Iterative and Noniterative Recovery of Missing Samples for 1-D Band-Limited Signals

Paulo J. S. G. Ferreira

Dep. de Electrónica e Telecomunicações / IEETA
Universidade de Aveiro
3810-193 Aveiro Portugal

E-mail: pjf@det.ua.pt
URL: http://www.ieeta.pt/~pjf

To appear in *Sampling Theory and Practice*, F. Marvasti (editor), Plenum, 2000.

$Id: main.tex,v 1.17 2000/03/10 19:20:51 pjf Exp $

Contents

Chapter 5

Iterative and Noniterative Recovery of Missing Samples for 1-D Band-Limited Signals

5.1 Introduction

This chapter discusses a class of iterative and noniterative methods for the reconstruction of partially known band-limited signals. Band-limiting is equivalent to the vanishing of a known subset of the samples of the Discrete Fourier Transform (DFT) of the signals. The signals are partially known in the sense that only a subset of its samples is supposed to be available for measurement. The problem is to determine the signal from the available samples. Although this is in essence a nonuniform sampling problem, the methods employed differ considerably from the "infinite series" approaches. This is due to the nature of the finite-dimensional spaces in which we work, for which the matrix formulations are the most natural tool.

The chapter is written as a tutorial. Although it is not a comprehensive review of the many methods that have been proposed for the solution of the missing data and interpolation problems, it focuses on a variety of possible methods, their interrelations, possible implementations, stability, and relative performance. Understanding the connections between the methods is particularly useful since it helps the reader in navigating through the literature, and it contributes to a better grasp of the subject and of the main options and algorithms.

The methods described can be readily extended to signal and image interpolation and extrapolation, superresolution, and other multivariate problems, but for simplicity and brevity we refrain from discussing the theory behind the multivariate problems. The next chapter of this book will cover some of the multi-dimensional issues.

There are four possible models for the band-limited missing data problem, which correspond to all possible combinations of discrete and continuous time and frequency. We will not insist on the L_2 or continuous time, continuous frequency problem, nor on its infinite-dimensional sampled versions (continuous time and discrete frequency, or discrete time and continuous frequency). The focus will be on finite-dimensional signals, characterized by finite (or periodic) time domain and frequency domain vectors. In the terminology of [38], this corresponds to the discrete-discrete model. Mathematically, we deal with functions $f : G \mapsto \mathbb{C}$ or $f : G \mapsto \mathbb{R}$, where G is the group of integers with addition modulo N, equipped with the counting measure. The Fourier transform in this group is the Discrete Fourier Transform (DFT).

5.1.1 Outline of the chapter

The structure of the chapter is as follows. Band-limited signals are introduced in Section 5.2, together with the terminology and notation that will be used throughout the chapter.

The iterative methods discussed in Section 5.3 operate on the entire signal (known data and estimated data) on a step-by-step basis. The discussion centers around the finite-dimensional Papoulis-Gerchberg algorithm and its connections with other similar or equivalent methods (alternating projections and the frame algorithm, for example).

It is shown that the iterations converge if the density of the sampling set exceeds a certain minimum value which naturally increases with the bandwidth of the data. Naturally, the minimum density is related to the Nyquist-Landau density. Upper and lower error bounds are given as a function of the number of iterations. These bounds are best-possible, that is, there exist signals for which the bounds are exact. These signals are characterized, and turn out to be related to certain eigenvalue problems.

Several other issues are examined, relaxation being among them. Briefly, relaxation is a technique whereby one goes from the iteration

$$x^{(i+1)} = Tx^{(i)} + v, \tag{5.1}$$

to the iteration

$$x^{(i+1)} = \mu T x^{(i)} + (1 - \mu) x^{(i)},$$

where $\mu \in \mathbb{R}$ is the relaxation parameter. We study the effect of μ on the convergence rate of the Papoulis-Gerchberg method, and the value that leads to the best convergence rate (assuming a given, fixed, sampling set).

The number of distinct sampling sets with the same density is usually large. In fact, if N is the total number of samples, and if K denotes the number of known samples, then the number of sampling sets with density K/N is given by the binomial coefficient $\binom{N}{K}$. One of the questions that is discussed in the chapter concerns the sampling sets with a given density K/N that lead to maximum or minimum convergence rates (or equivalently, to maximum or minimum numerical stability).

We will see that for low-pass signals the best convergence rates result when the distances among the missing samples are a multiple of a certain integer. The worst convergence rates generally occur when the missing samples are contiguous, as in the case of extrapolation problems. Scattered missing data usually leads to well-posed and stable reconstruction problems, whereas extrapolation leads to ill-posed, numerically difficult problems. This will be confirmed, not just qualitatively, but also quantitatively, in this section and also in Section 5.5.

In contrast to the methods studied in Section 5.3, the formulation described in Section 5.4 leads to (iterative or noniterative) algorithms of "minimum dimension". The dimensions of the vectors and matrices in the algorithm is given by the number of unknown data, whereas the iterative methods studied in Section 5.3 involve matrices and vectors of dimension N (the total number of data).

The relation between the minimum dimension methods and the iterative methods from Section 5.3 is explained. It is shown how and why minimum dimension formulations lead to more efficient iterative and noniterative solutions to the missing data problem. Both time domain and frequency domain minimum dimension formulations are described.

Any given recovery algorithm, no matter how efficient, is of limited practical use if it is not numerically stable. Stability issues are the subject of Section 5.5. They allow a proper understanding of the limitations of both iterative and noniterative methods. On one hand, the convergence rate of the linear stationary iterative method of first order (5.1), with iteration matrix T and iteration vector v, depends on the magnitude and distribution of the eigenvalues of T. On the other hand, the noniterative reconstruction methods that we will study require solving linear equations of the form $Ax = b$, whose numerical stability depends on the condition number of the matrix A, and hence on the magnitude of its extreme eigenvalues. The estimation of the eigenvalues of the iteration matrices and its dependence upon the sampling set is the core of Section 5.5.

The chapter closes with a brief comparison of the methods, and some hints concerning the choice of the algorithm for a specific problem.

5.2 Band-limited signals

The recovery of missing samples from an arbitrary signal is of course impossible. This does not represent a serious difficulty, since the signals that occur in the real world, far from being "arbitrary", often satisfy known constraints. For example, the values attained by the signal, in practice, are always bounded above and below by constants. Many other constraints convey useful information that can be used by the recovery algorithms, such as constraints on the number and magnitude of the derivatives of the signal, convexity, analyticity, duration, and so on. The constraint extensively used in this chapter is called band-limiting. Before defining it, it is necessary to introduce some concepts and notation.

The complex N-dimensional space is denoted by $\mathbb{C}^N$, with the usual inner product

$$\langle a, b \rangle := \sum_{i=0}^{N-1} a_i \, b_i^*$$

and norm

$$\|x\|^2 := \langle x, x \rangle.$$

A signal with N samples can be described by a N-dimensional complex vector x. The elements of the vector are denoted by $x(0), x(1), \ldots, x(N-1)$, and correspond to the samples of the signal.

The Discrete Fourier Transform (DFT) of a signal $x \in \mathbb{C}^N$ is the vector $X \in \mathbb{C}^N$ defined by

$$X(k) := \sum_{i=0}^{N-1} x(i) \, e^{-\mathtt{j} \frac{2\pi}{N} ik}$$

($\mathtt{j}$ denotes the imaginary unit, not to be confused with i or j, which usually denote integer indexes). Because the DFT is a linear map in $\mathbb{C}^N$, it can be expressed succinctly in matrix form,

$$X = Fx,$$

where F is the $N \times N$ matrix with elements

$$F_{ab} = e^{-\mathrm{j}\frac{2\pi}{N}ab}.$$

Let F_i denote the ith column of F. Obviously, $\langle F_i, F_i \rangle = \|F_i\|^2 = N$, regardless of the value of i. On the other hand, any two distinct columns F_i and F_j with $i \neq j$ are orthogonal, that is, $\langle F_i, F_j \rangle = 0$. The Fourier matrix F has orthogonal columns and rows, and the inverse transformation (IDFT) is

$$x(k) = \frac{1}{N} \sum_{i=0}^{N-1} X(i)\, e^{\mathrm{j}\frac{2\pi}{N}ik}.$$

This can also be written in matrix form

$$x = F^{-1}X,$$

where

$$F_{ij}^{-1} = \frac{1}{N} e^{\mathrm{j}\frac{2\pi}{N}ij}.$$

Note that $x(k)$ and $X(k)$ should be interpreted as periodic functions of k, with period N. This fact is often convenient in computations, and we will use it tacitly from now on.

A signal x is band-limited if its discrete Fourier transform X vanishes on some fixed set, that is, if

$$X(i) = 0, \quad i \in S,$$

where S is a fixed nonempty proper subset of $\{0, 1, \ldots, N-1\}$. The set of signals band-limited to a specific set S is a linear subspace of $\mathbb{C}^N$, of dimension equal to the cardinal of the complement of S. This dimension is often called the bandwidth of the signals. When the complement of S is the set of $2M+1$ elements

$$\{0\} \cup \underbrace{\{1, 2, \ldots, M\}}_{M \text{ elements}} \cup \underbrace{\{N-M, \ldots, N-1\}}_{M \text{ elements}},$$

or equivalently (by periodicity)

$$\{0, \pm 1, \pm 2, \ldots, \pm M\},$$

the band-limited signal is called a low-pass signal, band-limited to $2M+1$ harmonics. Such signals can be synthesized using only the $2M+1$ harmonics of lowest (positive and negative) frequency,

$$\begin{aligned} x(j) &= \frac{1}{N}\sum_{i=-M}^{+M} X(i)\, e^{\mathrm{j}\frac{2\pi}{N}ij} \\ &= \frac{1}{N}\sum_{i=-M}^{+M}\left[\sum_{k=0}^{N-1} x(k)\, e^{-\mathrm{j}\frac{2\pi}{N}ki}\right] e^{\mathrm{j}\frac{2\pi}{N}ij} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} x(k) \sum_{i=-M}^{+M} e^{\mathrm{j}\frac{2\pi}{N}i(j-k)} \\ &= \sum_{k=0}^{N-1} x(k)\frac{\sin[\pi(2M+1)(j-k)/N]}{N\sin[\pi(j-k)/N]}. \end{aligned} \tag{5.2}$$

Band-limited signals are smooth, and the smoothness increases as the bandwidth decreases. Figure 5.1 shows the practical meaning of this concept, and the effect of varying M. Note how the redundancy of the signal increases as the bandwidth decreases.

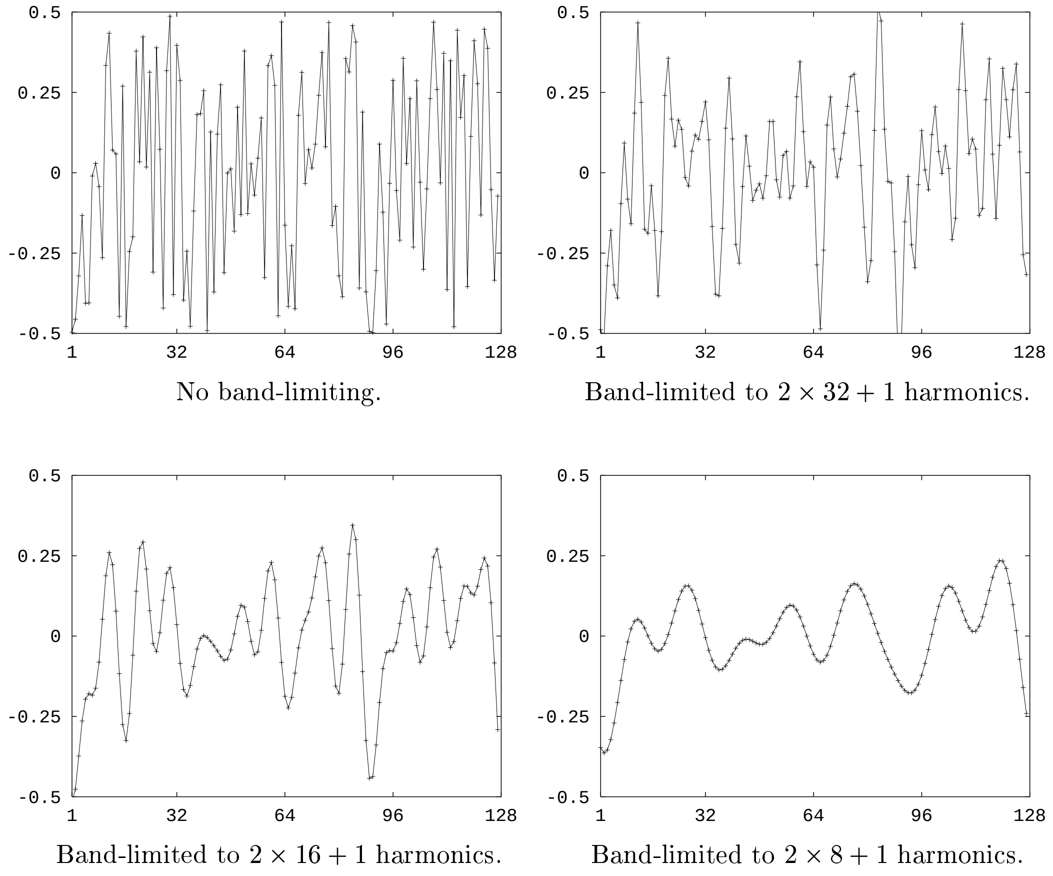
Equations (5.2) can also be written in matrix form. But before doing that, we need to introduce two linear operations defined on $\mathbb{C}^N$ which are of great importance in the context of missing data recovery.

The first operation is *sampling*. It maps a signal into another by setting to zero a certain subset of its samples. In matrix form, this corresponds to multiplication by a diagonal matrix D containing only zeros or ones. The diagonal of D will be called the sampling set associated with the sampling operation D, and D itself will be called a *sampling matrix*.

The second operation is *band-limiting*. Band-limiting is performed by a linear operator characterized by a matrix B of the form $B = F^{-1}\Gamma F$, where Γ is a sampling matrix other than the identity I, and F and F^{-1} denote the forward and inverse discrete Fourier transform matrices. Thus, according to this definition, the band-limiting operation in $\mathbb{C}^N$ is equivalent to sampling in the frequency domain.

