



Survey of the Ring

A Trilogy

Part 2: Pseudoinverting a Large,  
Almost Block Tridiagonal  
System

Leo Michelotti

Abstract

A procedure is developed for constructing the Moore-Penrose pseudoinverse of a large, almost block tridiagonal system whose null space is known. The system was motivated by a particular survey problem on a large ring of sites, and the application to its solution, including construction of the appropriate null space, is presented in detail. An analytic singular value decomposition is carried out for the case of a maximally symmetric ring.



TABLE OF CONTENTS

Abstract	i
1. The Problem	1
1.1 Conceptual Overview	5
1.2 Frames and Coordinates	10
1.3 Singular value decomposition	13
2. Symmetric Rings	20
3. Basis States over Null Space	25
4. Exact Solution for Sparse Matrices	30
Figure captions	37
Figures	38
Appendix	45
References	47

SECTION ONE: THE PROBLEM

The focal point of this memo is the development of an effective procedure for solving large, singular, linear systems defined by a particular class of sparse matrices --- those that are "almost block tridiagonal." The work was motivated originally by the need to process and interpret data from a survey carried out on the Fermilab Main Ring, but the procedure that evolved is sufficiently general to be used in a wider class of applications. We will begin with an overview of the original survey problem in order to motivate the direction of the mathematics that follows.<sup>†</sup>

A survey is carried out on a set of markers, its objective being to discover where they are. The position of each marker possesses three degrees of freedom. However, using (constant, local) gravity as a pointer, it is possible to decouple the "vertical" dimension from the "horizontal" ones and handle it separately, which is advantageous, since the vertical problem is a much easier one. (See Appendix A.) In the body of this memo we shall assume that this has been done and constrain the markers to lie in a (horizontal) plane. This is not a real restriction; the formalism can be generalized without difficulty to any number of dimensions.

"Ring" will refer to the set of all markers being surveyed, its "state" to the ordered set of their positions, and "state space" to the set of all possible states. The positions will be labelled  $P_k$

---

<sup>†</sup>Some of the material in the first two sections has been repeated from a previous internal memo for completeness. [1]

$k = 1 \dots N$  for a ring of  $N$  markers. The markers lie on the vertices of a polygon, which we shall take to be convex and, roughly, not much different from a circle. (These assumptions are not crucial.)

Orientation does not matter, but to be specific, assume that  $k$  increases clockwise. In our problem, both the ring and the survey data have a cyclic character, so it will be convenient to assume that the index set for the markers obeys a mod  $N$  arithmetic. As a consequence,

$$P_{k \pm N} = P_k.$$

The global state of the ring is to be estimated by patching together data from a number of local, relative measurements. The particular survey variables that were used --- sagittal offsets,  $D_k$ , and next-nearest neighbor (NNN) chords,  $a_k$  --- are illustrated in Fig. 1 along with their indicial labelling. A survey sample consists of a list

$$M = (D_1, a_1; D_2, a_2; \dots; D_N, a_N)$$

of  $2N$  real numbers. Sagittal offsets were deemed attractive as survey variables because they give direct information about transverse displacements, which have the greatest influence on closed orbits in the accelerator. NNN chords are logically their complementary variables, as they are orthogonal to the sagittas. Nonetheless, there is nothing sacred about this choice; these variables could be replaced or supplemented by alternatives such as nearest neighbor (NN) chords,  $x_k$ , or angles,  $\phi_k$ , between lines of sight. A critique of the survey process is not our concern here; rather, the question we address is, given these particular data, what can one do with them? We seek a procedure that will accept a data sample as input and write the state of the ring corresponding to it as output.

The first and obvious point to be made is that because all the measurements are internal to the ring, any states that can be transformed into one another by rigidly rotating and translating the ring in its plane produce identical survey data. States that can be connected by such isometric transformations, or isometries, clearly form an equivalence class, which will be called a configuration. Each configuration is a three dimensional submanifold of state space generated by rotations of the ring about some point and translations in two orthogonal directions. A state therefore possesses three degrees of freedom unobservable to the data, which means that configurations can be specified by  $2N-3$  numbers.

Let us consider a naive, straightforward approach for constructing the configuration associated with a data sample, if only to point out that it does not work. Begin by placing  $P_1$  (arbitrarily) at the origin. To keep things simple, assume that the NN chord  $x_1$  has been appended to the data and place  $P_2$  at  $(0, x_1)$ . With  $P_1$  and  $P_2$  in place, knowledge of  $(D_2, a_2)$  enables  $P_3$  to be positioned to within a reflection about the y-axis, which is taken care of provided we know the orientation of the ring. And so forth: the procedure is iterated around the ring, each  $P_k$  being placed by using its two predecessors and  $(D_{k-1}, a_{k-1})$ . When the last point,  $P_N$ , is laid, the job is finished.

This scheme would work if measurements could be carried out with infinite accuracy and precision. The problem is that the presence of errors will prevent the configuration from closing. This is observed as a discrepancy between values of the final, unused numbers  $(D_N, a_N)$  and  $(D_1, a_1)$  as measured and as predicted from the construction. The

discrepancy is not small; errors of about 0.1" in survey data will induce positional errors of meters in the placement of the last site, even in rings with only moderately large  $N$ .<sup>†</sup> Were it not for the requirement of closing the ring even in the presence of such errors, the procedure described above would be adequate. Because of it, we must try something else.

In starting the "naive" solution, a NN chord was added to the data sample in order to lay  $P_2$ . This was not done merely to expedite the proceedings. Neglecting closure amounts to deleting two successive sagittas and NNN chords from the data set --- in our example these were  $(D_N, a_N)$  and  $(D_1, a_1)$ . This leaves only  $2N-4$  numbers, which is one fewer than required to specify a configuration. This implies that without the NN chord and without closure there is a new continuous symmetry transformation, a way of deforming the ring without altering data. As illustrated in Fig. 2, it goes like this. (a) Consider  $P_1$  and  $P_3$  fixed. (b) Move  $P_2$  parallel to  $P_1P_3$ , so that  $(D_2, a_2)$  does not change. Of course, this changes the other variables, but we will correct for that. (c) Using  $P_2$  as a center, move  $P_4$  in a circle of radius  $a_3$  until  $D_3$  is correct. (d) Then, using  $P_3$  as center, move  $P_5$  in a circle of radius  $a_4$  until  $D_4$  is correct. (e) And so forth: continue adjusting the ring one marker at a time until all the survey data have been reproduced. The addition of a single NN chord breaks this symmetry, since sliding the points past one another would change its value. Closure provides the extra constraint, and a strong one at that, that allows one to solve for a configuration using only sagittas

---

<sup>†</sup>Partially, such a large effect is due to the survey's not being a redundant triangulation.

and NNN chords.

### 1.1 Conceptual Overview.

Before going into mathematical details, it will be useful to create a conceptual framework in which to visualize essential features of our problem. This will make it easier to separate its various levels; especially, we want (a) to draw clear distinctions between those processes which require human judgment and the purely mechanical ones and (b) to trace the connections between its linear and nonlinear computational aspects. In doing this, much of what will be said may seem pedantic, but this problem possesses characteristics which should be pointed out, as they significantly affect what one accepts as a "solution." Besides, since we ultimately want to write a procedure --- that is, an algorithm --- all the "obvious" things must eventually be spelled out in great detail anyway. It will be helpful to refer liberally to Fig. 3 during the discussion.

To begin with, we are dealing with two "spaces:" the first, call it  $W$ , is the state space of the ring; the second, call it  $M$ , is identical to  $R^{2N}$ , the space of all  $2N$ -tuples of real numbers. Both spaces are manifestly  $2N$  dimensional ---  $W$  is so because there are  $N$  markers in the ring, and by fiat, each is constrained to lie in a plane. An ideal survey can be thought of as mapping  $\phi: W \rightarrow M$  which assigns to each state the data that the survey would produce if the ring were in that state. Now, it is trivial but nonetheless important that the dimension of the data manifold, defined as the image of  $W$  under  $\phi$ ,  $\phi(W) = \{\phi(s), s \in W\}$ , is smaller than  $2N$ , because  $\phi$  is not a one-to-one mapping. More exactly, because configurations are three dimensional,

the data manifold is  $2N-3$  dimensional --- equal to the number of parameters required to fix a configuration.<sup>†</sup>

A procedure for processing survey data must essentially invert all of this: given some  $M$  on the data manifold, it must find a configuration, the class of states  $\phi^{-1}(M)$  that map into it under  $\phi$ . Of course, only a single member of this class need be presented explicitly, the others being obtainable from it via the isometries. The chosen element is thus a representative of its equivalence class, and the solution can be written formally:  $s = \text{rep}(\phi^{-1}(M))$ . Since all members of  $\phi^{-1}(M)$  are equivalent, any one of them can be chosen as its representative. To facilitate this decision, it is useful to single out one state,  $s^*$ , as "special" --- call it the design state. It serves as a zeroth order estimate of the state of the ring. If we now impose a metric,  $\rho_W$ , on  $W$ , then  $\text{rep}(\phi^{-1}(M))$  can be taken to be that member of  $\phi^{-1}(M)$  that is closest to  $s^*$ . The metric  $\rho_W$  can be chosen arbitrarily; generally, the choice reflects one's assessment of "penalties" or "costs" associated with interpreting markers as having moved from their design positions. The important point here is that it really does not matter what metric is used. All representatives of a class contain the same information; the solutions presented via different metrics will be equivalent.

---

<sup>†</sup>Note that  $W$  can be naturally given a vector space structure, but  $\phi(W)$  cannot. For example, although  $\phi(W)$  contains the point  $(0, 0, 0, \dots, 0)$ , corresponding to a degenerate ring in which all points have coalesced into one, it does not contain points with negative coordinates.

It is not sufficient to have a procedure that correctly maps the data manifold back into state space. Because of measurement errors, it is probable that the actual data one has to work with does not even lie in  $\phi(W)$ . Thus, we must enlarge the domain of the inversion to include these "off-shell" points as well. Because the problem no longer admits a legitimate solution, the analyst is free to choose any approach that makes sense, but the one that is (almost) invariably preferred follows the familiar "least squared error" strategy. First, one endows  $R^{2N}$  with a metric structure, say  $\rho_M$ ; then, given some  $M \in R^{2N}$  one finds that point  $M'$  on the data manifold that is closest to it;† finally, one defines  $w = \text{rep}_{s^*}(\phi^{-1}(M'))$  where the  $s^*$  subscript on rep makes explicit its dependence on the design. The usual "weighted least squares" metric, in which measurements variables are weighted according to their variances, is a natural one to use. It is usually adequate provided that measurement errors are unbiased and uncorrelated. However, unlike the situation in state space, different metrics here give rise to inequivalent solutions, so its choice has a real significance.

