

Least-squares wave front errors of minimum norm

Jan Herrmann

Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Massachusetts 02175

(Received 11 June 1979)

Requirements for adaptive optics and compensating imaging systems lead to wave front reconstruction problems which we formulate as generalized least-squares problems. For a given array of phase-difference measurements, we construct explicit and exact solutions for the least-squares wave front error. Of particular interest are solutions with minimum norm. Two different discretizations for the gradient are used and the reasons for the different results are given.

INTRODUCTION AND SUMMARY

Adaptive optics and compensating imaging control systems require methods that construct the phase of a wave front from measured gradients of the wave front. The noise in gradient measurements leads to a wave-front error that is related to the variance of the noise by the error propagator. The calculation of the error propagator is the main task of this paper.

We summarize the main points of our calculations. In Sec. I we present some analytical background which can be used to relate the least-squares phase error to the noise. We also point out that the irrotational part of the noise is part of the solution. In Sec. II we formulate two discretizations for the gradients of the phase in terms of an overdetermined system of linear equations. The rank of the matrix of this linear system is of crucial importance. The least-squares minimum-norm solutions are discussed in Sec. III and are used to calculate the mean-square phase error with minimum norm. In Sec. IV we show that many other least-squares solutions can be constructed; they differ from the minimum-norm solution by the amount of piston they contain. In Sec. V we show how to extend the system of linear equations to a nonsingular problem of full rank. The relation to network

theory is discussed in Sec. VI. We show that the minimum-norm solution cannot be realized by a linear and passive network. In Sec. VII we perform least-squares fits with polynomial expansions. We show that a least-squares fit to the phase gives different results from a least-squares fit to the gradient. In Sec. VIII we apply the developed theory to a square array from which the corners were removed. Finally in Sec. IX we analyze the properties of the curl operator and use it to estimate the mean-square gradient error.

I. ANALYTICAL BACKGROUND

In this introduction we present some useful background material in analytical form. The general problem is the determination of the phase of a wave front from its gradient. If the gradient is known and the phase is given at a point $\mathbf{r}_0 = (0,0)$ then the phase at any other point $r = (x,y)$ is given by

$$\phi(\mathbf{r}) = \int_C \nabla\phi \cdot d\mathbf{s} + \phi(\mathbf{r}_0), \quad (1)$$

where C is any curve connecting points \mathbf{r}_0 and \mathbf{r} .

In the presence of noise, the computed value of the phase can depend on the integration path, and Eq. (1) may not be

useful. A more general method to calculate the phase from the gradient field might be preferable.

The phase $\phi(x,y)$ and the measured gradient $g(x,y)$ are related in the presence of noise $\mathbf{n}(x,y)$ by

$$\nabla\phi + \mathbf{n} = \mathbf{g}, \quad (2)$$

an equation which can be satisfied in a least-squares sense (indicated by the circumflex)

$$\int (\nabla\hat{\phi} - \mathbf{g})^2 dx dy = \text{minimum}. \quad (3)$$

This is a variational problem that has the Euler equation¹

$$\nabla^2\hat{\phi} = \nabla \cdot \mathbf{g}, \quad (4)$$

which is an elliptic differential equation. In our case, since the gradients are known we therefore get a Neumann boundary-value problem.² The generalization to a weighted least-squares fit

$$\int W(x,y)(\nabla\hat{\phi} - \mathbf{g})^2 dx dy = \text{minimum}, \quad (5)$$

leads to the Euler equation

$$\nabla \cdot (W\nabla\hat{\phi}) = \nabla \cdot (W\mathbf{g}) \quad (6)$$

with a weight function $W(x,y)$.

The existence of the solution for the Neumann problem requires that the line integral over the boundary of the driving term in Eq. (4) vanishes. Owing to the divergence form in our case this condition is automatically satisfied. The solution of a Neumann problem for a bounded region is unique except for an additive constant.

The error in the least-squares solution is defined by

$$\epsilon = \hat{\phi} - \phi, \quad (7)$$

and it also satisfies Eq. (4):

$$\nabla^2\epsilon = \nabla \cdot \mathbf{n}. \quad (8)$$

We can assume a decomposition of the noise field into an irrotational and a solenoidal part. The irrotational part is included in the solution of Eq. (8), whereas the solenoidal part of the noise is eliminated by taking the divergence of the noise.

The purpose of the following calculations is the determination of the mean-square wave-front distortion $\langle \epsilon^2 \rangle$ averaged over an appropriate noise ensemble.

