Journal of Statistical Physics, Vol. 76, Nos. 1/2, 1994

Application of Generalized Linear Filters in Data Analysis

S. E. Barnes,[1,2] M. Peter,[2] L. Hoffmann,[2] A. A. Manuel, [2] and A. Shukla[2]

Received July 14, 1993

It is shown that a useful generalized linear filter W can be constructed from experimental data. The data are divided into many experiments and this ensemble is used to calculate the autocorrelation functions which appear in W. In turn, from this filter one determines a "Hamiltonian" $\mathscr{H}$. The eigenvectors and eigenvalues of this Hamiltonian are evaluated. For a "good" experiment there is one small eigenvalue, and the rest are ~ 1. The W so determined usefully reduces the noise in a new data set. The presence of two or more small eigenvalues indicates that the experimental data contains more than a single signal. The action of W on selected members of the ensemble, and/or new data sets, extracts the different signals with, again, a useful noise reduction. Both computer simulations and real positron annihilation data are used to illustrate these developments.

KEY WORDS: Generalized linear filters; data analysis; image processing.

1. INTRODUCTION

The idea of using a linear filter for the treatment of data originates with Wiener (see, e.g., ref. 1) and was developed almost half a century ago in the context of a passive electronic filter element for the treatment of radar signals. More recently, a generalization of this technique has been applied to the problem of the treatment of images (see, e.g., refs. 1 and 2). Here we present what we believe is a new application of these methods (for a different approach see refs. 3) for the analysis of large data sets accumulated

This paper is dedicated to Philippe Choquard on the occasion of this 65th birthday.

[1] Physics Department, University of Miami, Coral Gables, Florida 33124.

[2] Département de Physique de la Matière Condensée, Ecole de Physique, Université de Genève, 1211 Geneva 4, Switzerland.

0022-4715/94/0700-0679$07.00/0

in connection with, e.g., two-dimensional angular correlation positron annihilation experiments (see, e.g., the work of the Geneva group[4]).

As is the case for the experiments in question,[4] it is envisaged that a data set comprises a two-dimensional image $g_i(x, y)$. Here the subscript i labels the data set and the coordinates x, y locate a given pixel in that image. Dirac notation will be adopted, i.e., there is a vector $|g_i\rangle$ for which $g_i(x, y) = \langle x, y | g_i \rangle$ is its representation in the *pixel set*. The signal corresponds to $|f\rangle$ and the data set $|g_i\rangle = |f\rangle + |n_i\rangle$ is degraded by a particular example of the noise reflected by $|n_i\rangle$. Initially, the idea is to demand a filter W such that $W|g_i\rangle = |f\rangle$. Clearly such a filter minimizes (actually yields zero for) the average normalization $A = \overline{\langle e_i | e_i \rangle}$, where the "error" vector is $|e_i\rangle = W|g_i\rangle - |f\rangle$. If such a W can be found it is described as *ideal*, since it recovers the noise-free signal from a noisy data set. Of course, the criterion that the mean square error A be a minimum is not unique. The present formulation leads to a linear filter. The development, e.g., of a nonlinear filter is also possible, but if formulated in the present fashion, would correspond to some other minimal criterion.

In this analysis it is assumed, as is the case for the experiments of interest, that the noise power is known. In the context of the positron data, the average amplitude is the usual $\sim\sqrt{N_i}$, where N_i is the number of counts associated with a given pixel, i.e., in this case there is good model for the noise. In other contexts it might be that the noise can be measured separately, as in the original problem considered by Wiener. Whichever is the case, the knowledge of the noise is contained in an autocorrelation function $R_{nn} = \overline{|n_i\rangle\langle n_i|}$. There will be similar definitions for R_{ff} and R_{gg}.

Common sense dictates that an ideal filter can only be found under very special circumstances. Such a filter exists in the simplest case when it is assumed that the above-defined average is for an infinite ensemble of experiments. This simplest theory is presented in Section 1. Section 2 presents an analysis valid when a single W is sought but in fact there are two or more signals $|f_1\rangle$, $|f_2\rangle$,... contained in the data. It is shown that the same general formula for W can be used for the two cases and that therefore this filter can be used as an analytic tool by which to examine data sets for the presence of many signals or signals which fluctuate between experiments. In this context it is useful to introduce a "Hamiltonian" $\mathcal{H} = (1 - W^\dagger)(1 - W)$. This $\mathcal{H}$ has as many small eigenvectors as there are significant signals in the data set, i.e., for the present example, there are two eigenvalues which are approximately zero and the rest which correspond to the possible noise signals having eigenvalues of unity. The eigenvectors with small eigenvalues correspond to the average and difference of the signals.

In Section 3 attention is focused on the more realistic problem when,

in fact, there is only a finite set of experiments. In this section it is still assumed that the number of experiments is large but not infinite, i.e., account is taken of the fact that in general $|\bar{n}\rangle$, the average noise, is *not* negligible compared to the average signal. Surprisingly, there is still an ideal filter W which will remove all noise. Unfortunately, or perhaps evidently, the construction of such a filter requires precisely a foreknowledge of the signal itself and therefore is not of utility in the present experimental context. In fact, in the absence of *a priori* knowledge about the signal, the best which can be done is to form some average $|\bar{g}\rangle$ of the data. The construction of a filter which returns such an optimal average is described in this section.

Section 4 describes the truly realistic case, in the experimental context, which interests the present authors. There are a large number of pixels, typically an array of a few hundreds by a few hundreds, but only a small number $N_e \sim 10$ experiments. In the present case it is not possible to compare the noise power with that which would be reflected by an R_{nn} which would pertain for the classical case of an infinite ensemble. Instead the noise power is roughly equally distributed over a limited number of signal degrees of freedom. Also discussed in this section is the more practical problem of matrix and array size. Arrays of the size mentioned above would imply matrices such as R_{gg} which would have dimensions of the order of 10^8, which is simply not practical. It is shown how the problem can be reduced so that the matrices are never larger than $N_e \times N_e$.

In each section the theory is illustrated. In Sections 1–3 the examples are computer generated. Section 4 has examples drawn from experimental positron annihilation data. Finally Section 5 contains the discussion and our conclusions.

