns 0 when:

1. A is not invertible,
2. the error between successive approximations increases,
3. a predetermined number of iterations are completed.

An example of case 2 is given at the end of Appendix B. Otherwise, the algorithm ceases when the error between successive solutions is sufficiently small.

Algorithms build_triangulation() and find_adj() are used in each iteration.

Running time: $O(L^* J^3)$, where L is the number of iterations performed.

Algorithm main

Input: An original approximation B and $\mu_{M+1}, \ldots, \mu_J$.

Effect: Finds the solution B corresponding to $u_1, \ldots, u_J$ and $\mu_{M+1}, \ldots, \mu_J$. So long as elements of B lie above S, the basic steps of each iteration are as follows:

$S1$: first build_triangulation() and find_adj() are called to determine the adjacencies of b_i $i = M+1, \ldots, J$.

$S2$: for each $i = M+1, \ldots, J$ a new z-value for b_i is computed using newz() and stored in $z2_i$. The z-values of the interior vertices are unchanged.

$S3$: at the end the z-value of each $b_i(z)$ is set equal to $z2_i$ and the process is repeated.

After K steps the running time is $O(K^* NF^* J)$. When all elements in B lie on S, newton() is called to accelerate the convergence. If newton() is unsuccessful, the first process is repeated a predetermined number of times to get a better approximation.

Running time: $O(K^* NF^* J) + O(L^* J^3)$, where K is the number of iterations of $S1$–$S3$ performed and L is the number of iterations newton() makes before finding a solution.

Appendix B

Numerical Examples

We present here four examples where the initial domain Ω is a unit disk in R^2 centered at the origin. Two methods are used to find the solution corresponding to $b_1, \ldots, b_n$ and $\mu_{N+1}, \ldots, \mu_n$. The first is the algorithm described in Sect. 3. In the second we proceed as in the first algorithm until $(u_{N+1}, (z_m)_{N+1}), \ldots, (u_n, (z_m)_n)$ lie on S_{z_m} at which time we attempt to use the Newton algorithm from Appendix A to find the true solution.

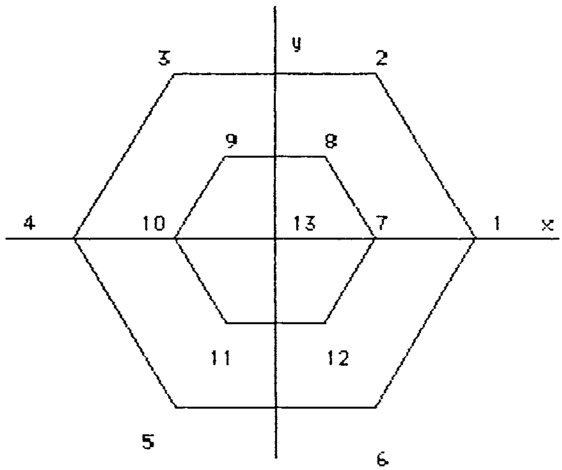

Fig. 1

In Table 1, m denotes the number of runs of the first algorithm, where a run means a complete sweep through all nodes one time. In all examples the initial approximation was constructed using only the data on the boundary. In Table 1 we give the errors $|z_m - \bar{z}|_\infty$ and $|z_m - z_{m-1}|_\infty$ for different number of runs of the first algorithm, where z_m is the approximation after m steps and $\bar{z}$ is the true solution. After Table 1 we discuss what happens when we use the second algorithm.

Example 1. The function is $z=z^2+y^2$. The grid is on Fig. 1. The points 1, 2, 3, 4, 5, 6 lie on unit circle $x^2+y^2=1$, the points 7, 8, 9, 10, 11, 12 are on unit circle $x^2+y^2=\frac{1}{2}$, and the point 13 is the origin $(0,0)$. The two hexagons are homothetic and the angle between two consecutive rays on which the vertices lie is $\pi/3$.

Example 2. The function is $z=\tan(x^2+y^2)$. The grid is the same as in Example 1.

Example 3. The function is $z=x^2+y^2$. The grid is on Fig. 2 There are four homothetic hexagons whose vertices are correspondingly at distance $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ and 1 from the origin. As in Examples 1 and 2, the consecutive rays, on which the vertices lie, form an angle $\pi/3$.