This complete the specification of a solution procedure, at least on the descriptive level. To recapitulate quickly, the survey problem is formally stated thusly: given  $M \in R^{2N}$ , find  $M' \in \phi(W)$  and  $s \in W$  such that

---

† In principle there may be more than one; in practice, this is extremely improbable.

$$\rho_M (M', M) = \inf_{M'' \in \phi(W)} \rho_M (M'', M)$$

$$s = \text{rep}_{S^*} (\phi^{-1} (M')).$$

A solution is presented as the pair (s, M').

We must now consider how to go about computing (s, M') in a practical way. As with a great many other nonlinear problems, the first step is to linearize it, in this case by expanding  $\phi$  to first order about the design state. Symbolically, if we think of  $\phi$  as written in a Taylor series about  $s^*$ ,

$$\phi: \text{data} = \text{data}^* + \left. \frac{\delta(\text{data})}{\delta(\text{state})} \right|_* \delta(\text{state}) + \dots,$$

then  $d\phi$  is the first term, mapping differential variations  $\delta(\text{state})$  into differential variations  $\delta(\text{data})$ ,

$$d\phi: \delta(\text{data}) \equiv \left. \frac{\delta(\text{data})}{\delta(\text{state})} \right|_* \delta(\text{state}).$$

This is shown graphically in Fig. 3.  $d\phi$  is a linear mapping from the state space onto the tangent space over the data manifold at  $M^* = \phi(s^*)$ . In contemporary notation,

$$d\phi: W \rightarrow T_{M^*} \phi(W).$$

By introducing local coordinates,  $\delta(\text{data})$ ,  $\delta(\text{state})$ , and  $d\phi$  are represented as  $2N \times 1$ ,  $2N \times 1$ , and  $2N \times 2N$  matrices,  $m, w$ , and  $\Omega$ , and this mapping gives rise to the local equation

$$m = \Omega w. \tag{1}$$

We shall call  $\Omega$  the geodetic matrix of the design.

The linear problem has the same structure as the nonlinear one but a different realization. The equivalence classes are now (hyper) planes parallel to  $T_{S^*} \phi^{-1}(M^*)$ , and the representative of each class is the point of its closest approach to  $s^*$ . As before, and for the same reasons, Eq. (1) will usually have no solution for  $w$ . Differential isometries imply that all vectors in  $T_{S^*} \phi^{-1}(M^*)$  are annihilated by  $\Omega$ . In matrix terminology, the subspace of states satisfying  $\Omega w = 0$  is called the null space of  $\Omega$ , written null( $\Omega$ ). It is obvious that null( $\Omega$ )

is itself a sub-vector space of  $W$  --- that is, it is closed under the usual vector space operations.<sup>†</sup> Its orthogonal space in  $W$ ,  $D(\Omega) = (\text{null}(\Omega))^\perp$ , is called the domain of  $\Omega$ . To complete the terminology, the range of  $\Omega$  is identified with  $T_{M^*}\phi(W)$ ,

$$R(\Omega) \cong T_{M^*}\phi(W) \cong d\phi(W) \cong \{\Omega w, w \in W\}.$$

(We are not being purists and will not make clear distinctions between equality and morphism. They are unimportant here. The notation could be cleaned up if need be.) The linear problem statement is essentially unchanged from the non-linear one: given  $m$ , find  $w \in D(\Omega)$  and  $m' \in R(\Omega)$  such that

$$|m' - m| = \inf_{m'' \in R(\Omega)} |m'' - m|. \quad (2)$$

The big difference is that this is now a linear problem, and its solution is hopefully more amenable to computation.

The condition (2) can be written algebraically

$$\Omega^T \Omega w = \Omega^T m. \quad (3a)$$

In addition, the condition " $w \in D(\Omega)$ " can be written algebraically,

$$\forall e \in \text{null}(\Omega): e^T w = 0. \quad (3b)$$

The state  $w$  that solves the system (3) is the image of a linear mapping of  $M$  back into  $W$ . It can be represented by a matrix equation  $w = \Omega^\ddagger m$ . The matrix  $\Omega^\ddagger$  is called the Moore-Penrose pseudoinverse of  $\Omega$ . [2].<sup>††</sup> In particular, if basis states are chosen in such a way that  $\Omega$  is partitioned

$$\Omega = \left( \begin{array}{c|c} \Lambda & 0 \\ \hline 0 & 0 \end{array} \right) \quad (4)$$

<sup>†</sup>In algebraic terms, thinking of  $\Omega$  as a morphism between vector spaces, null( $\Omega$ ) is identified with its kernel.

<sup>††</sup>This notation is not standard. Mathematicians indicate a pseudoinverse as  $\Omega^\dagger$ , but to a physicist this means a Hermitian conjugate --- which mathematicians symbolize as  $\Omega^H$ .

where  $A$  is non-singular, then  $\Omega^{\dagger}$  is given by

$$\Omega^{\dagger} = \left( \begin{array}{c|c} A^{-1} & 0 \\ \hline 0 & 0 \end{array} \right) \quad (5)$$

as can be checked by substitution into (3).

By solving the linearized survey problem, we can derive a state, say  $s_0 = s^* + w$ , that is an approximate solution to the original, non-linear problem, but one that is usually not good enough. Suppose that  $\phi: s_0 \rightarrow M_0'$ . It would be fortunate indeed if  $M_0' = M'$ , but this is not to be expected. Nonetheless, it is possible to obtain the complete non-linear solution  $(s, M')$  by employing a variant of Newton's method for finding zeroes of an arbitrary, differential function, that is, iterating the linear algorithm. At each iteration, the pseudoinverse solution of the previous one is used to specify a new "design" state, which is then used to set up the next linear problem. The sequence of linear solutions generated by these iterations will converge to  $(s, M')$  --- provided that it converges at all, which is normally the case --- because it is a fixed point of the iterative scheme. Iterations are stopped when changes in the state from one to the next are small enough to be considered negligible or become comparable to computational error.

## 1.2 Frames and Coordinates.

To do calculations a coordinate grid must be introduced over  $W$ . The obvious approach would be to use a global Euclidean frame in the plane of the ring, but this would turn out to be a little awkward for calculations. Instead, we use the design state,  $s^*$ , as a reference. Hereafter, the term "site" shall refer to the position of a marker for a ring in the design state. At each site place a local  $(u, v)$  frame  $F_k$

oriented so that one axis, say the  $v$  axis, is aligned along  $\overrightarrow{P_{k-1} P_{k+1}}$ . The other, the  $u$  axis, is orthogonal to it and points outward, as shown in Fig. 4. An arbitrary state of the ring is then coordinatized as the vector

$$w = (u_1, v_1; u_2, v_2; \dots u_N, v_N),$$

where  $(u_k, v_k)$  represents

the (local) Euclidean coordinates of  $P_k$  relative to  $F_k$ . Note the vector space structure over the reals. Any real linear combination of two of these represents another legitimate state. The point  $(0, 0, 0, \dots 0)$  represents  $s^*$ , the design.

It will be useful to introduce a block notation for vectors and matrices. We shall use a subscripted semicolon to separate "outer" from "inner" indices: outer indices will refer to the sites; inner, to the coordinates. This induces a partitioning of the  $2N$  dimensional state vector into  $N$  two dimensional blocks  $w_{1;k} = u_k, w_{2;k} = v_k$ . If an index (normally the inner one) is suppressed, then the corresponding sub-array is represented. Suppression is done either by blanking or inserting a dot into an index's location. Thus,

$$w_{;k} = w_{.;k} = \begin{pmatrix} u_k \\ v_k \end{pmatrix} = \underline{\text{col}} (u_k, v_k). \quad (\text{Here "col" serves as a reminder that } w \text{ is a } 2N \times 1 \text{ column matrix, although it is written as a row matrix.})$$

Matrices are partitioned in the naturally induced manner:

$\Omega_{ij;k\ell}$ ,  $i, j = 1 \dots 2, k, \ell = 1 \dots N$ . Thus,  $\Omega_{;k\ell}$  represents a  $2 \times 2$  block,  $\Omega_{.j;k\ell}$  a  $2 \times 1$  array, and  $\Omega_{i.;k\ell}$  a  $1 \times 2$  array.

Evaluation of the geodetic matrix relative to the local frames  $\{F_k\}$  was carried out in UPC-140 [1]; here we shall content ourselves with quoting the result.

$$\Omega_{;ik} = \tilde{\omega}^{(i)} \delta_{i-1 k} + Q^{(i)} \delta_{ik} + \omega^{(i)} \delta_{i+1 k} \quad (6)$$

$$\omega^{(k)} = \begin{pmatrix} \cos\theta_{k+1} & -\sin\theta_{k+1} \\ \sin\theta_{k+1} & \cos\theta_{k+1} \end{pmatrix},$$

$$\tilde{\omega}^{(k)} = \begin{pmatrix} (\bar{\alpha}_k/\alpha_k) \cos\theta_k & (\bar{\alpha}_k/\alpha_k) \sin\theta_k \\ \sin\theta_k & -\cos\theta_k \end{pmatrix},$$

$$Q^{(k)} = \begin{pmatrix} -1/\alpha_k & 0 \\ 0 & 0 \end{pmatrix}.$$

The variables  $\theta_k$  and  $\alpha_k$  are defined in Fig. 5:  $\theta_k$  is the (absolute value of the) angle through which  $F_{k-1}$  must be turned in order to align it parallel to  $F_k$ ;  $\alpha_k$  is a dimensionless ratio of distances, and  $\bar{\alpha}_k = 1 - \alpha_k$ . Note that  $\omega(k)$  is a rotation matrix that takes  $F_k$  into  $F_{k+1}$ . Let us introduce  $2 \times 2$  frame arrays

$$F_{;k} = (\underline{u}_k \quad \underline{v}_k) \quad (7)$$

where  $\underline{u}_k$  and  $\underline{v}_k$  represent the local frame vectors relative to some global frame. Then,

$$F_{;k+1} = F_{;k} \omega(k) \quad (8)$$

is the connection between two successive frame arrays.

The linear system (1) written in block form becomes

$$m_{;k} = \tilde{\omega}^{(k)} w_{;k-1} + Q^{(k)} w_{;k} + \omega^{(k)} w_{;k+1},$$

$$m_{;k} = \begin{pmatrix} -dD_k/\alpha_k \\ da_k \end{pmatrix},$$

Note that the sagittal differentials have been rescaled according to  $\alpha_k$ . This was done to make  $\tilde{\omega}^{(k)}$  orthogonal for symmetric rings

( $\alpha_k = \bar{\alpha}_k = 1/2$ ). Note also that the only length scale to appear comes from the measurement differentials themselves.

The matrix  $\Omega$  has been set up so that its pseudoinverse is correct for the Euclidean metrics

$$\rho_M: |m-m'| \equiv \left( \sum_{k=1}^{2N} (m_k - m'_k)^2 \right)^{\frac{1}{2}}$$

$$\rho_W: |w-w'| \equiv \left( \sum_{k=1}^{2N} (w_k - w'_k)^2 \right)^{\frac{1}{2}}$$

If one wants to change  $\rho_M$  to another quadratic metric represented by a metric tensor  $g$ , then the condition  $\min_w |\Omega w - m|$  goes over to

$$\min_w |g^{\frac{1}{2}}(\Omega w - m)| = \min_w |g^{\frac{1}{2}}\Omega w - g^{\frac{1}{2}}m|. \text{ If } g^{\frac{1}{2}} \text{ is block diagonal, then the}$$

structure of  $\Omega$  is unchanged, and the problem is essentially the same as before. One merely replaces  $\Omega \rightarrow g^{\frac{1}{2}}\Omega$ ,  $m \rightarrow g^{\frac{1}{2}}m$ . The metric over  $W$  has no effect on the content of the final results, as was already noted.