1. SIMPLEST CASE: ONE SIGNAL, MANY EXPERIMENTS

In this section we consider the simplest case, namely the generalized linear filter which corresponds to an infinite number of experiments. The aim is to find a filter which extracts, in an optimal fashion, a signal corresponding to a vector $|f\rangle$ from the ith experiment in which the data corresponding to $|g_i\rangle = |f\rangle + |n_i\rangle$ is the signal degraded by the presence of the noise reflected by $|n_i\rangle$. What is required is a filter W such that $|e_i\rangle = W|g_i\rangle - |f\rangle$ is a minimum in the appropriate sense. More specifically, the filter is required to minimize

$$A = \overline{\langle e|e\rangle} \tag{1.1}$$

where $\bar{X}$ implies the ensemble average of X over a very large (infinite in the limit) number of experiments. Since it is only the noise which is to change

from experiment to experiment, this is equivalent to an average over all possible noise signals. Evidently, $|e\rangle = (W-1)|f\rangle + W|n\rangle$, whence

$$A = \overline{\langle f|(1-W^\dagger)(1-W)|f\rangle} - \overline{\langle f|W^\dagger(1-W^\dagger)|n\rangle} - \overline{\langle n|W^\dagger(1-W)|f\rangle} + \overline{\langle n|W^\dagger W|n\rangle} \tag{1.2}$$

Since only the noise changes between the different members of the ensemble, this can be simplified to

$$\begin{aligned} A &= \langle f|(1-W^\dagger)(1-W)|f\rangle - \langle f|(1-W^\dagger)W\overline{|n\rangle} \\ &\quad - \overline{\langle n|}W^\dagger(1-W)|f\rangle + \overline{\langle n|W^\dagger W|n\rangle} \\ &= \langle f|(1-W^\dagger)(1-W)|f\rangle + \overline{\langle n|W^\dagger W|n\rangle} \end{aligned} \tag{1.3}$$

where in writing the last line it has been observed that, e.g., $\overline{|n\rangle} \to 0$ because the average is over an infinite ensemble of random noise signals.

The next step is to expand this equation using an appropriate complete set of states. To this end, the *pixel* set $|x, y\rangle$ is introduced. The data comprise a two-dimensional image. It is *not* implied that the pixels are independent. It might be that if an experimental event implies a finite intensity at x, y there is also an intensity in the neighboring pixels. In this basis set $g(x, y) = \langle x, y|g\rangle$ corresponds to the actual image which represents the data set. With this definition,

$$A = \mathrm{Tr}[(1-W^\dagger)R_{ff}(1-W) + W^\dagger R_{nn}W] \tag{1.4}$$

involves the matrix

$$\langle x, y|R_{ff}|x', y'\rangle = \langle x, y|f\rangle\langle f|x', y'\rangle \tag{1.5}$$

while

$$\langle x, y|R_{nn}|x', y'\rangle = \overline{\langle x, y|n\rangle\langle n|x', y'\rangle} \tag{1.6}$$

and where now W also denotes a matrix in the pixel vector space.

Demanding that the variance $\delta A = 0$ for, e.g., every $\delta W^\dagger$ results in the equation

$$\delta A = 0 = \mathrm{Tr}[\delta W^\dagger(R_{ff}(1-W) + R_{nn}W)] \tag{1.7}$$

which implies that $(R_{ff}(1-W) + R_{nn}W) = 0$, i.e., the optimal linear filter

$$W = R_{ff}(R_{ff} + R_{nn})^{-1} \tag{1.8}$$

This is written in terms of the signal autocorrelation operator $R_{ff} = |f\rangle\langle f|$, which is not accessible experimentally. However, this is related to the similar autocorrelation functions

$$R_{gg} = \overline{|g\rangle\langle g|} \tag{1.9}$$

for the data and R_{nn} for the noise. Using again $\overline{|n\rangle} = 0$, valid for an infinite ensemble, it is trivial that

$$R_{ff} = R_{gg} - R_{nn} \tag{1.10}$$

which represents a statement that the "power" in each pixel of the signal is equal to that in the data minus the noise power.

It is assumed possible to calculate from a model and/or determine in a separate experiment the noise autocorrelation function, whence combining (1.8) and (1.10) yields the filter W.

This filter W is an ideal filter in the sense that, given a noisy data set $|g_1\rangle = |f\rangle + |n_1\rangle$, the result

$$W|g_1\rangle \to |f\rangle \tag{1.11}$$

i.e., is noise-free. The intensity of the result reflects the strength of the signal in the data.

There exists a "Hamiltonian" which follows from W. It is observed that $A = \langle f|\,(1 - W^\dagger)(1 - W)\,|f\rangle + \overline{\langle n|\,W^\dagger W\,|n\rangle}$ is the sum of two definite positive terms, since each is the normalization integral of some vector. Since W minimizes A, it follows that $|f\rangle$ is a vector which minimizes the expectation value of the Hamiltonian

$$\mathscr{H} = (1 - W^\dagger)(1 - W) \tag{1.12}$$

Since $\mathscr{H}$ is manifestly Hermitian, it follows that $|f\rangle$ is the ground-state vector for this Hamiltonian. The eigenvalue

$$f_0 = \frac{1}{(1 + \langle f|\,R_{nn}\,|f\rangle)^2} \tag{1.13}$$

corresponding to $|f\rangle$, reflects the signal-to-noise ratio in a typical member of the data set. Further, the remaining eigenvalues are all unity, i.e., there is one small eigenvector corresponding to the signal and many which are close to unity and have eigenvectors which reflect noise signals.

On occasion, it is desired to find the best fit to the data of a set of models $|m_1\rangle$, $|m_2\rangle$, $|m_3\rangle$, $|m_4\rangle$,.... Because of the well-known Hamiltonian

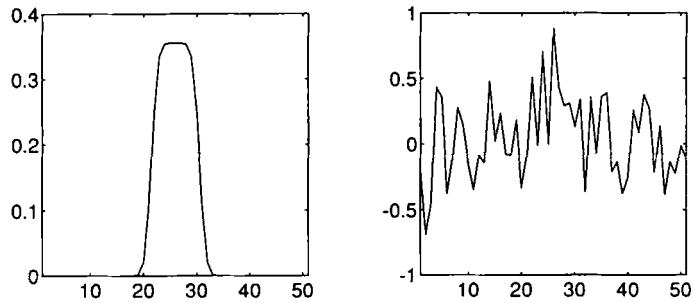

Fig. 1. Left: The noise-free signal; right: a typical data set. The signal-to-noise ratio is roughly unity.

minimal principle, it follows that the model with the smallest expectation value,

$$e = \frac{\langle m_i | \mathscr{H} | m_i \rangle}{\langle m_i | m_i \rangle} \tag{1.14}$$

is the closest to $|f\rangle$.

The developments presented in this section are illustrated in Figs. 1 and 2. To the left in Fig. 1 is shown the noise-free signal, while to the right is a typical data set. The signal-to-noise ratio is roughly unity. In Fig. 2 is shown, from left to right, the filtered signal and the eigenvectors of $\mathscr{H}$ with the lowest and the next lowest eigenvalues. The lowest eigenvalue is zero with numerical precision and almost unity for what is evidently a noise signal on the right.

Of course, the result that W is an ideal filter, with only a single small eigenvalue for $\mathscr{H}$, corresponds to an idealized situation which cannot be realized in practice. However, these results remain approximately valid in a sense which will be elaborated upon below.

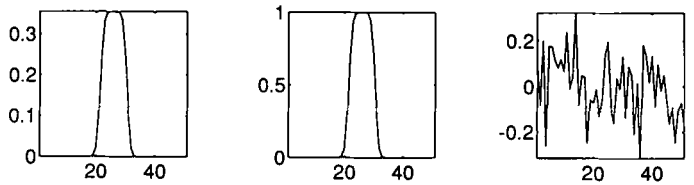

Fig. 2. Left to right: The filtered signal and the eigenvectors of $\mathscr{H}$ with the lowest and the next lowest eigenvalues. The lowest eigenvalue is zero with numerical precision and almost unity for what is evidently a noise signal on the right.

