Table 2. *Results of second calculation of B and s*

$[(\sin \theta)/\lambda]_{\max}$	B	s	Points
0·60	0·85	1·18	7
0·65	0·71	1·19	8
0·70	0·89	1·12	9

calculated dividing the range of $(\sin^2 \theta)/\lambda^2$ into seven, eight and nine equidistant zones respectively. A straight line was interpolated with the help of the method of least squares into each of the diagrams and the values for s and B in Table 2 were obtained.

If the temperature factors are so large that the effective interpenetration of the atoms is considerable, one should use (18) instead of (17). This is equivalent to replacing b in (17), (21), (22) and (23) by $b + B$. Hence, in (21) and (23) the errors resulting from thermal interpenetration of the atoms decrease, but the errors from series termination increase with b.

If one wants to find s and $\boldsymbol{\beta}_\mu$, $\mu = 1, 2, \ldots, p$, or s and $\boldsymbol{\beta}$ it seems appropriate first to determine s and B by (21) and (23) and then to refine this result by (15) or (16) and finally by (12).

The author wants to thank Professor M. M. Woolfson for many helpful discussions and Drs M. Irvin and W. Horst for helpful comments on the manuscript. The work was supported by the Deutsche Forschungsgemeinschaft.

References

COCHRAN, W. (1954). *Acta Cryst.* **7**, 503–504.
HARKER, D. (1953). *Acta Cryst.* **6**, 731–736.
KARTHA, G. (1953). *Acta Cryst.* **6**, 817–820.
KROGH-MOE, J. (1956). *Acta Cryst.* **9**, 951–953.
MAGDOFF, B. S., CRICK, F. H. C. & LUZZATI, V. (1956). *Acta Cryst.* **9**, 156–162.
NORMAN, N. (1957). *Acta Cryst.* **10**, 370–373.
ROLLETT, J. S. & DAVIES, D. R. (1955). *Acta Cryst.* **8**, 125–128.
ROTHBAUER, R. (1975). *Neues. Jahrb. Mineral. Monatsh.* **3**, 121–134.
ROTHBAUER, R. (1976). *Acta Cryst.* A**32**, 169–170.
ROTHBAUER, R. (1977*a*). *Verh. Dtsch. Phys. Ges.* **12**, 194.
ROTHBAUER, R. (1977*b*). *Acta Cryst.* A**33**, 365–367.
ROTHBAUER, R., ZIGAN, F. & O'DANIEL, H. (1967). *Z. Kristallogr.* **125**, 317–351.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
WASER, J. (1955). *Acta Cryst.* **8**, 731.
WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
WOOLFSON, M. M. (1958). *Acta Cryst.* **11**, 277–283.

Acta Cryst. (1978). A**34**, 533–541

Extrapolative Filtering.
I. Maximization of Resolution for One-Dimensional Positive Density Functions

BY DOUGLAS M. COLLINS

Department of Chemistry, Texas A&M University, College Station, Texas 77843, *USA*

(*Received* 10 *January* 1977; *accepted* 24 *January* 1978)

A one-dimensional formalism based on extrapolative filtering can lead to electron-density maps at more than twice the resolution displayed by maps obtained by straightforward Fourier synthesis of structure factors. A worked example illustrates the computations for a hypothetical one-dimensional structure.

1. Introduction

Norbert Wiener's *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (1949) is of forbidding difficulty to most physical scientists. Originally published under military classification (Wiener, 1949, p. v), its importance was ironically acclaimed as those with access termed it 'The Yellow Peril' for the color of its binding (Bode & Shannon, 1950). Motivated by the need to simplify Wiener's work and relate it in a more obvious way to physical problems, Bode & Shannon (1950) presented a simplified development of Wiener's principal results in terms of electric-circuit theory. However helpful their work may have been, it does not appear to have provided an immediate impetus to scientific applications involving discrete time series, possibly because in the case of discrete time it is difficult to follow the mathematics intuitively. After some years, Enders Robinson (1967) put together an elementary account of discrete filters and included Bode & Shannon's presentation of Wiener's work in a still simpler form adapted to discrete time series. But the problem of satisfying the intuition remained, and as late as 1971, Lacoss wrote of a result similar to the one presented in this paper, 'Why this might be a reasonable estimator ... is certainly not obvious'. Although extrapolative filtering is novel in crystallography, and some crystallographers also may find, as the author did, that its results are not intuitively obvious, the mathematical operations themselves are routine in crystallographic practice.

For the present application, however, we are concerned with extrapolation not in discrete time but in sampled reciprocal space in order to obtain high-resolution density functions from low-resolution data. While such an effort may appear self-contradictory, it can succeed through the imposition of a model as in the determination of precise position coordinates by least-squares refinement of atomic models. It is our goal, however, to formulate a general rather than a particular model so that one can compute an extrapolated high-resolution estimate of a density function using only a low-resolution set of structure factors (amplitudes and phases) and universally applicable constraints. This is a basic need of protein crystallographers – never mind, for now, what 'high' and 'low' resolution may mean – and such an ability could be of great value in working any problem for which it is difficult to formulate a discrete-atom model. In the main body of the paper we shall present some general principles and apply them in a simple one-dimensional illustration; the n-dimensional problem will be dealt with in another paper. A more detailed development is given in the Appendix where the presentation has been designed for emphasis of some important features of the mathematical background and to correspond with the relevant literature. There will be no attempt to supply proofs as they can be found in the cited literature.

2. A high-resolution electron-density estimator

A direct approach to our problem would be to find and apply a general method of estimating data (structure factors, both amplitudes and phases) beyond the bounds of an existing data record. This we shall do, but in such a way that the final result will correspond to an implicit estimation rather than an explicit expansion of a data record. Also, in order to impose positivity of electron density, we shall assume

$$\rho_x = |g_x|^2 > 0. \qquad (2.1)$$

This assumption certainly requires that an electron-density function be positive at every sampled point. But more than this, we shall interpret F_h, the Fourier transform of ρ_x, to be the (complex) autocorrelation of G_h, the transform of g_x. Even though it will never be necessary to evaluate any G_h, the entire set exists for any real non-negative electron-density function (Papoulis, 1973). Consequently a formalism may be based on extrapolation of G_h to higher orders. If, hypothetically speaking, values for G_k were available as data,[†] then extrapolated values at higher orders could be included in computation of g_x and $|g_x|^2$ would be an estimator of electron density of an extrapolated higher resolution.