Example 4. The function is $z=\tan(x^2+y^2)$. The grid is the same as in Example 3.
In the Table 2 we give data concerning the second algorithm. The first number, m is the number of times the first algorithm must be iterated so that successive errors are smaller than 5×10^{-6} The second number, p, is the number of runs of the first algorithm before the Newton scheme is attempted. Next is the number K of times the Newton step is repeated to find a solution. The third and fourth numbers are the approximation errors at the time when the Newton scheme begins.

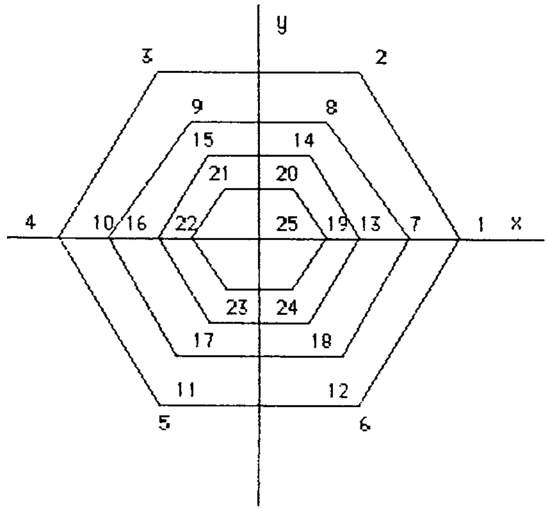

Fig. 2

Table 1

	$m=5$	$m=10$	$m=15$	$m=20$	$m=25$
Example 1:					
$\lvert z_m - z_{m-1}\rvert_\infty$	5.3×10^{-2}	7.4×10^{-3}	1.1×10^{-3}	1.6×10^{-4}	2.4×10^{-5}
$\lvert z_m - \bar{z}\rvert_\infty$	1.1×10^{-1}	1.6×10^{-2}	2.4×10^{-3}	3.5×10^{-4}	5.2×10^{-5}
Example 2:					
$\lvert z_m - z_{m-1}\rvert_\infty$	8.3×10^{-2}	1.1×10^{-2}	1.6×10^{-3}	2.2×10^{-4}	3.2×10^{-5}
$\lvert z_m - \bar{z}\rvert_\infty$	1.7×10^{-1}	2.3×10^{-2}	3.3×10^{-3}	4.7×10^{-4}	6.8×10^{-5}
	$m=10$	$m=20$	$m=30$	$m=40$	$m=50$
Example 3:					
$\lvert z_m - z_{m-1}\rvert_\infty$	2.7×10^{-2}	6.6×10^{-3}	2.0×10^{-3}	6.0×10^{-4}	1.9×10^{-4}
$\lvert z_m - \bar{z}\rvert_\infty$	1.8×10^{-1}	5.2×10^{-2}	1.6×10^{-2}	4.8×10^{-3}	1.5×10^{-3}
Example 4:					
$\lvert z_m - z_{m-1}\rvert_\infty$	2.9×10^{-2}	5.4×10^{-3}	1.2×10^{-3}	2.6×10^{-4}	5.9×10^{-5}
$\lvert z_m - \bar{z}\rvert_\infty$	1.6×10^{-1}	3.2×10^{-2}	7.2×10^{-3}	1.6×10^{-3}	3.7×10^{-4}

Table 2

	m	p	K	$\lvert z_p - z_{p-1}\rvert_\infty$	$\lvert z_p - \bar{z}\rvert_\infty$
Example 1:	31	2	5	1.9×10^{-1}	3.8×10^{-1}
Example 2:	31	3	5	2.2×10^{-1}	3.7×10^{-1}
Example 3:	82	5	7	6.4×10^{-2}	3.8×10^{-1}
Example 4:	68	8	5	4.5×10^{-2}	2.3×10^{-1}

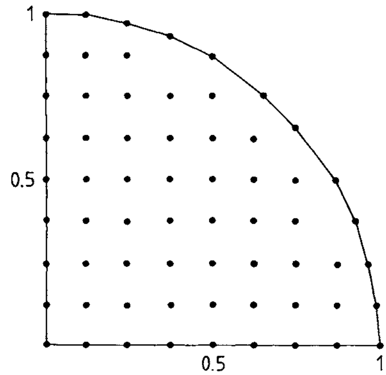

Fig. 3.