II. DISCRETIZATIONS

For the discretizations of Eq. (2) we use $K = N^2$ phase points arranged in a square array. The finite difference forms for the gradient and the divergence can be expressed in a variety of ways. We present in this paper results following the discretization used by Hudgin³ and Fried.⁴ For all cases the relation between the phase, written as a linear array x , and the gradient, a linear array g , can be written as

$$Ax = g, \quad (9)$$

with a matrix $A(M \times K)$ which relates the array $x(K)$ to the array of M gradient measurements $g(M)$. For Hudgin's discretization the matrix A is given by

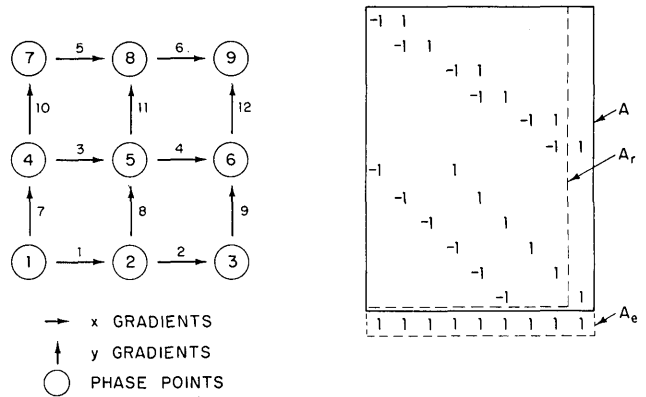


FIG. 1. Hudgin's discretization for $N = 3$, $M = 12$, and the corresponding matrix A , the reduced matrix A_r for $x_0 = 0$, and the extended matrix A_e .

$$\left. \begin{aligned} A[q + (p-1)(N-1), q + (p-1)N] &= -1 \\ A[q + (p-1)(N-1), 1 + q + (p-1)N] &= 1 \end{aligned} \right\} \begin{aligned} p &= 1, N \\ q &= 1, N-1 \end{aligned} \quad (10)$$

$$\left. \begin{aligned} A[q + M/2, q] &= -1 \\ A[q + M/2, q + N] &= 1 \end{aligned} \right\} q = 1, M/2$$

with all remaining elements equal to zero. The dimensions M and K for a specified value of N are

$$M = 2N(N-1), \quad K = N^2, \quad (10a)$$

and the rank of the matrix A is

$$R(A) = N^2 - 1. \quad (10b)$$

In Fig. 1 we show the discretization used by Hudgin and the resulting matrix for $N = 3$.

The discretization used by Fried leads to the matrix given by

$$\begin{aligned} A[m, n] &= -1 & A[m+1, n] &= -1 \\ A[m, n+M] &= 1 & A[m+1, n+M] &= -1 \\ A[m, n+NM] &= -1 & A[m+1, n+NM] &= 1 \\ A[m, n+NM+M] &= 1 & A[m+1, n+NM+M] &= 1 \end{aligned} \quad (11)$$

$$\begin{aligned} m &= 1 + 2(q-1) + 2(p-1)(N-1) \\ n &= [q + (p-1)N]M \\ p &= 1, N-1 \\ q &= 1, N-1 \end{aligned}$$

with all remaining elements equal to zero. The dimensions M and K are

$$M = 2(N-1)^2, \quad K = N^2, \quad (11a)$$

and the rank of the matrix is

$$R(A) = N^2 - 2. \quad (11b)$$

An example for Fried's case for $N = 3$ is shown in Fig. 2. In Hudgin's case the difference between the number of columns in the matrix A and its rank is 1, corresponding to the property that the phase is determined only up to a constant by its gradient. In Fried's case the discretization leads to two sets of disconnected phase points, with two arbitrary constants, one for each set. This fact is expressed by the rank deficiency of

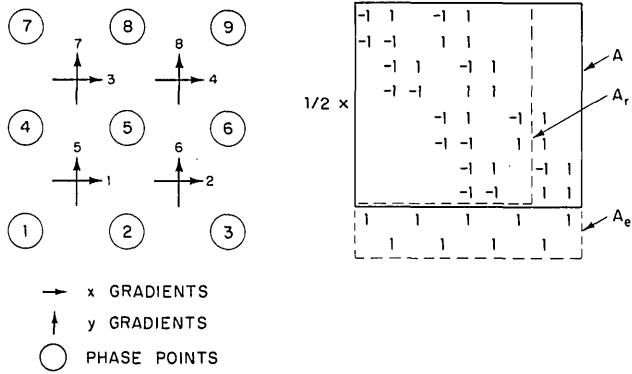


FIG. 2. Fried's discretization for $N = 3$, $M = 8$, and the corresponding matrix A , the reduced matrix A_r for $x_8 = x_9 = 0$, and the extended matrix A_e .