We now observe that equation (5.2) can be expressed, in matrix form, as $x = Bx$, where $B = F^{-1}\Gamma F$, with the main diagonal Γ being

$$[1, \underbrace{1, \ldots, 1}_{M\text{ times}}, 0, \ldots, 0, \underbrace{1, \ldots, 1}_{M\text{ times}}].$$



No band-limiting. Band-limited to $2 \times 32 + 1$ harmonics.

Band-limited to $2 \times 16 + 1$ harmonics. Band-limited to $2 \times 8 + 1$ harmonics.

Figure 5.1: Band-limiting a signal to successively smaller bandwidths. The number of data samples (hence the maximum number of nonzero harmonics) is $N = 128$. The dimension of the subspace of signals band-limited to $2M+1$ harmonics is $2M+1$, and hence $N-(2M+1)$ data samples are redundant.

The relation $B = F^{-1}\Gamma F$ translates into

$$\begin{aligned} B_{ij} &= \frac{1}{N}\sum_{k=-M}^{M} e^{-\mathrm{j}\frac{2\pi}{N}(i-j)k} \\ &= \frac{\sin[\pi(2M+1)(i-j)/N]}{N\,\sin[\pi(i-j)/N]}, \end{aligned} \tag{5.3}$$

which shows that $x = Bx$ is indeed the matrix form of (5.2). The signals that satisfy $x = Bx$ are low-pass signals and B is a low-pass filter matrix. Note the similarity between the previous equation and

$$f(t) = \int_{-\infty}^{+\infty} f(x)\frac{\sin w(t-x)}{\pi(t-x)}\,dx$$

which is satisfied by any $f \in L_2(\mathbb{R})$ if and only if its Fourier transform vanishes almost everywhere outside $[-w, w]$. A signal belonging to $L_2(\mathbb{R})$ is band-limited if its Fourier transform vanishes outside a compact (closed and bounded) set. When this set is an interval of the form $[-w, w]$ the signal is called a low-pass signal. Bandpass signals are characterized by a Fourier transform that vanishes outside $[-b, -a] \cup [a, b]$.

We are now ready to formulate the missing data problem, around which this chapter turns.

Problem 1 *Let $x \in \mathbb{C}^N$ be a low-pass signal with $2M+1$ nonzero harmonics. Given the observation $y = Dx$, where D is a sampling matrix, how can one recover x from y? Is this problem well-posed?*

There is a trivial necessary condition that D has to satisfy: the number of samples of x that D preserves should be at least equal to $2M + 1$, the dimension of the subspace of low-pass, band-limited signals with $2M + 1$ nonzero harmonics. The problem becomes under-determined if this condition does not hold, and exact recovery ceases to be possible. Although this is also an interesting situation and there are several ways of dealing with it, we will restrict ourselves to the exactly or over-determined case.

5.3 Iterative methods

5.3.1 The Papoulis-Gerchberg iteration

One of the methods for the solution of the missing data problem is the Papoulis-Gerchberg algorithm. This is an iterative method, illustrated in Fig. 5.2. Each iteration of the algorithm consists of two steps: filtering, and resampling.

The filtering step imposes the frequency domain or DFT-domain constraints about the data, that is, the fact that they are band-limited to $2M+1$ nonzero harmonics. This low-pass filtering produces an improved approximation $b^{(i)}$ from the previous approximation $x^{(i)}$, according to

$$b^{(i)} = Bx^{(i)}. \tag{5.4}$$

The next step restores the time domain knowledge about x, that is, the known samples, which were preserved by D but which were changed by the preceding filtering step. Resampling, or reinserting, the known data maps the output of the previous step, $b^{(i)}$, into $x^{(i+1)}$, which will be the input to the next iteration, that is

$$x^{(i+1)} = Rb^{(i)}. \tag{5.5}$$

We have denoted by R the resampling operator, defined by

$$Ra := (I - D)a + y,$$

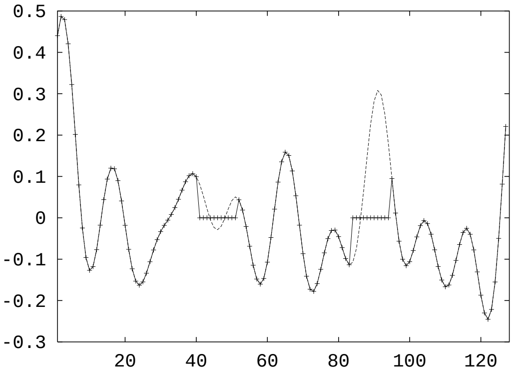
for any $a \in \mathbb{C}^N$. Recall that $y = Dx$, and that y is known. We can now write down the rule through which one proceeds from iteration i to iteration $i+1$, by combining (5.4) and (5.5). This leads to

$$x^{(i+1)} = (I - D)Bx^{(i)} + y, \tag{5.6}$$

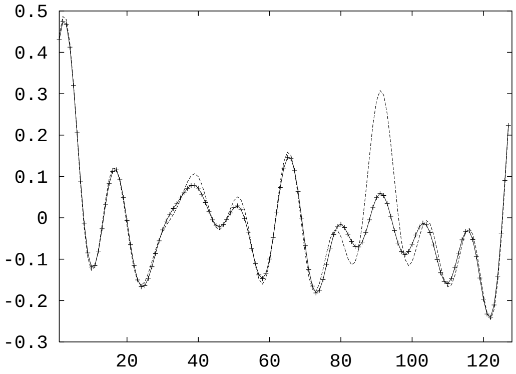
This algorithm is the finite-dimensional, or discrete-discrete version of the Papoulis-Gerchberg extrapolation algorithm, adapted to the missing data problem, originally studied by Papoulis [34] and Gerchberg [19] in a different context, for the continuous-continuous $L_2(\mathbb{R})$ extrapolation problem. See also [4].

5.3.2 Alternating projections, POCS, and the frame algorithm

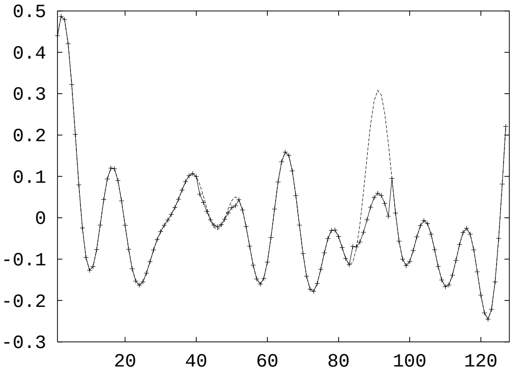
The Papoulis-Gerchberg iteration may also be described as an alternating projection algorithm [52], a class of algorithms based on successive projections onto subspaces.



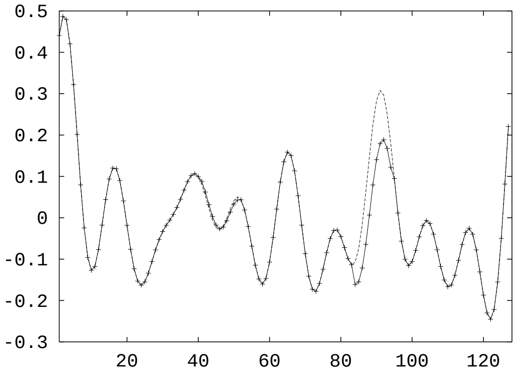
Replacing the unknown samples with zeros.



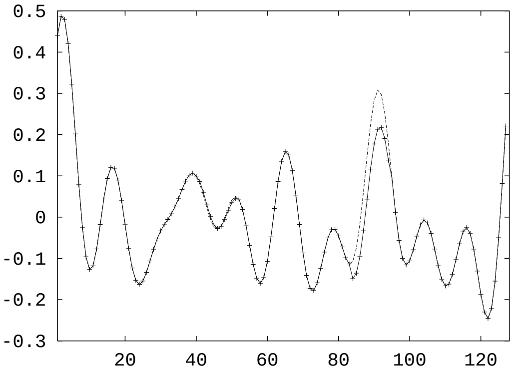
The signal obtained after filtering.



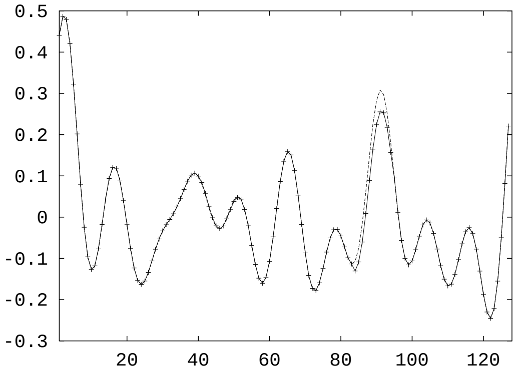
Resampled signal, after 10 iterations.



Filtered signal, after 10 iterations.



Resampled signal, after 30 iterations.



Filtered signal, after 50 iterations.

Figure 5.2: The Papoulis-Gerchberg method: at each iteration the signal is resampled (that is, the known samples are reinserted) and then filtered. For better comparison, the original signal has been superimposed on the reconstructed signal.

The alternating projection algorithms, in turn, are special cases of the method known as Projections Onto Convex Sets (POCS). They can be obtained from POCS by taking all the convex sets to be subspaces. For more details on POCS see [42], [53].

A countable set of elements f_n ($n \in J$) belonging to a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ is a *frame* if there exist two constants α and β such that

$$\alpha \|g\|^2 \leq \sum_{n \in J} |\langle g, f_n \rangle|^2 \leq \beta \|g\|^2.$$

for every element g of that Hilbert space. The Papoulis-Gerchberg algorithm is also closely related to the frame algorithm, a procedure through which the signal g can be recovered from the data $\langle g, f_n \rangle$. A detailed tutorial exposition of the connections between iterative signal reconstruction and frames, eigenvalue problems, and energy concentration, can be found in [15].

5.3.3 Constrained iterative restoration, relaxation

The iteration (5.6) can also be derived using the general framework for constrained iterative restoration proposed in [41]. In this particular case, we take the sampling and band-limiting operations to be, respectively, the distortion and constraint mentioned in [41]. The original signal satisfies $x = Bx$ and the observed signal is $y = Dx$, for appropriate matrices B and D. This suggests the equation

$$\begin{aligned} x &= Bx + \mu(y - Dx) \\ &= \mu y + (I - \mu D)Bx, \end{aligned}$$

where μ is any fixed constant, and I denotes the identity matrix. To solve this equation we consider the sequence of successive approximations

$$x^{(i+1)} = T_1 \, x^{(i)}, \tag{5.7}$$

where the operator T_1 is defined by

$$T_1(\cdot) := \mu y + (I - \mu D)B(\cdot). \tag{5.8}$$

Note that when $\mu = 1$ the iteration (5.7) reduces to (5.6). Conversely, it is easy to introduce relaxation on the basic iteration (5.6), leading to (5.7).

It is possible to obtain a sequence of band-limited successive approximations by taking the distortion to be sampling followed by a band-limiting operation. This gives

$$x^{(i+1)} = \mu By + (I - \mu BD)x^{(i)} =: T_2\, x^{(i)} \tag{5.9}$$

Yet another possibility is to apply the band-limiting operator to (5.7). This leads to

$$x^{(i+1)} = \mu By + B(I - \mu D)x^{(i)} =: T_3\, x^{(i)}. \tag{5.10}$$

It is well-known that such iterative methods converge if and only if the spectral radius (the largest eigenvalue in absolute value) of the respective iteration matrices[1] is strictly less than unity. See, for example, [49], [54].

Since the iteration matrices of the methods (5.7) and (5.10) are the transpose of each other, and (5.9) is equivalent to (5.10) if the first approximation is band-limited ($x^{(0)} = Bx^{(0)}$), it is sufficient to consider (5.7).

It will be shown that this method leads to a sequence of approximations that converges to the required solution, provided that suitable conditions are imposed upon D, B and μ. Under the same conditions the other two iterations will also converge to the same solution, provided that $x^{(0)} = Bx^{(0)}$ in the case of (5.9).

The major drawback of the Papoulis-Gerchberg iteration and similar methods is related to its convergence rate. The precise rate critically depends on the distribution of the missing data, and can be very slow. We will see that the minimum dimension formulations, discussed in Section 5.4, allow much more efficient algorithms.

5.3.4 Convergence analysis

In this section we will examine the convergence of (5.7). It turns out that if the density of the known samples is above the minimum density the iteration converges regardless of the distribution of the missing samples. The convergence was established in [25] for extrapolation problems, in which the known samples are contiguous, using a technique based on matrix partitioning. The general case requires different tools, and was treated in [7].

There are several possibilities for the analysis of (5.7). As in [7], we use a few well-known results in fixed-point theory. This technique is often used in

[1] By definition, the iteration matrix of the iteration $x^{(i+1)} = Ax^{(i)} + b$ is the matrix A.

signal restoration [41],[48] and does not require the linearity of the operators involved. In fact, it is easy to see that other possibly non-linear nonexpansive constraints can be incorporated in the algorithm without disturbing the convergence.

Denote the set of subscripts of the known samples of x by

$$S_t = \{i_1, i_2, \ldots, i_s\},$$

and its complement by $\bar{S}_t$. The observed signal y satisfies $y = Dx$, where D is the sampling matrix that corresponds to S_t, that is,

$$D_{ii} = \begin{cases} 1, & i \in S_t \\ 0, & i \in \bar{S}_t. \end{cases}$$

By definition, the *density* of the sampling matrix D is s/N, s being the cardinal of S_t.

Let the set of subscripts of the known vanishing DFT coefficients of x be denoted by S_f, and its complement in $\{0, 1, \ldots, N-1\}$ by $\bar{S}_f = \{k_1, k_2, \ldots, k_q\}$.

The process of band-limiting a signal x with the band-limiting matrix B is equivalent to setting to zero a subset of its DFT coefficients X. More precisely,

$$X(i) = 0, \quad i \in S_f.$$

By definition, the *bandwidth* of B is q/N, q being the cardinal of $\bar{S}_f$. For low-pass signals, both S_f and $\bar{S}_f$ are contiguous sets (modulo N).

Lemma 1 *Let D be a sampling operator with density d, and let B be a band-limiting operator with bandwidth w. If $d \geq w$, $x = (I - D)x$ and $x = Bx$, then x is the zero vector.*

Proof:

The hypothesis $x = (I - D)x$ means that $x(i) = 0$ for all $i \in S_t$, that is, s samples of x vanish. The N equations $X = Fx$ reduce to

$$X = \sum_{i \notin S_t} F_i \, x(i),$$

where F_i denotes the ith column of F.