Suppose G_h can be estimated by

$$\hat{G}_h = \sum_{k=o}^{n-1} G_{h-1-k} A_k, \qquad (2.2)$$

which defines A_k (as a Wiener–Kolmogorov extrapolation filter, see Appendix). Subtraction of (2.2) from the identity $G_h = G_h$ gives the related equation

$$\varepsilon_h = G_h - \hat{G}_h = \sum_{k=0}^{n} G_{h-k} C_k, \qquad (2.3)$$

which defines C_k (as an extrapolation-error filter, see Appendix), and

$$C_0 = 1, \quad C_1 = -A_0, \ldots, \quad C_n = -A_{n-1}. \qquad (2.4)$$

C_k may be evaluated formally by minimization of

$$\frac{1}{L} \sum_h \left| \varepsilon_h - \sum_{k=0}^{n} G_{h-k} C_k \right|^2 \qquad (2.5)$$

with respect to C_s^*; L is the period of both g_x and ρ_x. The resulting equations are

$$\frac{1}{L} \sum_h \sum_{k=0}^{n} G_{h-k} G_{h-s}^* C_k = \frac{1}{L} \sum_h \varepsilon_h G_{h-s}^*; \quad s = 0, 1, \ldots, n. \qquad (2.6)$$

Because of (2.1)

$$\frac{1}{L} \sum_h G_{h-k} G_{h-s}^* = F_{s-k}, \qquad (2.7)$$

and (2.6) is the same as

$$\sum_{k=0}^{n} F_{s-k} C_k = \frac{1}{L} \sum_h \varepsilon_h G_{h-s}^*; \quad s = 0, 1, \ldots, n. \qquad (2.8)$$

The right-hand side of (2.8) may be evaluated for $s > 0$ by substituting (2.2) into (2.3) and minimization of

$$\frac{1}{L} \sum_h \left| G_h - \sum_{k=0}^{n-1} G_{h-1-k} A_k \right|^2, \qquad (2.9)$$

the variance of ε_h, with respect to $A_{s'}^*$. The resulting equations are

$$\frac{1}{L} \sum_h \sum_{k=0}^{n-1} G_{h-1-k} G_{h-1-s'}^* A_k = \frac{1}{L} \sum_h G_h G_{h-1-s'}^*;$$
$$s' = 0, 1, \ldots, n-1; \qquad (2.10)$$

[†] Hereafter in this section, h is used to call the reader's attention to a complete coefficient set which has no definite bounds; the subscripts k and s are used to emphasize a specifically delimited set of coefficients.

or

$$\sum_{k=0}^{n-1} F_{s'-k}A_k = F_{s'+1};\quad s'=0,1,\ldots,n-1; \tag{2.11}$$

which may be rearranged to give

$$F_{s'+1} - \sum_{k=0}^{n-1} F_{s'-k}A_k = 0;\quad s'=0,1,\ldots,n-1; \tag{2.12}$$

and, in view of (2.4),

$$\sum_{k=0}^{n} F_{s-k}C_k = 0;\quad s=1,2,\ldots,n. \tag{2.13}$$

Thus the right-hand side of (2.8) may be written as

$$\frac{1}{L}\sum_h \varepsilon_h G^*_{h-s} = \beta_s, \tag{2.14}$$

where $\beta_s = 0$ when $s > 0$, and the significance of β_0 is to be determined presently. In consequence of (2.14), equations (2.8) may be written in matrix notation as

$$\mathbf{FC} = \boldsymbol{\beta}. \tag{2.15}$$

Premultiplication by $\mathbf{C}\dagger$, the Hermitian transpose of $\mathbf{C}$, yields the Hermitian form

$$\mathbf{C}\dagger\mathbf{FC} = \mathbf{C}\dagger\boldsymbol{\beta}, \tag{2.16}$$

and from (2.4) and (2.14) $\mathbf{C}\dagger\boldsymbol{\beta}$ is readily confirmed to be β_0. Karle & Hauptman (1950) have shown that for real non-negative ρ_x, the left-hand side of (2.16) is real and non-negative, hence β_0 is real and non-negative. Moreover, from (2.3), (2.7), and (2.16) we have

$$\frac{1}{L}\,\boldsymbol{\varepsilon}\dagger\boldsymbol{\varepsilon} = \frac{1}{L}\,\mathbf{C}\dagger\mathbf{G}\dagger\mathbf{GC}, \tag{2.17}$$

$$\frac{1}{L}\,\mathbf{C}\dagger\mathbf{G}\dagger\mathbf{GC} = \mathbf{C}\dagger\mathbf{FC}, \tag{2.18}$$

$$\mathbf{C}\dagger\mathbf{FC} = \beta_0 = \frac{1}{L}\,\boldsymbol{\varepsilon}\dagger\boldsymbol{\varepsilon}, \tag{2.19}$$

and β_0 is seen to be σ_n^2, the variance of ε_h. Consequently, equations (2.8) may be cast in the form

$$\begin{bmatrix} F_0F_1^*\ldots F_n^* \\ F_1F_0 \\ \vdots \\ \vdots \\ F_nF_{n-1}\ldots F_0 \end{bmatrix} \begin{bmatrix} C_0=1 \\ C_1 \\ \vdots \\ \vdots \\ C_n \end{bmatrix} = \begin{bmatrix} \sigma_n^2 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \tag{2.20}$$

and it is this matrix equation which is to be solved for C_k.