Changing it will not alter  $\phi^{\ddagger}(M)$ , but it can alter  $d\phi^{\ddagger}(M)$ ;

that is, different sequences of linear solutions may be generated, but they should converge to the same limit.

### 1.3 Singular Value Decomposition.

One method frequently employed for calculating the pseudoinverse of a matrix is singular value decomposition (SVD). Although this approach is not particularly efficient when  $N$  is large --- primarily because of such things as size and efficiency of the algorithm --- the basic ideas involved will be important later for suppressing undesirable features introduced into the solution by random noise in the data. Because SVD is not commonly used by physicists, it is worthwhile to go through it in some detail. To begin with, we must

define what is meant by the singular values and singular vectors of an operator.

Definition: Let  $E$  and  $F$  be vector spaces,  $\Omega: E \rightarrow F$  be a linear map (operator), and  $\Omega^\dagger$  its adjoint. Then,  $u \in E$  and  $v \in F$  are singular vectors of  $\Omega$  with singular value  $\lambda \in \mathbb{R}^+$  if and only if  $\Omega u = \lambda v$  and  $\Omega^\dagger v = \lambda u$ .

We will denote this association as  $\Omega: u \xrightarrow{\lambda} v$ . The language of the definition was employed so as to make it as general as possible. Note that singular values are real and non-negative. The basic theorem on SVD, whose very simple proof is offered below, is an extension to arbitrary matrices of the principal axis theorem, familiar from quantum mechanics, which states that any (finite) Hermitian matrix can be diagonalized by a unitary transformation. [3]<sup>†</sup> Essentially, it states that all operators possess singular vectors and that they can be chosen to form complete orthonormal bases in  $E$  and  $F$ . In this way, the action of any operator can be decomposed into a direct sum of real one-dimensional actions. The proof is motivated by the observation that if  $\Omega: u \xrightarrow{\lambda} v$ , then it follows that  $\Omega^\dagger \Omega u = \lambda^2 u$ . The trick is to take the converse. We will treat only the finite dimensional case.

THEOREM: (Singular value decomposition.) Let  $E$  and  $F$  be finite dimensional vector spaces, with  $\dim(E) = n$  and  $\dim(F) = m$ , and let  $\Omega: E \rightarrow F$  be a linear mapping. There exist complete orthonormal bases

---

<sup>†</sup>More generally, any normal matrix --- one that commutes with its adjoint --- is unitary equivalent to a diagonal matrix.

$B_E \equiv \{u_i \mid i = 1 \dots n\} \subseteq E$ ,  $B_F \equiv \{v_i \mid i = 1 \dots m\} \subseteq F$ , such that, for each  $i = 1 \dots n$ , either  $u_i \in \text{null}(\Omega)$  or  $\Omega:u_i \xrightarrow{\lambda_i} v_i$ , for some  $\lambda_i \in \mathbb{R}^+$ .

PROOF: Because  $E$  is finite dimensional, the Hermitian operator  $\Omega^\dagger \Omega: E \rightarrow E$  possesses a complete (in  $E$ ) orthonormal set of eigenvectors whose eigenvalues are real. Further, since  $\Omega^\dagger \Omega$  is positive semi-definite, the eigenvalues must be non-negative. Define  $B_E$  to be this complete set, and define  $\lambda_i$  by the relation  $\Omega^\dagger \Omega u_i = \lambda_i^2 u_i$ ,  $\lambda_i \in \mathbb{R}^+$ .

If  $\lambda_i = 0$ , then  $u_i \in \text{null}(\Omega)$ . For each  $u_i \in \text{null}(\Omega)$  define

$v_i \equiv (1/\lambda_i) \Omega u_i$ . Then  $\Omega^\dagger v_i = (1/\lambda_i) \Omega^\dagger \Omega u_i = \lambda_i u_i$ , so that  $\Omega:u_i \xrightarrow{\lambda_i} v_i$ .

Finally, the scalar products  $(v_k, v_m)$  are given by

$$\begin{aligned} (v_k, v_m) &= \frac{1}{\lambda_k \lambda_m} (\Omega u_k, \Omega u_m) \\ &= \frac{1}{\lambda_k \lambda_m} (u_k, \Omega^\dagger \Omega u_m) \\ &= \delta_{km}, \end{aligned}$$

proving that  $\{v_k \mid k = 1 \dots p\}$ ,  $p = \dim(E) - \dim(\text{null}(\Omega))$ , forms an orthonormal set. By itself, this set already spans a subspace,  $\Omega E = \text{range}(\Omega) \subseteq F$ . If necessary, it can be completed to form  $B_F$  by appending a set of orthonormal vectors spanning the orthogonal complement  $(\Omega E)^\perp$ . -QED-

Note that the only reason for making  $E$  finite dimensional was to assure that the eigenvectors of  $\Omega^\dagger \Omega$  spanned  $E$ . If this is added as an extra condition in the hypothesis of the theorem, then  $E$  can be an infinite dimensional space.

There are two simple corollaries to this theorem:

Corollary 1: In terms of representations, if  $\Omega$  is a  $m \times n$  complex matrix, then there exist  $m \times m$  and  $n \times n$  unitary matrices  $V$  and  $U$  which

satisfy  $\Omega = V\Omega_D U^\dagger$ , where (a) all off-diagonal elements of  $\Omega_D$  are zero, and (b) all diagonal elements are non-negative.

Proof: Simply fill the columns of U and V with the appropriate representation of the singular vectors,  $U_{\bullet j} \equiv u_j$ ,  $V_{\bullet j} \equiv v_j$ . Unitarity follows from their orthonormality and completeness. Of course, the non-zero elements of  $\Omega_D$  are the singular values  $\lambda_i$ , and  $\Omega u_i = \lambda_i v_i \leftrightarrow \Omega U = V\Omega_D$ . -QED-

Corollary 2: If  $\Omega$  is real, then its singular vectors can also be taken to be real.

Proof: It is easily seen that since  $\lambda = \lambda^*$ ,  $(\Omega:u \xrightarrow{\lambda} v)$  implies  $(\Omega:u^* \xrightarrow{\lambda} v^*)$  when  $\Omega^* = \Omega$ . Thus, since the mapping is linear,  $\Omega:\text{Re}(u) \xrightarrow{\lambda} \text{Re}(v)$  and  $\Omega:\text{Im}(u) \xrightarrow{\lambda} \text{Im}(v)$ . At least one of this pair must be non-trivial, so that we can use it (or them) as the real singular vector(s) associated with  $\lambda$ . -QED-

The content of the SVD theorem is appealing. To employ a descriptive, physical terminology, we can think of the singular vectors as modes and the associated singular values as excitation factors. Then each mode in E uniquely excites a single mode in F with a "strength"  $\lambda$ . Further, because the modes are real, they can be associated in our problem with real, infinitesimal displacements of the ring.

By inverting the one dimensional relations, SVD can be used to calculate the pseudoinverse of a mapping. Clearly,  $\Omega_D^\dagger$  is easy to construct, using Eqs. (4) and (5). Then  $\Omega^\dagger = U\Omega_D^\dagger V^\dagger$  as can be verified by using the defining relations (3). As a procedure for getting  $\Omega^\dagger$ ,

this is not very efficient: one usually does not diagonalize a matrix in order to invert it. However, it does provide useful insights into the effects of measurement errors on the solution. Note that

$\Omega^{\dagger} : v_i \xrightarrow{\lambda_i^{-1}} u_i$ . The fact that the reciprocal of the singular values appear means that random noise uniformly distributed in data space will preferentially excite those modes in state space corresponding to the smallest singular values of  $\Omega$ . This turns out to be an important consideration. In our particular problem a few such "low order" modes are excited very strongly. The solutions thus tend to contain substantial coherent displacements which reflect nothing more than random noise in the data. These also represent modes of displacement to which the beam is least sensitive, and so it is not necessary to have detailed information about them. It is usually desirable to decrease their influence on the solution by either eliminating or suppressing them.

The difficulty with applying SVD, or any other "library" algorithm, to our problem is that the geodetic matrix tends to be large: in the Fermilab Main Ring, there are 204 survey markers which means that  $\Omega$  contains 166,464 elements even if we confine our attention to the horizontal plane alone. In addition, with matrices that large one can legitimately begin to worry about the accuracy of any general purpose algorithm, as well as the time involved. (Think of it in this way: we are essentially trying to invert a 408x408 singular matrix!)

There are two saving features which enable us to carry out computations in a civilized manner. First,  $\Omega$  is sparse: of its  $(2N)^2$  elements, only  $9N$  are not zero. The pseudoinverse is itself not sparse, of course, but we are not interested in it so much as in the pseudoinverse solution,  $w$ . (This is analogous to routines for solving nonsingular linear systems, like the Gauss-Seidel algorithm, in which one does not explicitly compute the inverse of the coefficient matrix.) The question then is whether any particular algorithm conserves sparseness in its intermediate steps. The difficulty with general purpose SVD procedures --- the Golub-Reinsch[4] algorithm, for example --- is that they do not; intermediate steps require  $O(N^2)$  storage locations. The second saving feature is that we can explicitly construct null( $\Omega$ ): it is spanned by vectors representing translations and rotations of the ring. This gives us an enormous advantage over the general case, and it would be wasteful not to take advantage of it.

Suppression of low order modes is a separate problem. In order to suppress these modes we must first compute them. The same difficulties regarding size aggravate the calculations in the large  $N$  limit. We are also somewhat hampered by the need to compute singular vectors with smallest, rather than largest, singular values. Only a few modes are required, however, and it seems reasonable to try a perturbation theoretic approach, starting from the solutions for symmetric rings to be developed in Sec. 2. Restricted eigenvalue problems for sparse matrices have been studied elsewhere.[7] Although we shall not address the question here, it is an important one which cannot be ignored in arriving at sensible answers from the data.

The major part of this memo deals with an algorithm that was developed to pseudoinvert the system (1) in a way that conserves sparseness in its intermediate steps, so that storage requirements grow like  $N$  rather than  $N^2$ . It turns out that there are two or three ways of actually doing this,<sup>†</sup> but subsidiary objectives of accuracy and speed dictated the choice that is presented here. Of course, we also require the algorithm to be completely general so that no a priori assumptions are made concerning the configuration of the ring. (This precludes being satisfied with the solution presented in the next section.)

<sup>†</sup>For example, an iterative refinement[2] beginning from  $\Omega^{\dagger} \sim \Omega^{\dagger}$ . It converged too slowly to be useful. Faster versions exist, but they are not sparse.