On the other hand, $x = Bx$ means that $X(i) = 0$ for all $i \in S_f$, and so $N - q$ samples of X vanish. Thus,

$$\sum_{i \notin S_t} F_{ki}\, x(i) = 0, \quad k \in S_f. \tag{5.11}$$

The submatrix of F that appears in (5.11) is $(N-q) \times (N-s)$ with $(N-q) \geq (N-s)$ since $d \geq w$ (and so $s \geq q$).

Since $\bar{S}_f$ is contiguous, there is an integer r such that $k_i = (i+r) \bmod N$, and the submatrix of F that appears in (5.11) can be reduced to a Vandermonde matrix with linear independent columns. Thus, the unique solution to (5.11) is given by $x(i) = 0$, and x must be the zero vector. $\square$

Let X be a metric space, with metric $d(\cdot, \cdot)$. Recall that a possibly nonlinear operator $A : X \mapsto X$ is *nonexpansive* when

$$d(Ax, Ay) \leq d(x, y),$$

for all $x, y \in X$. In a normed space with norm $\|\cdot\|$, this can be converted to

$$\|Ax - Ay\| \leq \|x - y\|.$$

An operator $A : X \mapsto X$ is *strictly nonexpansive* if it is nonexpansive and

$$d(Ax, Ay) = d(x, y) \quad \Rightarrow \quad x = y.$$

If X is a normed space and A is strictly nonexpansive then

$$\|Ax - Ay\| < \|x - y\|$$

for all distinct pairs x, y of elements of X.

If the operator A is linear and nonexpansive then

$$\|Ax\| \leq \|x\|$$

for all $x \in X$. Therefore, the norm of a linear nonexpansive operator cannot exceed unity. In $\mathbb{C}^N$, a linear nonexpansive operator is a matrix with spectral norm bounded by one. Therefore, its spectral radius is also bounded by one, and consequently none of the eigenvalues can exceed unity in absolute value.

The point $x \in X$ is a *fixed point* of the operator $A : X \mapsto X$ if $Ax = x$. For example, the fixed points of ideal low-pass filters are low-pass signals. A nonexpansive operator may have any number of fixed points, but a strictly nonexpansive operator clearly has at most one.

Lemma 2 *Let $0 \leq \mu \leq 2$. Then, the operator T_1 defined by (5.8) is nonexpansive.*

Proof: Consider

$$A_\mu := (I - \mu D)B. \tag{5.12}$$

Checking (5.8), we see that T_1 is nonexpansive if and only if A_μ is. The nonexpansiveness of the projection operator B in the Euclidean norm is obvious. Thus, T_1 is nonexpansive if $I - \mu D$ is, which happens if and only if $0 \leq \mu \leq 2$. Otherwise there would be an eigenvalue of the diagonal matrix $I - \mu D$ greater than one in absolute value. □

Lemma 3 *If $0 \leq \mu \leq 2$, the equation $x = T_1 x$, with T_1 defined by (5.8), has at least one solution.*

Proof: T_1 is nonexpansive, hence continuous. It maps a non-empty, compact and convex subset of $\mathbb{C}^N$ into itself (consider the set of all signals with norm less than or equal to a given bound). Any continuous operator that maps a non-empty compact and convex set into itself has the fixed-point property [41], [43], and the result follows. □

Theorem 1 *Let D be a sampling operator with density d, and let B be a band-limiting operator with bandwidth w. If*

$$0 < \mu < 2, \tag{5.13}$$

$$d \geq w, \tag{5.14}$$

the operator T_1, defined by (5.8), is strictly nonexpansive, and the equation $x = T_1 x$ has one and only one solution. This solution is given by

$$x = \lim_{i \to \infty} T_1^i x^{(0)},$$

where the initial approximation $x^{(0)} \in \mathbb{C}^N$ is arbitrary.

Proof: Assume, to the contrary, that (5.13) and (5.14) hold but T_1 is *not* strictly nonexpansive. This means that

$$\|A_\mu v\| := \|(I - \mu D)Bv\| = \|Bv\| = \|v\|$$

for some $v \neq 0$. We will show that this assumption leads to a contradiction. The last equality implies the vanishing of $V(i)$ for $i \in S_f$, and thus $Bv = v$. This reduces the second equality to

$$\|(I - \mu D)v\| = \|v\|.$$

But

$$\begin{aligned}
\|(I - \mu D)v\|^2 &= (1-\mu)^2 \sum_{i \in S_t} v(i)^2 + \sum_{i \in \bar{S}_t} v(i)^2 \\
&= \|v\|^2 - \mu(2-\mu) \sum_{i \in S_t} v(i)^2
\end{aligned}$$

which implies the vanishing of the $v(i)$ for $i \in S_t$, since, by hypothesis, $0 < \mu < 2$ and so $\mu(2-\mu) > 0$. By Lemma 1, v must be the zero vector, a contradiction. Thus, T_1 must be strictly nonexpansive, and a strictly nonexpansive mapping cannot have more than one fixed point. □

5.3.5 Upper and lower error bounds

The iterations (5.7) or (5.6) are examples of iterative methods in $\mathbb{C}^N$ of the general form

$$x^{(i+1)} = Tx^{(i)} + v,$$

where T and v are the iteration matrix and vector. Assume that this iteration converges to a vector x such that $x = Tx + v$. The error at the ith iteration is

$$e^{(i)} = x^{(i)} - x,$$

and the errors at iterations i and $i+1$ are related by

$$e^{(i+1)} = Te^{(i)}.$$

Clearly, the behavior of $e^{(i)}$ is crucial when the iterative method is used to approximate x.

We will now consider this issue for the iteration (5.7). Upper and lower bounds for the reconstruction error at any given iteration will be given, assuming that the relaxation parameter is $\mu = 1$. The bounds are best-possible, in the sense that there are initial error vectors $e^{(0)}$ for which they are exact. The initial error vectors for which the error bounds are attained will be identified.

Recall that, in agreement with (5.12), the matrix A_1 is given by

$$A_1 := (I - D)B.$$

Let v denote an eigenvector of $A_1^H A_1 = B(I-D)B$, pertaining to the eigenvalue λ, that is,

$$A_1^H A_1 v = B(I-D)Bv = \lambda v. \tag{5.15}$$

The subscript ^{H} denotes conjugate transpose. We claim that $A_1 v$ is an eigenvector of A_1 pertaining to the same eigenvalue, and thus

$$A_1^2 v = \lambda A_1 v.$$

To check that this is true, left-multiply (5.15) by $(I - D)B$ and use the idempotency of B.

It is now easy to derive bounds for the error $e^{(k)}$ at iteration k of the algorithm. It follows from $e^{(k)} = A_1 e^{(k-1)}$ that

$$\max_{\|e^{(0)}\|=1} \|e^{(1)}\|^2 = \max_{\|e^{(0)}\|=1} \|A_1 e^{(0)}\|^2 = \lambda_{\mathsf{max}}$$

and

$$\min_{\|e^{(0)}\|=1} \|e^{(1)}\|^2 = \min_{\|e^{(0)}\|=1} \|A_1 e^{(0)}\|^2 = \lambda_{\mathsf{min}},$$

where λ_{min} and λ_{max} denote the smallest and largest of the eigenvalues of $A_1^H A_1$. The maximum and the minimum are attained when $e^{(0)}$ equals v_{max} or v_{min}, the eigenvectors of $A_1^H A_1$ that correspond to the eigenvalues λ_{max} and λ_{min}, respectively.

Now, we have already seen that $A_1 v_{\mathsf{max}}$ and $A_1 v_{\mathsf{min}}$ will also be eigenvectors of A_1, and therefore

$$e^{(2)} = A_1^2 v_{\mathsf{max}} = \lambda_{\mathsf{max}} A_1 v_{\mathsf{max}}$$

if $e^{(0)} = v_{\mathsf{max}}$, and

$$e^{(2)} = A_1^2 v_{\mathsf{min}} = \lambda_{\mathsf{min}} A_1 v_{\mathsf{min}}$$

if $e^{(0)} = v_{\mathsf{min}}$. By induction it is now clear that

$$\lambda_{\mathsf{min}}^k \|e^{(1)}\| \le \|e^{(k+1)}\| \le \lambda_{\mathsf{max}}^k \|e^{(1)}\|. \tag{5.16}$$

The upper bound is attained if $e^{(1)} = v_{\mathsf{max}}$, the lower bound if $e^{(1)} = v_{\mathsf{min}}$.

In many cases one takes $x^{(0)}$, the initial approximation, to be the zero vector. This leads to $x^{(1)} = y$, and $e^{(k)} = A_1^k x$. Taking $x = v_{\mathsf{max}}$, we get the lowest possible asymptotic convergence rate. In this sense, the v_{max} are the signals which result in the worst possible algorithm performance. A similar argument, regarding best possible performance, holds for v_{min}. Also note that v_{max} and v_{min}, being eigenvectors of $A_1^H A_1$, depend on B and D, and hence on the bandwidth and sampling set. In particular, changing the sampling set may lead to dramatic changes in the convergence rate.

For examples of v_{max} and v_{min} for particular problems see Fig. 5.3 and Fig. 5.4. It is instructive to compare the effect of the sampling set and the signal on the performance of the algorithm. Roughly speaking, well distributed sampling sets lead to values of λ_{max} reasonably below unity, and thus to favorable convergence rates. In those cases λ_{min} is also not much smaller than λ_{max}, and hence the worst and best convergence rates do not differ appreciably.

The opposite situation occurs when the sampling set has large gaps. In that case, there exist band-limited signals whose energy is well concentrated in the sampling set, but also other band-limited signals well concentrated on its complement. The signal v_{min} is an (extreme) example of the former class, and v_{max} of the latter. In practice, as the gaps increase λ_{max} tends to unity and the convergence rate becomes increasingly slower (the iterations do not reduce the magnitude of the components of the error signal proportional to v_{max}).

When B is a low-pass matrix and the sampling set corresponds to an extrapolation problem (that is, the known data are contiguous), the analysis that we have made reduces to the singular value analysis of the operator DB and to the periodic discrete prolate spheroidal sequences (P-DPSS) discussed in [24]. In fact, setting $M = (I - D)B$ in $M^H Mx = \lambda x$ leads to

$$B(I - D)Bv = \lambda v, \tag{5.17}$$

which is equivalent to

$$BDBv = (1 - \lambda)v \tag{5.18}$$

since any solution of (5.17) must be band-limited (left multiply (5.17) by B). Equation (5.18) is equivalent to the equation used to define the P-DPSS in [24]. The eigenvectors of (5.17) v_{min} and v_{max}, which correspond to the smallest and largest eigenvalues λ_{min} and λ_{max} in (5.17), reduce to P-DPSS in the case of extrapolation problems.

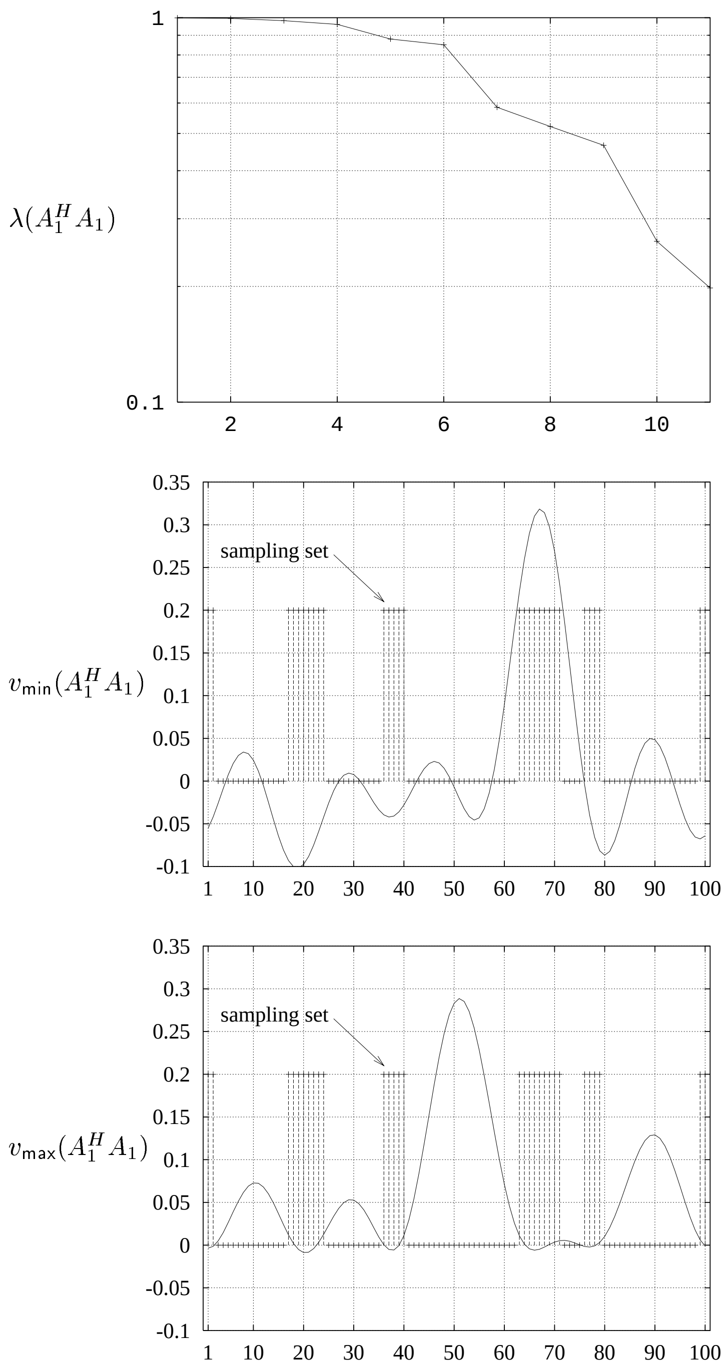


Figure 5.3: The eigenvalues and the eigenvectors v_{min} and v_{max} of the matrix $A_1^H A_1 = B(I-D)B$. The extreme eigenvalues dictate the best and worst (geometric) convergence rates in (5.16). The easiest signal to handle is v_{min}, the most difficult is v_{max} (for this bandwidth and sampling set).

