

# Reduction Method for Duplicated Times

George B. Rybicki

3 May 1995

Consider the linear equation

$$(\mathbf{\Phi} + \mathbf{N})\mathbf{x} = \mathbf{y}, \quad (1)$$

which plays an important role in the fast methods proposed by Rybicki and Press (1995; hereafter RH). Here  $\mathbf{x}$  and  $\mathbf{y}$  are  $n$ -vectors, and  $\mathbf{N}$  is the  $n \times n$  diagonal “noise” matrix, with components

$$N_{ij} = N_i \delta_{ij}, \quad (2)$$

and  $\mathbf{\Phi}$  is the  $n \times n$  correlation matrix of the “signal”, with components

$$\Phi_{ij} = \Phi(t_i, t_j) = \langle s(t_i)s(t_j) \rangle. \quad (3)$$

For convenience, the times  $t_i$ ,  $i = 1, 2, \dots, n$  are assumed to be given in (or sorted into) non-decreasing order,  $t_1 \leq t_2 \leq \dots \leq t_n$ . The object is to solve for the unknown  $\mathbf{x}$  for a given  $\mathbf{y}$

A problem arises in solving Eq. (1) when some of the times  $t_i$  are duplicated. Although the inverse  $(\mathbf{\Phi} + \mathbf{N})^{-1}$  can still exist, providing the elements of  $\mathbf{N}$  are nonzero, the inverse  $\mathbf{T} = \mathbf{\Phi}^{-1}$  does not exist, and the formal solution  $\mathbf{x} = \mathbf{T}(\mathbf{1} + \mathbf{N}\mathbf{T})^{-1}\mathbf{y}$ , suggested in RH, cannot be used to effect a fast solution.

The purpose of this note is to show how to solve Eq. (1) by reducing it to one of the same form, but one of lower dimensionality in which all duplicate times are eliminated. This reduction works for matrices  $\mathbf{\Phi}$  of the general form (3), not just for the exponential matrices treated by RH.

# 1 The Reduction

There can be a number of sets of duplicated times, each with its own multiplicity. Let us concentrate on one such set consisting of  $m$  duplicated times, say,  $t_k = t_{k+1} = \dots = t_{k+m-1}$ . Then the  $m$  rows  $k, k+1, \dots, k+m-1$  of  $\Phi$  are identical, and so are the corresponding  $m$  columns (this shows that  $\Phi$  is singular).

Eq. (1) can now be substantially simplified by use of simple row and column operations. Subtracting the  $k$ th row from rows  $k+1$  through  $k+m-1$ , all elements of  $(\Phi + \mathbf{N})$  in these rows are eliminated, except for the  $k$ th through  $k+m-1$ st columns. Other rows of the matrix are not affected. Rows  $k+1$  through  $k+m-1$  of the right hand side are changed from  $y_i$  to  $y_i - y_k$ .

A similar set of column operations can be effected by replacing the variable  $x_k$  by the new variable

$$\bar{x}_k = x_k + x_{k+1} + \dots + x_{k+m-1}, \quad (4)$$

keeping all other variables the same. This eliminates all elements of the matrix in columns  $k+1$  through  $k+m-1$ , except for rows  $k$  through  $k+m-1$ . Other columns are not affected, nor is the right hand side.

The effect of these operations is to partially decouple the system, so that the set of variables outside the range  $k$  to  $k+m-1$  couple only with themselves and with  $\bar{x}_k$ . The main coupling to be resolved is between the new variable  $\bar{x}_k$  and the set of variables  $x_{k+1}$  through  $x_{k+m-1}$ . The relevant set of equations consists of the system

$$\begin{array}{cccccc} F_{k+1}x_{k+1} & +N_kx_{k+2} & +\dots & +N_kx_{k+m-1} & = & N_k\bar{x}_k + y_{k+1} - y_k \\ N_kx_{k+1} & +F_{k+2}x_{k+2} & +\dots & +N_kx_{k+m-1} & = & N_k\bar{x}_k + y_{k+2} - y_k \\ \vdots & \vdots & \ddots & \vdots & = & \vdots \\ N_kx_{k+1} & +N_kx_{k+2} & +\dots & +F_{k+m-1}x_{k+m-1} & = & N_k\bar{x}_k + y_{k+m-1} - y_k, \end{array} \quad (5)$$

where  $F_l = N_k + N_l$ , plus the single equation

$$\dots + (\Phi_{kk} + N_k)\bar{x}_k - N_kx_{k+1} \dots - N_kx_{k+m-1} + \dots = y_k. \quad (6)$$

The dots here refer to terms containing unknowns other than  $\bar{x}_k$  and  $x_{k+1}$  through  $x_{k+m-1}$ . The system (4) and Eq. (5) can be solved for the variables