2. SEVERAL SIGNALS, MANY EXPERIMENTS

Because of its nature, it is possible that an experiment yields, in a nonrandom fashion, data sets which correspond to more than one signal. The signal vectors corresponding to these possibilities are $|f_1\rangle$, $|f_2\rangle$, etc., and the experiment gives the results $|g_i\rangle = |f_1\rangle + |n_i\rangle$, $|g_i\rangle = |f_2\rangle + |n_i\rangle$, etc., with probabilities p_1, p_2, etc. That these data sets differ in a nonrandom fashion implies, in the present context, that, e.g., $|f_1\rangle - |f_2\rangle$ is *at the most* an improbable noise signal. In writing the above it has been assumed that the noise signal is of the same statistical character for the two signals.

Consider the case when there are only two possible signals. Given that the difference between the signals is larger than the noise and that a suitable filter W exists, then it is clear that if, for the ith experiment,

$$|g_i\rangle = |f_1\rangle + |n_i\rangle \tag{2.1}$$

the difference

$$W|g_i\rangle - |f_1\rangle \tag{2.2}$$

will have a smaller normalization than

$$W|g_i\rangle - |f_2\rangle \tag{2.3}$$

It is desired that W minimizes the average of the *smaller* of these normalizations which reflect the "error." Generally, this smaller difference is

$$|d_1\rangle = W|n_i\rangle - (1 - W)|f_1\rangle \tag{2.4}$$

or

$$|d_2\rangle = W|n_i\rangle - (1 - W)|f_2\rangle \tag{2.5}$$

and the quantity to minimize is

$$A = p_1\overline{\langle d_1|d_1\rangle} + p_2\overline{\langle d_2|d_2\rangle} \tag{2.6}$$

where the average is over all noise signals, and where again the probabilities of the signals are p_1 and p_2. It is straightforward to derive the result that, again,

$$W = R_{ff}R_{gg}^{-1} \tag{2.7}$$

where now more generally

$$R_{ff} = \sum_i |f_i\rangle\, p_i\, \langle f_i| \tag{2.8}$$

and where still

$$R_{ff} = R_{gg} - R_{nn} \tag{2.9}$$

is determined from the data, given the noise spectrum.

For the special case of two signals with equal weight, it is possible to write

$$|f_1\rangle = |\bar{f}\rangle + \alpha\,|d\rangle \tag{2.10}$$

and

$$|f_2\rangle = |\bar{f}\rangle - \alpha\,|d\rangle \tag{2.11}$$

where it is assumed that both $|\bar{f}\rangle$ and $|d\rangle$ are normalized vectors. It follows that

$$R_{ff} = |\bar{f}\rangle\langle\bar{f}| + \alpha^2\,|d\rangle\langle d| \tag{2.12}$$

It is observed again for this particular case that $|\bar{f}\rangle$ and $|d\rangle$ are automatically orthogonal. However, *in general* it is true that

$$R_{ff} = \sum_i F_i\,|F_i\rangle\langle F_i| \tag{2.13}$$

where F_i and $|F_i\rangle$ are the eigenvalues and orthonormal eigenvectors of R_{ff}. Such a decomposition is *always* possible since R_{ff} is Hermitian.

The filter constructed in this fashion remains essentially "ideal." To a very good approximation, the filter $W = |\bar{f}\rangle\langle\bar{f}| + |d\rangle\langle d|$, and so, given some signal, e.g.,

$$|g_i\rangle = |f_1\rangle + |n_i\rangle = |\bar{f}\rangle \pm \beta\,|d\rangle + |n_i\rangle \tag{2.14}$$

it follows that

$$W\,|g_i\rangle = |\bar{f}\rangle \pm \beta\,|d\rangle \tag{2.15}$$

i.e., the signal with the noise eliminated. Notice that the present formulation leads to the same filter, (2.7), independent of the number of signals anticipated.

The formalism also leads again to a Hamiltonian $\mathscr{H} = (1 - W^\dagger)(1 - W)$. The data now contain n signals and the corresponding number of small eigenvalues. As before, the remaining eigenvalues, reflecting noise signals, have values ~ 1.

This $\mathscr{H}$ represents a way in which to analyze a data set for the presence of more than a single signal without *a priori* knowledge that this is in fact the case.

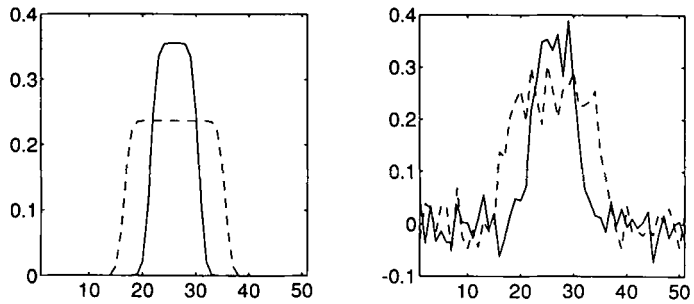

Fig. 3. Left: The two signals; right: two typical data sets, i.e., the signal plus noise.

It is possible to use essentially the same scheme in the case when each data set is a fluctuating mixture of the two signals, i.e.,

$$|g_i\rangle = u_i^2 |f_1\rangle + v_i^2 |f_2\rangle + |n_i\rangle \tag{2.16}$$

with $v_i^2 + u_i^2 = 1$. The two signals might, e.g., be recovered by filtering many data sets and retaining only the extreme cases.

The results of this section are illustrated by Figs. 3–5. To the left of Fig. 3 are shown the two signals, while to the right are two typical data sets, i.e., the signal plus noise. Figure 4 shows from left to right the three eigenvectors of $\mathscr{H}$ with the lowest eigenvalues. On the left the vector with the smallest eigenvalue is seen to correspond to the average signal. The center is a vector which corresponds to the difference, while the third eigenvalue corresponds to the noise signal. In Fig. 5 are shown the two signals recovered by filtering typical members of the data set.

These figures reflect a real potential application of the present methods. A filter can be constructed which embodies the collection of data for some number of possible experimental outcomes. The filter will

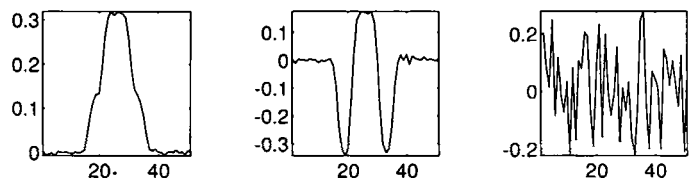

Fig. 4. Left to right: The three eigenvectors of $\mathscr{H}$ with the lowest eigenvalues. On the left the vector with the smallest eigenvalue is seen to correspond to the average signal. The center is a vector which corresponds to the difference, while the third eigenvalue corresponds to a noise signal.