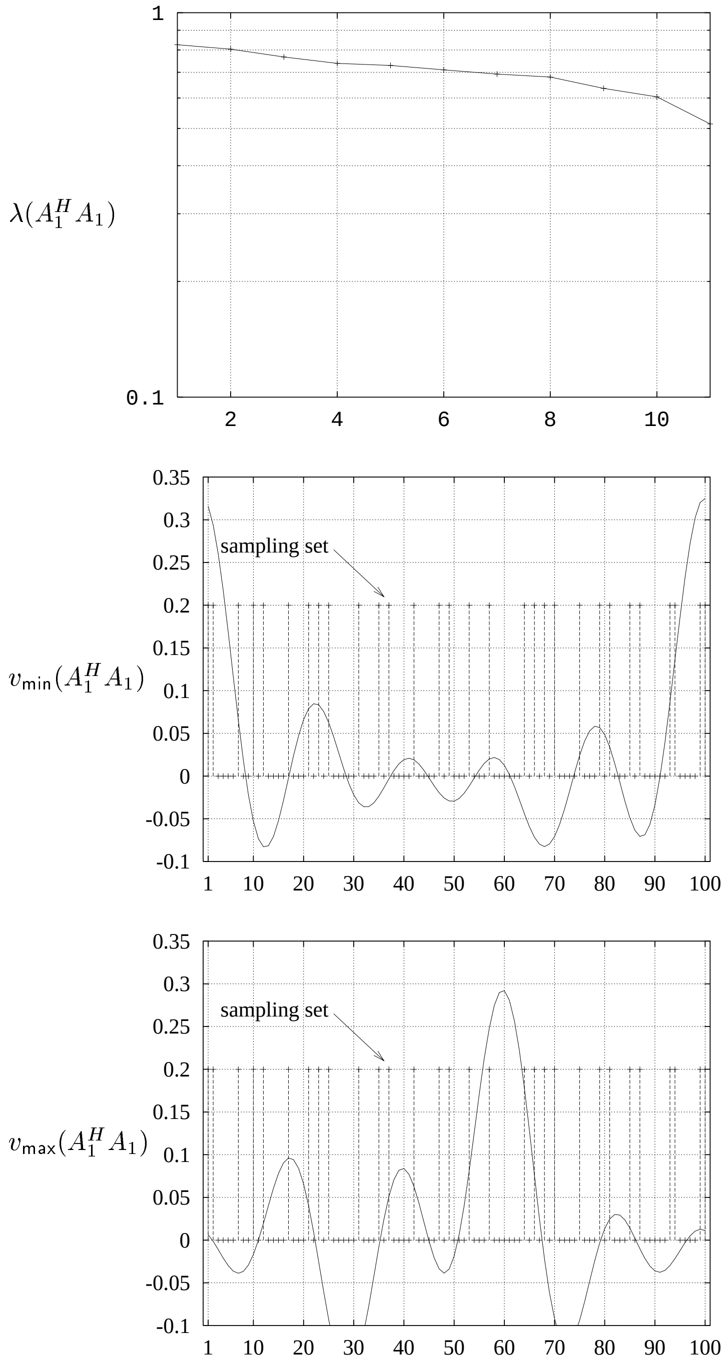


Figure 5.4: The eigenvalues and the eigenvectors v_{min} and v_{max} of the matrix $A_1^H A_1 = B(I-D)B$. The extreme eigenvalues dictate the best and worst (geometric) convergence rates in (5.16). The easiest signal to handle is v_{min}, the most difficult is v_{max} (for this bandwidth and sampling set). Compare with Fig. 5.3.

5.3.6 Other results

There are several other questions that can be asked concerning the iterative method (5.7). For example, one may be interested in the norm and spectral radius of the matrix $A_1 = (I - D)B$, or in the effect of the relaxation parameter μ upon the convergence rate of the algorithm. It is not difficult to see that the optimal value of μ is

$$\mu_{\mathsf{opt}} = \frac{2}{2 - \lambda_{\mathsf{max}}},$$

where λ_{max} is the largest eigenvalue of the matrix $A_1^H A_1 = B(I - D)B$. For details, see [7].

We have mentioned that the sampling sets that minimize or maximize the asymptotic convergence rate of the algorithm, for a given sampling density, depend on the B matrix. It is shown in [7] that for low-pass data, for example, the missing sample positions that give best-possible sampling sets define a grid with a certain spacing. Thus, the optimum sampling strategy is, in a sense, uniform sampling. On the other hand, the worst possible sampling sets are contiguous. This may not be the case for data that are neither low-pass nor high-pass.

Let us consider a number of missing samples greater than two, and let s/N be the density of the sampling matrix, and $N - s$ the number of unknown samples. What can we say about the best and worst possible sampling sets?

It can be shown [7] that the squared norm of $B(I - D)B$ is given by

$$\sup_{\|x\|=1} \left(\sum_{i \notin S_t} \langle b_i, x \rangle^2 + (1 - \mu)^2 \sum_{i \in S_t} \langle b_i, x \rangle^2 \right), \tag{5.19}$$

where b_i denotes the ith column of B.

There are two extreme situations. The best situation arises when all the vectors b_i with $i \notin S_t$ are orthogonal, or very nearly so. The worst possible situations occur when the b_i with $i \notin S_t$ are close to linearly dependent.

Assume for a moment that the sampling set is such that the vectors b_i, with $i \notin S_t$, are orthogonal. The squared norm of the b_i equals q/N, where q is the number of nonzero harmonics, or the cardinal of the set $\bar{S}_f$, in the notation introduced in Section 5.3.4. Therefore, one has

$$x = \frac{N}{q} \sum_{i \notin S_t} \langle b_i, x \rangle b_i$$

for any x belonging to the subspace spanned by the b_i, $i \notin S_t$. Thus

$$\|x\|^2 = \frac{N}{q} \sum_{i \notin S_t} \langle b_i, x \rangle^2,$$

or, equivalently,

$$\sum_{i \notin S_t} \langle b_i, x \rangle^2 = \|x\|^2 \frac{q}{N}.$$

It follows from (5.19) with $\mu = 1$ that

$$\|(I - D)B\|^2 = \sup_{\|x\|=1} \frac{q}{N} \|x\|^2 = \frac{q}{N},$$

that is, $(I-D)B$ will be a convergent matrix with squared norm q/N. In this case, the norm is independent of the number of missing samples, and equals the value that would be obtained for a single missing sample. For these "ideal" sampling sets, the recovery of one, two, or more missing samples can be achieved with a matrix whose norm is determined only by the bandwidth of the data. Also, the asymptotic convergence rate of the reconstruction algorithm would be independent of the number of missing samples. We are assuming, of course, that the conditions of Theorem 1 are satisfied, that is, that the sampling set has sufficiently high density.

To find out more about the "ideal" sampling sets, we need to see under what conditions, if any, the columns b_i of the matrix B can be orthogonal to each other. For signals with a contiguous set of nonzero harmonics, and in particular for low-pass signals, the elements of B are given by (5.3). Since $B^2 = B$, it follows that $\langle b_i, b_j \rangle = B_{ij}$, and so

$$\langle b_i, b_j \rangle = \frac{\sin[\pi(2M+1)(i-j)/N]}{N \sin[\pi(i-j)/N]}.$$

One may assume, without loss of generality, that one of the missing samples is sample number zero. But because

$$\langle b_i, b_0 \rangle = \frac{\sin[\pi(2M+1)i/N]}{N \sin[\pi i/N]},$$

orthogonality is only possible if the remaining missing samples belong to the set

$$\left\{ \frac{N}{2M+1}, \frac{2N}{2M+1}, \frac{3N}{2M+1}, \ldots \right\}$$

assuming of course that $2M + 1$ divides N. These quantities determine the possible positions of the missing samples which correspond to optimum sampling sets. If $2M+1$ does not divide N, they still originate sampling sets with near-optimum convergence properties.

5.4 Minimum dimension formulations

The iterative reconstruction algorithms mentioned in Section 5.3 are useful in connection with many reconstruction problems. They have been used, for example, in the estimation of sinusoids from incomplete time series [33], or in the removal of lipid artifacts in ^{1}H spectroscopic imaging [22].

There are, however, classes of problems for which these algorithms do not seem well suited. The extrapolation of low-pass signals is an example. In that case, and more generally whenever there are long contiguous gaps of missing samples, the convergence rate falls down to very low values.

The minimum dimension formulations presented in this section lend themselves to better and faster algorithms. The ideas behind these methods were put forward in several works, including [8], [12], which address the finite-dimensional case, and [11], [18], which consider finite-dimensional images. The study of the continuous-time case was initiated with [6].

The computational requirements necessary to complete one iteration of the Papoulis-Gerchberg algorithm do not decrease as the number of unknown samples decreases. The algorithm operates on N-dimensional vectors, and the computational burden per iteration is dominated by the filtering operation, which requires $O(N \log_2 N)$ arithmetic operations using fast Fourier transform (FFT) algorithms. It seems redundant to operate on sets of N samples when only a smaller number is unknown.

The minimum dimension methods lead to sets of linear equations for the unknown quantities, with as many equations as unknown samples (for the time domain formulation), or as many equations as unknown DFT coefficients (for the frequency domain formulation). Hence, the system matrix, in both cases, has minimum dimension. In the time domain case the equations yield the unknown samples directly, whereas in the frequency domain one obtains coefficients of the DFT of the signal.

The basic idea behind the minimum dimension method can also be applied to image interpolation problems, and to continuous-time missing data problems. A signal $f : \mathbb{R} \to \mathbb{R}$ (for example of finite-energy) can be recovered from its samples $f(k)$, $k \in \mathbb{Z}$ (for simplicity we are normalizing the sampling period to unity) if its Fourier transform is zero outside the interval $[-r/2, r/2]$, with $0 \le r \le 1$. The case $r = 1$ corresponds to sampling at the critical Nyquist-Landau density. When $r < 1$ there is oversampling, and the relevant reconstruction formula is

$$f(t) = r \sum_{k=-\infty}^{+\infty} f(k) \operatorname{sinc} r(t-k).$$

Theoretically (barring stability issues), it is possible to reconstruct f even if any finite number n of samples $f(i)$ are unknown [3], [6], [26], [27]. Let the unknown samples be

$$f(j_1), f(j_2), \ldots, f(j_n).$$

Setting $t = j_i$ for $i = 1, 2, \ldots, n$, we get n equations

$$f(j_i) = r \sum_{k=-\infty}^{+\infty} f(k) \operatorname{sinc} r(j_i - k),$$

and by splitting the summation in two, one over the unknown samples, another over the known samples, we get

$$f(j_i) = r \sum_{k=1}^{n} f(j_k) \operatorname{sinc} r(j_i - j_k) + h_i.$$

In matrix form, this can be written as

$$u = Su + h,$$

with the obvious interpretations for u, S, and h. The elements h_i of the vector h can be computed because they depend only on known data. By solving these n equations we get the n unknown samples (but see [3], [9] for some crucial stability and implementation issues). Note that the essential steps of the method are:

- Obtaining an equation of the form $x = Tx$, where x is the vector of all samples and T is a specific linear or affine operator.
- Obtaining n linear equations for the n unknown samples, by splitting x into known and unknown samples.

We are already aware of an equation of the required form, namely, equation (5.2). This solves the first part of the problem for the time domain case. For the frequency domain formulation, it will be necessary to determine the operator T such that $X = TX$, X being the DFT of x.

We start with the time domain formulation, expressing the vector x of the time domain samples of the signal as $x = u + v$, where

$$v_i = \begin{cases} x_i, & i \in S_t, \\ 0, & i \notin S_t. \end{cases}$$

Recall the notation introduced in Section 5.3.4: S_t is the set of known samples, and therefore the vector v is known, whereas u is not. Defining the diagonal matrix

$$D^u_{ii} = \begin{cases} 0, & i \in S_t, \\ 1, & i \notin S_t, \end{cases}$$

it is seen that $u = D^u u$ and $D^u v = 0$.

We now deal with the frequency domain case. The vector X of the samples of the DFT of the signal x is written as $X = G + H$, where

$$H_i = \begin{cases} X_i, & i \in S_f, \\ 0, & i \notin S_f. \end{cases}$$

In accordance with the notation of Section 5.3.4, S_f denotes the set of known DFT coefficients. The vector H is known, but G is not. Defining the diagonal matrix

$$D^g_{ii} = \begin{cases} 0, & i \in S_f, \\ 1, & i \notin S_f, \end{cases}$$

it is seen that $G = D^g G$, and $D^g H = 0$.

Theorem 2 *The signal $x \in \mathbb{C}^N$ that solves problem 1 of Section 5.2 satisfies $x = Bx + h$, and its DFT satisfies $X = TX + V$. The matrices B and T are circulant projections, with $B = F^{-1} D^g F$ and $T = F D^u F^{-1}$. Explicitly,*

$$\begin{aligned} B_{ab} &= \frac{1}{N} \sum_{p \notin S_f} e^{\mathrm{j} \frac{2\pi}{N} (a-b)p}, \\ T_{ab} &= \frac{1}{N} \sum_{p \notin S_t} e^{-\mathrm{j} \frac{2\pi}{N} (a-b)p}. \end{aligned}$$

Proof: The decompositions $x = u + v$ and $X = G + H$ introduced above lead to

$$\begin{aligned} x &= F^{-1}X = F^{-1}G + F^{-1}H \\ &= F^{-1}D^gG + h = F^{-1}D^g(X - H) + h \\ &= F^{-1}D^gX + h = F^{-1}D^gFx + h. \end{aligned}$$

Setting $B = F^{-1}D^gF$ yields the equation $x = Bx + h$. Similarly,

$$\begin{aligned} X &= Fx = Fu + Fv \\ &= FD^uu + V = FD^u(x - v) + V \\ &= FD^ux + V = FD^uF^{-1}X + V, \end{aligned}$$

and setting $T = FD^uF^{-1}$ we get $X = TX + V$. $\square$

We now face a slight difficulty. The equations $x = Bx + h$ and $X = TX + V$ cannot be solved for x or X, because the matrices $I - B$ and $I - T$ are singular. However, this obstacle can be circumvented.