2. These properties pose no problem in our treatment of Eq. (9).

III. LEAST-SQUARES MINIMUM-NORM SOLUTIONS

The least-squares minimum-norm solution of Eq. (9) is unique and is⁵⁻⁷

$$x = A^+g, \quad (12)$$

where the Moore-Penrose generalized inverse is written as A^+ . Solutions of linear equations are discussed in the Appendix. The generalized inverse can be evaluated in a variety of ways. We start with an elementary method using a partitioning of the matrix A .

We define four submatrices that make up the matrix A

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad (13)$$

where A_1 is a full rank square matrix, its rank equal to the rank of A , while A_2 , A_3 , and A_4 are matrices of appropriate dimensions. The Moore-Penrose inverse of A is then⁵

$$A^+ = [A_1 A_2]^T U^T \begin{bmatrix} A_1 \\ A_3 \end{bmatrix}^{-1} T, \quad (14)$$

with

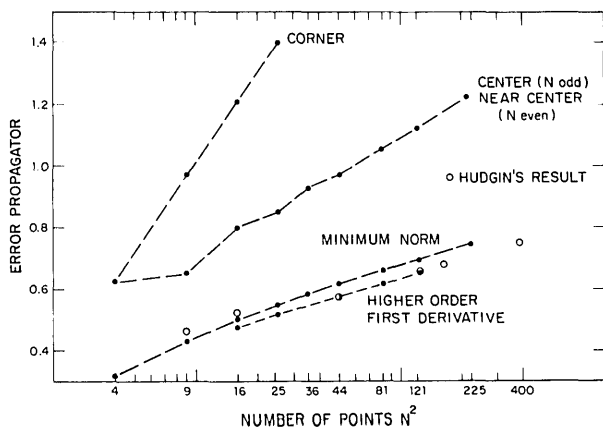


FIG. 3. Error propagator for Hudgin's discretization as function of number of points N^2 for a variety of conditions.

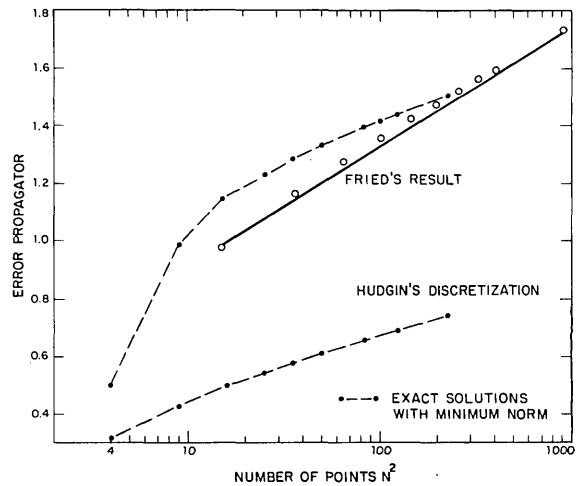


FIG. 4. Error propagator for Fried's and Hudgin's discretization for the minimum-norm solutions and the values from Ref. 4.

$$U = \left([A_1 A_2] A^T \begin{bmatrix} A_1 \\ A_3 \end{bmatrix} \right)^{-1}, \quad (14)$$

where T indicates the transpose of a matrix.

The minimum-norm least-squares error is then given by

$$\epsilon^2 = \sigma^2 S / N^2, \quad (15)$$

where σ^2 is the mean-square error of the gradient measurements and S is the sum of the squares of all elements in the matrix A^+ . In the derivation of Eq. (15) we assumed that the noise is uncorrelated between the different gradient measurements and that each gradient has the same mean-square error. These are the same statistical properties of the noise as in Refs. 3 and 4. The error propagator, defined by $E_p = S / N^2$, is plotted in Figs. 3 and 4 as a function of the number of points N^2 .

The structure of the matrix A makes it easy to perform the required partitioning in Eq. (13) and to calculate the generalized inverse using Eq. (14). The general procedure to calculate the generalized inverse uses first the singular-value decomposition^{5,7} of the matrix A

$$A = USV^T, \quad (16)$$

where U and V are orthogonal matrices and S is a diagonal matrix containing the singular values of A . The generalized inverse is then given by

$$A^+ = VS^{-1}U^T. \quad (17)$$

A FORTRAN program to perform these operations is available.⁸

IV. OTHER LEAST-SQUARES SOLUTIONS

The phase is determined by the gradient field only up to a constant (or two constants in Fried's case). This permits us to construct particular least-squares solutions of Eq. (9) by setting the phase equal to zero at a particular point (at two points for Fried's case). As an example, one can take the center or the corner value of the phase to be zero. In this way the matrix A becomes a full rank matrix A_r , reduced by one

column for Hudgin's case and two columns for Fried's case. The matrix relating the phase to all gradients is then given by

$$B_r = \begin{bmatrix} (A_r^T A_r)^{-1} A_r^T \\ 0 \end{bmatrix}, \quad (18)$$

where 0 is a one-row matrix of zeros for Hudgin's case and a two-row matrix of zeros for Fried's case. The least-squares errors corresponding to this solution are shown in Fig. 3 for some center values of the phase equal to zero or some corner values equal to zero. These solutions are obviously not minimum-norm solutions.