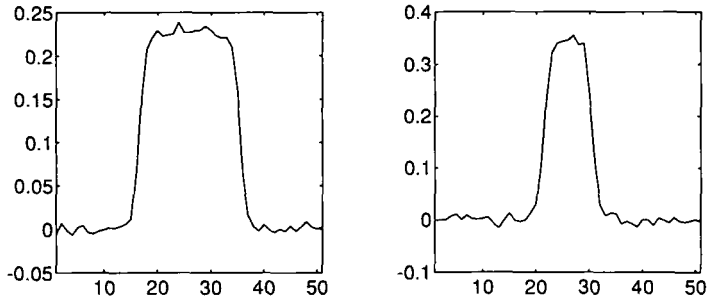

Fig. 5. The two signals recovered by filtering typical members of the data set.

accurately return the appropriate signal stripped of its noise. It is a scheme which uses the *a priori* knowledge gained in many previous experiments to perform a useful noise reduction.

3. MANY SIGNALS, LIMITED ENSEMBLE

It is rarely possible in interesting situations to perform a limiting large number of experiments. It is therefore appropriate to modify the analysis for the case in which there remains an important average noise $|\bar{n}\rangle$.

Consider the simplest case when only a single signal is expected. It is relatively straightforward to repeat the analysis of Section 1, retaining the average noise signal. The resulting *ideal* filter is

$$W = |f\rangle\langle\bar{g}|\, R_{gg}^{-1} \tag{3.1}$$

However, in the absence of any *a priori* knowledge of $|f\rangle$, it is impossible to construct this filter. It is intriguing that such a knowledge would not be necessary if it remained the case that $|f\rangle$ was the ground state for $\mathscr{H} = (1 - W^{\dagger})(1 - W)$, since there are any number of ways to find, e.g., numerically, such a ground state. Unfortunately it is not possible to establish the required minimal principle for this case.

In fact, on general grounds, it is clear that for a given ensemble $\{|g_i\rangle\}$ of data, and in the absence of any *a priori* knowledge except for a model for the noise power, the filtered signal can only be some weighted average of the set $\{|g_i\rangle\}$, i.e., the filtered signal is

$$|\bar{g}\rangle = \sum_i w_i |g_i\rangle \equiv |f\rangle + |\bar{n}\rangle \tag{3.2}$$

were the weights must obey $\sum_i w_i = 1$. The filter should yield $|\bar{g}\rangle$ as a result, hence the error vector is defined to be $|e_i\rangle = W|g_i\rangle - |\bar{g}\rangle$. It is easy to show that

$$R_{gg} = R_{\bar{g}\bar{g}} + R_{nn} \tag{3.3}$$

where

$$R_{\bar{g}\bar{g}} = |\bar{g}\rangle\langle\bar{g}| \tag{3.4}$$

plays the role of R_{ff} in the earlier development. The filter is

$$W = R_{\bar{g}\bar{g}} R_{gg}^{-1} \tag{3.5}$$

However, the optimal weights $\{w_i\}$ have yet to be determined. Substituting the result for W into the expression for A, it is found that the minimum for A corresponds to a maximum for a matrix element of the noise power matrix:

$$\langle\bar{g}|\, R_{nn}\, |\bar{g}\rangle \tag{3.6}$$

where this matrix is now defined by

$$R_{nn} = \sum_i |\delta n_i\rangle\, w_i\, \langle\delta n_i| \tag{3.7}$$

with $|\delta n_i\rangle = |n_i\rangle - |\bar{n}\rangle$, and where $|\bar{n}\rangle$ was defined by (3.2). Using this definition, we find that the quantity to be maximized is

$$S = \sum_i w_i\, |\langle\bar{g}|\,\delta n_i\rangle|^2 \tag{3.8}$$

subject to the constraint that $\sum w_i = 1$. Introducing a Lagrange multiplier α and taking the variation with respect to each w_i results in

$$\alpha = -\frac{d}{dw_1}\left(w_1\, |\langle\bar{g}|\,\delta n_1\rangle|^2\right) = -\frac{d}{dw_2}\left(w_2\, |\langle\bar{g}|\,\delta n_2\rangle|^2\right) = \cdots \tag{3.9}$$

Given that the data sets differ only in their noise signals, and that there are sufficient pixels in the image that the total noise power in the image is the same for each example of the noise, then the functional dependence $|\langle\bar{g}|\,\delta n_i\rangle|^2 \equiv F(w_i)$ must be the same independent of i. It follows that

$$w_1 = w_2 = w_3 = \cdots \tag{3.10}$$

i.e., the expected result that the best expression for $|\bar{g}\rangle$ is simply the plain average over all the data sets. An example when the data sets contain noise with different powers will be discussed in the next section.

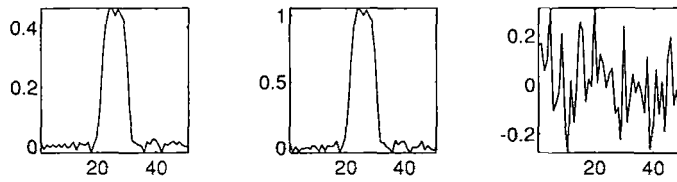

Fig. 6. The signal and noise treated are similar to those shown in Fig. 1. Left: A typical filtered signal; center: the vector with the lowest eigenvalue; right: the noise signal associated with the second smallest eigenvalue. Observe that now the filtered signal remains degraded by the average noise.

Given the specification of the filter for many signals given at the beginning of Section 2, the derivation is identical with the simple replacement $|f\rangle \rightarrow |\bar{g}\rangle$, i.e., the same result (2.7) obtains, and the discussion of that section applies.

Figure 6 illustrates the developments of this section, again using computer simulations. The signal and noise are similar to those shown in Fig. 1. The panels of Fig. 6 correspond, on the left, to a typical filtered signal, in the center to the vector with the lowest eigenvalue, and to the right to the noise signal associated with the second smallest eigenvalue. It is to be observed that now the filtered signal remains degraded by the average noise, and the separation between the single small (~ 0) and large (~ 1) eigenvalues is less extreme.

4. SEVERAL SIGNALS, SEVERAL EXPERIMENTS

This is perhaps the case of most interest. The idea is that, although it might be expected that the experiment corresponds to a single signal, the analysis will admit the possibility that, in fact, there are nonrandom fluctuations from experiment to experiment and/or that data contain rather a mixture of different signals.

The important difference envisaged in this section is that there are far too few experiments for the convergence of R_{nn} to that expected for a very large ensemble. For example, if the experiment yields a 100 by 100 array of pixels, then the convergence necessary for the equation $R_{gg} = R_{\bar{g}\bar{g}} + R_{nn}$ would require many times $100 \times 100 = 10^4$ experiments, which typically is not practical, or even desirable. The situation envisaged in the present analysis is that there are perhaps $N_e = 10$ experiments for an array of this size. It *is* assumed that the number of pixels is large. With this assumption, it *is* the case that, for each experiment, the noise power Tr R_{nn} converges very well to its average value p_n.