The use for C_k follows closely upon transformation of (2.3) according to

$$\frac{1}{L}\sum_h \varepsilon_h \exp\{-2\pi i h x\} = \frac{1}{L}\sum_h \sum_{k=0}^{n} G_{h-k}C_k \exp\{-2\pi i h x\}. \tag{2.21}$$

If on the right-hand side we set $h' = h - k$, then drop the prime, (2.21) becomes

$$\frac{1}{L}\sum_h \varepsilon_h \exp\{-2\pi i h x\} = \frac{1}{L}\sum_h G_h \exp\{-2\pi i h x\} \sum_{k=0}^{n} C_k \exp\{-2\pi i k x\}, \tag{2.22}$$

or

$$\xi_x = g_x \sum_{k=0}^{n} C_k \exp\{-2\pi i k x\}. \tag{2.23}$$

Let us assume that the error series, ε_h, is uncorrelated, or a series of random errors; in the Appendix this intuitively reasonable assumption is seen to be valid. Consequently, the autocorrelation of ε_h is zero everywhere but at the origin, where it has the value $\beta_0 = \sigma_n^2$. It follows that in

$$|\xi_x|^2 = \frac{1}{L}\left[\sum_{h'}\left(\frac{1}{L}\sum_h \varepsilon_h \varepsilon^*_{h-h'}\right)\exp\{-2\pi i h' x\}\right], \tag{2.24}$$

the right-hand side is non-zero only for $h' = 0$, and

$$|\xi_x|^2 = \frac{1}{L}\,\sigma_n^2. \tag{2.25}$$

It is then clear from (2.23) that

$$|g_x|^2 \left|\sum_{k=0}^{n} C_k \exp\{-2\pi i k x\}\right|^2 = \sigma_n^2/L, \tag{2.26}$$

and, since $\rho_x = |g_x|^2$, an obvious rearrangement yields as the principal result of this paper, the electron-density estimator

$${}^{c}\rho_x = \frac{\sigma_n^2/L}{\left|\sum_{k=0}^{n} C_k \exp\{-2\pi i k x\}\right|^2}. \tag{2.27}$$

It is implicit in the rearrangement of (2.26) that the sum over k is non-zero for every x. If this were not the case, then either $|g_x|$ would be infinite for some x, or σ_n^2 would be zero. The latter possibility is easily eliminated by use of suitably small values of n in (2.20) (Goedkoop, 1950), and the former may be ruled out on the grounds that $|g_x| = (\rho_x)^{1/2}$.

There are a number of ways to see that ${}^{c}\rho_x$ should be of higher resolution than ρ_x, the transform of F_k. Likely the most straightforward is to consider the hypothetical computation of ρ_x from G_h by usual methods. Transformation of G_k, a bounded set of coefficients, would yield g_x at limited resolution and $\rho_x = |g_x|^2$, though non-negative, also would be of limited resolution. On the other hand, computation of ${}^{c}\rho_x$ would require explicit use of the *complete* set, G_h. Clearly, if it were possible to use all G_h explicitly, G_h would be used to compute not ${}^{c}\rho_x$ but an exact g_x, and $\rho_x = |g_x|^2$ also would be exact. But ${}^{c}\rho_x$ can be computed by implicit use of the *complete* set, G_h, because it enters the computation only as its autocorrelation, which does not need to be calculated for it is experimentally accessible in excellent approximation. Even so, ${}^{c}\rho_x$ is only an estimate of true ρ_x and as n is allowed to increase to cause every ε_h (thus σ_n^2) $\to 0$ and ${}^{c}\rho_x \to$ true ρ_x, the right-hand side of (2.27) $\to 0/0$ and the whole formalism breaks down; this breakdown occurs for n in the neighborhood of N, the number of atoms in a unit cell (Goedkoop, 1950). Nevertheless, for n somewhere in the neighborhood of N, the formal criterion of (relatively) small σ_n^2 leads to the expectation that ${}^{c}\rho_x$ will provide a good high-resolution estimate of true ρ_x, whereas the same structure factors employed in a conventional Fourier synthesis would yield a result of quite limited resolution. From this point of view, the superiority of resolution in ${}^{c}\rho_x$ over that in the conventional estimation of ρ_x may be considered as at least partially due to implicit use of the *complete* set G_h. The related general electron-density model given by (2.1) is crucial to our results, but has no apparent direct effect upon the resolution in ${}^{c}\rho_x$. It should be noted that the implicit estimation referred to in the beginning of this section relates to F_k as unmeasured members of a data record may be estimated by conventional transformation of ${}^{c}\rho_x$.

3. An example

Consider the artificial one-dimensional structure described in Table 1. This structure was constructed to contain typical atoms separated by typical interatomic distances so that it could serve as a realistic simulation of crystallographic application, insofar as that is possible in one dimension. Theoretical structure factors were calculated for atoms free of thermal motion and they are listed in Table 2; these structure factors were used without further modification in all the calculations.

Table 1. *Atomic position parameters and interatomic distances for a hypothetical one-dimensional structure*

Atom	x	Interatomic distance (Å)	Atom	x	Interatomic distance (Å)
C(1)	0·01094		N(8)	0·45312	
		1·5			1·3
O(2)	0·03438		C(9)	0·47344	
		1·2			15·1
C(3)	0·05312		O(10)	0·70938	
		1·4			1·3
N(4)	0·07500		C(11)	0·72969	
		1·4			3·8
C(5)	0·09688		N(12)	0·78906	
		10·1			1·5
O(6)	0·25469		C(13)	0·81250	
		11·3			12·7
C(7)	0·43125		C(1)	1·01094	
		1·4			—
N(8)	0·45312			Axis length = 64·0	

The actual calculations are exceedingly simple. It is necessary only to set up the square array of (2.20) for any desired n, set $\sigma_n^2 = 1{\cdot}0$, and solve for C_k. The result is scaled so that $C_0 = 1{\cdot}0$, thus σ_n^2 has its proper value, and the desired electron-density estimate follows at

