

PHASE RETRIEVAL AS A NONLINEAR  
ILL-POSED PROBLEM

*G.N. Newsam*

INTRODUCTION

Phase retrieval is a common problem in many branches of physics, such as optics or crystallography, but a great deal of difficulty has been encountered in the construction of numerical solutions to it. This arises because the problem is ill-posed; this talk briefly describes the two sources of ill-conditioning, nonuniqueness and discontinuous dependence of the solution on the data, and considers the implication for numerical algorithms for the problem's solution.

The prototypical phase retrieval problem occurs when a light beam passes through a small aperture  $B$  and then falls on a flat screen  $A$ . Classical optics states that the wavefront at  $A$  is the Fourier transform of the wavefront across the aperture, so that knowledge of the wavefront in the plane of  $A$  allows reconstruction of it at  $B$ . Measurement of the intensity of the beam on the screen is an easy task; measurement of the phase is quite a different matter and is usually impossible. This leads to the following mathematical model problem of phase retrieval

Given the measured modulus  $m(s)$  of a function  $g(s)$  on a bounded set  $A \subset \mathbb{R}^N$ , where  $g(s)$  is the Fourier transform of a function  $G(w)$  with support contained in the bounded set  $B \subset \mathbb{R}^N$ , find

- i. the phase of  $g(s)$  on  $A$
- ii. the extension of  $g(s)$  to all of  $\mathbb{R}^N$
- iii. the function  $G(w)$  on  $B$ .

Since an arbitrary phase may be assigned to a given wavefront without altering its modulus, it is necessary to show that the condition that the beam has passed through a bounded aperture severely limits the class of possible phases. Fortunately this follows from the next theorem

PALEY-WEINER THEOREM [3]:  $g(s)$  is the Fourier transform of a function  $G(w)$  with bounded support iff  $g(s)$  is an entire function of exponential growth.

However, even with this restriction, the mathematical model is still an ill-posed problem. Hadamard defined a problem to be ill-posed if the solution failed to either

- i. exist
- ii. be unique
- iii. depend continuously on the data

Phase retrieval fails all three conditions.

Nonexistence, although ominous at first sight, is, in fact, a non-issue. Since  $g(s)$  is analytic, so is  $m(s)$ ; thus any nonanalytic perturbation of the modulus through measurement error produces a mathematical model with no exact solution. However, as the data is collected from an actual wavefront, a real solution must exist. Thus the mathematical model may be adjusted with confidence to produce approximate solutions even in the presence of nonanalytic errors.

However nonuniqueness and discontinuous dependence create real difficulties requiring special attention. The first is due to the nonlinear projective nature of the modulus operator, and may be characterized using the theory of analytic functions. The second occurs as information on  $g(s)$  outside the bounded set  $A$  is not available; it is manifested in the presence of a compact linear operator, the finite Fourier transform, whose inversion requires techniques from the theory of linear ill-posed problems.

## NONUNIQUENESS

Nonuniqueness in phase retrieval turns out to be an elaboration of the following basic idea. A function  $g(s)$  is an analytic function of the complex variable  $s$  iff the conjugate function  $g^*(s^*)$  is also analytic. Thus, if  $g(s)$  may be factored into the product  $g_1(s)g_2(s)$  of analytic functions, then  $\tilde{g}(s) \equiv g_1(s)g_2^*(s^*)$  is also analytic; and for real  $s$  (i.e.  $s = s^*$ )

$$(2.1) \quad |g(s)| = |g_1(s)| \cdot |g_2(s)| = |g_1(s)| \cdot |g_2^*(s^*)| = |\tilde{g}(s)|$$

All solutions to phase retrieval may be related in this manner.

In one dimension a complete characterization of the solution set is possible. If  $g$  is any solution, then, as a consequence of the Paley-Weiner Theorem, it has a Hadamard factorization, i.e.

$$(2.2) \quad g(s) \equiv e^{\alpha + \beta s} s^m \prod_{k \in \mathbb{N}} \left(1 - \frac{s}{s_k}\right)$$

where  $\alpha$  and  $\beta$  are constants,  $\mathbb{N}$  is the set of natural numbers,

and  $\{s_k\}_{k \in N}$  is the set of zeroes of  $g$ . Viewing this as a factorization, and noting that  $(1-s^*/s_k)^* = (1-s/s_k^*)$  leads to Akutowicz's results [1,2]: all possible alternative solutions  $\tilde{g}$  must have the form

$$(2.3) \quad \tilde{g}(s) \equiv e^{i\theta} e^{\alpha+\beta s} s^m \prod_{k \in \Lambda} \left(1 - \frac{s}{s_k}\right) \prod_{k \in N-\Lambda} \left(1 - \frac{s}{s_k^*}\right)$$

where  $\theta$  is an arbitrary real number and  $\Lambda$  is any subset of  $N$ . Furthermore if  $B$  is an interval  $[a,b]$  then any function of this form is a solution. Since the set of all subsets  $\Lambda$  of  $N$  is uncountable, there is a real continuum of possible solutions.