Theorem 3 *Let the number of unknown time domain samples (the cardinal of $\bar{S}_t$) be denoted by n. Let P be the $n \times n$ principal submatrix of B, obtained by deleting from B all rows and columns whose indices belong to S_t. Denote by $c \in \mathbb{C}^n$ the vector with elements*

$$c_i = \sum_{j \in S_t} B_{ij}v_j + h_i \quad (i \notin S_t).$$

Then, the vector $y \in \mathbb{C}^n$ formed by the n time domain unknowns x_i $(i \notin S_t)$ satisfies

$$y = Py + c. \tag{5.20}$$

Proof: The N equations $x = Bx + h$ are redundant in the sense that there are only n unknown time domain samples x_i. Therefore, we look for a subset of n equations. Since $x = u + v$, it follows that $u + v = Bu + Bv + h$, that is,

$$u_i + v_i = \sum_{j=0}^{N-1} B_{ij}u_j + \sum_{j=0}^{N-1} B_{ij}v_j + h_i.$$

Considering only $i \notin S_t$, and noting that $u_i = 0$ for all $i \in S_t$ whereas $v_i = 0$ for $i \in S_f$, we get

$$u_i = \sum_{j \notin S_t} B_{ij} u_j + \sum_{j \in S_t} B_{ij} v_j + h_i \quad (i \notin S_t),$$

which is (5.20). $\square$

Theorem 4 *Let the number of unknown DFT coefficients (the cardinal of $\bar{S}_f$) be denoted by r. Let Q be the $r \times r$ principal submatrix of T, obtained by deleting from T all rows and columns whose indices belong to S_f. Denote by $d \in \mathbb{C}^r$ the vector with elements*

$$d_i = \sum_{j \in S_f} T_{ij} H_j + V_i \quad (i \notin S_f).$$

Then, the vector $z \in \mathbb{C}^r$ formed by the r unknown DFT harmonics X_i $(i \notin S_f)$ satisfies

$$z = Qz + d. \tag{5.21}$$

Proof: One can appeal to duality and the reversal of the roles of the time and frequency domains. It is also possible to consider the N equations $X = TX + V$, and substituting $X = G + H$ to obtain $G + H = TG + TH + V$, that is,

$$G_i + H_i = \sum_{j=0}^{N-1} T_{ij} G_j + \sum_{j=0}^{N-1} T_{ij} H_j + V_i.$$

Thus, just as in the case of Theorem 3,

$$G_i = \sum_{j \notin S_f} T_{ij} G_j + \sum_{j \in S_f} T_{ij} H_j + V_i \quad (i \notin S_f),$$

which is (5.21). $\square$

The following result [8] can be used to show that the matrices $I - P$ and $I - Q$, corresponding to the equations given in Theorem 3 and 4, are nonsingular, provided that the sampling sets fulfill the necessary density conditions. Recall that $I - P$ and $I - Q$ are submatrices of the singular circulant projection matrices $I - B$ and $I - T$ mentioned in Theorem 2.

Theorem 5 *Let the matrix C be $n \times n$, nonnegative definite, circulant, with z contiguous zero eigenvalues. Then every principal submatrix of C of order $p \leq n - z$ is positive definite.*

Proof: We argue by contradiction. Assume to the contrary that there exists a $n \times n$ nonnegative definite circulant matrix C with singular principal submatrices of order $p = n - z$.

This, in turn, would imply the existence of a nonzero vector $v \in \mathbb{C}^n$ satisfying $v_i = 0$ for all i belonging to a certain set $\mathcal{T}$ containing at least k distinct integers, and such that $v^H C v = 0$.

But $v^H C v$ can be rewritten as $V^H \Gamma V$, where V is the DFT of v, $V = Fv$, and the matrix $\Gamma = FCF^H$ is diagonal. Hence, $v^H C v = 0$ would imply $V_i = 0$ for all i belonging to a certain contiguous set $\mathcal{F}$ with $n - k$ elements. Because of this, v would have to satisfy

$$\sum_{i \notin \mathcal{T}} F_{ki}\, v_i = 0, \quad (k \in \mathcal{F}). \tag{5.22}$$

The matrix of this set of $n - z$ equations for no more than $n - z$ unknowns v_i has linear independent columns when $\mathcal{F}$ is contiguous (the matrix will be Vandermonde if it is square). The linear independence would imply $v_i = 0$ for all $i \notin \mathcal{T}$, leading to the conclusion that v is the zero vector, a contradiction.
□

The minimum dimension equations in the frequency domain appear in [21], and have been used as starting points for efficient methods for signal and image reconstruction [45],[46]. The similarity (duality) between the time domain formulations found in [8] and the frequency domain formulation was discussed in [12]. The time domain equations were applied to the recovery of clipped electrocardiographic signals in [5]. Several other approaches and applications have been discussed. See, for example, [1],[3],[24],[28]–[32],[35]–[40],[47],[50],[51].

5.4.1 Examples and comparisons

The minimum dimension formulations, in the time or frequency domain, allow a variety of reconstruction algorithms (iterative and noniterative). In this section we discuss some of the possibilities, giving examples to illustrate and confirm the main findings.

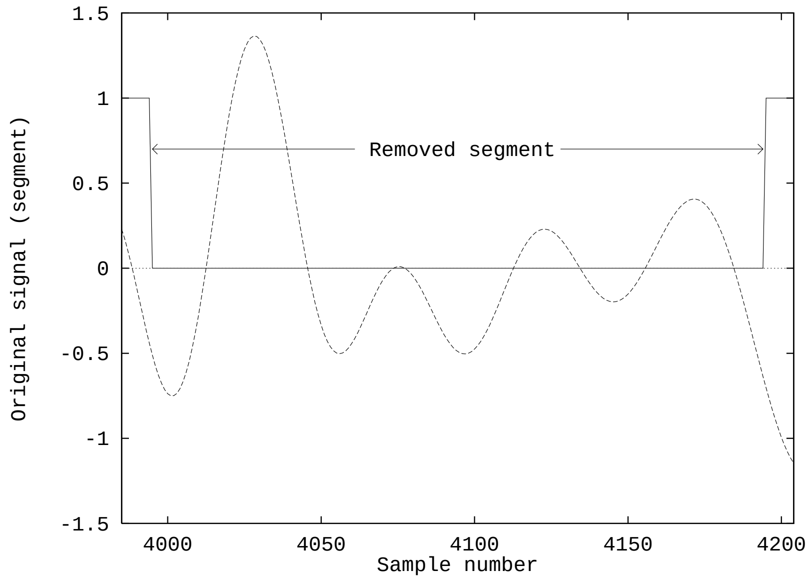


Figure 5.5: A detail of the data for problem A. We have $M = 200$, and the number of (contiguous) missing samples is $n = 200$.

We will start by considering the reconstruction problem illustrated in Fig. 5.5 (problem A). For this problem, we have $M = 200$ and $n = 200$. Therefore, the time domain equations are of dimension $n = 200$, whereas the frequency domain equations involve vectors of dimension $2M + 1 = 401$.

The basic method is the Papoulis-Gerchberg iteration. The error evolution is illustrated in Fig. 5.6, which also depicts the reconstruction result. Note that the error is plotted against CPU time, not against the number of iterations. This was done in order to make the comparisons between the methods simpler. In fact, a method that achieves a certain error in L iterations is not necessarily faster than a method that needs more than L iterations to achieve the same error, because the necessary CPU time per iteration may differ between the methods.

The performance of the Papoulis-Gerchberg method can be easily improved using the minimum dimension formulations. For example, taking the minimum dimension equations in the time domain (5.20) and iterating one obtains the method

$$y^{(i+1)} = Py^{(i)} + c,$$

denoted by SIT (simple iteration, time domain) in Fig. 5.7. The error evolution and the reconstructions depicted in Fig. 5.7 confirm that the method

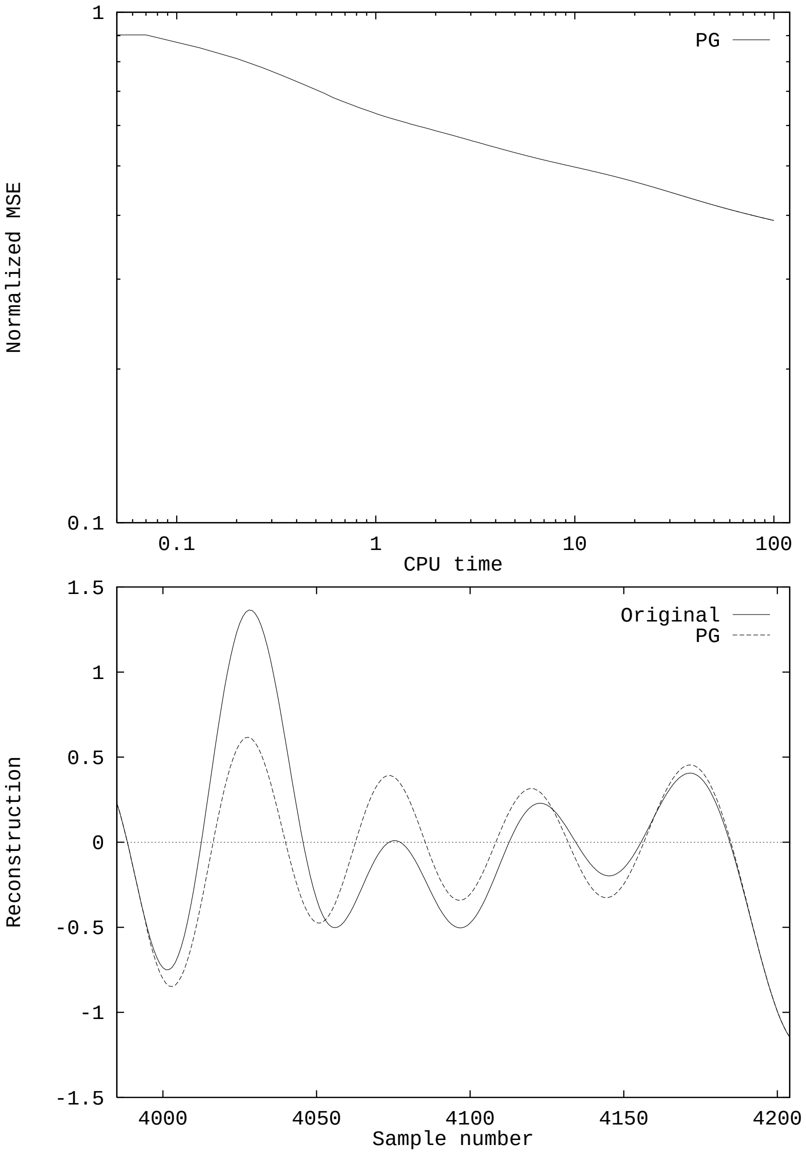


Figure 5.6: Results using the Papoulis-Gerchberg algorithm. The mean-square error is plotted as a function of time, not as a function of the number of iterations.

is slightly more efficient. Each iteration requires one matrix-vector multiplication of size n, as opposed to $N \gg n$ in the Papoulis-Gerchberg algorithm. Thus, the iterations will in general be faster.

Any of the classical iterative methods (Jacobi, JOR, Gauss-Seidel, SOR, SSOR, among others) for the solution of linear systems can be applied, often leading to significant improvements in the convergence rate. A discussion of some of these methods can be found in [8]. Some results are given in Fig. 5.7.

When the iteration matrix is Toeplitz, as is the case for example when there are contiguous losses, the matrix-vector multiplications can be implemented using the FFT, leading to further speed-ups. The effect of performing the Toeplitz matrix-vector multiplications in the SIT method with the FFT is also demonstrated in Fig. 5.7. However, for the sampling set in the example, the convergence rate remains slow. Fortunately, there are much better alternatives.

First of all, the equations can be solved using standard *noniterative* techniques. There are two options: one is to start from the time domain equations, a particularly interesting option when the number of unknown samples n is small compared with the number of nonzero harmonics $2M+1$. Another is to start from the frequency domain equations, which lead to more efficient methods when $2M+1$ is sufficiently small compared with n.

Because the matrices of the minimum dimension equations are Hermitian and positive definite, Cholesky decomposition is a natural possibility (other methods, such as Gaussian elimination, LU decomposition, SVD, are of course also possible). Refer to Fig. 5.8 for examples using Cholesky decomposition.

Recall that when solving a set of equations $Ax = b$ using the Cholesky or LU decomposition methods, the factorization of A has to be performed only once, even if there are multiple right-hand sided vectors b that have to be processed. This is important since the time necessary to perform the decomposition dominates the total CPU time. This is clearly shown in Fig. 5.8. After computing the factorization of the matrix, the solution to linear problems involving the same matrix can be computed in negligible time.

Some further comments are in order. The frequency domain equations involve complex numbers, and the implementations of the frequency domain methods have to be carried out using complex arithmetic. Despite this, if $2M+1$ is sufficiently small compared with n, the frequency domain methods can be preferable. The opposite is of course true when n is sufficiently small.

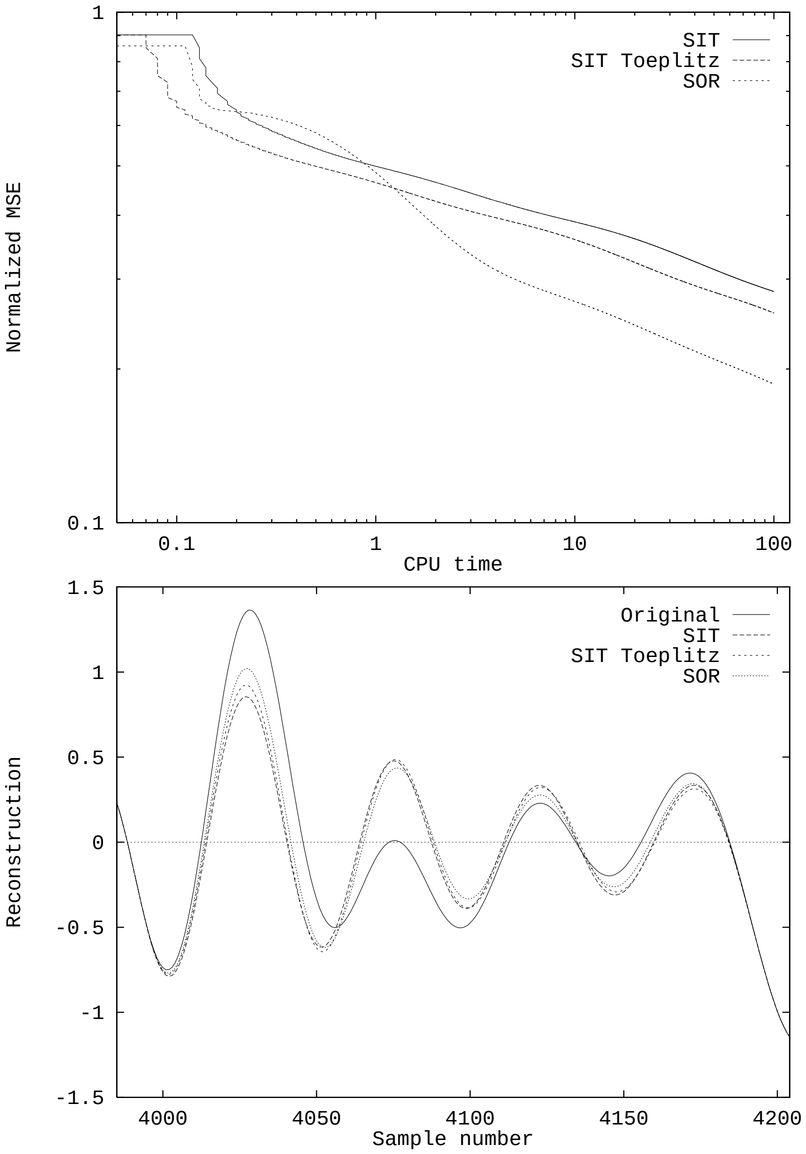


Figure 5.7: Results obtained with some of the iterations to which the minimum dimension equations in the time domain lead. The mean-square error is again plotted as a function of time.