The other matter to be dealt with in this section is more practical in nature. With 10^4 total pixels the matrices R_{gg} and R_{nn}, etc., *if* expressed in the pixel basis, are of dimension $(10^4)^2 = 10^8$, which is very large even by current computing standards. Such large matrices are, in fact, not necessary and not useful. It has already been observed that the filtered signal cannot be other than a linear combination of the various data vectors $|g_i\rangle$ and this implies that the dimensions of the matrices need not exceed the number of data sets N_e.

In order to form a basis set, the autocorrelation function

$$R_{gg} = \overline{|g_i\rangle\langle g_i|} = \sum_{i=1}^{N_e} w_i |g_i\rangle\langle g_i| = \sum_{i=1}^{N_e} G_i |G_i\rangle\langle G_i| \tag{4.1}$$

for the data is constructed. The expression to the right corresponds to the result of diagonalizing the previous, experimentally determined expression. The $|G_i\rangle$ are a set of N_e orthonormal eigenvectors of R_{gg}, while G_i are the corresponding eigenvalues. Necessarily these vectors are a complete set with respect to the data vectors, i.e.,

$$|g_i\rangle = \sum_{n=1}^{N_e} d_n^i |G_n\rangle \tag{4.2}$$

In this data basis set, the specific data set $|g_i\rangle$ is represented by a vector $(d_1^i, d_2^i, \ldots, d_{N_e}^i)$. Clearly *any* relevant vector can be expressed in a similar fahsion. Since there are very few members of the data set as compared with the number of (nonzero) pixels, the noise set $|\delta n_i\rangle$ are, to a very good approximation, independent vectors each of which, to a not so good approximation, has the same norm $\langle \delta n_i | \delta n_i \rangle \sim p_n$, i.e., roughly the same noise power spectrum. Even if there are a few ($< N_e$) effective signals, most of the vectors $|G_n\rangle$ are normalized examples of the noise and the corresponding $G_n \sim P_n$.

For simplicity of presentation, it is assumed that the data contain only a single valid signal. The model noise matrix is

$$R_{nn} = \sum_{i=2}^{N_e} p_n |G_i\rangle\langle G_i| \tag{4.3}$$

where it is understood that G_1 has the largest eigenvalue. (Notice the lower limit of the sum.) As a practical matter it is convenient to set $p_n = \min\{G_i\}$ and to use, as a quality measure, the ratio $\min\{G_i\}/p_n^{\text{model}}$. The value p_n^{model} is that calculated from a noise model or obtained from an independent experiment. However, there remains a complication. Given the rather small number of data sets N_e, the correct relationship is $R_{gg} = R_{\bar{g}\bar{g}} + R_{nn} - |\bar{n}\rangle\langle\bar{n}|$.

The problem is that $|\bar{n}\rangle$ is not known except for a good estimate for its noise amplitude, which is $\sim\sqrt{N_e}$ times smaller than the noise in a typical data set. It is implied that the eigenvalues for the $i>1$ should lie in a range $1 \to 1-1/\sqrt{N_e}$, i.e., if $N_e \sim 10$, there is to be expected an $\sim 30\%$ variation in the noise eigenvalues. This will be seen in the illustrations below.

The matrix operations necessary for the construction of W and $\mathscr{H}$ are then trivial and can be performed analytically. Specifically, in the $|G_n\rangle$ basis set, R_{gg} is diagonal with a diagonal comprising the eigenvalues $\{G_i\}$, while R_{nn} is similar but with a constant diagonal $\sim p_n$. It follows, ignoring the corrections due to $|\bar{n}\rangle\langle\bar{n}|$, that $R_{\bar{g}\bar{g}}$ has a diagonal $\{G_i - p_n\}$, while W has the diagonal $\{1-p_n/G_i\}$, and the diagonal of $\mathscr{H}$ reflects the eigenvalues $\{(p_n/G_i)^2\}$. Clearly the eigenvectors for all of these mathematical objects are the $|G_i\rangle$ and the only real computational effort required is to determine the coefficients in the decomposition

$$|G_i\rangle = \sum_{n=1}^{N_e} c_n^i (w_n)^{1/2} |g_n\rangle \tag{4.4}$$

These c_n^i are determined by observing that the $|G_i\rangle$ are eigenvectors of R_{gg}, i.e.,

$$R_{gg}|G_i\rangle = \sum_{j=1}^{N_e}\sum_{k=1}^{N_e} (w_j)^{1/2}|g_j\rangle(w_j)^{1/2}\langle g_j|\, g_k\rangle(w_k)^{1/2} c_k^i = G_i|G_i\rangle \tag{4.5}$$

which implies that the column vector $(c_k^i; k=1, N_e)^T$ is an eigenvector of the symmetric matrix

$$M(j,k) = (w_j)^{1/2}\langle g_j|\, g_k\rangle(w_k)^{1/2} \tag{4.6}$$

constructed from the overlap between the various data sets and the weight factors w_i introduced in Section 3. The eigenvectors written in terms of pixels, with the coordinates (x, y), $G_i(x, y)$, are obtained from

$$G_i(x,y) = \sum_{n=1}^{N_e} c_n^i (w_n)^{1/2} g_n(x,y) \tag{4.7}$$

The filtered result is

$$Ws(x,y) = \sum_{i=1}^{N_e}\sum_{j=1}^{N_e} c_j^i (1-p_n/G_i)(w_j)^{1/2}\langle g_j|s\rangle\, g_j(x,y) \tag{4.8}$$

where $|s\rangle$ is some new data vector to be filtered, and $\langle g_j|s\rangle$ are the overlaps with the already specified data sets. None of the numerical calculations involves matrices larger than N_e by N_e and the largest objects to be stored

are the N_e data sets $g_i(x, y)$. The last task in connection with this filter is to calculate the weights w_i. These involve the overlaps $\langle \bar{g} | \delta n_i \rangle$ and must be calculated self-consistently. With conditions specified at the beginning of this section, the overlap is $\langle f | \delta n_i \rangle \sim 0$, since the sign of the elements of δn_1 are random and there are many pixels. Further, since there are few examples of the noise and again since there are many pixels, $\langle \delta n_j | \delta n_i \rangle \sim 0$ if $i \neq j$. Statistically the only part of the overlap which can have any appreciable magnitude is $w_i \langle \delta n_i | \delta n_i \rangle \equiv w_i p_n^i$, where p_n^i is the noise power for the data set $|g_i\rangle$. It follows that $\alpha = -(d/dw_i)\, w_i^3 (p_n^i)^2 = -3 w_i^2 (p_n^i)^2$, whence

$$w_i = \left(\overline{\frac{1}{p_n}}\right)^{-1} \frac{1}{p_n^i} \tag{4.9}$$

where

$$\left(\overline{\frac{1}{p_n}}\right)$$

is the harmonic average of the noise power.