$x_l$  in terms of  $\bar{x}_k$  and the quantities  $y_l$  and  $N_l$  for  $k \leq l \leq k+m-1$ . Defining the quantities  $\bar{N}_k$  and  $\bar{y}_k$  by

$$\begin{aligned}\bar{N}_k^{-1} &= \sum_{l=k}^{k+m-1} N_l^{-1} \\ \bar{N}_k^{-1} \bar{y}_k &= \sum_{l=k}^{k+m-1} N_l^{-1} y_l,\end{aligned}\tag{7}$$

the result can be expressed in the simple form

$$x_l = (\bar{N}_k \bar{x}_k + y_l - \bar{y}_k)/N_l, \quad k \leq l \leq k+m-1.\tag{8}$$

Using this equation, we can eliminate variables  $x_{k+1}$  through  $x_{k+m-1}$  from Eq. (6), and we obtain

$$\dots + (\Phi_{kk} + \bar{N}_k) \bar{x}_k + \dots = \bar{y}_k.\tag{9}$$

We note that Eq. (9) plus all the other equations in rows outside the range  $k$  to  $k+m-1$  form a system of  $n-m+1$  equations of precisely the original form without the duplicated times, except for the replacements  $x_k \rightarrow \bar{x}_k$ ,  $y_k \rightarrow \bar{y}_k$ , and  $N_k \rightarrow \bar{N}_k$ . Once this system is solved, the original variables  $x_k$  through  $x_{k+m-1}$  can be determined by use of Eq. (8).

Since the altered system is of the same general form as the original, the same reduction can be applied to it to eliminate a second set of duplicate times, and a third, etc., until all duplicate times are eliminated. By relabeling the variables, it is possible to use the same method to solve the reduced system as is used to solve a system without duplicate times. Since the additional manipulations implied in Eqs. (7) and (8) add at most only  $O(n)$  operations, the reduction does not spoil any fast methods.

## 2 Further Developments

### 2.1 Determinants

For some statistical applications it is necessary to compute the determinant of the matrix  $\Phi + \mathbf{N}$ . We show here how this can be done in terms of the reduced problem. After a little consideration of the reduced problem, it can

be seen that this determinant differs from that of the reduced problem by the determinant of the  $(m - 1) \times (m - 1)$  matrix  $\mathbf{K}_k$  on the left hand side of Eq. (5), that is,

$$\det(\Phi + \mathbf{N}) = \det \mathbf{K}_k \det(\Phi + \mathbf{N})_{\text{reduced}}. \quad (10)$$

The matrix  $\mathbf{K}_k$  can be written

$$\mathbf{K}_k = \mathbf{D}_k + N_k \mathbf{E} \mathbf{E}^T = (\mathbf{1} + N_k \mathbf{E} \mathbf{E}^T \mathbf{D}_k^{-1}) \mathbf{D}_k, \quad (11)$$

where  $\mathbf{D}_k$  is the  $(m - 1) \times (m - 1)$  diagonal matrix  $(D_k)_{ij} = N_{k+j-1} \delta_{ij}$  and  $\mathbf{E}$  is the column vector with unity elements  $E_i = 1$ . Using the general result  $\det(\mathbf{1} + \mathbf{A} \mathbf{B}) = \det(\mathbf{1} + \mathbf{B} \mathbf{A})$ , valid for rectangular matrices  $\mathbf{A}$  and  $\mathbf{B}$  for which  $\mathbf{A} \mathbf{B}$  and  $\mathbf{B} \mathbf{A}$  are square, we obtain the final result

$$\begin{aligned} \det \mathbf{K}_k &= \det(\mathbf{1} + N_k \mathbf{E}^T \mathbf{D}_k^{-1} \mathbf{E}) \det \mathbf{D}_k \\ &= \left( 1 + N_k \sum_{l=k+1}^{k+m-1} N_l^{-1} \right) N_{k+1} N_{k+2} \cdots N_{k+m-1} \\ &= \bar{N}_k^{-1} \prod_{l=k}^{k+m-1} N_l. \end{aligned} \quad (12)$$

When there are several sets of duplicate times, the correction to the determinant will be a product of determinants of the form (12), one for each set.

In practice, it is more convenient to work with the logarithms of the determinant, since these are less likely to cause overflow. Thus we write Eq. (12) in the equivalent form

$$\log \det \mathbf{K}_k = -\log \bar{N}_k + \sum_{l=k}^{k+m-1} \log N_l \quad (13)$$

## 2.2 Evaluation of $\chi^2$

One of the important statistical uses of the fast methods is to calculate the quantity  $\chi^2$ , defined by

$$\chi^2 = \mathbf{y}^T (\Phi + \mathbf{N})^{-1} \mathbf{y}. \quad (14)$$

By Eq. (1)  $\chi^2$  can be written simply

$$\chi^2 = \mathbf{y}^T \mathbf{x} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i. \quad (15)$$

The evaluation can be done straightforwardly by solving the system (1) by the reduction method to get  $\mathbf{x}$ . However, it is of some interest to see how this  $\chi^2$  relates to the reduced problem.