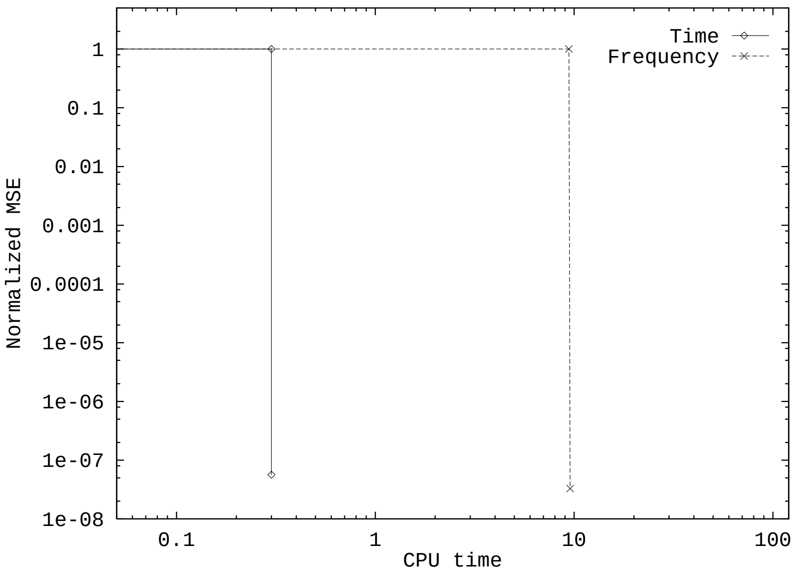


Figure 5.8: Two noniterative methods: Cholesky factorization applied to the minimum dimension equations, in both the time and frequency domains. The initial and final mean-square errors are plotted *versus* time. The horizontal lines correspond to the time spent in the Cholesky factorization itself, the remaining (and by comparison negligible) time corresponds to the solution of the equations. The reconstruction result is not shown (it is indistinguishable from the original).

There is no "faster" method in general.

The structure of the matrices should also be taken in consideration. The known frequency domain samples are contiguous (the data are low-pass), and hence the frequency domain equations are always Toeplitz. This is not necessarily the case with the time domain equations: the matrix can be Toeplitz, but for sparse or irregular distributions of unknown samples the Toeplitz structure will in general be absent.

For problem A, $n < 2M+1$, and the two matrices P and Q that appear in the minimum dimension equations (5.20) (time) and (5.21) (frequency) are Toeplitz. Furthermore, P is real and Q complex. Therefore, the approaches based on the time domain equations should be preferable. The results, as we have seen, confirm this.

The conjugate gradient (CG) method also leads to efficient ways of solving the equations. The basic iteration can be easily modified to explore the structure of the matrix, whenever it exists, or to use preconditioning (see [20] for details).

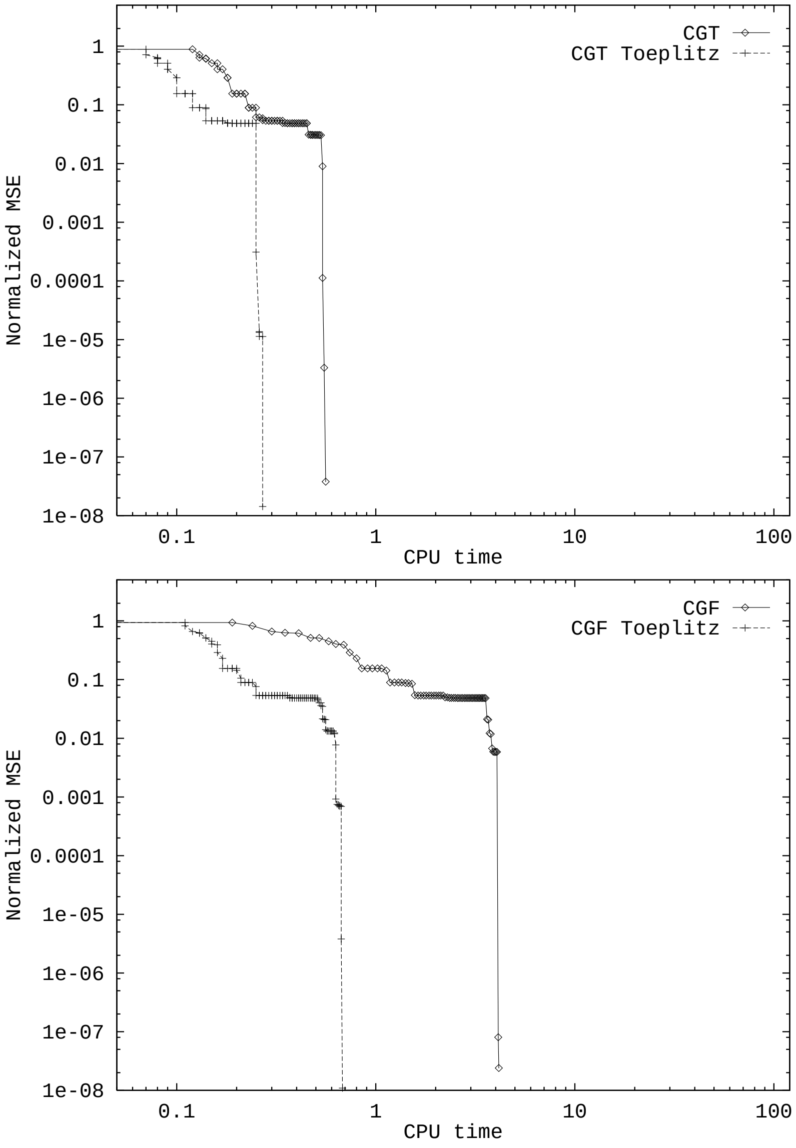


Figure 5.9: The error evolution as a function of time for iterations based on the conjugate gradient method, for the time domain (CGT) and frequency domain (CGF) equations, with and without taking advantage of the Toeplitz character of the matrix.

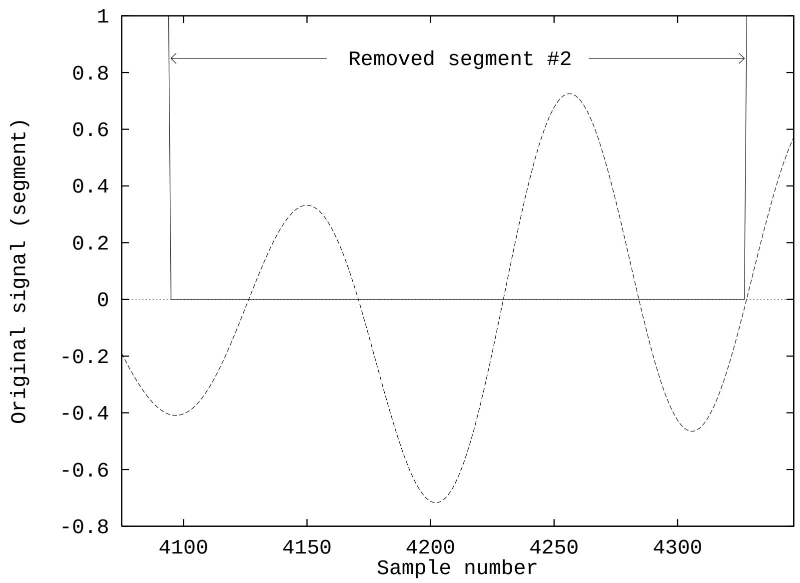


Figure 5.10: A detail of the data for problem B. In this case, $M = 100$, but there are $n = 700$ unknown samples, distributed in three segments of approximately equal length. The figure depicts the central segment.

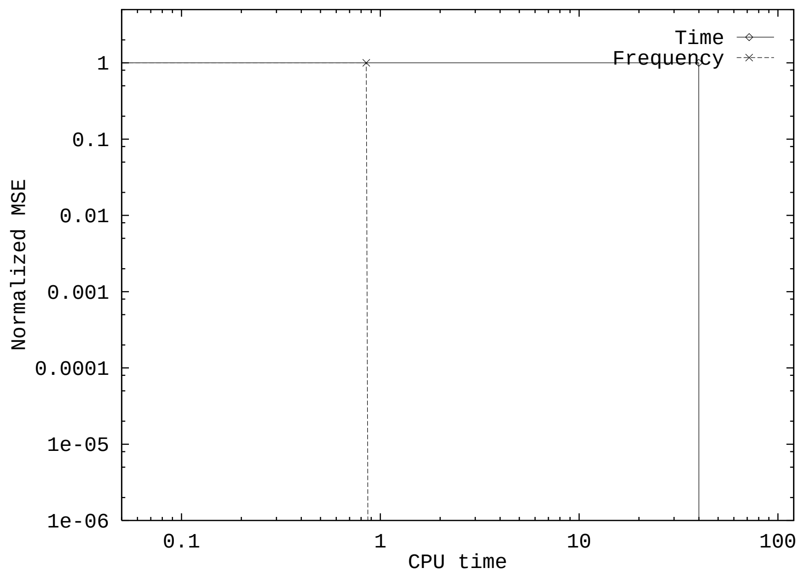


Figure 5.11: Two noniterative methods: Cholesky factorization applied to the minimum dimension equations, in both the time and frequency domains. The initial and final mean-square errors are plotted *versus* time. Compare with Fig. 5.8.

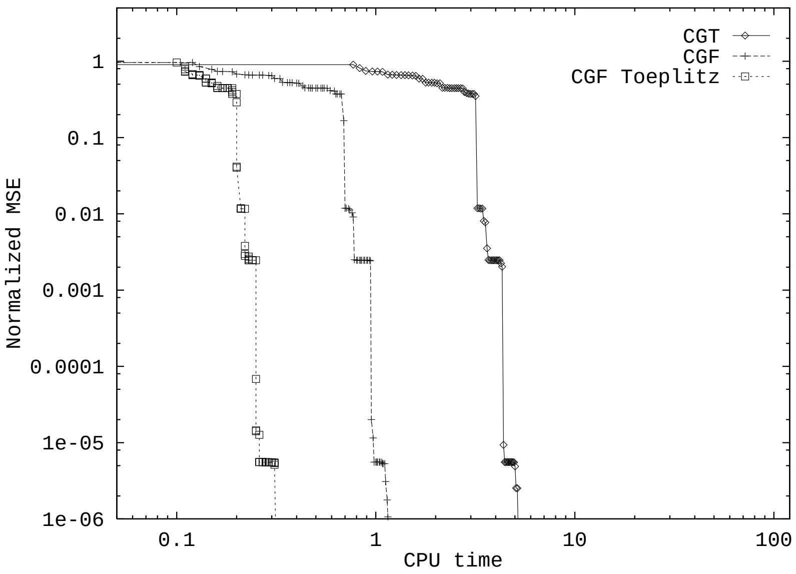


Figure 5.12: The error evolution as a function of time for iterations based on the conjugate gradient method, for the time and frequency domain equations. In the frequemcy domain case, the matrix is Toeplitz and it is possible to accelerate the algorithm using the FFT. Compare with Fig. 5.9.

There are classes of problems for which CG leads to very interesting performance. The combination of the conjugate gradient technique with the equations given in [21], as discussed in [45], [46], leads to a very attractive algorithm. The equivalence between the minimum dimension formulation in the frequency domain and the equations given in [21], [44] was established in [12].

It is of course possible to apply the CG method to either the time or frequency domain equations. This is demonstrated in Fig. 5.9. Is it better to use the time domain or the frequency domain equations? Again, there is not a unique answer, nor a "better" choice. One should at least consider the relative sizes of n and $2M + 1$, taking into account the need for complex arithmetic in the frequency domain case, and the structure of the matrices (Toeplitz or non-Toeplitz).

To demonstrate this, consider a second problem (problem B). In this case, $M = 100$ and $n = 700$, and the unknown samples are distributed in three separate segments of approximately equal length, one of which is depicted in Fig. 5.10. Therefore, the frequency domain equations have lower dimensionality, and the matrix of the time domain equations is not Toeplitz (although it has block structure, we did not take advantage of it). In this case, it turns out that the methods based on the frequency domain equations are preferable. This can be confirmed in Fig. 5.11 and Fig. 5.12.

5.5 Stability analysis

The stability of the interpolation problem critically depends on the sampling set. Therefore, for both iterative and noniterative methods, assessment of the sampling sets is a crucial task. Several results concerning this topic have been published.

References [6], [9] deal with stability issues related to the interpolation of continuous-time band-limited signals from incomplete sets of samples. The approach in [2] is based on the minimization of the maximum of the mean-squared error with respect to the sampling set, whereas [3] studies, among other issues, the effect of noise and out-of-band components on the reconstruction.

Some stability aspects in the context of finite-dimensional signals were already discussed in section 5.3, largely based on the methods employed in [7]. Other issues related to the stability of the band-limited reconstruction problem, also as a function of the sampling set, were studied in [13]. A review of several stability and eigenvalue problems in reference to the minimum dimension methods, including the 2-D case, is available in [10]. The interplay between eigenvalue problems, energy concentration, and the stability of the signal reconstruction problem, is discussed (using the language of frames) in [15].

As already seen in Section 5.3, for contiguous pass-bands (including low-pass and high-pass signals), equidistant losses lead to interpolation processes with optimum stability, and contiguous losses lead to ill posed problems. These conclusions do not necessarily hold for, say, bandpass signals.