This is the expected result for the weights w_i. For example, consider the case where each pixel is the accumulation of a certain number of counts and there is no other appreciable source of noise. If the total number of counts for a given normalized data set is N_i, then the noise divided by the signal goes as $1/\sqrt{N_i}$ and the corresponding power is $p_n^i \sim 1/N_i$. It follows

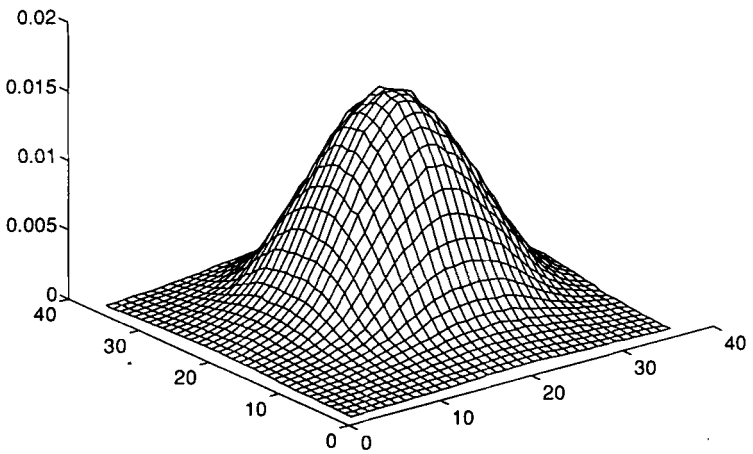

Fig. 7. A typical data set.

that a given normalized data set must be weighted by N_i, the total number of counts it contains. This finally implies that $|\bar{g}\rangle$ is simply the result obtained by accumulating all of the counts into one data set.

Figures 7–15 illustrate an analysis of experimental data for an untwinned sample of the high-temperature superconductor $YBa_2Cu_3O_{7-\delta}$. The total number of useful counts is $\sim 10^9$ and these have been divided among nine different experiments. The number of counts was not the same for each experiment and hence the formalism with the weight factor w_i was necessary. Figure 7 shows a typical data set. Figure 8 plots the entries in the eigenvector $\{c_n^i\}$ of the matrix M which corresponds to the lowest eigenvalue. This reflects the weights w_i and illustrates that, while they are comparable, they are not all the same. It is observed that data set 2 had the lowest number of counts and also the lowest weight in this vector. This same eigenvector in the pixel basis is shown in Fig. 9. This does not differ from the data averaged with the weights w_i. Figure 10 illustrates a typical eigenvector corresponding to the noise. Because of the large number of pixels, such a vector is difficult to represent in a useful fashion. To create Fig. 10, the actual result, which could not be reproduced in a printable form, was averaged using a Gaussian smoothing function. The eigenvalues

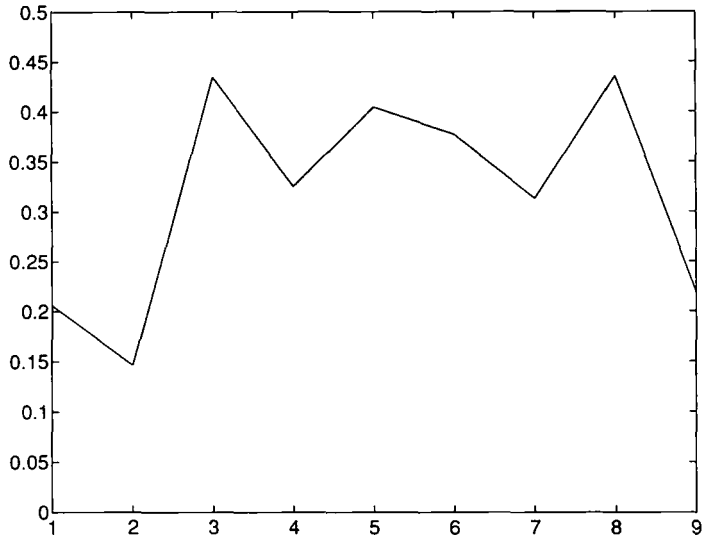

Fig. 8. The entries in the eigenvector $\{c_n^i\}$ of the matrix M which corresponds to the lowest eigenvalue. This reflects the weights w_i and illustrates that, while they are comparable, they are not all the same. It is observed that data set 2 had the lowest number of counts and also the lowest weight in this vector.

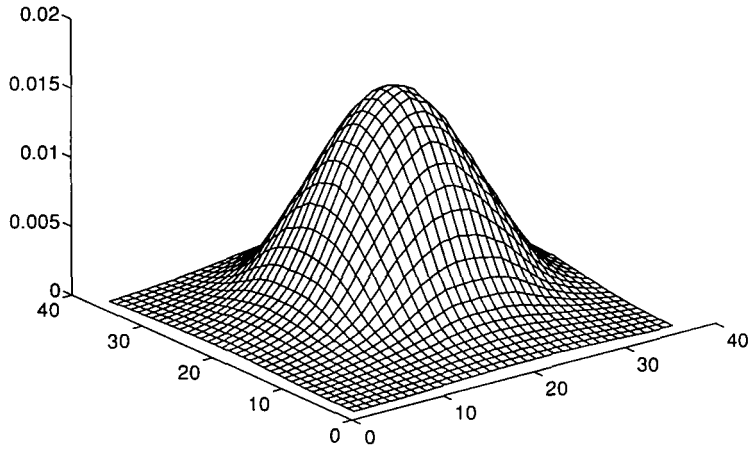

Fig. 9. This same eigenvector as in Fig. 8, but now in the pixel basis.

of $\mathscr{H}$ for this example were 1.00, 0.96, 0.95, 0.85, 0.82, 0.77, 0.67, 0.65, 0.00. It has been verified that all the eigenvectors, except that associated with the smallest eigenvalue, have the appearance of noise, as for the example, Fig. 10. The spread in the eigenvalues for the noise is $\sim 33\% \sim 1/\sqrt{9}$, as would be expected for nine data sets.

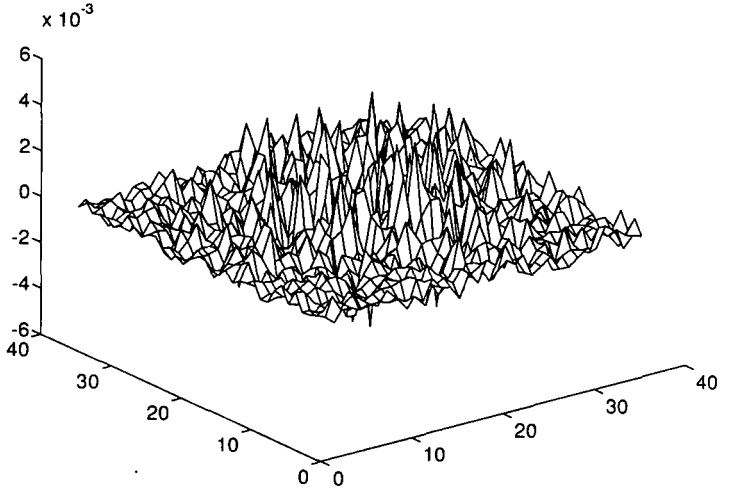

Fig. 10. A typical eigenvector corresponding to the noise. Because of the large number of pixels, such a vector is difficult to represent in a useful fashion. To create this figure the actual result, which could not be reproduced in a printable form, was averaged using a Gaussian smoothing function.