The result implies that all solutions are generated by flipping zeroes  $s_k$  of a particular solution to their conjugates  $s_k^*$ . An extension of these arguments to higher dimensions [6, Appendix D] shows that if  $g(s_1, \dots, s_n)$  and  $\tilde{g}(s_1, \dots, s_n)$  are both solutions to phase retrieval, then the set  $Z_{\tilde{g}}$  of zeroes of  $\tilde{g}$  must be contained in the union of the sets  $Z_g$  and  $(Z_g)^*$ , and vice versa. The zeroes of an entire function of  $n$  variables form an analytic manifold of dimension  $n-1$ . If  $n = 1$  then this manifold must be a collection of disjoint points, thus it is possible to flip any part to its conjugate without destroying the analyticity of the whole manifold. However, for  $n \geq 2$ ,  $Z_g$  will be a collection of connected analytic components; flipping an arbitrary subset to its conjugate will most likely create a nonanalytic set  $Z_{\tilde{g}}$  which cannot be associated with an alternative analytic solution  $\tilde{g}$ . Furthermore, in higher dimensions almost all sets  $Z_g$  will consist of a single connected analytic component; this is analogous to the result that almost all polynomials of two or more variables are irreducible,

i.e. cannot be factored. No proper subset of such a manifold may be flipped to its conjugate without destroying analyticity; therefore almost all solutions will be unique.

Unfortunately the analysis used to resolve questions of uniqueness has not yet produced an associated numerical algorithm for recovery of the phase of  $g$ . Any method of generating even an approximate phase would be of great help, as experience has shown that any iterative algorithm for recovery of  $G(w)$  usually fails to converge unless started close to a known solution. Recently Bates, Fright and Garden [4,5] have obtained some promising results in this area which suggest the following open questions

1. In one dimension, given  $g(0)$ ,  $g'(0)$  and  $|g|$ , does there exist a second order o.d.e. that may be integrated up to give the phase of  $g$  on  $A$ .
2. In two dimensions, given  $g(0,0)$  and  $|g|$ , does there exist a p.d.e. that may be integrated up to give the phase of  $g$  on  $A$ .

#### DISCONTINUOUS DEPENDENCE OF THE SOLUTION ON THE DATA

The source of discontinuous dependence is best seen in a formal description of the problem, therefore the following notation is introduced.  $F$  denotes the Fourier transform,  $P_{A(B)} : L^2(\mathbb{R}^N) \rightarrow L^2(A(B))$  denotes the projection operator

$$(3.1) \quad (P_A g)(s) = \begin{cases} g(s) & s \in A \\ 0 & s \notin A \end{cases}$$

The composite operator  $P_A F P_B : L^2(A) \rightarrow L^2(B)$  is termed the finite Fourier transform.

In this notation the mathematical model of phase retrieval requires solution of the system of equations

$$(3.2) \quad (P_A F P_B g)(s) = g(s)$$

$$(3.3) \quad |g(s)| = m(s)$$

Even if the phase of  $g$  is known on  $A$ , Eq. 3.2 must still be solved and this requires inversion of the linear operator  $P_A F P_B$ . Inversion of the full Fourier transform  $F: L^2(\mathbb{R}^N) \rightarrow L^2(\mathbb{R}^N)$  is well known to be a stable procedure as  $F$  is a unitary operator. However, as information is lost in the truncations  $P_A$  and  $P_B$ , inversion of the finite Fourier transform is no longer stable. Indeed, as  $P_A F P_B$  is a compact linear operator, its inversion is the prototypical linear ill-posed problem. This problem underlies phase retrieval and is unavoidable in any method of solution; for example consider the linear system that results after applying Newton's method to Eqs. 3.2 and 3.3.