Table 2. *Structure factors*; $F_h = A_h + iB_h$

h	A_h	B_h	h	A_h	B_h
0	86·97	0·0	17	−7·28	9·71
1	13·52	−1·78	18	−7·05	−6·36
2	8·24	6·15	19	2·17	−8·07
3	2·00	48·24	20	3·91	4·22
4	32·88	10·82	21	3·07	2·96
5	−3·69	36·87	22	−6·60	−3·56
6	−18·40	−0·59	23	7·48	−3·78
7	3·79	8·22	24	7·43	6·10
8	−15·88	−8·49	25	−4·43	−0·84
9	−2·98	25·18	26	−5·52	2·33
10	−2·63	6·34	27	9·24	−0·74
11	21·41	−19·22	28	12·17	2·60
12	−10·29	−0·46	29	−3·94	2·47
13	−7·56	21·31	30	−1·97	1·56
14	−0·44	12·58	31	6·64	4·71
15	5·34	−11·38	32	1·92	11·49
16	−4·14	3·60			

Table 3. *The extrapolation-error filter derived from data at* 4·6 Å *resolution*; $C_h = A_h + iB_h$

For this filter, the error-series variance is $\sigma_{14}^2 = 2{\cdot}94$.

h	A_h	B_h	h	A_h	B_h
0	1·000	0·0	8	−0·691	−0·447
1	−0·416	0·146	9	0·624	0·686
2	0·068	−0·190	10	0·261	0·168
3	0·132	−1·417	11	0·558	−0·253
4	−0·975	0·449	12	0·485	0·155
5	0·242	−1·155	13	−0·246	0·071
6	−0·598	0·642	14	−0·087	−0·457
7	−0·528	1·536			

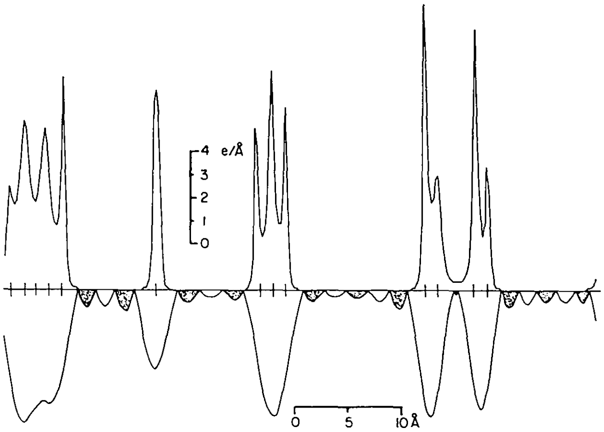

Fig. 1. Electron density functions based on data at 4·6 Å resolution. See text for details.

Table 4. *The extrapolation-error filter derived from data at* 2·0 Å *resolution*; $C_h = A_h + iB_h$

For this filter, the error-series variance is $\sigma_{32}^2 = 2\cdot14$.

h	A_h	B_h	h	A_h	B_h
0	1·000	0·0	17	0·996	0·357
1	−0·542	0·253	18	−0·784	−0·073
2	0·051	−0·276	19	0·124	−0·293
3	0·278	−1·341	20	0·786	0·366
4	−0·948	0·682	21	−1·180	0·369
5	0·364	−1·498	22	0·776	−0·689
6	−0·194	0·905	23	0·058	0·498
7	−0·901	1·397	24	−0·262	−0·290
8	−0·752	−0·998	25	0·417	−0·548
9	0·935	0·780	26	0·144	0·568
10	−0·701	0·149	27	−0·649	−0·397
11	1·066	−0·501	28	0·187	0·012
12	0·303	0·894	29	−0·116	0·068
13	−0·729	0·101	30	−0·205	−0·014
14	0·577	−0·597	31	0·221	−0·008
15	−0·099	0·696	32	−0·068	0·137
16	−0·330	−0·814			

once from (2.27). The minimum structure factor set which resolves 12 of the ideal 13 peaks includes structure factors with h in the range 0–14 and corresponds to data at a resolution of 4·6 Å (minimum interplanar spacing). The values of C_k and σ_{14}^2 for the 4·6 Å data set are given in Table 3. The corresponding density functions, ρ_x and $^C\rho_x$, were computed on grids of 256/period and are given in Fig. 1.

The minimum structure factor set which resolves all 13 of the ideal 13 peaks includes structure factors with h in the range 0–32 and corresponds to data at a resolution of 2·0 Å. The values of C_k and σ_{32}^2 for the 2·0 Å data set are given in Table 4. The corresponding density functions, ρ_x and $^C\rho_x$, were computed on grids of 256/period and are given in Fig. 2. In Figs. 1 and 2 the function $^C\rho_x$ is plotted above the zero line with $^C\rho_x$ increasing toward the top; from (2.27) it is clear that $^C\rho_x$ must always be positive. On the other hand, ρ_x is negative in some regions, so to plot it below the zero line and increasing toward the bottom, it was necessary to plot absolute values and identify the negative regions by shading. In both figures the positions of the 13 atoms are marked by vertical hash marks across the zero line.

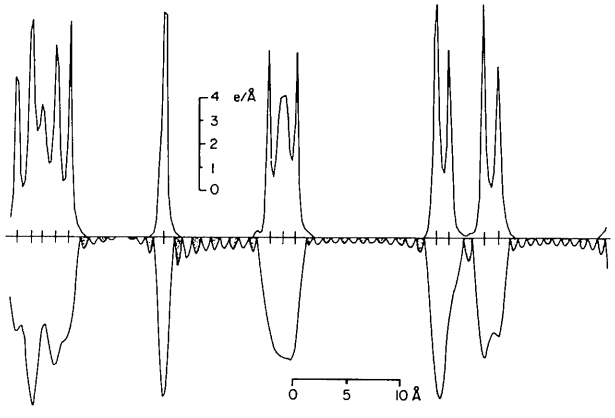

Fig. 2. Electron density functions based on data at 2·0 Å resolution. See text for details.