The least-squares solution $x = B_r g$ with the matrix B_r in Eq. (18) is a solution of the equation

$$A_r^T A_r x = A_r^T g, \quad (19)$$

which is a finite difference approximation to Eq. (4) and is identical to Hudgin's Eq. (28). Hudgin's derivation of this equation was performed with considerable detour into Fourier space. Equation (19) is called the normal equation⁵ of the least-squares problem $A_r x_r = g$.

It is possible to construct from B_r the minimum-norm solution B using the condition $\sum x = 0$. This leads for Hudgin's case to

$$B = \begin{bmatrix} B_r - C \\ -C \end{bmatrix}, \quad (20)$$

where the linear array C is given by

$$C_m = (1/N^2) \sum_{n=1}^{K-1} B_{nm}, \quad m = 1, M.$$

For Fried's case two linear arrays have to be introduced corresponding to the fact that there are two sets of disconnected phase points.

We summarize now the numerical results for the error propagator. In Fig. 3 we plot it as functions of N^2 for Hudgin's case. We show the minimum-norm solution, a solution that assumes the center phase put to zero (for N even, we use a point closest to the center), and a corner phase set to zero. In addition we show the values calculated by Hudgin.³ (We corrected for a missing square-root sign in Hudgin's paper; his noise coefficients have to be squared to give the correct error propagator for the variance.)

The discretization used in the construction of the matrix A in Eq. (10) is based on the approximation for the first derivative

$$f' = \frac{1}{h} (f_{n+1} - f_n) + O(h^2),$$

where h is the spacing between the points. A higher-order approximation is given by

$$f' = \frac{1}{h} \left(-\frac{1}{24} f_{n+2} + \frac{9}{8} f_{n+1} - \frac{9}{8} f_n + \frac{1}{24} f_{n-2} \right) + O(h^4).$$

An appropriate change in the matrix A leads then to a least-squares solution with a slightly smaller error propagator, also shown in Fig. 3.

The difference between the least-squares errors and the least-squares minimum-norm error is due to the piston error term. The piston is an unwanted parameter in the solution

for the wave front and has only some indirect effects. If the piston is not removed the control system has to include the dynamic range of the piston error and the effect on the system will depend on the size and spectrum of the piston error. A phase observed at a single point would include the piston noise.

In Fig. 4 we plot the error propagator for the minimum-norm solution for Fried's case. For large N Fried's solution approaches the correct values. The deviation for small N is probably due to the approximation introduced by Fried following his Eq. (22). We also plot the error propagator for the minimum-norm solution for Hudgin's case in order to show the effect of the different discretizations.

V. NONSINGULAR LEAST-SQUARES MINIMUM-NORM SOLUTION

It is possible to extend system Eq. (9) to a full rank problem which gives the unique least-squares minimum-norm solution. Using the requirement $\sum x = 0$, which removes the piston, we can extend Eq. (9) to

$$A_e x = g_e. \quad (21)$$

For Hudgin's case the dimension of A_e is $(M+1) \times K$. By adding a row of ones

$$A_s(q) = 1, \quad q = 1, K$$

to the matrix A we get

$$A_e = \begin{bmatrix} A \\ A_s \end{bmatrix}, \quad (22)$$

and adding one zero to the array g , we get g_e .

For Fried's case the dimension of A_e is $(M+2) \times K$ constructed by adding a two-row matrix A_s to A as in Eq. (22) with

$$A_s[1, 1 + 2p] = 1 \quad p = 0, (N^2 - 1)/2$$

$$A_s[2, 2 + 2p] = 1 \quad q = 0, (N^2 - 3)/2 \quad \text{for odd } N,$$

$$A_s[1, 1 + 2(q-1) + N(p-1) + \text{mod}(p+1, 2)] = 1$$

$$A_s[2, 1 + 2(q-1) + N(p-1) + \text{mod}(p, 2)] = 1 \quad \text{for even } N.$$

These extended systems lead to the normal equation

$$A_e^T A_e x = A_e^T g_e = A^T g, \quad (23)$$

with the least-squares minimum-norm solution

$$x = (A_e^T A_e)^{-1} A^T g, \quad (24)$$

which is identical to the solution equation (12).