In addition to the apparent, nearly isotropic, approximately Gaussian structure seen in Fig. 9, there is a small ridge which passes over the summit. In spite of this ridge the data for this crystal orientation should have a symmetry which permits the signal to be added to itself after a rotation by 180 deg. Such an averaging procedure has been simulated by performing reflections about the x or y axis of half of the data sets. This is simpler, for computational reasons, than the correct procedure, which would be to double the effective number of data sets by adding as new sets the original ones modified by such a reflection. If the symmetry is good, there should remain only a single small eigenvalue. If this is verified for a reflection about the x axis, this *is not* the case for a reflection about the y axis, due to a small bias introduced by the nonsymmetric sample-source geometry along this direction, which is made evident by the two new eigenvectors, with small eigenvalues, shown in Figs. 11 and 12. These have again been smoothed for ease of reproduction. The lowest two eigenvectors were 0.00 and 0.22 or 0.00 and 0.28 for the two possible reflections.

To illustrate the analysis of two signals, the data sets 5 and 6 were rotated by 90 deg. This should bring the signal due to the ridge into evidence as a new vector with a small eigenvalue. The eigenvector with the smallest eigenvalue is identical in appearance to Fig. 9. The second eigenvector is shown in Fig. 13; again smoothing has been used. The corresponding eigenvalues were 0.00 and 0.0156. An almost identical result to Fig. 13 can be obtained by taking the difference between each data set and the same data set rotated by 90 deg and then averaging over all data sets,

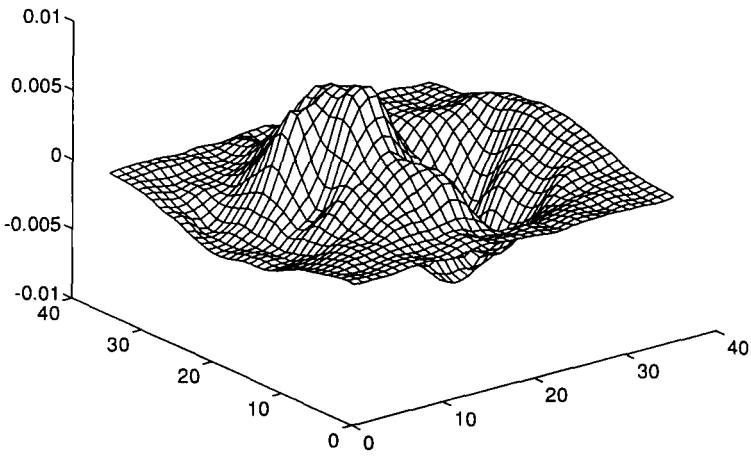

Fig. 11. The eigenvector with the second smallest eigenvector when half the data have undergone a reflection.

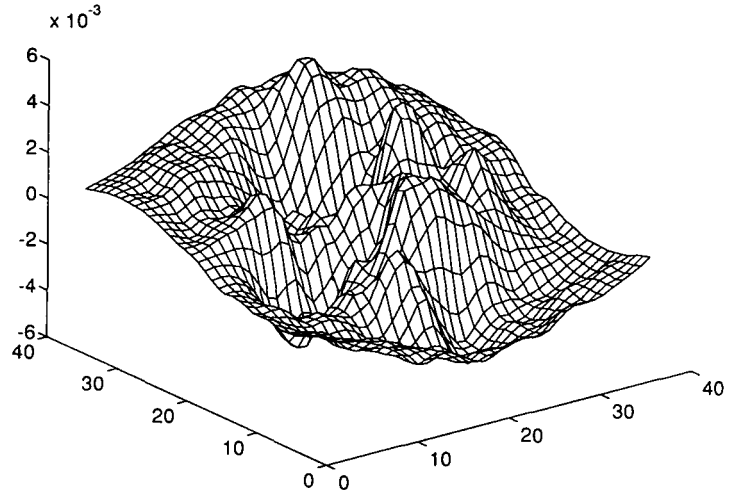

Fig. 12. Similar to Fig. 11, but for the other nonequivalent symmetry reflection.

i.e., this signal corresponds precisely to the anisotropy in the signal due to the ridge. It is of the form of the ridge rotated by 90 deg and subtracted from the unrotated equivalent. Notice, however, that this anisotropy in the data is made evident, in the filter, by the *addition* of data sets via the construction of R_{gg}. Because the anisotropy in the signal is very small, it is

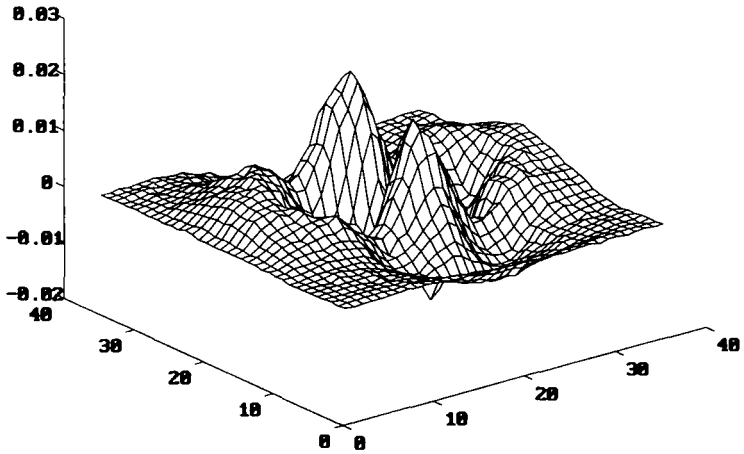

Fig. 13. The second eigenvector for the case when data sets 5 and 6 have undergone a 90 deg rotation. Again smoothing has been used. The corresponding eigenvalue is 0.0156.

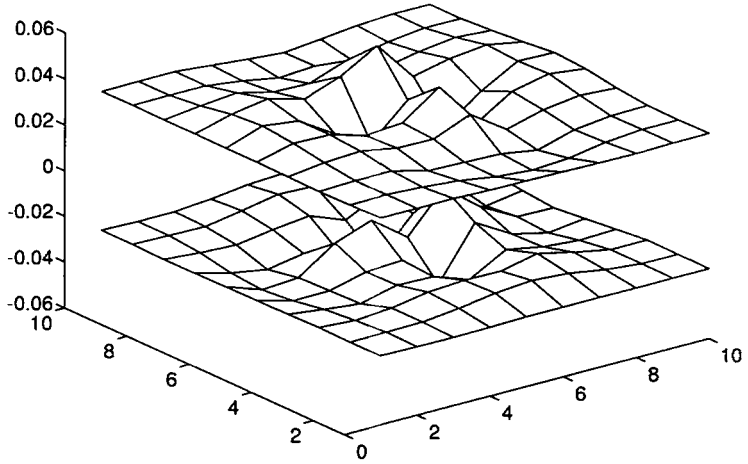

Fig. 14. To show that the ridge *is* differently oriented, here the anisotropy is shown in the *filtered* data sets. The top graph is for data set 5, while the lower is for data set 3. The difference in sign evident in this graph shows that the filter does return the smoothed signal of a ridge with the correct orientation.