Use of the C series to estimate electron density can result in a substantial increase of resolution over that provided by straightforward Fourier synthesis of structure factors. This is clearly seen in the figures which show that resolution in $^C\rho_x$ is greater by a factor of ~2–3 than resolution in ρ_x. It should not be supposed, however, that resolution can be arbitrarily increased by indiscriminant choice of large n in (2.20 and 2.27). Inclusion of too many terms leads to spurious detail and, as Ulrych & Bishop (1975) show, even false major maxima. This problem can be seen in our example by carefully comparing the profiles for C(7), N(8), C(9) in the two figures. Although Fig. 2 is based on more than twice as much data as is Fig. 1, in Fig. 1 the profiles for C(7), N(8), C(9) have better shape and, in fact, are better resolved though their positions are somewhat better in Fig. 2. At present, a value for n is chosen empirically; a first value of $n < N$, the number of atoms/period, may be increased so long as the matrix in (2.20) remains positive definite and $^C\rho_x$ does not become unreasonable.

4. Conclusion

In the general context of improving electron-density maps of large biological molecules, two objectives led to the present work. The two, estimation of unmeasured data by extrapolation of existing data and electron-density constraint by imposition of a general model, are somehow related through the dual-basis relationship of direct and reciprocal space. However, we have achieved the objectives simultaneously without concern for their relationship by use of the Wiener–Kolmogorov extrapolation theory and inclusion in the formalism of an assumption concerning the electron-density function. The assumption is that the electron density is the squared modulus of some underlying (complex) function, and it was employed as its reciprocal-space equivalent, namely that the structure factors are equispaced samples of an autocorrelation function. It is remarkable that the final equations (2.20 and 2.27) do not make explicit use of any extrapolation, nor is any function underlying the electron density ever calculated, although the transform of C_k may be proportional to the inverse of some such function.

In the beginning of this paper, reference was made to deriving high-resolution results from low-resolution data, where 'high' and 'low' remained indefinite terms. The example of the preceding section and some of the geophysical literature (*e.g.* Lacoss, 1971) empirically imply that 'high', in terms of resolution, means an increase in the range ~2–5 times 'low'. From our example of a crystallographic application, it appears that 'low' refers to resolutions below true atomic resolution on the joint grounds that Fig. 2 shows some degradation of peak shape, and ever larger matrices in (2.20) are certain to become ill-conditioned. In this connection it should be noted that to avoid matrix singularity at order $n = N + 1 = 14$ (Goedkoop, 1950), the example was worked with natural rather than normalized structure factors.

There are many other questions which need answering but they are of little consequence next to the question of the formalism's applicability in n dimensions. This has yet to be determined and will be the subject of another paper.

Professor Terry W. Spencer of the Texas A&M University Department of Geophysics provided invaluable assistance without which this work could not have gone forward. This work was supported in part by grant GM19455 of the National Institute of General Medical Sciences and by a National Institutes of Health Biomedical Support Grant administered by the Texas A&M University Office of University Research.

APPENDIX *A*

The WK linear predictor*

The general problem of predicting (or estimating) data not yet measured has been solved (Wiener, 1949) and its solution is the central result of the Wiener–Kolmogorov (WK, hereafter) prediction theory for stationary series. The key to practical application of the WK theory lies in the term 'stationary,' which, in reference to time series, means that the statistical properties of the series do not change with time. And if a series is stationary and of finite mean-square modulus, then its autocorrelation will exist and will be a function of definite finite value (Wiener, 1949, § 0.8). Here the crystallographic application comes into view, for we shall not inquire concerning the stationarity of any series, but where there exists a definite autocorrelation function, the formalism of the WK prediction theory is assumed to be applicable. For ease of exposition, the remainder of this section, and most of the next, will deal with time series.

Consider the discrete time series x_t known for past time. Wiener's general problem is to find a_s such that the expression (Wiener, 1949, § 2.6)

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{N} \left| x_{t+\alpha} - \sum_{s=0}^{\infty} x_{t-s} a_s(\alpha) \right|^2 \qquad (A.1)$$

is a minimum. In (A.1), $a_s(\alpha)$ is the WK linear predictor (extrapolation filter) for prediction span α, that is, given the values of x up to time t, there is a least-squares optimum predicted value for $x_{t+\alpha}$ and it is given by the inner sum of (A.1). Expression (A.1) is only of formal value for there is little purpose to predicting values already known, but the concept is employed as follows. Assume x_t to be a real-valued stationary series (hereafter we shall assume all series to be stationary). Straightforward minimization of the expectation in time,

$$I = E\left\{ \left[x_{t+\alpha} - \sum_{s=0}^{\infty} x_{t-s} a_s(\alpha) \right]^2 \right\}, \qquad (A.2)$$

with respect to the operator coefficients, $a_s(\alpha)$, gives the equations

$$\sum_{s=0}^{\infty} a_s(\alpha)\varphi_{\tau-s} = \varphi_{\tau+\alpha}, \quad \tau = 0, 1, 2, \ldots; \qquad (A.3)$$

where φ_τ is the autocorrelation of x_t. Equations (A.3) could be solved by ordinary methods to any practicable order of approximation to provide values for $a_s(\alpha)$ and a formally satisfactory solution to the simple prediction problem. Wiener's book-length treatment of the prediction problem includes the details necessary for mathematical rigor, but for the sake of insight we shall follow the more transparent development of Robinson (1967) as based on the work of Bode & Shannon (1950). Nevertheless, one outstanding feature of the WK linear predictor, or extrapolation filter, is clear in (A.3). An extrapolation filter depends neither on the choice of origin nor on explicit values for the time series, but only on its autocorrelation function.