VI. RELATION TO NETWORK THEORY

It is useful to formulate the least-squares equations in terms of network theory if a hardware implementation is considered.³ The simplest case is given by Hudgin's discretization where the matrix A of Eq. (10) can be directly interpreted as the transpose of the complete incidence matrix⁹ of a network for which the phase points are used as nodes and the lines connecting these nodes are used as branches. Putting one node voltage to zero corresponds to removing one row of the matrix A . This leads to the matrix A_r , which can be interpreted as the transpose of the reduced incidence matrix of a

network. For our case we can assume that the branch-admittance matrix is proportional to the unit matrix. Equation (19) is then the system of node equations with the node-admittance matrix $A_r^T A_r$. The driving term of Eq. (19) is the node current-source vector which is given by the sum of the currents incident at the corresponding nodes.

The solution arrived at in this way does not have minimum norm. We have shown that the least-squares minimum-norm solution can be determined from Eq. (23). We can again interpret $A_r^T A_r$ as an extended node-admittance matrix. However, a realization by a linear and passive network is not possible, because the extended node-admittance matrix is not a diagonally dominant matrix.¹⁰ In order to get the minimum-norm solution an active element has to be added to the network.

In the terminology of network theory⁹ Hudgin's discretization leads to a connected graph for which at least one connected subgraph exists. These subgraphs are called trees. Fried's discretization leads to an unconnected graph, which contains a set of trees, one for each of the separate parts, and is called a forest.

VII. LEAST-SQUARES FIT AND POLYNOMIAL EXPANSIONS

It is sometimes useful to represent a phase by the coefficients of an expansion in a set of basic functions P_i (e.g., Zernike polynomials)

$$\phi(x,y) = \sum c_i P_i(x,y). \quad (25)$$

The polynomials do not have to be orthogonal, but they do have to be linearly independent on the set of points for which they are evaluated.

We construct a matrix $P[K,L]$ with the dimension K given by the number of points (x_k, y_k) and the dimension L given by the number of basis functions. Equation (25) can then be written

$$\phi = Pc, \quad (26)$$

with the phase array $\phi[K]$ and the coefficient array $c[L]$.

We analyze now two methods to treat the two equations

$$g = A\phi, \quad (27a)$$

$$\phi = Pc. \quad (27b)$$

The first equation relates the measured gradients to the phase, usually in the presence of noise. The second equation relates the phase to the coefficients of an expansion. In general each represents an overdetermined linear system (more equations than unknowns).

In the first method we use a least-squares solution (with minimum norm) to Eq. (27a)

$$\phi = A^+g, \quad (28a)$$

and a least-squares solution to Eq. (27b)

$$c = P^+\phi. \quad (28b)$$

The coefficients c are then related to the gradient g by

$$c = P^+A^+g. \quad (29)$$

In the second method we substitute the phase expansion Eq. (27b) into the gradient Eq. (27a)

$$g = APc, \quad (30)$$

and obtain a least-squares solution

$$c_g = (AP)^+g, \quad (31)$$

which inserted into Eq. (27b) leads to the phase

$$\phi_g = P(AP)^+g. \quad (32)$$

The two solutions, Eqs. (29) and (31), as well as Eqs. (28a) and (32) are in general *not* the same, and we indicate this by a subscript g .

A simple example showing the difference between the two methods is presented in Fig. 5. For three gradient values, four phase values, and a single polynomial (tilt), we show the relevant matrices and their generalized inverses.

In the first method a least-squares fit gives the minimum variance of the phase, whereas the second method gives a minimum variance of the gradient. The first method using Eq. (28a) maximizes the Strehl ratio, whereas the second method using Eq. (32) minimizes the second moment of a beam propagated, in a geometrical optics sense, to a focal plane. The main computational difference is the evaluation of the generalized inverse A^+ for the first method. The matrix A can be very large. The second method requires at most an inversion of a matrix with dimension equal to the number of polynomials required.

VIII. RESULTS FOR A SPECIAL CASE

A. Error propagator coefficient

In this section we analyze some interesting properties for a special case which consists of a 9×9 array from which the three points in each corner are removed. Additional modifications include a central obscuration and a phase error calculation based on a larger number of points used for the gradient measurements than for the evaluation of the phase. The array used is shown in Fig. 6, it has 69 phase points and 120 gradients.