SECTION TWO: SYMMETRIC RINGS

There is a class of designs for which the linear problem is exactly solvable in closed form<sup>†</sup> via singular value decomposition, those consisting of a number,  $N$ , of sites symmetrically placed on a circle. We will denote these designs as  $C_N$ , just to give them a name. That they are so easy to handle, not surprisingly, has nothing to do with details of the problem; rather, it is a direct consequence of their high degree of symmetry. Before attacking the more general case, we will pause to go through this exercise. (Some of this section has already been presented, but from a different perspective, in UPC-140. [1])

It is easily checked from the explicit representation in Eq. (6) that the geodetic matrix  $\Omega$  of a  $C_N$  ring is block circulant,<sup>††</sup> that is,

$$\forall k, i, j : \Omega_{i+k, j+k} = \Omega_{i, j}.$$

It turns out that any such matrix can be

<sup>†</sup>The words "closed form" are ambiguous. The position that the procedure to be outlined in Sec. 4 is a "closed form" solution is certainly defensible: it is finite; it is exact; it includes no functions more complicated than sines and cosines. Yet, probably for psychological reasons, one thinks of it as a "numerical" rather than an "analytic" solution.

<sup>††</sup>Notice the usefulness here of representing  $\Omega$  relative to local frames. If an overall, global frame had been employed, this frame would itself break the symmetry,  $\Omega$  would not be a block circulant matrix --- although it would be unitarily equivalent to one --- and the analyses here and in Sec. 3 would be less transparent.

block diagonalized by a discrete Fourier transform. [5] To see this, define the unitary matrix  $S$  in block form,  $S_{;km} = N^{-1/2} l \eta^{km}$ ;  $\eta = \exp(2\pi i/N)$ ,  $l = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ . It then follows that

$$\begin{aligned}
 [S^\dagger \Omega S]_{;km} &= \sum_{\ell, j} \left( \frac{1}{\sqrt{N}} \eta^{*\ell k} \right) \Omega_{; \ell j} \left( \frac{1}{\sqrt{N}} \eta^{jm} \right) & (9) \\
 &= \frac{1}{N} \sum_{\ell, j} \eta^{-\ell k} \Omega_{; 0 \ j-\ell} \eta^{jm} \\
 &= \frac{1}{N} \sum_{\ell, q} \eta^{-\ell k} \Omega_{; 0q} \eta^{(q+\ell)m} \\
 &= \frac{1}{N} \left( \sum_{\ell} \eta^{\ell(m-k)} \right) \left( \sum_q \Omega_{; 0q} \eta^{qm} \right) \\
 &= \delta_{mk} \sum_q \Omega_{; 0q} \eta^{qm} \\
 &\equiv \delta_{mk} 2W(m) .
 \end{aligned}$$

The last line serves to define the matrix  $W(m)$ ; the factor of 2 has been introduced for convenience later. (Of course, the modular convention on site indices means that  $\Omega_{; 0q} = \Omega_{; Nq}$ .) Using Eq. (6) for  $\Omega_{; 0q}$ , evaluated with  $\alpha_k = 1/2$  and  $\theta_k = 2\pi/N$  for all  $k$ ,  $W(m)$  can be written explicitly,

$$W(m) = \begin{pmatrix} cc_m - 1 & -i ss_m \\ sc_m & i cs_m \end{pmatrix};$$

$$\begin{aligned}
 c &= \cos(2\pi/N) , & s &= \sin(2\pi/N) , \\
 c_m &= \cos(2\pi m/N) , & s_m &= \sin(2\pi m/N) .
 \end{aligned}$$

According to the discussion in the previous section, singular values of  $\Omega$  are to be identified with eigenvalues of  $\Omega^\dagger \Omega$ . Using Eq. (9) it is possible to decompose

$$\Omega^\dagger \Omega = S \left\{ \bigoplus_{m=0}^{N-1} 4W^\dagger(m)W(m) \right\} S^\dagger,$$

where  $\bigoplus$  signifies a matrix direct sum. It thus suffices to calculate the eigenvalues of  $W^\dagger(m)W(m)$  for all  $m = 0 \dots N-1$ , which is done easily enough. Let  $\lambda^2$  signify an eigenvalue of  $W^\dagger W$ . We then have the development

$$W^\dagger(m)W(m) = \begin{pmatrix} c_m^2 - 2cc_m + 1 & iss_m \\ -iss_m & s_m^2 \end{pmatrix} ;$$

$$\lambda_{\pm}^2(m) = \frac{1}{2} [T_m \pm \sqrt{T_m^2 - 4D_m}] ;$$

$$\begin{aligned} T_m &= \text{Tr} [W^\dagger(m)W(m)] \\ &= 2(1 - cc_m), \end{aligned}$$

$$\begin{aligned} D_m &= \det [W^\dagger(m)W(m)] \\ &= |\det W(m)|^2 \\ &= s_m^2 (c_m - c)^2 . \end{aligned}$$

With a little simplification the result is written

$$\lambda_{\pm}^2(m) = (1 - cc_m) \pm \sqrt{s^2 s_m^2 + c_m^2 (c_m - c)^2} \quad (10)$$

The corresponding singular values of  $\Omega$  are  $2\sqrt{\lambda_{\pm}^2(m)}$ . We note in passing that

$$\lambda^2(m) = \lambda^2(-m). \quad (11)$$

The expression (10) simplifies considerably in the limit  $N \rightarrow \aleph_0$ ,  $m \rightarrow \aleph_0$ ,  $m/N$  fixed.<sup>†</sup>

$$\begin{aligned} \lambda^2 &\approx \begin{cases} \sin^2(2\pi m/N) \\ 4 \sin^4(\pi m/N) \end{cases} , \\ 2\sqrt{\lambda^2} &\approx \begin{cases} 2 |\sin(2\pi m/N)| \\ 4 \sin^2(\pi m/N) \end{cases} . \end{aligned}$$

These are sketched in Fig. 6. There is a small region around  $m/N = 0$  in which these approximations are not valid, and that has been shaded

---

<sup>†</sup>This notation heralds an unprecedented achievement in hairsplitting. Aleph-null ( $\aleph_0$ ) represents the cardinality of the set of integers, and it always seemed clear to me that if an integer variable "approached infinity" it should approach aleph-null. A similar comment can be made about upper limits of infinite sums.

as a reminder. This limiting behavior means that all large  $C_N$  rings are basically the same, in a scaled way, except for very low harmonics.

The null space of  $\Omega$  is non-trivial provided that at least one of the singular values is zero. From Eq. (10),  $\lambda_-(0) = \lambda_-(1) = \lambda_-(-1) = \lambda_-(N/2) = 0$ , the last value occurring only when N is even. The first three --- vanishing at harmonic numbers 0 and  $\pm 1$  --- are related to the isometries, as was shown explicitly in UPC-140.[1] The last one is not; it represents an accidental, differential symmetry. Its corresponding singular vector (see below) represents alternating azimuthal displacements of the markers --- effectively looking as though a sub-ring consisting of every other site were rigidly rotated through an infinitesimal angle. (See Fig. 7.)

Finally, let us write out the singular vectors, which clearly will be associated with the eigenvectors of  $W^\dagger W$ . First define two dimensional vectors  $f_\pm(m)$  according to

$$W^\dagger(m)W(m) f_\pm(m) = \lambda_\pm^2(m) f_\pm(m),$$

or more explicitly

$$f_\pm(m) \sim \begin{pmatrix} iss_m \\ \lambda_\pm^2(m) - (c_m - c)^2 - s^2 \\ -iss_m \end{pmatrix} \sim \begin{pmatrix} \lambda_\pm^2(m) - s_m^2 \\ -iss_m \end{pmatrix}.$$

(We will ignore normalization here; once a singular vector is obtained, normalizing is a trivial step.) Now define  $g_\pm(m)$  to be the  $2N$  dimensional vector obtained by putting  $f_\pm(m)$  in the "correct" block.

$$g_\pm(m)_{;k} = \delta_{mk} f_\pm(m).$$

The singular vectors of  $\Omega$  are then given by  $e_\pm(m) = Sg_\pm(m)$ , as can be confirmed easily:  $\Omega^\dagger \Omega(Sg_\pm(m)) = S[\oplus_m 4W^\dagger(m)W(m)]g_\pm(m) = 4 \lambda_\pm^2(m) (Sg_\pm(m))$ .

Explicitly writing  $e_\pm(m)$  in block form,

$$\begin{aligned} \operatorname{Re}[e_{\pm}(m)]_{;k} &\sim \begin{pmatrix} -s_{km} s s_m \\ c_{km} [\lambda_{\pm}^2(m) - (c_m - c)^2 - s^2] \end{pmatrix} \\ \operatorname{Im}[e_{\pm}(m)]_{;k} &\sim \begin{pmatrix} c_{km} s s_m \\ s_{km} [\lambda_{\pm}^2(m) - (c_m - c)^2 - s^2] \end{pmatrix} \end{aligned} \quad (12)$$

The real and imaginary parts are offered as the singular vectors in accordance with Corollary 2 of Sec. 1.3. Of course, any linear combination would also serve.

At this point we shall drop the analysis of symmetric rings, although there are a good many more details that could be worked out, and return to the main theme, the construction of solutions for large, arbitrarily configured designs.

SECTION THREE: BASIS STATES OVER NULL SPACE

An important part of the procedure to be developed in Sec. 4 is the specification of a basis over  $\text{null}(\Omega)$  by constructing a maximal set of linearly independent vectors satisfying the null space condition,

$$\Omega e = 0 \quad (13)$$

If  $e$  is partitioned conformably to  $\Omega$ , then this can be written in block form

$$\tilde{\omega}(k)e_{;k-1} + Q(k)e_{;k} + \omega(k)e_{;k+1} = 0 \quad (14)$$

There must be at least three such vectors: two corresponding to translations in orthogonal directions, and the third to a rotation of the ring about some center. From consideration of  $C_{2N}$  designs we know that at least in these special cases there is one more, associated with alternating azimuthal displacements around the ring. The first thing that must be done is to convince ourselves that this exhausts the possibilities, that the dimensions of  $\text{null}(\Omega)$  never exceeds four. Fortunately, that is not difficult.

THEOREM: For all matrices  $\Omega$  given by Eq.(6),  $\dim(\text{null}(\Omega))$  is either three or four.

PROOF: Let  $D$  symbolize  $\dim(\text{null}(\Omega))$ . As noted above,  $D \geq 3$  is obvious, since  $\text{null}(\Omega)$  must contain the isometries. Now, reverting to the block format of Eq.(14),

$$e_{;k} = -\omega^{-1}(k-1) [\tilde{\omega}(k-1)e_{;k-2} + Q(k-1)e_{;k-1}] .$$

It is easy to see that the set of states which satisfy this equation for  $k = 3 \dots N$  is a vector space of dimension four, since  $e_{;1}$  and  $e_{;2}$  are arbitrary. But, it must contain  $\text{null}(\Omega)$ , which means that  $D$  can be at most four. -QED-

The key question that remains is whether other designs besides  $C_{2n}$  possess four dimensional null spaces. Numerical experiments

support the idea that none do, but as yet there is no satisfactory proof. For now, we shall adopt this as a conjecture and continue.