A well behaved time series certainly may be written as a convolution of the type

$$x_t = \sum_{s=-\infty}^{\infty} b_s \varepsilon_{t-s}, \qquad (A.4)$$

where ε is a white noise series, a series which by definition possesses an autocorrelation function that is non-zero only at the origin. The Fourier transform of such an autocorrelation function is a positive constant, say $|\sigma|^2$. In any case $X(f)$, the Fourier transform of the physically observed series x_t, surely exists, and just as surely so does

$$B(f) = X(f) \exp\{-[i\theta(f) + \tfrac{1}{2} \ln |\sigma|^2]\}, \qquad (A.5)$$

where we require that $\theta(f)$, though arbitrary, be real and finite. Rearrangement of (A.5) and subsequent transformation then gives (A.4). An obvious but

---

* There appears to be a controversy concerning the attribution of prediction theory to either Wiener or Kolmogorov. Wiener (1949, p. 59) asserted that their work was parallel but independent and, in agreement with this, prediction theory is attributed to both men in the present work, though only Wiener's work is cited.

important consequence of (A.5) is that

$$|X(f)|^2 = |\sigma|^2 \, |B(f)|^2. \qquad (A.6)$$

Thus it is clear that of the many ways to represent x according to (A.4), for each case the autocorrelation of b must equal the autocorrelation of x except for a constant factor.

Since the WK extrapolation filter depends only on the autocorrelation of x, and b must have the same autocorrelation, except for a constant factor, we anticipate that any extrapolation will depend on b, and ε will contain the information which distinguishes x from every other series possessing the same autocorrelation. To emphasize the correspondence between x and ε, it is desirable to rewrite (A.4) in the form

$$x_t = \sum_{s=-\infty}^{t} \varepsilon_s b_{t-s}, \qquad (A.7)$$

if possible. If b_t is a causal or one-sided function, that is, if $b_t = 0$ when $t < 0$, then (A.7) is not different from (A.4). So the problem in writing (A.7) is to determine the conditions under which phase angles can be found and assigned to $|B(f)|$ such that $B(f)$ is the transform of a causal or one-sided function. Satisfaction of the Paley–Wiener condition (Papoulis, 1962, § 10.5) is necessary and sufficient, which for our purposes means it is required that

$$|B(f)| > 0. \qquad (A.8)$$

If (A.8) is satisfied, then an admissible statement of x at time $t + \alpha$ is

$$x_{t+\alpha} = \sum_{s=-\infty}^{t+\alpha} \varepsilon_s b_{t+\alpha-s}, \qquad (A.9)$$

and upon distribution of terms,

$$x_{t+\alpha} = \sum_{s=-\infty}^{t} \varepsilon_s b_{t+\alpha-s} + \sum_{s=t+1}^{t+\alpha} \varepsilon_s b_{t+\alpha-s}. \qquad (A.10)$$

Evidently, the deconvolution of x implied in (A.7) is the crucial step in solving the prediction (extrapolation) problem. For, given the series ε_t and b_t, trivial manipulation can yield the first summation of (A.10) which is the desired estimate of $x_{t+\alpha}$; the second sum of (A.10) is the unpredictable part of $x_{t+\alpha}$ and is to be ignored.

Because future values of x_t are unknown, ε_t must be computed using an equation analogous to (A.7), namely,

$$\sum_{s=-\infty}^{t} x_s b_{t-s}^{\text{inv}} = \varepsilon_t, \qquad (A.11)$$

where b_t^{inv} is the Fourier transform of $1/B(f)$. Just as in (A.7) b_t must be a one-sided function, so must b_t^{inv} be a one-sided function and with this requirement b_t is uniquely specified. To see this, let us at first identify $\exp\{2\pi i f\}$ with z, then write the z-transform of b_t,

$$B(z) = \sum_{t=-\infty}^{\infty} b_t z^t, \qquad (A.12)$$

which is immediately recognizable as a Laurent series about the origin of the complex plane if the constraint $|z| = 1$ is removed. It is assumed that b_t is well behaved in the sense that its Fourier transform, $B(f)$, is given by a convergent series; consequently, the region of convergence of the series (A.12) includes at least the unit circle. Moreover, b_t must be a one-sided function; so $B(z)$ is given by an ordinary power series,

$$B(z) = \sum_{t=0}^{\infty} b_t z^t, \qquad (A.13)$$

and its absolute convergence within the unit circle is assured (Churchill, 1960). A similar analysis can be applied to $1/B(z)$ to demonstrate that it too must be free of singularities within the unit circle. Now if both $B(z)$ and $1/B(z)$ are free of singularities within the unit circle, and both satisfy the Paley–Wiener condition, then both must be free of zeros and singularities on and within the unit circle. Such a function is a 'minimum-phase function' and, apart from an arbitrary constant factor, is uniquely determined by its amplitude (Bode & Shannon, 1950; Robinson, 1967; Papoulis, 1962, § 10.3).

It remains to compute the minimum phase function, $B(z)$, and there are a number of ways to perform this computation. For the finite case Fejér's (1915) factorization is given by Wiener (1949, § 2.6) and developed in detail by Robinson (1967). Although this factorization is not practical for crystallographic applications it is conceptually straightforward. Consider $B^*(1/z)B(z)$, which, on the unit circle, differs from $|X(f)|^2$ only by a constant factor. If there are $N + 1$ terms in the summation of (A.13), then we may write

$$B(z) = (\alpha_1 + \beta_1 z)(\alpha_2 + \beta_2 z) \ldots (\alpha_N + \beta_N z), \qquad (A.14)$$

$$B^*(1/z) = (\alpha_1^* + \beta_1^* z^{-1})(\alpha_2^* + \beta_2^* z^{-1}) \cdots (\alpha_N^* + \beta_N^* z^{-1}), \qquad (A.15)$$

$$z^N B^*(1/z)B(z) = (\alpha_1 + \beta_1 z)(\beta_1^* + \alpha_1^* z)(\alpha_2 + \beta_2 z)$$
$$(\beta_2^* + \alpha_2^* z) \cdots (\alpha_N + \beta_N z)(\beta_N^* + \alpha_N^* z). \qquad (A.16)$$

Now (A.16) is a proper polynomial and may be set equal to zero whereupon it is clear that its roots are in pairs,

$$\left(-\frac{\alpha_i}{\beta_i}, -\frac{\beta_i^*}{\alpha_i^*}\right). \qquad (A.17)$$

One member of (A.17) lies inside the unit circle and the other lies outside the unit circle. The desired minimum-phase function, $B(z)$, is obtained by forming the product of the N factors from the right-hand side of

(A.16) which correspond to roots outside the unit circle.