In this section we address the stability problem using a minimum dimension formulation. Our aim is to supply accurate bounds for the eigenvalues of the interpolation matrices, in the spirit of [13]. The simplest bounds can be found in negligible time (they require only four arithmetic operations!). This dramatically simplifies the assessment of the sampling sets for this class of interpolation problems, yielding precise, quantitative numerical stability measures. It also explains the striking staircase-like behavior of the eigenvalues, illustrated in Fig. 5.13, which also demonstrates the effectiveness and

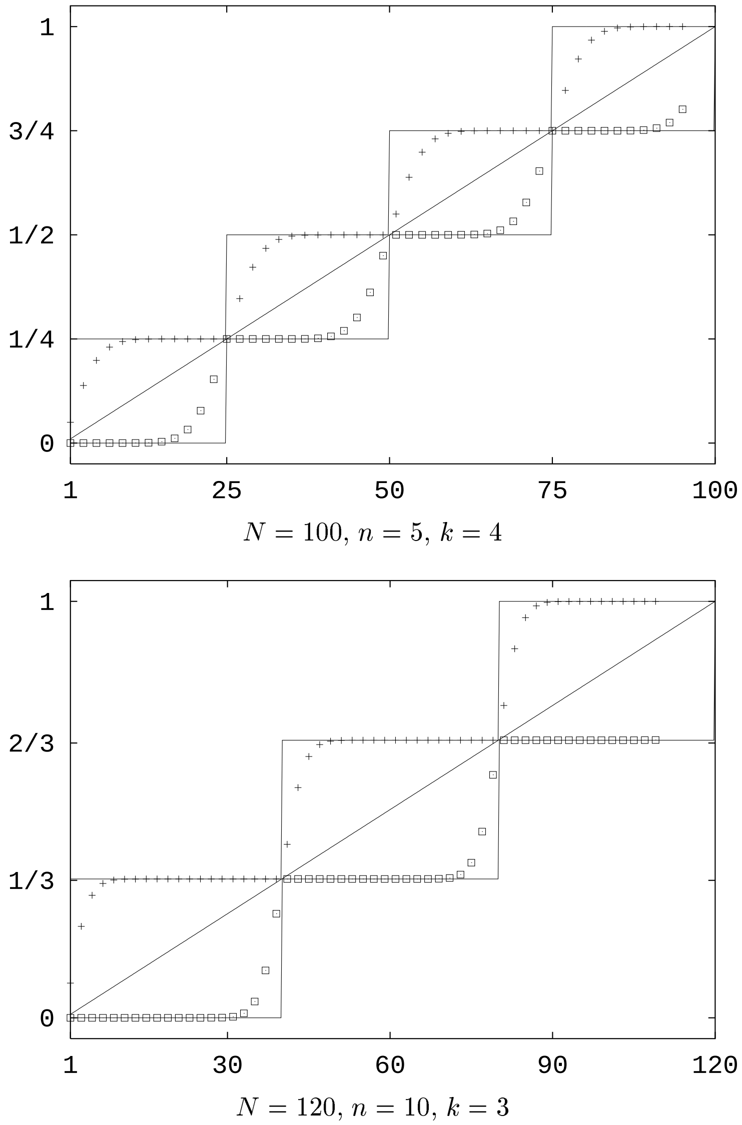


Figure 5.13: The extreme eigenvalues of S and the bounds predicted by Theorem 7, plotted as functions of the bandwidth $2M+1$. The separation line $(2M+1)/N$ predicted by Theorem 6 is also shown. The total number of samples is N, the number of unknown data is n, the gap between unknown samples is k. The bounds become increasingly tighter as N and n increase.

accuracy of the bounds.

As seen in the previous sections, a band-limited signal $x \in \mathbb{C}^N$ can be reconstructed from a subset of its samples. As in section 5.4, the unknown samples are denoted by

$$x(j_1), x(j_2), \ldots, x(j_n). \tag{5.23}$$

If the density of the known samples is above the critical value, the minimum dimension formulation in the time domain (see Theorem 3) leads to

$$u = Su + h, \tag{5.24}$$

which can then be solved for u, the vector of the n unknown samples (5.23). Recall that the $n \times 1$ vector h depends only on known data, and that S is a $n \times n$ Hermitian matrix. If the signal is low-pass with $2M + 1$ harmonics, the elements of S are

$$\begin{aligned} S_{pq} &= \frac{1}{N} \sum_{\ell=-M}^{M} e^{\mathrm{j}\frac{2\pi}{N}(j_p - j_q)\ell} \\ &= \frac{\sin[\pi(2M+1)(j_p - j_q)/N]}{N \sin[\pi(j_p - j_q)/N]}. \end{aligned} \tag{5.25}$$

when $p \neq q$. The diagonal elements S_{ii} equal $(2M + 1)/N$, which is the normalized bandwidth of the data, denoted by w from now on.

In the following, v denotes an arbitrary nth dimensional column vector, with elements $v_1, v_2, \ldots, v_n$. The quadratic form associated with the matrix S is

$$\begin{aligned} v^H S v &= \sum_{p=1}^{n} \sum_{q=1}^{n} v_p^* v_q \left[\frac{1}{N} \sum_{\ell=-M}^{M} e^{\mathrm{j}\frac{2\pi}{N}(j_p - j_q)\ell} \right] \\ &= \frac{1}{N} \sum_{\ell=-M}^{M} \left| \sum_{p=1}^{n} v_p e^{-\mathrm{j}\frac{2\pi}{N} j_p \ell} \right|^2. \end{aligned}$$

We use the notation

$$v^H S v = \frac{1}{N} \sum_{\ell=-M}^{M} |\phi(\ell)|^2, \tag{5.26}$$

where

$$\phi(\ell) = \sum_{p=1}^{n} v_p e^{-\mathrm{j}\frac{2\pi}{N} j_p \ell}.$$

The set of locations of the unknown samples is denoted by U,

$$U = \{j_1, j_2, \ldots, j_n\},$$

and to emphasize the dependence of S and its eigenvalues on U we write $S(U)$ whenever convenient, or, even more explicitly, $S(\{j_1, j_2, \ldots, j_n\})$. Thus, for example, $S(\{0, 2, 7\})$ is the 3×3 matrix which arises in the recovery of samples $x(0), x(2)$ and $x(7)$ of a band-limited signal. Note that the matrix S is a principal $n \times n$ submatrix of the low-pass $N \times N$ matrix B defined by (5.3). Thus, in this notation, $S(\{0, 1, \ldots, N-1\})$ reduces to B.

It follows from (5.26) that the eigenvalues of S are nonnegative. In fact, the spectrum (set of eigenvalues) of the $N \times N$ low-pass matrix B is, as we have seen, $\{0, 1\}$. Therefore, because S is a $n \times n$ principal submatrix of B, all of its eigenvalues belong to the interval $[0, 1]$. This follows from the interlacing inequalities [23].

We know that $I - S$ will be nonsingular if the number of known samples, $N - n$, exceeds the dimension of the subspace of low-pass signals with $2M + 1$ nonzero harmonics. Or, in other words, if the density of the known samples exceeds the normalized bandwidth $w = (2M + 1)/N$ of the data. This is also true for other classes of band-limited signals, provided that the associated matrix B has contiguous eigenvalues (this is the case for high-pass signals, for example). Under these conditions, the largest eigenvalue of S cannot be equal to one. However, a lot more can be said concerning the dependence of the eigenvalues of S on the sampling set U.

Theorem 6 *The minimum eigenvalue λ_{min} of S belongs to $[0, w]$, and the maximum eigenvalue λ_{max} to $[w, 1]$, with $w = (2M + 1)/N$.*

Proof: We have seen that the eigenvalues belong to $[0, 1]$. On the other hand,

$$n\lambda_{\mathsf{min}} \leq \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} S_{ii} = nw \leq n\lambda_{\mathsf{max}},$$

that is, $\lambda_{\mathsf{min}} \leq w \leq \lambda_{\mathsf{max}}$. □

The following result holds when U is a subset of an arithmetic progression, and gives more precise information concerning the eigenvalues. We use the following notations: $\lfloor x \rfloor$, for the largest integer smaller than or equal to x, and $\lceil x \rceil$, for the smallest integer greater than or equal to x.

Theorem 7 *If $U = \{j_1 k, j_2 k, \ldots, j_n k\}$, and N/k is an integer, the smallest and largest eigenvalues λ_{min} and λ_{max} of the matrix $S(U)$ defined by (5.25) satisfy*

$$\frac{\lfloor kw \rfloor}{k} \leq \lambda_{\mathsf{min}} \leq \lambda_{\mathsf{max}} \leq \frac{\lceil kw \rceil}{k},$$

where $w = (2M+1)/N$.

Proof: This is true when $k = 1$ because the eigenvalues lie in $[0, 1]$. Therefore, assume that $k > 1$ and consider the matrix S as a function of M. Clearly, S will be diagonal if and only if $(2M+1)k/N$ is an integer. The smallest value of $2M+1$ for which this can happen is

$$2M + 1 = \frac{N}{k}. \tag{5.27}$$

Under the conditions stated, there is an integer M that solves this equation, and $S_{ii} = (2M+1)/N = 1/k$. But, when (5.27) holds, one has

$$v^H S v = \frac{1}{k} \sum_{i=1}^{n} |v_i|^2,$$

for any nth dimensional column vector v. This means that

$$\sum_{\ell=-M}^{M} |\phi(\ell)|^2 = \frac{1}{k} \sum_{i=1}^{n} |v_i|^2. \tag{5.28}$$

Now, let $0 < 2M+1 < N$. Under the hypothesis placed on U, ϕ is periodic with period N/k. The interval $[-M, M]$ contains at least

$$\left\lfloor \frac{2M+1}{N/k} \right\rfloor$$

and at most

$$\left\lceil \frac{2M+1}{N/k} \right\rceil$$

periods of ϕ. This and (5.28) implies

$$v^H S v = \sum_{\ell=-M}^{M} |\phi(\ell)|^2 \geq \frac{\lfloor kw \rfloor}{k} \sum_{i=1}^{n} |v_i|^2,$$

$$v^H S v = \sum_{\ell=-M}^{M} |\phi(\ell)|^2 \leq \frac{\lceil kw \rceil}{k} \sum_{i=1}^{n} |v_i|^2.$$

The conclusion now follows because the smallest and largest eigenvalues of S are respectively equal to the minimum and maximum values assumed by $v^H S v$, subject to $\|v\| = 1$. □

The bounds can be found in negligible time (they require only a total of four arithmetic operations, including the rounding operation and the computation of w). Nevertheless, they are quite accurate, and can be combined with those of Theorem 6, as seen in Fig. 5.13.

We now turn to less regular sets U. Assume that there is a simple way of estimating the eigenvalues of $S(U_1)$ and $S(U_2)$, for some sets U_1 and U_2. Is it possible to combine those estimates to obtain information concerning the eigenvalues of the larger matrix $S(U_1 \cup U_2)$? The following result is based on this idea.

Theorem 8 *Let $U_{n-1} = \{j_1, j_2, \ldots, j_{n-1}\}$ and $U_n = U_{n-1} \cup \{j_n\}$. Denote by S_n and S_{n-1} the $n \times n$ and $(n-1) \times (n-1)$ reconstruction matrices associated with the sets U_n and U_{n-1}, respectively. Then,*

$$\lambda'(S_{n-1}) - \|v\| \leq \lambda(S_n) \leq \lambda'(S_{n-1}) + \|v\|, \tag{5.29}$$

where $v = [v_1, v_2, \ldots, v_{n-1}]$ and

$$v_k = \frac{\sin[\pi(2M+1)(j_k - j_n)/N]}{N \sin[\pi(j_k - j_n)/N]}, \quad 1 \leq k \leq n-1. \tag{5.30}$$

The $\lambda'(S_{n-1})$ denote the sequence of the eigenvalues of S_{n-1}, together with w, sorted by increasing order, and $\lambda(S_n)$ denotes the eigenvalues of S_n, similarly ordered.

The proof appears in [13]. The result can also be useful when $U = U_1 \cup U_2$, U_1 being of the type addressed by Theorem 7, and U_2 arbitrary. The simplest possible case arises when U_2 has a single element (consider, for example, $U = \{0, 4, 12, 17\}$).

Corollary 1 *The smallest and largest eigenvalues $\lambda_{\mathsf{min}}(S_n)$ and $\lambda_{\mathsf{max}}(S_n)$ of S_n satisfy*

$$\begin{aligned}
\lambda_{\mathsf{min}}(S_n) &\geq \lambda_{\mathsf{min}}(S_{n-1}) - \|v\|, \\
\lambda_{\mathsf{max}}(S_n) &\leq \lambda_{\mathsf{max}}(S_{n-1}) + \|v\|.
\end{aligned}$$

Computing $\|v\|$ requires only $O(n)$ operations, but it can be seen that $\|v\|^2 \leq w(1-w)$, where, as always, $w = (2M+1)/N$ for low-pass signals. Several other variants and related results are discussed in [13].

The problem of obtaining bounds when the elements of the set U are subject to much weaker restrictions is less simple. Results in terms of the minimum separation distance can be found in [14], [16], [17], in the discrete-discrete case and also in the context of L_2 signals.

5.5.1 Examples

The bounds for the eigenvalues can be applied in different contexts.

The estimation of the optimum relaxation parameter of the Papoulis-Gerchberg and similar algorithms is an example. As mentioned in Section 5.3.6, the optimum value of μ depends on the eigenvalues of certain matrices. For ill-posed or large problems, determination of the eigenvalues may be difficult or time-consuming. On the other hand, iteration with a sub-optimum relaxation parameter leads to comparatively low convergence rates.

The bounds lead to approximate values for the eigenvalues of S in (5.24). They also lead to estimates for the condition number of $I - S$, which give a quantitative measure of the feasibility of the reconstruction problem itself. See, for example, [13].

A simple example is given in Fig. 5.14 and Fig. 5.15. Note that in the situation depicted in Fig. 5.14, only eight data samples were unknown in a block of $N = 1024$, but the reconstruction error was still significant.

The situation in Fig. 5.15 appears to be hopeless, since the bandwidth and block length are the same, yet the number of samples to interpolated has increased by a factor of 25. However, the set U is more sparse, and the eigenvalue bounds predict a condition number of at most

$$\kappa(I-S) = \frac{1-\lambda_{\mathsf{min}}}{1-\lambda_{\mathsf{max}}} \leq \frac{0.5}{0.25} = 2.$$

The small condition number explains the unnoticeable interpolation error seen in Fig. 5.15.