Let us now adopt the notation that the reduced problem is

$$(\bar{\Phi} + \bar{\mathbf{N}})\bar{\mathbf{x}} = \bar{\mathbf{y}}, \quad (16)$$

which is of dimension  $\bar{n}$ . The reduced quantity analogous to  $\chi^2$  is

$$\bar{\chi}^2 = \sum_{i=1}^{\bar{n}} \bar{x}_i \bar{y}_i. \quad (17)$$

Let us assume that there is one set of duplicated times. Then  $\chi^2$  and  $\bar{\chi}^2$  are related by

$$\chi^2 = \bar{\chi}^2 + \sum_{l=k}^{k+m-1} x_l y_l. \quad (18)$$

Using Eq. (8) to substitute for  $x_l$  and with some manipulation, we obtain

$$\chi^2 = \bar{\chi}^2 + \chi_k^2 \quad (19)$$

where

$$\chi_k^2 = \sum_{l=k}^{k+m-1} N_l^{-1} (y_l - \bar{y}_k)^2. \quad (20)$$

That is, the original  $\chi^2$  is equal to the reduced  $\bar{\chi}^2$  plus a “local” contribution  $\chi_k^2$ , due to the dispersion of the  $y_l$  values about their local mean  $\bar{y}_k$ , inversely weighted by the appropriate mean square noise values  $N_l$ . In general, there will be a sum of local correction terms of this sort from each set of duplicated times.

It is interesting to note that for Gaussian processes the expectation value of  $\bar{\chi}^2$  is equal to the reduced number of variables, and it is the local contributions that bring the expectation value of  $\chi^2$  up to the full number of variables  $n$ .

### 2.3 Reduced Data Sets

The analysis of large data sets (i.e., with large dimension  $n$ ) can be computationally expensive. It is therefore of considerable importance to find ways to reduce the dimensionality of a data set, while retaining as much of its essential information as possible.

In this subsection we consider how the preceding formulas can be used to construct reduced data sets. We restrict the discussion to cases where the data is to be used only for calculating  $\chi^2$  and  $\log \det(\mathbf{\Phi} + \mathbf{N})$ . For example, these are the essential quantities in the method of Press, Rybicki and Hewitt (1992a,b) for determining the time delay of gravitational lenses.

The reduction proposed here is based on the assumption that some of the times  $t_i$  associated with the data, while not strictly at duplicate times, can be grouped into “clusters” having almost the same times. This might occur, for example, if astronomical observations of an object are taken over many years, but a number of the observations may have been taken within relatively short intervals, short in comparison to the correlation time of the signal. A reduced data set might be constructed which contains only one representative observation from each night (or maybe each hour; or maybe each week).

The question then arises, how best to characterize this reduced, representative data. The formulas of the preceding subsections give the answer, namely, for each group of times the new data set consists of the reduced values  $\bar{y}_k$  and  $\bar{N}_k$ , defined by Eqs. (7), as well as the associated local contribution  $\chi_k^2$ , defined by Eq. (20) and the local contribution to the logarithm of the determinant, given by Eq. (13). This constitutes the desired reduced data set.

This reduced data set is very much like the old one in structure, except it now has the new quantities of local  $\chi^2$  and log-determinant  $\log \det \mathbf{K}_k$  attached to each point. We adopt the convention that the original set also had such local quantities associated with each point, call them  $\chi_l^2$  and  $\log \det \mathbf{K}_l$ , but they were all zero. Then we can rewrite Eq. (20) as

$$\chi_k^2 = \sum_{l=k}^{k+m-1} N_l^{-1} (y_l - \bar{y}_k)^2 + \sum_{l=k}^{k+m-1} \chi_l^2 \quad (21)$$

and Eq. (13) as

$$\log \det \mathbf{K}_k = -\log \bar{N}_k + \sum_{l=k}^{k+m-1} \log N_l + \sum_{l=k}^{k+m-1} \log \det \mathbf{K}_l \quad (22)$$

With this slight change, it can be shown that the reduction procedure is “associative,” that is, if one goes through a series of reductions of the data, the result is independent of the intermediate stages and depends only on the final time groupings. It is easy to see that this is true for the quantities  $\bar{y}_k$  and  $\bar{N}_k$ , owing to the form of Eqs. (7). It is also true for the formulas (21) and (22), but we shall not give the proof here.

Note that the reduction achieved here is not necessarily in the storage requirement, since two extra quantities are now included per time. Of much greater importance is the reduction in the number of independent times  $n$ , since the computational times for the data analysis depend strongly on this.