To see that the Newton scheme alone will not always find a solution we use the grid whose boundary lies on the unit circle and whose first quadrant is given in Fig. 3. We let μ_i be obtained from the hessian $h(u)$, $u=(x, y)$, of $\tan(x^2+y^2)$ by multiplying $h(u_i)$ by a number representing the area of a polygon ω_i containing u_i. The sets $\omega_1, \ldots, \omega_n$ satisfy the properties $\Omega=\bigcup_i \omega_i$ and $u_j \in \omega_i$ if and only if $i=j$. After 27 runs of the first algorithm, the Newton scheme was begun. After 4 repetitions, however, the Newton scheme began to diverge. Five further runs of the first algorithm produced an approximation to the solution which, when used by the Newton scheme, converged after 6 steps.

References

1. Aleksandrov, A.D.: The Dirichlet problem for the equation $\det(z_{ij})=\phi(z_1, \ldots, z_n, z, x_1, \ldots, x_n)$. I. Vestnik Leningrad Univ. **1** 5–24 (1958) (in Russian)
2. Aleksandrov, A.D.: Convex polyhedra. GITTL, M.-L. 1950 (in Russian); German transl., Berlin, Akademie-Verlag 1958
3. Arnason, G.: A convergent method for solving the balance equation. J. Meteorol. **15** 220–225 (1957)
4. Bakelman, I.: Geometric methods for solving elliptic equations. Nauka 1965 (in Russian)
5. Busemann, H.: Convex surfaces, New York, Interscience Publishers, 1958
6. Caffarelli, L., Nirenberg, L., Spruck, J.: The Dirichlet problem for nonlinear second order elliptic equations. I. Monge-Ampère equation. Comm. Pure Appl. Math. **37** 369–402 (1984)
7. Cheng, S.Y., Yau, S.T.: On the regularity of the Monge-Ampère equation $\det(\partial^2 u/\partial x_i \partial y_j) = F(x, u)$. Comm. Pure Appl. Math. **30** 41–68 (1977)
8. Courant, R., Hilbert, D.: Methods of mathematical physics. Vol. II, New York Interscience Publishers, Wiley 1962
9. Haltiner, G.J.: Numerical weather prediction, New York, Wiley 1971
10. Kasahara, A.: Significance of non-elliptic regions in balanced flows of the tropical atmosphere (preprint). Boulder, Colorado, National Center for Atmospheric Research, 1981
11. Nirenberg, L.: The Weyl and Minkowski problems in differential geometry in the large. Comm. Pure Appl. Math. **6** 337–394 (1953)
12. Oliker, V.: On the linearized Monge-Ampère equations related to the boundary value Minkowski problem and its generalizations. In: Gherardelli, F. (ed.) Monge-Ampère Equations and Related Topics, Proceedings of a Seminar held in Firenze, 1980, pp. 79–112 Roma 1982

13. Ortega, J.M., Rheinboldt, W.C.: Iterative solutions of nonlinear equations in several variables. New York, Academic Press 1970
14. Pogorelov, A.V.: Deformation of convex surfaces. GITTL, M.-L. 1951 (in Russian) (see esp. Ch. II, Sect. 4)
15. Pogorelov, A.V.: The Minkowski multidimensional problem, Moscow, Nauka 1975 (in Russian); Engl. transl., New York, Wiley J. 1978
16. Rheinboldt, W.: On M-functions and their application to nonlinear Gauss-Seidel iterations and to network flows. J. Math. Anal. Appl. **32** 274–307 (1970)
17. Rheinboldt, W.: Methods for solving systems of nonlinear equations. Regional Conference Series in Applied Math. # 14, SIAM, Philadelphia 1974
18. Shuman, F.G.: Numerical methods in weather prediction: I. The Balance equation. Monthly Weather Review **85** 329–332 (1957)
19. Stoker, J.J.: Nonlinear elasticity. Ch. 5. New York, Gordon and Breach 1968
20. Swart, G.: Finding the convex hull facet by facet. Algorithms **6**, 17–48 (1985)
21. Volkov, Y.A.: An estimate for the change of solution of the equation $f(z_1, \ldots, z_n) \det(z_{ij}) = h(x_1, \ldots, x_n)$ in terms of the change of its right hand side. Vestnik Leningrad Univ. **13** 5–14 (1960) (in Russian)
22. Westcott, B.S.: Shaped reflector antenna design. Research Studies Press Ltd., Letchworth, Hertforedshire, England 1983

Received November 30, 1987/August 30, 1988