We calculated the wave front error propagation factor for a variety of conditions summarized in Table I. The whole array has 69 points and 120 gradients. The first modification leads to a system where the edge points are not used; it consists of 45 points for which the phase is calculated from all 120 gradients. Another option is the elimination of the center point, corresponding to a system with a central obscuration where one point and four gradients are removed. The error propagator coefficient for these cases is presented in Table I.

B. Tilts of phase fronts

We restrict the polynomial expansion for the special case to the determination of x tilt and y tilt. The coefficients for the tilt can be determined from a least-squares gradient fit

$$c_g = (AP)^+g, \quad (33)$$

or from a least-squares phase fit

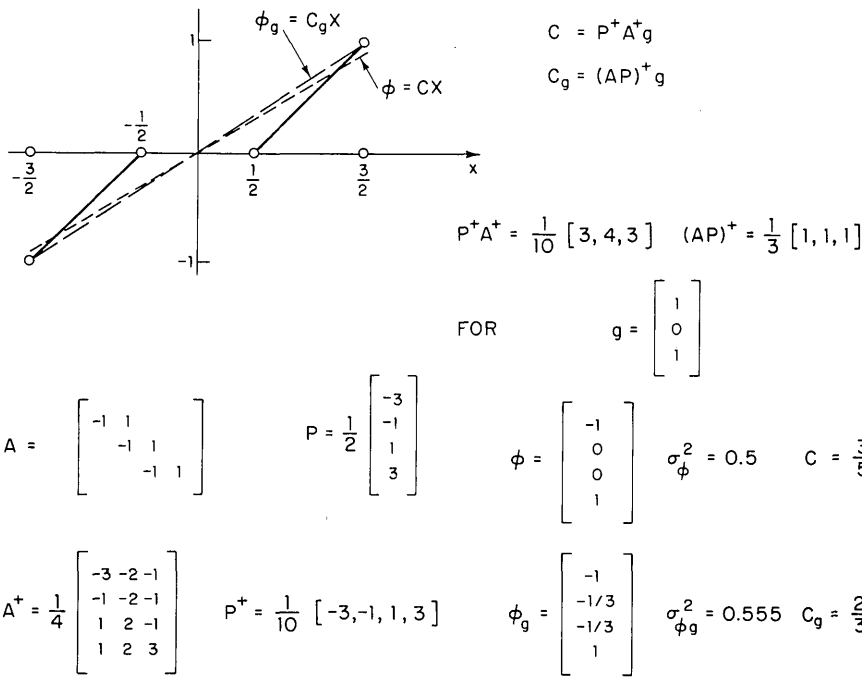


FIG. 5. Example showing the difference between least-squares phase fit and least-squares gradient fit.

$$c = P^+ A^+ g. \quad (34)$$

For tilt alone the matrix $(AP)^+$ reduces to an expression,

$$\begin{bmatrix} c_{xg} \\ c_{yg} \end{bmatrix} = \left(\frac{2}{M} \right) \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} g_x \\ g_y \end{bmatrix}, \quad (35)$$

showing that the tilts are given by the average of the corresponding gradients.

The tilts from a least-squares phase fit are considerably more complicated. We show the matrix $P^+ A^+$ in Table II for a full 9×9 array and for the special array with the corners cut off. The matrix elements are arranged in such a way that they correspond to the gradients as shown in Fig. 6. The tilts of the full array are a weighted average of the gradients, where the weights depend on the position in the array. For the special array the x tilt depends on the y gradients, an interaction absent in the full square array.

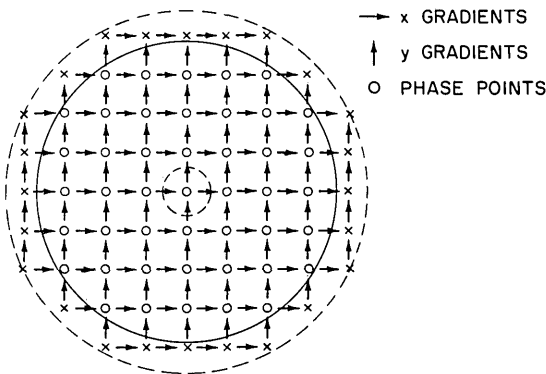


FIG. 6. Special 9×9 array with corners cut off. The outer circle contains 120 gradients and 69 phases. The inner circle contains 45 phases. A central obstruction removes four gradients and one phase.

IX. ANALYSIS USING THE CURL OPERATOR

A. Construction of an irrotational gradient field

It is possible to eliminate the solenoidal part of the noise of the gradient measurement and reduce the system equation (9) to a consistent system of linear equations. If we specify the piston (e.g., by $\phi_1 = 0$), a simple addition of the curl-free gradients can be performed, approximating the line integral in Eq. (1).