CONJECTURE: If  $\dim(\text{null}(\Omega)) = 4$ , then the design is  $C_{2n}$ , for some  $n$ .

Our next objective is to construct the basis states that correspond to isometries. Consider translations first. For a rigid translation of the ring, the displacements  $\delta P_k$  are identical, but the coordinates  $e_{;k}$  will differ because each  $\delta P_k$  is referenced to its own local frame. Since  $e_{;k}$  contains the coordinates of  $\delta P_k$  relative to  $F_k$ ,  $e_{;k+1}$  is related to  $e_{;k}$  simply by the rotation that takes  $F_k$  into  $F_{k+1}$  (see Eq. (8)):

$$e_{;k+1} = \omega^T(k) e_{;k} . \quad (15)$$

There are two linearly independent translations in the horizontal plane, so we are free to choose  $e_{;1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ , to construct a basis over the translation sector of  $\text{null}(\Omega)$ .

It now can be demonstrated formally that  $e$  satisfies Eq. (13). Substitute Eq. (15) into the block form (14) to get

$$\Omega e_{;k} = [\tilde{\omega}(k)\omega(k-1) + Q(k) + \omega(k)\omega^T(k)] e_{;k} . \quad (16)$$

Now,  $\omega(k)$  is orthogonal, and  $Q(k)$  is given by (6), which allows us to write

$$\begin{aligned} Q(k) + \omega(k)\omega^T(k) &= Q(k) + 1 \\ &= \begin{pmatrix} 1 & -1/\alpha_k & 0 \\ 0 & 0 & 1 \end{pmatrix} . \end{aligned}$$

Further,

$$\begin{aligned} \tilde{\omega}(k)\omega(k-1) &= \begin{bmatrix} (\bar{\alpha}/\alpha)c & (\bar{\alpha}/\alpha)s \\ s & -c \end{bmatrix}_{(k)} \begin{pmatrix} c & -s \\ s & c \end{pmatrix}_{(k)} \\ &= \begin{pmatrix} (\bar{\alpha}/\alpha)_{(k)} & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

Therefore, the rhs of (16) must be identically zero,

$$[\tilde{\omega}(k)\omega \dots \omega^T(k)] = \begin{bmatrix} \bar{\alpha}_k + \alpha_k - 1 & 0 \\ \alpha_k & 0 \\ 0 & 0 \end{bmatrix} \quad (17)$$

$$= 0 ,$$

which completes the demonstration.

Now consider rotations. To construct the corresponding basis vector it is necessary to select some global frame whose origin will be used as the center of rotation. Let  $(P_{1;k}, P_{2;k})$  be the position coordinates of  $P_k$  relative to this frame. Under an infinitesimal rotation, the coordinates of the displacement  $\delta P_k$  will be given by

$$\delta \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}_{;k} = \begin{pmatrix} P_2 \\ -P_1 \end{pmatrix}_{;k} \cdot \delta\theta = J \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}_{;k} \cdot \delta\theta , \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

where  $\delta\theta$  is the

differential rotation angle. To get  $e_{;k}$ ,  $\delta P_k$  must be referenced to  $F_k$ . To do this, it is convenient to use the frame matrix defined in (7)

$$e_{;k} = F_{;k}^T J p_{;k} \quad (18)$$

where the unimportant  $\delta\theta$  has been dropped from consideration.

The formal proof of Eq. (13) is similar to the one that was given for translations, although a little more involved. First note that  $\omega(k)$  is the rotation matrix connecting frames  $F_k$  and  $F_{k+1}$ . Combine Eq. (8) with Eq. (18) to get the development

$$\begin{aligned}
 \Omega e]_{;k} &= \tilde{\omega}(k) \omega(k-1) F_{;k}^T J_{P;k-1} \\
 &+ Q(k) F_{;k}^T J_{P;k} \\
 &+ \omega(k) \omega^T(k) F_{;k}^T J_{P;k+1} \\
 &= (-Q(k) - 1) F_{;k}^T J_{P;k-1} \\
 &+ Q(k) F_{;k}^T J_{P;k} \\
 &+ F_{;k}^T J_{P;k+1} \\
 &= F_{;k}^T J(p_{;k+1} - p_{;k-1}) \\
 &+ Q(k) F_{;k}^T J(p_{;k} - p_{;k-1}) \\
 &= \begin{bmatrix} u^T \\ v^T \end{bmatrix}_{;k} J(p_{;k+1} - p_{;k-1}) \\
 &+ \begin{bmatrix} -\frac{1}{\alpha_k} & u^T \\ \alpha_k & \end{bmatrix}_{;k} J(p_{;k} - p_{;k-1}) .
 \end{aligned}$$

[Eq. (17) was used in going from the second to the third line.] Consider the bottom row first. By definition of the frame  $F_k$ ,  $p_{;k+1} - p_{;k-1} \sim v_{;k}$ . Therefore,  $v_{;k}^T J(p_{;k+1} - p_{;k-1}) \sim v_{;k}^T Jv_{;k} = 0$ .

To evaluate the first row, we use (see Fig. 5)

$$(p_{;k} - p_{;k-1}) - \alpha_k (p_{;k+1} - p_{;k-1}) \sim u_{;k}$$

to get

$$\begin{aligned}
 u_{;k}^T J[(p_{;k+1} - p_{;k-1}) - \frac{1}{\alpha_k} (p_{;k} - p_{;k-1})] \\
 \sim -\frac{1}{\alpha_k} u_{;k}^T J u_{;k} \\
 = 0.
 \end{aligned}$$

This completes the demonstration that  $\Omega e]_{;k} = 0$ , and since  $k$  is arbitrary, Eq. (13) is verified.

Equations (15) and (18) amount to procedures for constructing a basis over null( $\Omega$ ) for all designs except  $C_{2N}$ , if our conjecture regarding this matter is correct. (To handle  $C_{2N}$ , simply adjoin the extra null vector.)

SECTION FOUR: EXACT SOLUTION FOR SPARSE MATRICES

The solution procedure written below takes advantage of sparseness so that its memory requirements grow linearly with N, rather than quadratically. The context of the problem is no longer important; that is, the specific definitions of Eq. (6) can be forgotten; only the structure of  $\Omega$  need be assumed. The matrices  $\tilde{\omega}$ , Q, and  $\omega$  are square but can be of arbitrary size. It is necessary to assume that  $\omega$  and  $\tilde{\omega}$  are not singular.

Our strategy follows from the idea that the pseudoinverse solution can be identified with the least squared error fit to data that contains no component in null( $\Omega$ ). These conditions were expressed algebraically in Eqs.(3), repeated here.

$$\Omega^T \Omega w = \Omega^T m \equiv b, \tag{19a}$$

$$\forall e \in B: e^T w = 0. \tag{19b}$$

Rules for computing B, a basis over null( $\Omega$ ), have been detailed in Sec. 3.

The first observation to be made is that  $\Omega^T \Omega$  is itself sparse, although it contains five block columns in each block row, as opposed to only three in  $\Omega$ . Define the set of 2x2 matrices Z(i,r), i = 1...N, r = -2...+2, so that

$$\Omega^T \Omega_{;ik} = \sum_{r=-2}^{+2} Z(i,r) \delta_{i+r k}$$

Because  $\Omega^T \Omega$  is symmetric,  $Z(i,-r) = Z^T(i-r,r)$ , so that only those matrices for non-negative r need be computed and stored. Using Eq. (6) it is easy to develop the expressions for Z(i,0...2).

$$\begin{aligned} Z(i,0) &= \tilde{\omega}^T(i+1)\tilde{\omega}(i+1) + Q^T(i)Q(i) + \omega^T(i-1)\omega(i-1), \\ Z(i,1) &= \tilde{\omega}^T(i+1)Q(i+1) + Q^T(i)\omega(i), \\ Z(i,2) &= \tilde{\omega}^T(i+1)\omega(i+1). \end{aligned}$$

Now, the obvious thing to do is to recast Eq's. (19a) in terms of these Z's. Partitioning the arrays b and w into block form, this is written  $\sum_{r=-2}^{+2} Z(i,r)w_{;i+r} = b_{;i}$ . By rearranging terms a little, using the assumption that  $\omega$  and  $\tilde{\omega}$  are not singular, we can further express this as

$$w_{;k} = z^{-1}(k-2,2) [b_{;k-2} - \sum_{r=-2}^{+1} Z(k-2,r)w_{;k-2+r}]. \quad (20)$$

To ease the notation slightly, let us define the new arrays

$$f_{;k} \equiv z^{-1}(k-2,2)b_{;k-2},$$

$$z(k,j) \equiv -z^{-1}(k-2,2) Z(k-2,2-j), \quad j = 1, \dots, 4,$$

so that Eq.(20) takes on the form

$$w_{;k} = f_{;k} + \sum_{j=1}^4 z(k,j)w_{;k-j}. \quad (21)$$

This represents progress: Eq. (21) is transparently a fourth order recursion for the  $w_{;k}$ 's. All that remains is to solve for the "seeds"  $w_{;1\dots 4}$ ; the rest of the solution can then be generated from them.†

In order to develop a linear system for the seeds alone, we will construct a set of  $2 \times 1$  arrays  $d_{;k}$  and  $2 \times 2$  matrices  $x_{;k\ell}$ ,  $k=1\dots N$ ,

---

†Of course, any four consecutive w's could serve as the seeds, because of the cyclic conditions. Choosing  $w_{;1\dots 4}$  is arbitrary.

$\ell=1\dots 4$ , such that

$$\forall k=1\dots N : w_{;k} = d_{;k} + \sum_{\ell=1}^4 x_{;k\ell} w_{;\ell} \quad (22)$$

(The key point here is not that such matrices exist but that one can define them constructively. For example, it would be possible to set all  $x_{;k\ell} = 0$  and  $d_{;k} = w_{;k}$ , but that would not be a constructive algorithm, and therefore not useful.) To begin, for  $k=1\dots 4$  set

$$\begin{aligned} d_{;k} &= 0 \\ x_{;k\ell} &= \mathbf{1} \delta_{k\ell}, \quad \ell=1\dots 4. \end{aligned} \quad (23)$$

The rest are defined so as to assure the validity of Eq. (21) for  $k=5\dots N$ . Substituting from Eq. (22) into Eq. (21),

$$\begin{aligned} d_{;k} + \sum_{\ell=1}^4 x_{;k\ell} w_{;\ell} &= f_{;k} + \sum_{j=1}^4 z(k,j) w_{;k-j} \\ &= f_{;k} + \sum_{j=1}^4 z(k,j) \left( d_{;k-j} + \sum_{\ell=1}^4 x_{;k-j\ell} w_{;\ell} \right) \\ &= \left[ f_{;k} + \sum_{j=1}^4 z(k,j) d_{;k-j} \right] \\ &\quad + \sum_{\ell=1}^4 \left[ \sum_{j=1}^4 z(k,j) x_{;k-j\ell} \right] w_{;\ell} , \end{aligned}$$

which provides the rules

$$d_{;k} = f_{;k} + \sum_{j=1}^4 z(k,j) d_{;k-j} , \quad (24)$$

$$x_{;k\ell} = \sum_{j=1}^4 z(k,j) x_{;k-j\ell} .$$

To be formally complete, this development should be traversed backwards, beginning from Eq's. (23) and (24), to prove inductively that Eq. (22) is indeed satisfied.