This factorization to obtain the minimum-phase function, $B(z)$, completes the solution of the WK extrapolation problem except for routine manipulations. Given the physically observed (finite) time series, x_t, Fourier transformation yields $X(f)$ and, of course, $|X(f)|^2$, which may differ from $|B(f)|^2$ by a constant factor. The foregoing factorization yields, on the unit circle, $B(f)$ such that its inverse $1/B(f)$ exists and is well behaved. Thus the series b_t and ε_t can be calculated by transformation of $B(f)$ and $X(f)/B(f)$, and the first sum of (A.10) can be formed to give the desired estimate of $x_{t+\alpha}$. Alternatively, $a_t(\alpha)$, the extrapolation filter itself, could be calculated from its transform, $A(\alpha, f) = B'(\alpha,f)/B(f)$. $B'(\alpha,f)$ is the transform of the one-sided function b'_t, where $b'_0 = b_\alpha$, $b'_1 = b_{\alpha+1}, \ldots$.

APPENDIX B

The extrapolation-error filter

The development of the preceding section is useful in understanding the nature of an extrapolation filter. But the formalism requires that we achieve our goal by explicit estimation of future (unmeasured) data, an unnecessarily cumbersome procedure. It turns out, for our purposes, that it is more efficient to use an extrapolation-error filter than an extrapolation filter. Here we shall follow Peacock & Treitel's (1969) matrix formulation and development of the extrapolation-error filter.

But first let us be assured that it is reasonable to work with filters of finite length. In the Fejér factorization of the preceding section it was assumed that the series b_t was finite and had $N+1$ members. Whether or not N is exceedingly large from an experimental point of view is of no consequence. For if $B(z)$ is a minimum-phase function, then b_t is minimum delay and for any different b'_t corresponding to $B'(z)$ with one or more of its roots inside the unit circle [refer to (A.14–A.17)],

$$\sum_{t=0}^{M} |b_t|^2 \geq \sum_{t=0}^{M} |b'_t|^2; \quad M = 0, 1, \ldots, N. \quad (B.1)$$

Robinson (1967) discusses this remarkable fact in detail and Oppenheim & Schafer (1975) give an outline for its proof. The minimum-phase function, $B(z)$, is unique not only as it was derived to have all its roots outside the unit circle, but also in this sense. That of all possible sequences which yield the known values $|B(f)|^2$, the minimum-delay sequence b_t, whose z-transform is the minimum-phase function $B(z)$, will give the best possible estimate in the mean of $|B(f)|^2$ after truncation at any given $t = M < N$. So whatever the practical limitations of length may be, optimum extrapolation is ensured by use of minimum-delay sequences for filters. Of course this does not mean that short filters are better than long ones, or even that they are useful at all. But it does mean that a well chosen short filter may be very much better than a poorly chosen long one, and that there is a best filter of given length. Furthermore, in the context of data records, it is most improbable that any experimental series is minimum delay, so in view of (A.6 & B.1) it is probable that a good estimate of $|B(f)|^2$ will require significantly fewer terms than make up $|X(f)|^2$.

If equation (A.2) is rewritten with $n-1$ as the upper limit of s, minimization with respect to the filter elements again gives (A.3) but with the limit on s and $\tau = 0, 1, \ldots, n-1$. When in addition $\alpha = 1$, we have

$$\begin{bmatrix} \varphi_0 \varphi_1 \cdots \varphi_{n-1} \\ \varphi_1 \varphi_0 \cdots \varphi_{n-2} \\ \cdot \\ \cdot \\ \varphi_{n-1} \varphi_{n-2} \cdots \varphi_0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \cdot \\ \cdot \\ \varphi_n \end{bmatrix}, \quad (B.2)$$

which is the same as Peacock & Treitel's (1969) equation (9). Now if a_t is the unit extrapolation filter such that the estimate of x_{t+1} is

$$\hat{x}_{t+1} = \sum_{s=0}^{n-1} x_{t-s} a_s \simeq x_{t+1}, \quad (B.3)$$

then the (unit) extrapolation-error filter is γ_t, where

$$\varepsilon_t = x_t - \hat{x}_t = \sum_{s=0}^{n} x_{t-s} \gamma_s \quad (B.4)$$

and

$$\gamma_0 = 1, \quad \gamma_1 = -a_0, \quad \gamma_2 = -a_1, \ldots, \quad \gamma_n = -a_{n-1}. \quad (B.5)$$

Apart from the arbitrary scale of (A.11), equations (A.11) and (B.4) are completely analogous. Moreover, because in (A.11) b_t^{inv} must be minimum delay, γ_t also must be minimum delay, or at least very nearly so. Thus, γ_t is in the mean the best n-term approximation of b_t^{inv}, apart from a constant factor. Let it be observed that this may not be the case if the square matrix in (B.2) is not positive definite, or γ_t is not exactly minimum delay, or the Fourier transform of φ_t is not greater than zero at every point.