We introduce a matrix Q , corresponding to a discrete form of the curl operator

$$\left. \begin{aligned} Q[p, q] &= 1 \\ Q[p, q + N - 1] &= -1 \end{aligned} \right\} \begin{aligned} p &= 1, (N - 1)^2 \\ q &= 1, (N - 1)^2 \end{aligned}$$

$$\left. \begin{aligned} Q[q + (p - 1)(N - 1), M/2 + q + (p - 1)N] &= -1 \\ Q[q + (p - 1)(N - 1), M/2 + 1 + q + (p - 1)N] &= 1 \end{aligned} \right\}$$

$$p = 1, N - 1 \quad q = 1, N - 1. \quad (36)$$

TABLE I. Wave-front error for the 69 phase-point system.

Array	Mode ^a	K^b	M^c	E_p
Whole	C	69	120	1.025
Inside	C	45	120	0.8539
Whole	M	69	120	0.6421
Inside	M	45	120	0.5054
With central obscuration				
Whole	C	68	116	1.218
Inside	C	44	116	1.098
Whole	M	68	116	0.6740
Inside	M	44	116	0.5540

^aMode: C = centered, phase is set to zero at center point (or next to center if center point omitted); M = minimum norm (piston removed).

^bNumber of phase points.

^cNumber of gradients.

TABLE II. Matrix for least-squares phase fit of tilt ($\times 10^4$).

Tilt coefficients: $\begin{bmatrix} c_x \\ c_y \end{bmatrix} = \begin{bmatrix} (P^+A^+)_{xx} & (P^+A^+)_{xy} \\ (P^+A^+)_{yx} & (P^+A^+)_{yy} \end{bmatrix} \begin{bmatrix} g_x \\ g_y \end{bmatrix}$.								
Full Array,								
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	$(P^+A^+)_{xx}$
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	$(P^+A^+)_{xy} = 0$
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	
74	130	167	185	185	167	130	74	
Special Array								
		121	170	170	121			
		141	168	191	191	168	141	
116	159	191	210	210	191	159	116	
101	163	202	221	221	202	163	101	$(P^+A^+)_{xx}$
98	164	204	224	224	204	164	98	
101	163	202	221	221	202	163	101	
116	159	191	210	210	191	159	116	
		121	170	170	121			
		-68	-22	0	22	68		
-61	-42	-19	0	19	42	61		
-10	-25	-21	-11	0	11	21	25	10
-4	-7	-6	-3	0	3	6	7	4
4	7	6	3	0	-3	-6	-7	-4
10	25	21	11	0	-11	-21	-25	-10
		61	42	19	0	-19	-42	-61
		68	22	0	-22	-68		

The matrix Q has the dimension $(N - 1)^2 \times M$ and an example for $N = 3$ is shown in Fig. 7. The curl matrix Q has full row rank.

The curl-free gradient g_c is determined from the measured gradient by subtracting a solenoidal gradient g_s ,

$$g_c = g - g_s. \quad (37)$$

The solenoidal gradient has to satisfy

$$Qg_s = Qg, \quad (38)$$

which is an underdetermined system with $(N - 1)^2$ equations for $M = 2N(N - 1)$ unknown values of g_s . We select the solution with minimum norm of g_s ,

$$g_s = Q^+Qg. \quad (39)$$

The curl-free gradient is then

$$g_c = (I - Q^+Q)g = Rg, \quad (40)$$

which leads to the overdetermined but consistent system

$$A\phi = g_c. \quad (41)$$

If we turn to a reduced system (e.g., $\phi_1 = 0$) and use all x gradients but only one column of y gradients we get a nonsingular system of $N^2 - 1$ equations for the same number of unknown phases

$$A_{rd}\phi_r = g_{rd}, \quad (42)$$

where the subscript rd indicates the appropriately reduced arrays. The matrix A_{rd} is constructed from the matrix A by eliminating the first column and $(N - 1)^2$ rows. The solution of Eq. (42) is then simply

$$\phi_r = A_{rd}^{-1}g_{rd} = S_L g_{rd}, \quad (43)$$

where we call S_L the inverse of A_{rd} because it is the discrete equivalent of the line integral in Eq. (1).