A linear system for the seeds is now obtained by demanding that Eq. (21) be satisfied for  $k=1\dots 4$ . Substitute from Eq. (22) into Eq. (21) for  $w_{;k-j}$ , transfer all terms containing w's to the lhs, absorb  $w_{;k}$  into the sum over  $\ell$ , and interchange the double sums. The result is

$$\begin{aligned} \forall k=1\dots 4: \quad & \sum_{\ell=1}^4 \left[ \delta_{k\ell} \mathbf{1} - \sum_{j=1}^4 z(k,j) x_{;k-j} \right] w_{;\ell} \\ & = f_{;k} + \sum_{j=1}^4 z(k,j) d_{;k-j} . \end{aligned}$$

This is a system of eight equations in eight unknowns. However, the null space conditions (19b) have not yet been taken into account, so the system is singular: it possesses multiple solutions. To select the one that we desire, it is expedient to substitute from (22) into (19b) to obtain

$$\sum_{\ell=1}^4 \left( \sum_{k=1}^N e_{;k}^T x_{;k\ell} \right) w_{;\ell} = - \sum_{k=1}^N e_{;k}^T d_{;k} = -e^T d.$$

According to the conjecture of the previous section, there will usually be only three such equations yielding a system of eleven consistent equations in eight unknowns. To solve this last system one can delete three equations and invert an eight dimensional subsystem or find the pseudoinverse solution of the full set, using standard algorithms --- a procedure that is more symmetric than the other in that it does not single out equations for deletion. Because the eleven equations are

consistent, there is one and only one solution, and it is equal to their pseudoinverse solution. In fact, if we write the seed system symbolically as

$$\underline{\text{lhs}} \cdot \underline{w}_{;1\dots 4} = \underline{\text{rhs}}, \quad (25)$$

then its (unique) solution can be written

$$\underline{w}_{;1\dots 4} = (\underline{\text{lhs}}^T \underline{\text{lhs}})^{-1} \underline{\text{lhs}}^T \underline{\text{rhs}}$$

in accordance with the usual way of calculating least-squared-error fits.

This completes the description of the procedure. If computer memory were all that one had to be concerned with, our task would be finished, apart from programming. Unfortunately, the accuracy of the solutions constructed by the above procedure degrades rather quickly with N. This is undoubtedly connected with the observation that the condition number of the system (25) tends to be large, typically  $\approx 10^8$ . Interestingly, this is much larger than the condition number of the original system (19), which is typically  $\approx 10^6$ .

(Here, the condition number is taken to be the ratio of maximum to minimum non-zero singular values. Obviously, the actual condition number of a singular system is infinite, but that reflects degrees of freedom that we are not interested in.) This is an interesting phenomenon in its own right, but for our purposes it is primarily a nuisance. For  $N \approx 200$ , the procedure as outlined calculates the pseudoinverse only to about ten percent accuracy on a 60 bit machine.

Fortunately, it is not difficult to improve accuracy without losing sparseness. This can be done analogously to a common method for iteratively refining the solutions of non-singular systems. First we show that corrections can be made to obey the original system (19). Let  $w$  represent the true solution and  $\tilde{w}$  an approximate solution

obtained by numerical procedure, or whatever; let  $\delta w$  be the difference:  $\tilde{w} = w + \delta w$ . Neither  $w$  nor  $\delta w$  are known, but it is possible to compute  $\delta b$ , according to

$$\delta b \equiv \Omega^T \Omega \tilde{w} - b, \quad (26)$$

and we clearly must have  $\Omega^T \Omega \delta w = \delta b$ . It is not a priori true that  $\delta w$  will be orthogonal to null( $\Omega$ ), but this condition can be forced by projection: that is, we replace  $\tilde{w}$  with its projection on  $D(\Omega)$ ,

$$\tilde{w} := \tilde{w} - \sum_{e \in B} (e^T \tilde{w}) e \quad (27)$$

(This requires an orthonormal basis,  $B$ .) The combined equations (26) and (27) are equivalent to the original system, so the same procedure can be used to solve for the correction. By iterating, the solution can be computed to machine accuracy of computing  $\delta b$ .

A schema for doing all of this is easily written. Let prog[ $w, \Omega, b$ ] be the procedure described in this section, or any other procedure for solving (19). Then iterative refinement is accomplished as follows.

$$\underline{a}: w := w - \sum_{e \in B} (e^T w) e ;$$

$$\delta b := \Omega^T \Omega w - b ;$$

$$\text{prog}[\delta w, \Omega, \delta b] ;$$

$$w := w - \delta w ;$$

goto a.

Iterations are stopped when  $|\delta b|$  is small enough; about five or six are usually sufficient.

This memo is not meant to be a programming note, but it is worth mentioning that, according to Eq. (24), only five of the  $x$ 's and  $d$ 's

need be stored at any one time: the four "previous" ones and the "current" one. This suggests that pushdown stacks and recursive construction of the matrix "lhs" in (25) might be an efficient programming approach, and the program written to implement this procedure was indeed developed along those lines.

Figure Captions

Fig. 1 Survey variables. (Only part of a ring is illustrated.) Sagittal offset  $D_k$  and next-nearest neighbor chord  $a_k$  are indexed to correspond with site  $k$ . Other variables, such as nearest neighbor chords  $x_k$  or angles  $\phi_k$  also may be measured and used as checks on the primary data, but they are not considered to be part of the survey data set.

Fig. 2 Non-isometric symmetry of a cut ring. A ring can be "cut" by removing two sagitta and chord measurements from the data set. Doing so creates a new symmetry transformation, one that is not an isometry. Without a closure condition, the remaining  $2N-4$  measurements are not sufficient to fix a configuration.

Fig. 3 State estimation viewed as a geometry problem. (a) The complete, non-linear problem. An ideal measurement is modelled as a mapping  $\phi$  of  $\mathcal{W}$ , the state space, into  $\mathcal{M} = \mathbb{R}^{2N}$ , the space of real  $2N$ -tuples. An actual set of measurement data,  $M$ , generally will not lie on the data manifold,  $\phi(\mathcal{W})$ , because of errors. (b) The linear problem. As a computational method, the problem is linearized about a design state,  $s^*$ , and solved using the tangent spaces  $T_{M^*} \phi(\mathcal{W})$  and  $T_{s^*} \phi^{-1}(\mathcal{W})$ . The solutions of a sequence of such linear problems converge quickly to the solution of the original, nonlinear one.

Fig. 4 Local frames and the state of the ring. (a) A frame,  $F_k$ , is placed at each site of the design state in such a way that its axes are aligned along the directions of the (orthogonal) survey variables. (b) Positions of the markers that were surveyed are assigned coordinates relative to these local frames.

Fig. 5 Defining the variables  $\alpha_k$  and  $\theta_k$ .  $\alpha_k$  is a ratio of two lengths and must satisfy  $0 \leq \alpha_k \leq 1$ ; usually  $\alpha_k \approx 1/2$ .  $\theta_k$  is the (unsigned) angle through which  $F_k$  must be rotated in order to align with  $F_{k-1}$ . ( $F_{k-1}$  is  $F_k$  parallel translated to site  $k-1$ .)

Fig. 6 Large  $N$  limit of singular values for a symmetric ring. The limiting form is not a good approximation for  $m/N \approx 0$ .

Fig. 7 Accidental symmetry.  $C_{2n}$  rings possess an accidental differential symmetry corresponding to alternating longitudinal displacements of the sites. Variations in the data are of second order in the displacements.