Rather than find the extrapolation-error filter by (B.2) and (B.5), Peacock & Treitel (1969) showed that it could be obtained directly from a simple modification of (B.2). Their modification is in two steps. First, subtract the right-hand side of (B.2) from both sides. Second, add to the system the equation

$$\sum_{s=0}^{n} \varphi_s \gamma_s = \beta. \quad (B.6)$$

Now, instead of (*B*.2) which gives the unit extrapolation filter, we have

$$\begin{bmatrix} \varphi_0\,\varphi_1 \cdots \varphi_n \\ \varphi_1\,\varphi_0 \cdots \varphi_{n-1} \\ \vdots \\ \vdots \\ \varphi_n\,\varphi_{n-1} \cdots \varphi_0 \end{bmatrix} \begin{bmatrix} \gamma_0 = 1 \\ \gamma_1 \\ \vdots \\ \vdots \\ \gamma_n \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \qquad (B.7)$$

which gives the (unit) extrapolation-error filter. An interpretation of (*B*.6) can be found by multiplying both sides of (*B*.4) by x_{t+p} and determining the expectation in time as follows.

$$E\{x_{t+p}\,\varepsilon_t\} = E\left\{\sum_{s=0}^{n} x_{t+p}\,x_{t-s}\,\gamma_s\right\}. \qquad (B.8)$$

The right-hand side of (*B*.8) is

$$\sum_{s=0}^{n} \varphi_{p+s}\,\gamma_s \qquad (B.9)$$

where we wish to consider only the case $p = 0$. Let us assume that γ_t is minimum delay, or equivalently, that its Fourier transform is minimum phase. Then γ_t^{inv} exists, and (*B*.4) can be rearranged to give x_{t+p} and the left-hand side of (*B*.8) can be approximated by

$$E\left\{\sum_{s=0}^{n} \varepsilon_{t+p-s}\,\gamma_s^{\text{inv}}\,\varepsilon_t\right\}. \qquad (B.10)$$

It is readily verified that $\gamma_0^{\text{inv}} = 1$, so when $p = 0$ (*B*.10) becomes

$$E\{\varepsilon_t\,\varepsilon_t\} + E\{\varepsilon_{t-1}\,\varepsilon_t\,\gamma_1^{\text{inv}}\} + \ldots E\{\varepsilon_{t-n}\,\varepsilon_t\,\gamma_n^{\text{inv}}\}. \qquad (B.11)$$

It is clear that the only non-zero term in (*B*.11) is the first and it is σ_n^2, the variance of ε_t as given by (*B*.4). The equality of β and σ_n^2 is considered in detail by McDonough (1974). In the geophysical literature this quantity is also referred to as 'prediction error' (Burg, 1972) and 'prediction-error power' (Ulrych & Bishop, 1975).

Equation (*B*.4) can be transformed to give

$$\frac{1}{L}\sum_{t=-\infty}^{\infty} \varepsilon_t \exp\{-2\pi itf\} = \frac{1}{L}\sum_{t=-\infty}^{\infty}\sum_{s=0}^{n} x_{t-s}\,\gamma_s \exp\{-2\pi itf\}, \qquad (B.12)$$

where *f* is in fractions of its period, *L*. If on the right-hand side we set $t' = t - s$, then drop the prime, (*B*.12) becomes

$$\frac{1}{L}\sum_{t=-\infty}^{\infty} \varepsilon_t \exp\{-2\pi itf\} = \frac{1}{L}\sum_{t=-\infty}^{\infty} x_t \exp\{-2\pi itf\} \sum_{s=0}^{n} \gamma_s \exp\{-2\pi isf\}, \qquad (B.13)$$

or

$$\xi(f) = X(f) \sum_{s=0}^{n} \gamma_s \exp\{-2\pi isf\}. \qquad (B.14)$$

Because $|\xi(f)|^2$ is the transform of the autocorrelation of ε_t,

$$|\xi(f)|^2 = \sigma_n^2/L = |X(f)|^2 \left|\sum_{s=0}^{n} \gamma_s \exp\{-2\pi isf\}\right|^2, \qquad (B.15)$$

and

$$|X(f)|^2 = \frac{\sigma_n^2/L}{\left|\sum_{s=0}^{n} \gamma_s \exp\{-2\pi isf\}\right|^2}. \qquad (B.16)$$

Equation (*B*.16) expresses Burg's maximum-entropy spectral-density estimate for x_t (Lacoss, 1971) and is identical to the spectral-density estimate for x_t modeled as an order-*n* autoregressive process (McDonough, 1974). The correspondence and distinctives of extrapolative-filtering, autoregressive, maximum-entropy, and even maximum-likelihood calculations are discussed by Ulrych & Bishop (1975); at present, their meaning and significance in crystallographic applications is not certain.

It is clear, however, that with the assumption given by (2.1), and upon renaming variables as

$$\begin{aligned} x_t &\to G_h, \\ \varphi_t &\to F_k, \\ X(f) &\to g_x, \\ a_t &\to A_k, \\ \gamma_t &\to C_k; \end{aligned}$$

then (*B*.16) becomes the same as (2.27), the principal result of this paper.

References

Bode, H. W. & Shannon, C. E. (1950). *Proc. IRE*, **38**, 417–425.

Burg, J. P. (1972). *Geophysics*, **37**, 375–376.

Churchill, Ruel V. (1960). *Complex Variables and Applications*, p. 138. New York: McGraw-Hill.

Fejér, L. (1915). *J. Reine Angew. Math.* **146**, 53–82.

Goedkoop, J. A. (1950). *Acta Cryst.* **3**, 374–378.

Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.

Lacoss, R. T. (1971). *Geophysics*, **36**, 661–675.

McDonough, R. N. (1974). *Geophysics*, **39**, 843–851.

Oppenheim, A. V. & Schafer, R. W. (1975). *Digital Signal Processing*, p. 371. Englewood Cliffs, New Jersey: Prentice-Hall.

Papoulis, A. (1962). *The Fourier Integral and its Applications*. New York: McGraw-Hill.

Papoulis, A. (1973). *IEEE Trans. Inf. Theory*, **19**, 9–12.

Peacock, K. L. & Treitel, S. (1969). *Geophysics*, **34**, 155–169.

Robinson, E. A. (1967). *Statistical Communication and Detection*. New York: Hafner.

Ulrych, T. J. & Bishop, T. N. (1975). *Rev. Geophys. Space Phys.* **13**, 183–200.

Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: John Wiley.