The piston of the phase ϕ_r can be removed as shown in Sec. III by introducing

$$S = \begin{bmatrix} -C \\ S_L - C \end{bmatrix}, \quad (44)$$

where C is the array of the column sums of S_L

$$C_j = \frac{1}{N^2} \sum_i (S_L)_{ij}. \quad (45)$$

The piston removed phase $\phi[N^2]$ can then be calculated from

$$\phi = SR_{rd}g, \quad (46)$$

where S is the $[N^2 \times (N^2 - 1)]$ matrix from Eq. (44), and R_{rd} is an $[(N^2 - 1) \times M]$ matrix obtained from the matrix R by eliminating the $(N - 1)^2$ rows corresponding to the gradients which are not needed if the solenoidal part of the gradient is removed. This phase is identical to the phase calculated with the generalized inverse

$$\phi = A^+g.$$

B. Use of the curl measurements

The curl measurements can be used to estimate the noise present in the gradient measurements. We make the assumption that the measurement noise values are statistically independent. The relation between the curl q and the gradient g

$$q = Qg \quad (47)$$

leads then to the relation between the two corresponding variances

$$\sigma_q^2 = (N - 1)^{-2} \sum |Q|^2 \sigma_g^2 = 4\sigma_g^2. \quad (48)$$

The noise propagation factor in this case is independent of the number of gradient values.

The variance of the solenoidal part of the gradient g_s is related to the variance of the gradient g by

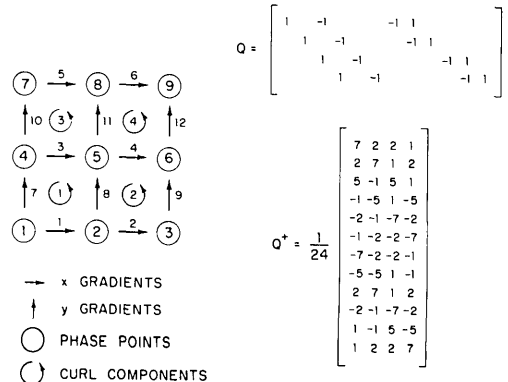


FIG. 7. Curl operator and its generalized inverse for $N = 3$.

$$\sigma_{gs}^2 = \frac{1}{M} \sum |Q+Q|^2 \sigma_g^2 = \frac{N-1}{2N} \sigma_g^2, \quad (49)$$

with a proportionality factor that approaches 1/2 for large N .

ACKNOWLEDGMENTS

The author wishes to acknowledge stimulating discussions with C. A. Primmerman of M.I.T./Lincoln Laboratory. This work was sponsored by the Advanced Research Projects Agency of the Department of Defense.

APPENDIX

Glossary of symbols:

$\| \cdot \|$, Euclidean or L_2 norm;
 A^+ , Generalized inverse of matrix A ;
 A^T , Transpose of matrix A .

Solutions of linear systems:

A system of linear equations

$$Ax = b$$

has the following solutions:

(i) For a square matrix with full rank,

$$x = A^{-1}b.$$

(ii) For a rectangular matrix $A[M \times N]$ with $M > N$, and with full column rank,

$$x = (A^T A)^{-1} A^T b$$

is the least-squares solution

$$\|Ax - b\| = \min.$$

(iii) For a rectangular matrix $A[M \times N]$ with $M < N$, and with full row rank,

$$x = A^T (A A^T)^{-1} b$$

is the least-norm solution.

(iv) For other matrices

$$x = A^+ b + (I - A^+ A)y,$$

with arbitrary y , is the least-squares solution

$$\|Ax - b\| = \min,$$

and for $y = 0$ one gets the least-squares minimum-norm solution

$$\|Ax - b\| = \min,$$

and

$$\|x\| = \min.$$

The generalized inverse A^+ can be defined by the Moore-Penrose conditions

$$A A^+ A = A, \quad A^+ A A^+ = A^+,$$

$$(A A^+)^T = A A^+, \quad (A^+ A)^T = A^+ A.$$

¹H. Margenau and G. M. Murphy, Inc., *The Mathematics of Physics and Chemistry* (Van Nostrand, New York, 1956), Chap. 6.

²G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961), Chap. 15.

³R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *J. Opt. Soc. Am.* **67**, 374-378 (1977).

⁴D. L. Fried, "Least-squares fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Am.* **67**, 370-375 (1977).

⁵A. Ben-Israel and N. E. Greville, *Generalized Inverses: Theory and Applications* (Wiley, New York, 1974), Chaps. 3 and 5.

⁶C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications* (Wiley, New York, 1971), Chap. 11.

⁷C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice Hall, Englewood Cliffs, New Jersey, 1974), Chap. 25.

⁸International Mathematical & Statistical Libraries, Inc., Sixth Floor GNB Building, 7500 Bellaire, Houston, Texas 77036.

⁹N. Balabanian and T. A. Bickart, *Electrical Network Theory* (Wiley, New York, 1969).

¹⁰W. H. Kim and H. E. Meadows, Jr., *Modern Network Analysis* (Wiley, New York, 1971), Chap. 3.