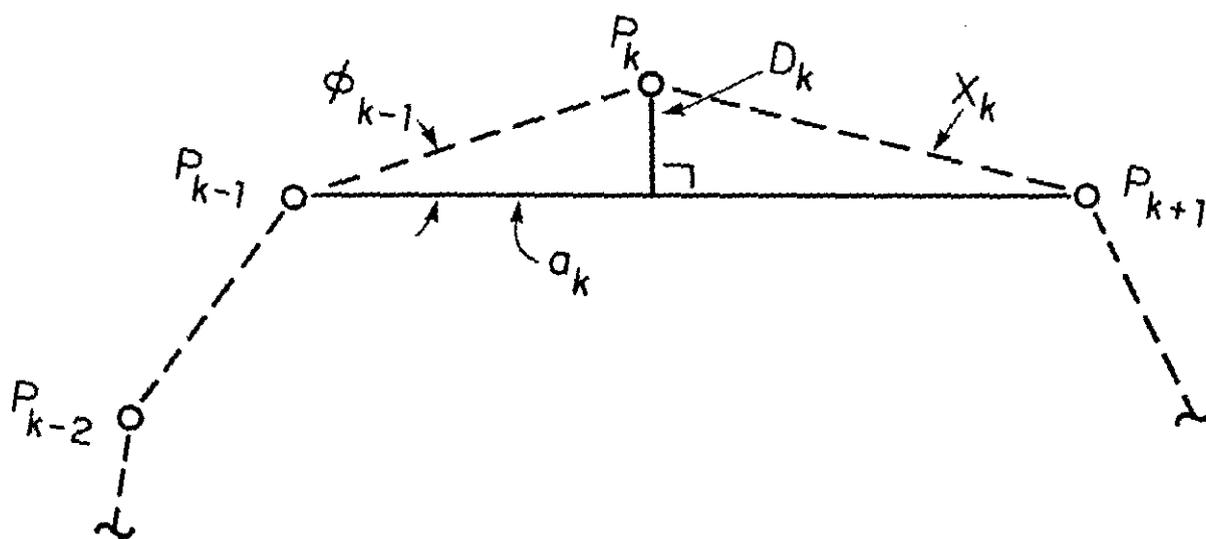


Figure 1

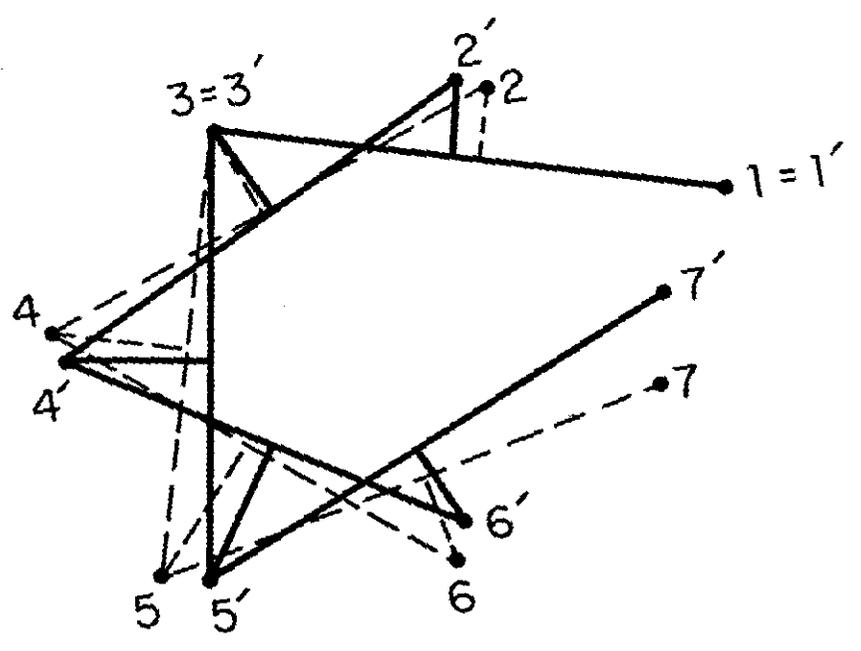
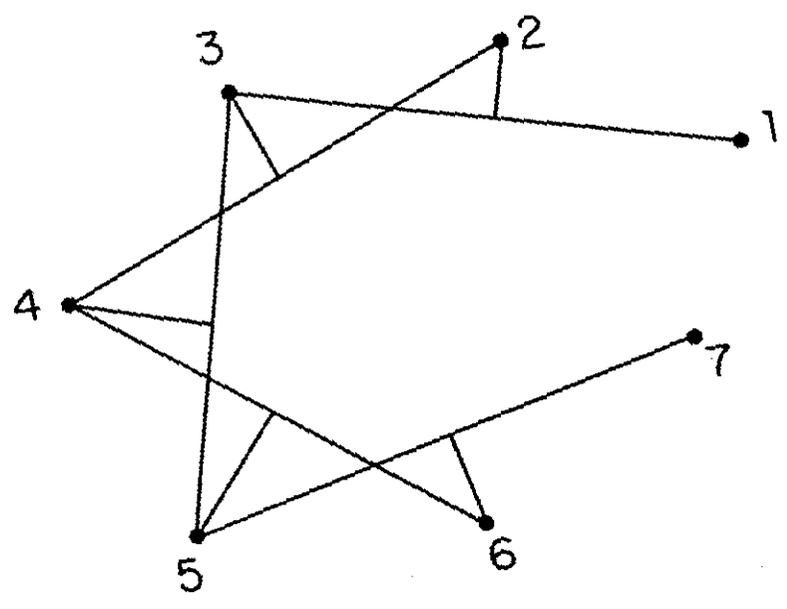


Figure 2

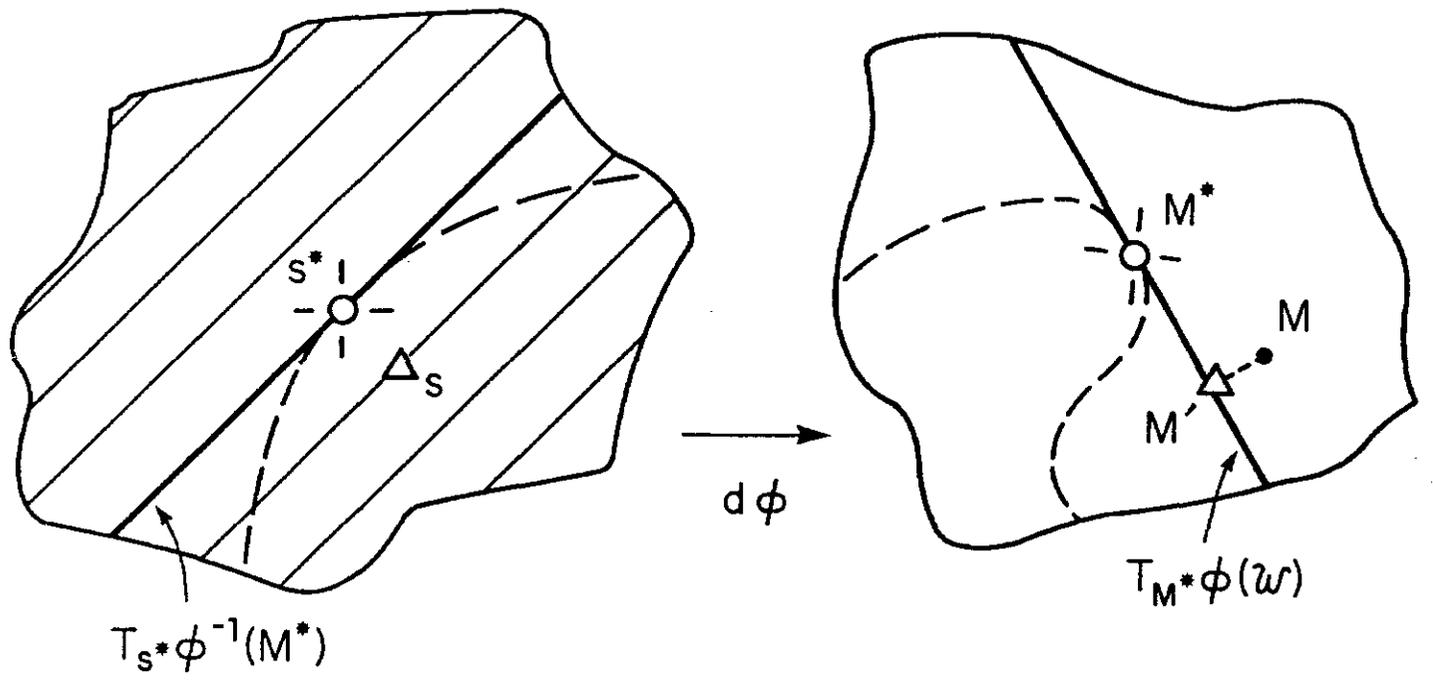
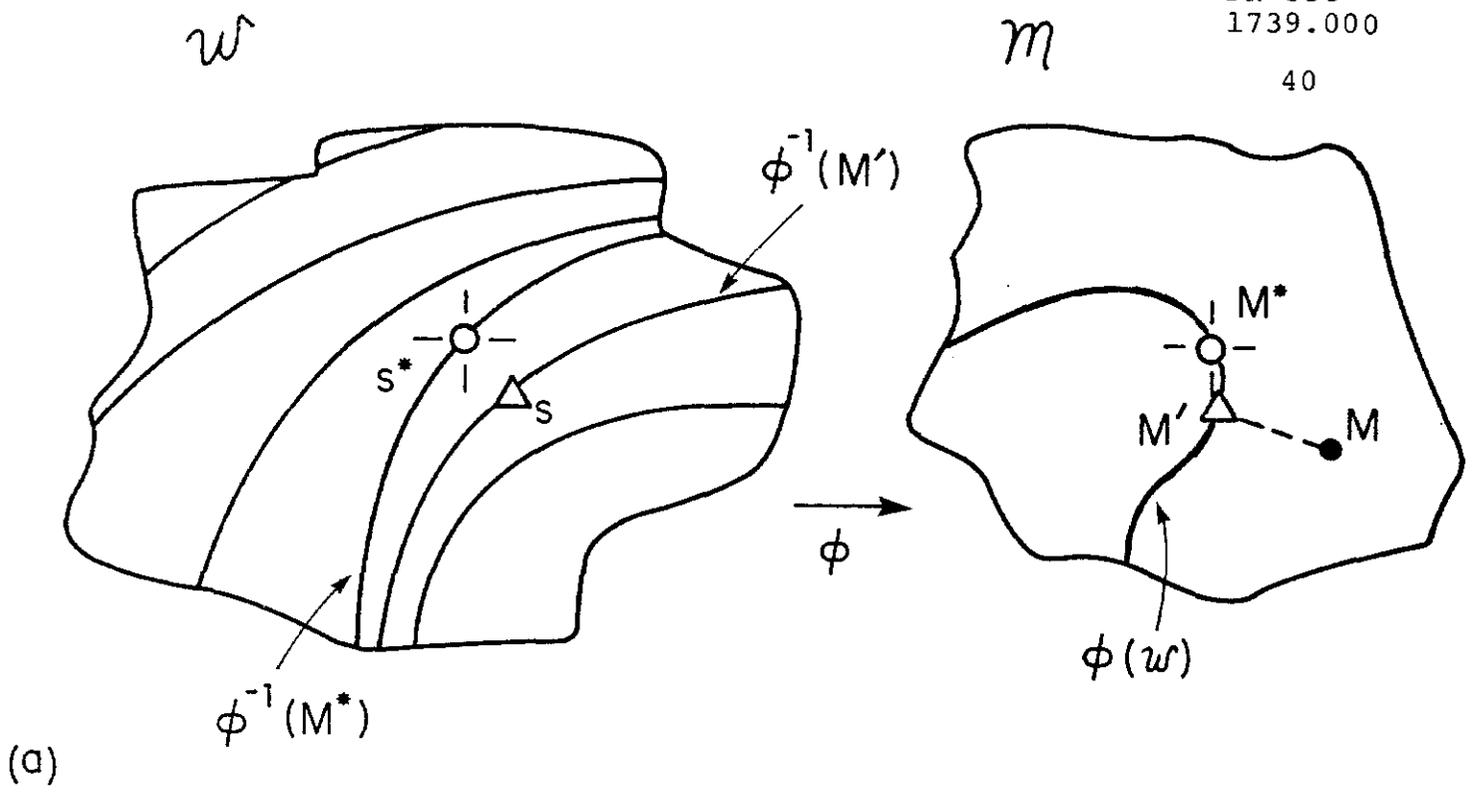