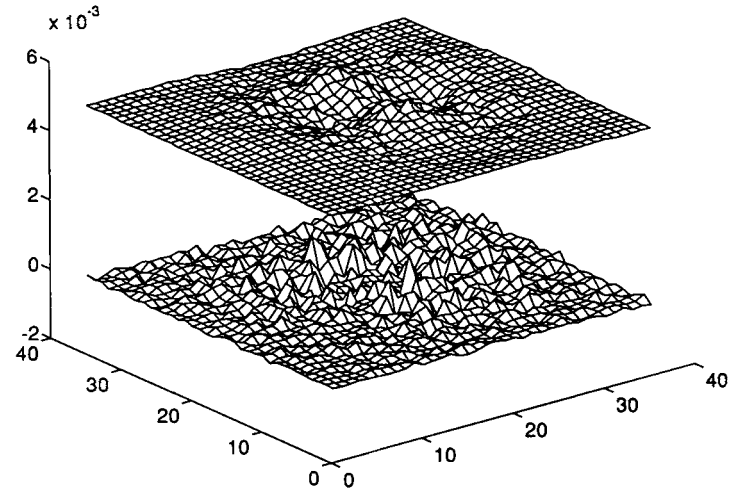

Fig. 15. The lower graph shows the unfiltered data, while the upper graph is the filtered equivalent; see the text. The noise reduction is apparent. (Again each example was smoothed to make reproduction possible. The actual noise signal is larger in both cases.)

difficult to show that, e.g., the filtered versions of the data 5 and 6 have the ridge differently oriented than the filtered versions of the other data sets. Each such filtered result is identical in appearance to Fig. 9. To show that the ridge *is* differently oriented, Fig. 14 shows the anisotropy in the *filtered* data sets. The top graph is for data set 5 while the lower is for data set 3. The difference in sign evident in this graph shows that the filter does return the smoothed signal of a ridge with the correct orientation.

The last illustration comprises constructing the filter with all the data sets *except* for data set 1. The filter W thus contains the prior experience of the other 8 similar data sets. It is used to filter the "new" data set 1. Since the noise is barely evident in the full signal, the effect on the anisotropic part of the result obtained by a subtraction after a rotation by 90 deg is again used for this illustration. The lower graph in Fig. 15 shows the unfiltered data, while the upper graph is the filtered equivalent. The noise reduction is apparent. (Again each example was smoothed to make reproduction possible. The actual noise signal is larger in both cases.)

5. DISCUSSION AND CONCLUSIONS

The present approach has an interesting relationship with the Karhunen–Loève transformation (see e.g., ref. 5). Given an ensemble of correlated data sets, this is the linear transformation which expresses these *correlated* sets in terms of a *noncorrelated* basis. The matrix which effects the transformation is precisely the autocorrelation matrix R_{gg} which appears in the present developments. If the aim is to encode, in an optimal fashion, these data, it is envisaged that the basis set given by the Karhunen–Loève transformation is to be truncated by some prescription. The truncation eliminates that part of the basis which has a small weight in each of the signals and thereby encodes the data with a minimum degradation of the information content. Such unimportant parts of the basis set correspond to eigenvectors of R_{gg} with a *small* eigenvalue.

The present approach might be thought of as a special extension of this transformation method, and represents a specific *optimal* resolution of the truncation problem. In the situation envisaged here there is a more or less large ensemble of data sets which contains one or a few significant experimental signals. Each member of the experimental ensemble comprises one or another of these possible signals degraded by the noise. It is assumed in the analysis that there is an *a priori* or *a posteriori* knowledge of the noise autocorrelation matrix R_{nn}. The presently derived filter W might be thought of as generalization of the Karhunen–Loève transformation which in a single step constructs the relevant noncorrelated basis set *and* realizes the optimal truncation for the present noise reduction exercise.

Since the object here is to reduce the noise, rather than encode the signal, the truncation is realized by including a suitable weight for each basis vector and, evidently, does not reduce the size of the vector space which is invoived.

In conclusion, a statistical theory for linear filters constructed from experimental data has been presented. The only *a priori* knowledge required to construct this filter is a model and/or a separate experiment for the noise power spectrum.

In the case when there are many but not an infinite number of experiments the prescription for W leads to a filter which when it acts on the results of any single experiment returns the signal with a strong reduction in noise. When there is only a single signal but many experiments, the associated Hamiltonian $\mathscr{H} = (1 - W^{\dagger})(1 - W)$ has only a single small eigenvalue and the corresponding eigenvector reflects this signal. The other eigenvalues are close to unity and correspond to typical examples of the noise. If instead two signals are present, $\mathscr{H} = (1 - W^{\dagger})(1 - W)$ has two small eigenvalues and the corresponding eigenvalues reflect the average and difference signals. In this case the same filter W projects out the correct signal from each data set suppressing the noise.

Using actual data from positron annihilation experiments, it has proved possible to illustrated the various possibilities. It is envisaged that in the future it will be useful to construct a filter which contains the prior knowledge drawn from all previous experiments on the same class of materials. For example, it might be that a filter contains all the available positron data for $YBa_2Cu_3O_{7-\delta}$ for a given orientation *but* with different values of the oxygen variable δ. When used to filter new data, the result will be the noise-reduced signal for the appropriate value of δ. The greater the prior knowledge contained in the earlier data sets, the greater the noise reduction. This has evident application for the quick analysis of new samples.

ACKNOWLEDGMENTS

The work was performed with the support of the Swiss National Fund for Scientific Research. S.E.B. wishes to thank the DPMC for their kind hospitality.

REFERENCES

1. A. Tikhonov and V. Arsenine, *Méthodes de résolution de problèmes mal posés* (Mir, Moscow, 1976).
2. W. K. Pratt, *IEEE Trans. Computers C* **21**:636 (1972).

3. A. Shukla, M. Peter, and L. Hoffmann, *Nucl. Instr. Meth. A* **335**:310 (1993); L. Hoffmann, A. Shukla, M. Peter, B. Barbiellini, and A. A. Manuel, *Nucl. Instr. Meth. A* **335**:276 (1993).
4. M. Peter, T. Jarlborg, A. A. Manuel, B. Barbiellini, and S. E. Barnes, *Z. Naturforsch. A* **48**:390 (1993); A. A. Manuel, B. Barbiellini, M. Gauthier, L. Hoffmann, T. Jarlborg, S. Massidda, M. Peter, W. Sadowski, A. Shukla, and E. Walker, *J. Phys. Chem. Solids* **54**:1223 (1993); B. Barbiellini, M. Gauthier, L. Hoffmann, T. Jarlborg, A. A. Manuel, S. Massidda, M. Peter, W. Sadowski, A. Shukla, and E. Walker, *Physica C* **209**:75 (1993).
5. M. Kunt, *Traité d'électricité*, Vol. XX (Presses Polytéchnique Romandes, 1984).