The finite Fourier transform has been extensively studied (for a review see [8]), the more important results are briefly reviewed here. First, it has a singular value decomposition  $\{\phi_i, \sigma_i, \psi_i\}_{i=1}^{\infty}$ , where  $\{\phi_i\}_{i=1}^{\infty}$  and  $\{\psi_i\}_{i=1}^{\infty}$  are analytic functions forming orthonormal bases for  $L^2(A)$  and  $L^2(B)$  (in addition, in one dimension they also form Sturm-Liouville systems). The  $\sigma_i$  are scalars and

$$(3.4) \quad P_A F P_B \psi_i = \sigma_i \phi_i \quad \sigma_i \geq 0 \quad \sigma_i \downarrow 0$$

The crucial result is that the  $\sigma_i$  are asymptotically distributed as

$$(3.5) \quad \sigma_i \sim \begin{cases} 1 & i < |A| \cdot |B| \\ e^{-\alpha i} & i > |A| \cdot |B| \end{cases}, \quad \alpha \text{ a constant}$$

These expressions are valid everywhere except on a narrow interval, centered on  $i = |A| \cdot |B|$  and of width proportional to  $|\partial A| \cdot |\partial B| \cdot \log((|\partial A| \cdot |\partial B|) / (|A| \cdot |B|))$ , over which the distribution changes smoothly from one asymptotic form to the other. Here  $|A|$  denotes the volume of  $A$ ,  $|\partial A|$  the area of the boundary of  $A$ .

Discontinuous dependence can now be clearly seen by considering the effects of errors on formal expansions of  $g$  and  $G$  as series of singular functions :

$$(3.6) \quad g = \sum_{i=1}^{\infty} a_i \phi_i \quad G = \sum_{i=1}^{\infty} b_i \psi_i$$

Eqs. 3.2 and 3.4 imply that  $b_i = a_i / \sigma_i$ . If  $a_i$  is perturbed to  $a_i + \epsilon$ , this induces a perturbation of  $b_i$  to  $b_i + \epsilon / \sigma_i$ . As  $i \rightarrow \infty$  the perturbation in  $g$  remains bounded, but the induced perturbation in  $G$  grows exponentially.

Thus, to avoid catastrophic amplification of errors, the expansion of  $G$  must be truncated to those values of  $i$  for which  $\sigma_i \sim 1$ . This restricts the solution space to the span of the first  $N$  singular functions, where  $N \sim |A| \cdot |B|$ .  $N$  is termed the essential dimension of the problem; loosely speaking, it is the maximum number of coefficients in any expansion of  $G$  that can be accurately recovered from Eq. 3.2 in the presence of noise. A precise definition is given in [6].

## IMPLICATIONS FOR NUMERICAL SOLUTIONS

The above results have important implications for the construction and performance of numerical algorithms for phase recovery. In one dimension, the nonuniqueness results imply the existence of uncountably many solutions. Thus, unless the initial guess is close to a particular solution, any iterative algorithm will dither among an infinite set of attraction points. Extra information must be obtained, if possible, to reduce the size of the possible solution set.

In higher dimensions solutions are most likely unique. However numerical algorithms used so far, such as the Gershberg and Saxton version of the iterated projection algorithm [7], still show poor convergence. This arises as the algorithms must still effectively invert the finite Fourier transform, and this will only be done in a stable manner if the solution space is restricted to the subspace spanned by the first  $|A| \cdot |B|$  singular functions. Thus only this smooth component of  $G$  may be accurately determined from the model.

Moreover numerical discretizations should be approximately the same size as the essential dimension. Much larger, and no additional accurate information is obtained for the extra work; much smaller, and accurate information is lost. In addition, discretizations should be tailored to the smooth singular functions, e.g. based on expansions in Legendre polynomials or Gaussian quadrature, and not on the fast Fourier transform.

However, even with the above improvements, phase retrieval is still a difficult problem to solve numerically. In particular there is still much work to be done on optimizing algorithm design, generating good initial guesses, and on incorporation of extra conditions that  $g$  or  $G$  may be known to satisfy.



## REFERENCES

- [1] E.J. Akutowicz, *On the determination of the phase of a Fourier integral - I*, Trans. Amer. Math. Soc., 83 (1956), 179-192.
- [2] E.J. Akutowicz, *On the determination of the phase of a Fourier integral - II*, Proc. Amer. Math. Soc., 8 (1957), 234-238.
- [3] R.P. Boas, *Entire Functions*, Academic Press, New York, 1954.
- [4] K.L. Garden and R.H.T. Bates, *Fourier phase problems are uniquely solvable in more than one dimension - II: one dimensional considerations*, Optik, 62 (1982), 131-142.
- [5] W.R. Fright and R.H.T. Bates, *Fourier phase problems are uniquely solvable in more than one dimension - III: computational examples for two dimensions*, Optik, 62 (1982), 219-230.
- [6] G.N. Newsam, *Numerical reconstruction of partially known transforms*, Ph.D. thesis, Harvard University, Cambridge, Mass., 1982.
- [7] W.O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*, Academic Press, New York, 1978.
- [8] D. Slepian, *Some comments on Fourier analysis, uncertainty and modelling*, SIAM Review, 25 (1983), 379-393.