Figure 3

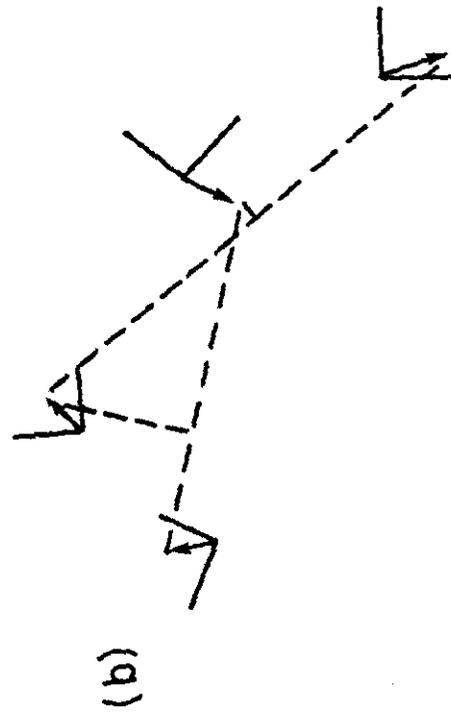
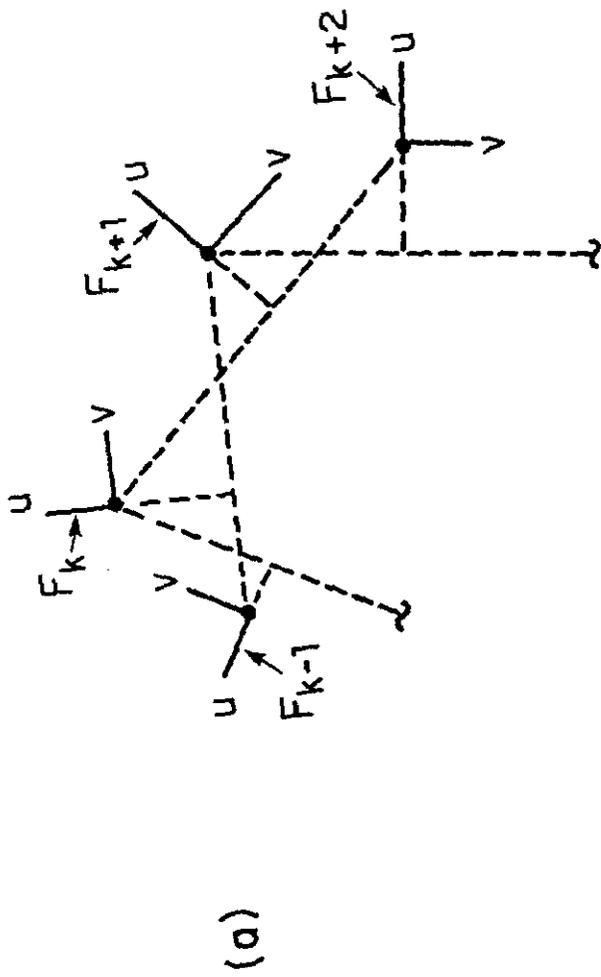


Figure 4

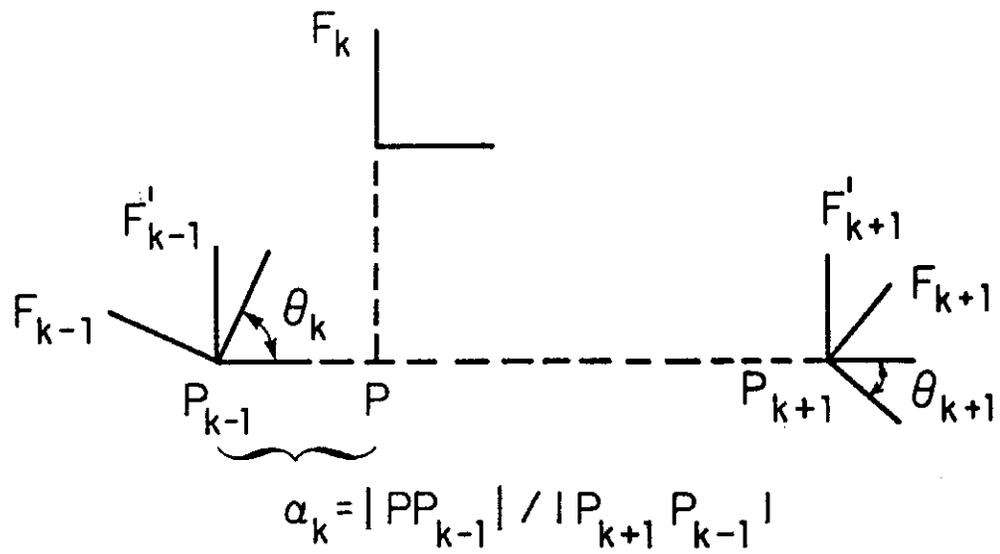


Figure 5

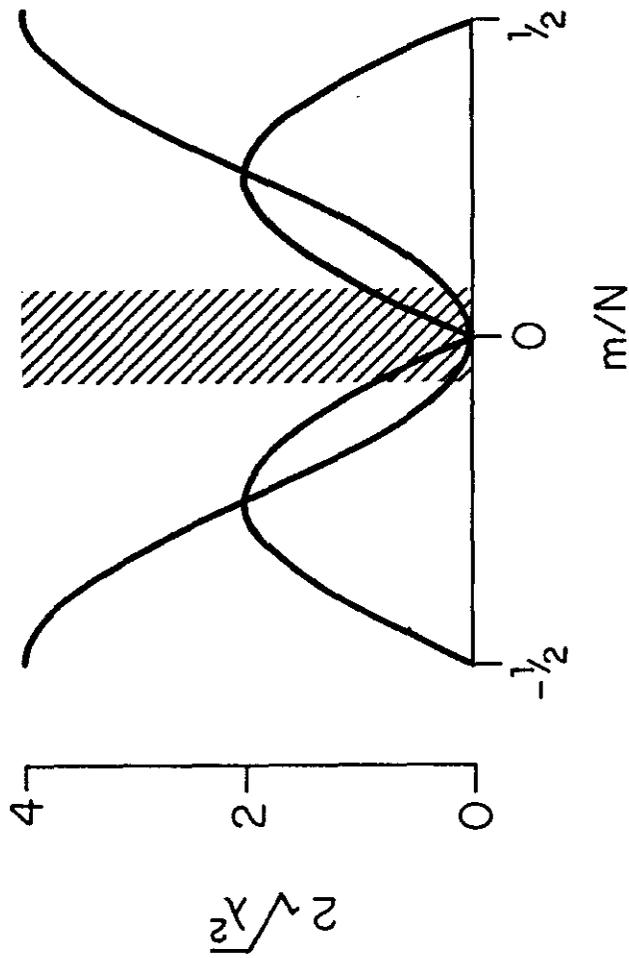


Figure 6

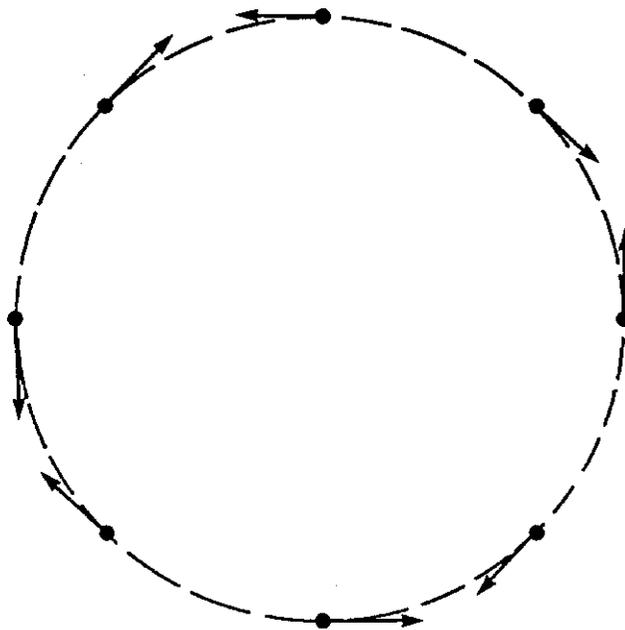


Figure 7

Appendix A.

The following one-dimensional problem serves as a simple yet enlightening example of employing singular value decomposition on a circulant matrix to pseudoinvert a singular system. Consider a collection of  $N$  points on the real axis, say  $x_k$ ,  $k = 0 \dots N-1$ . Suppose one is given as data the differences

$$h_k = x_{k+1} - x_k, \quad k = 0 \dots N-1.$$

(Assume the usual cyclic condition on the index set. In the absence of measurement errors the solution would be trivial, of course:

$$x_{k+1} = x_k + h_k, \quad k = 0 \dots N-2, \\ x_0, \text{ arbitrary.})$$

Write this as a matrix equation

$$\underline{h} = \underline{\Omega} \underline{x}, \\ \Omega_{km} \equiv \delta_{km} - \delta_{k+1, m}; \quad (\det \Omega = 0).$$

We seek the pseudoinverse solution,

$$\underline{x}^\dagger \equiv \underline{\Omega}^\dagger \underline{h}$$

Because  $\Omega$  is circulant it can be diagonalized by a discrete Fourier transform. Define the unitary matrix  $S$ ,

$$S_{km} \equiv \sqrt{1/N} \exp(2\pi i km/N), \quad k, m = 0 \dots N-1.$$

It is easy to verify that

$$(S^\dagger \Omega S)_{km} = \delta_{km} (e^{2\pi i m/N} - 1).$$

In this representation,

$$(S^\dagger \Omega^\dagger \Omega S)_{km} = \delta_{km} 4 \sin^2(\pi m/N)$$

so that the singular values of  $\Omega$  are given by

46

$$\lambda_m = 2 \left| \sin(\pi m / N) \right|, \quad m = 0 \dots N-1.$$

The null space of  $\Omega$  is one dimensional; it corresponds to the  $m=0$  singular value and therefore is spanned by the vector

$$\underline{e} = S_{.0} = \sqrt{1/N} (1 \ 1 \ 1 \ \dots \ 1)^T.$$

Of course, the motion associated with  $\underline{e}$  is a uniform translation of all points along the real axis.

The solution  $\underline{x}^\dagger$  is then found by using

$$\begin{aligned} \underline{x}^\dagger &= S (S^T \underline{x}^\dagger) \\ &= S (S^T \Omega^\dagger S) S^T \underline{h} \end{aligned}$$

where

$$(S^T \Omega^\dagger S)_{km} = \delta_{km} \cdot \begin{cases} 0, & m=0 \\ (e^{2\pi i m / N} - 1)^{-1}, & m \neq 0. \end{cases}$$

This can also be written in a crudely algorithmic form as follows.

$$\begin{aligned} \underline{h} &:= S^T \underline{h}; \quad h_0 := 0; \\ \text{for } m = 1 \dots N-1: \quad h_m &:= (e^{2\pi i m / N} - 1)^{-1} h_m; \\ \underline{x}^\dagger &:= S \underline{h}. \end{aligned}$$

REFERENCES

- (1) Michelotti, L. , "Survey of the Ring, a Trilogy. Part 1: Magnet Positions, Matrix Partitions, and Moore-Penrose," Fermilab UPC-140, Sep 23 1980. (unpublished)
- (2) Ben-Israel, A. and T.N.E. Greville, Generalized Inverse: Theory and Applications, John Wiley and Sons, New York, N.Y., 1974.
- (3) Maclane, S. and G. Birkhoff, Algebra, The Macmillan Co., New York, N.Y., 1968.
- (4) Golub, G.H. and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," Numer. Math. 14, 403-420 (1970).
- (5) Davis, P.J., Circulant Matrices, John Wiley and Sons, New York, N.Y., 1979.

"Pseudoinverting a Large, Almost  
Block Tridiagonal System,"

by Leo Michelotti

page 8 (middle of page):

This is shown graphically in Fig. 3.  $d\phi$  is a linear mapping from the state space onto the tangent space over the data manifold at  $M^* = \phi(s^*)$ . In contemporary notation,

$$d\phi: \mathcal{W} \rightarrow T_{M^*} \phi(\mathcal{W}) .$$

By introducing local coordinates, ...