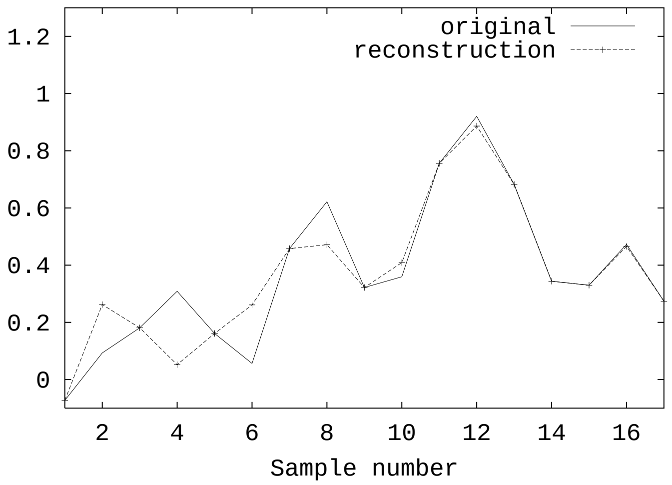


Figure 5.14: Interpolating eight data samples in a record of $N = 1024$ samples with normalized bandwidth $(2M + 1)/N = 0.7$, using noisy data. The figure depicts the original data and the interpolated samples. Only 17 samples out of $N = 1024$ are shown. The indexes of the interpolated samples were multiples of two, and the bounds predicted by Theorem 7 were 0 and 1.

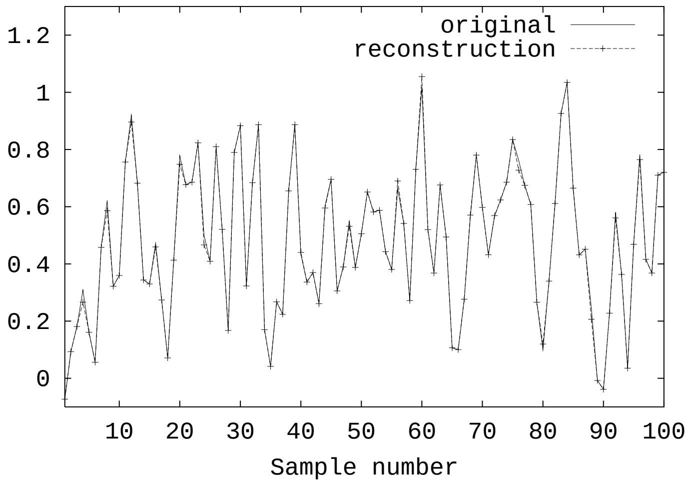


Figure 5.15: Interpolating 200 data samples in a record of $N = 1024$ samples with normalized bandwidth $(2M + 1)/N = 0.7$, using the same data used in Fig. 5.14. The figure depicts the original and reconstructed data. Only 100 samples out of $N = 1024$ are shown. The indexes of the interpolated samples were multiples of four, and the bounds predicted by Theorem 7 were 0.5 and 0.75.

5.6 Conclusions

In this chapter we have examined several iterative and noniterative methods for the reconstruction of partially known band-limited signals. The minimum dimension equations, in the time or frequency domains, were the starting point from which several possible implementations were derived and compared.

The minimum dimension matrices are Hermitian and nonnegative definite, and nonsingular if the density of the known data is above the Nyquist-Landau minimum density. Several possible strategies for dealing with the equations exist, and we have examined some of the several factors that the practitioner should consider when designing an algorithm for a specific problem. For the minimum dimension formulation in the time domain, the main issues and possibilities are:

- The matrices are real-valued, symmetric, $n \times n$. If the unknown data are consecutive, the matrices are also Toeplitz.

- The equations can be solved by simple iteration, or using iterative methods based on other additive splittings of the matrix (Jacobi, JOR, Gauss-Seidel, SOR, SSOR, block versions of these, etc.).

- If the matrix is Toeplitz, the Toeplitz-vector multiplications in the preceding methods can be computed using the FFT, leading to important speedups.

- The conjugate gradient method should also be considered. In this case, the matrix-vector products should of course be computed using the FFT whenever possible. Preconditioning can also have a favorable impact.

- Noniterative factorizations of the matrix (such as Cholesky decomposition, for example), or other noniterative techniques, can also be used.

The crucial points, in the frequency domain case, are as follows:

- In the low-pass case the matrices are always Toeplitz, of size $2M + 1 \times 2M + 1$. They are Hermitian and complex-valued, meaning that complex arithmetic and algorithms will have to be used.

- The equations can be solved using iterative techniques, such as Jacobi, JOR, Gauss-Seidel, SOR, SSOR, and so on. The necessary Toeplitz-vector multiplications can be carried out using the FFT.

- The preconditioned conjugate gradient method should be considered. Again, the matrix-vector products can be implemented using the FFT.

- Noniterative factorizations of the complex-valued matrix (say, Cholesky decomposition) or other standard noniterative techniques can also be used.

The algorithm for a given problem should be planned with these factors in mind. The time domain equations are of interest when the number of unknown samples n is small compared with the number of nonzero harmonics $2M+1$. The frequency domain equations may lead to more efficient methods when $2M+1$ is sufficiently small compared with n. The break-even points depend on the implementation details (factors such as cache size often lead to substantial performance differences).

When solving a set of equations using the Cholesky or LU decomposition methods, the factorization of the matrix has to be performed only once, even if there are multiple right-hand sided vectors that have to be processed. This is important since, as we have seen, the time necessary to perform the decomposition dominates the total CPU time. After computing the factorization of the matrix, the solution to linear problems involving the same matrix can be computed in negligible time. This can be successfully explored in image extrapolation or superresolution problems.

Bibliography

[1] S. D. Cabrera and T. W. Parks. Extrapolation and spectral estimation with iterative weighted norm modification. *IEEE Trans. Signal Processing*, 39(4):842–851, Apr. 1991.

[2] D. S. Chen and J. P. Allebach. Analysis of error in reconstruction of two-dimensional signals from irregularly spaced samples. *IEEE Trans. Acoust., Speech, Signal Processing*, 35(2):173–180, Feb. 1987.

[3] P. Delsarte, A. J. E. M. Janssen, and L. B. Vries. Discrete prolate spheroidal wave functions and interpolation. *SIAM J. Appl. Math.*, 45(4):641–650, Aug. 1985.

[4] H. G. Feichtinger and K. Gröchenig. Theory and practice of irregular sampling. In J. J. Benedetto and M. W. Frazier, editors, *Wavelets: Mathematics and Applications*, pages 305–363. CRC Press, Boca Raton, 1994.

[5] P. J. S. G. Ferreira. Fast iterative reconstruction of distorted ECG signals. In *Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume II, pages 777–778, Paris, France, Oct. 1992.

[6] P. J. S. G. Ferreira. Incomplete sampling series and the recovery of missing samples from oversampled band-limited signals. *IEEE Trans. Signal Processing*, 40(1):225–227, Jan. 1992.

[7] P. J. S. G. Ferreira. Interpolation and the discrete Papoulis-Gerchberg algorithm. *IEEE Trans. Signal Processing*, 42(10):2596–2606, Oct. 1994.

[8] P. J. S. G. Ferreira. Noniterative and faster iterative methods for interpolation and extrapolation. *IEEE Trans. Signal Processing*, 42(11):3278–3282, Nov. 1994.

[9] P. J. S. G. Ferreira. The stability of a procedure for the recovery of lost samples in band-limited signals. *Sig. Proc.*, 40(3):195–205, Dec. 1994.

[10] P. J. S. G. Ferreira. A class of eigenvalue problems in interpolation, extrapolation and sampling. In *SampTA'95, 1995 Workshop on Sampling Theory and Applications*, pages 125–136, Jurmala, Latvia, Sept. 1995.

[11] P. J. S. G. Ferreira. The stability of certain image restoration problems: quantitative results. In *Proceedings of the Second IEEE International Conference on Image Processing, ICIP-95*, volume II, pages 29–32, Washington, D.C., U.S.A., Oct. 1995.

[12] P. J. S. G. Ferreira. Interpolation in the time and frequency domains. *IEEE Sig. Proc. Letters*, 3(6):176–178, June 1996.

[13] P. J. S. G. Ferreira. The eigenvalues of matrices which occur in certain interpolation problems. *IEEE Trans. Signal Processing*, 45(8):2115–2120, Aug. 1997.

[14] P. J. S. G. Ferreira. The condition number of certain matrices and applications. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 99*, pages 2043–2046, Phoenix, Arizona, U.S.A., Mar. 1999.

[15] P. J. S. G. Ferreira. Mathematics for multimedia signal processing II — discrete finite frames and signal reconstruction. In J. S. Byrnes, editor, *Signal Processing for Multimedia*, pages 35–54. IOS Press, 1999. Presented at the NATO Advanced Study Institute "Signal Processing for Multimedia", Il Ciocco, Italy, July 1998. In press.

[16] P. J. S. G. Ferreira. Superresolution, the recovery of missing samples, and Vandermonde matrices on the unit circle. In *Proceedings of the 1999 Workshop on Sampling Theory and Applications, SampTA'99*, pages 216–220, Loen, Norway, Aug. 1999.

[17] P. J. S. G. Ferreira. Stability issues in error control coding in the complex field, interpolation, and frame bounds. *IEEE Sig. Proc. Letters*, 7(3), Mar. 2000. In press.

[18] P. J. S. G. Ferreira and A. J. Pinho. Errorless restoration algorithms for band-limited images. In *Proceedings of the First IEEE International*

Conference on Image Processing, ICIP-94, volume III, pages 157–161, Austin, TX, U.S.A., Nov. 1994.

[19] R. W. Gerchberg. Super-resolution through error energy reduction. *Optica Acta*, 21(9):709–720, 1974.

[20] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, 1989.

[21] K. Gröchenig. A discrete theory of irregular sampling. *Linear Algebra Appl.*, 193:129–150, 1993.

[22] C. I. Haupt, N. Schuff, M. W. Weiner, and A. A. Maudsley. Removal of lipid artifacts in ^{1}H spectroscopic imaging by data extrapolation. *Magn. Reson. Medic.*, 35(5):678–687, May 1996.

[23] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1990.

[24] A. K. Jain and S. Ranganath. Extrapolation algorithms for discrete signals with application in spectral estimation. *IEEE Trans. Acoust., Speech, Signal Processing*, 29(4):830–845, Aug. 1981.

[25] M. C. Jones. The discrete Gerchberg algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(3):624–626, June 1986.

[26] R. J. Marks II. Restoring lost samples from an oversampled band-limited signal. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(3):752–755, June 1983.

[27] R. J. Marks II. *Introduction to Shannon Sampling and Interpolation Theory.* Springer, Berlin, 1991.

[28] F. Marvasti, M. Hasan, M. Echhart, and S. Talebi. Efficient algorithms for burst error recovery using FFT and other transform kernels. *IEEE Trans. Signal Processing*, 47(4):1065–1075, Apr. 1999.

[29] F. Marvasti and T. J. Lee. Analysis and recovery of sample-and-hold and linearly interpolated signals with irregular samples. *IEEE Trans. Signal Processing*, 40(8):1884–1891, Aug. 1992.

[30] F. A. Marvasti, C. Liu, and G. Adams. Analysis and recovery of multidimensional signals from irregular samples using nonlinear and iterative techniques. *Sig. Proc.*, 36:13–30, 1993.

[31] F. M. Marvasti, P. M. Clarkson, M. V. Dokic, U. Goenchanart, and C. Liu. Reconstruction of speech signals with lost samples. *IEEE Trans. Signal Processing*, 40(12):2897–2903, Dec. 1992.

[32] M. Nafie and F. M. Marvasti. Implementation of recovery of speech with missing samples on a DSP chip. *Electron. Letters*, 30(1):12–13, Jan. 1994.

[33] P. S. Naidu and B. Paramasivaiah. Estimation of sinusoids from incomplete time series. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(3):559–562, June 1984.

[34] A. Papoulis. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Trans. Circuits Syst.*, 22(9):735–742, Sept. 1975.

[35] A. Papoulis and C. Chamzas. Detection of hidden periodicities by adaptive extrapolation. *IEEE Trans. Acoust., Speech, Signal Processing*, 27(5):492–500, Oct. 1979.

[36] L. C. Potter and K. S. Arun. Energy concentration in band-limited extrapolation. *IEEE Trans. Acoust., Speech, Signal Processing*, 37(7):1027–1041, Jul. 1989.

[37] J. L. C. Sanz and T. S. Huang. Discrete and continuous band-limited signal extrapolation. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(5):1276–1285, Oct. 1983.

[38] J. L. C. Sanz and T. S. Huang. Some aspects of band-limited signal extrapolation: Models, discrete approximations, and noise. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(6):1492–1501, Dec. 1983.

[39] J. L. C. Sanz and T. S. Huang. Unified Hilbert space approach to iterative least-squares linear signal restoration. *J. Opt. Soc. Am.*, 73(11):1455–1465, Nov. 1983.

[40] J. L. C. Sanz and T. S. Huang. A unified approach to noniterative linear signal restoration. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(2):403–409, Apr. 1984.

[41] R. W. Schafer, R. M. Mersereau, and M. A. Richards. Constrained iterative restoration algorithms. *Proc. IEEE*, 69(4):432–450, Apr. 1981.

[42] M. I. Sezan and H. Stark. Image restoration by the method of convex projections: Part 2 — applications and numerical results. *IEEE Trans. Med. Imag.*, 1(2):95–101, Oct. 1982.

[43] D. R. Smart. *Fixed Point Theorems.* Cambridge University Press, Cambridge, 1980.

[44] T. Strohmer. *Efficient Methods for Digital Signal and Image Reconstruction from Nonuniform Samples.* PhD thesis, Institut für Mathematik der Universität Wien, Nov. 1993.

[45] T. Strohmer. On discrete band-limited signal extrapolation. In M. E. H. Ismail, M. Z. Nashed, A. I. Zayed, and A. F. Ghaleb, editors, *Mathematical Analysis, Wavelets, and Signal Processing*, volume 190 of *Contemporary Mathematics*, pages 323–337, Providence, Rhode Island, 1995. American Mathematical Society.

[46] T. Strohmer. Computationally attractive reconstruction of bandlimited images from irregular samples. *IEEE Trans. Image Processing*, 6(4):540–548, Apr. 1997.

[47] B. J. Sullivan and B. Liu. On the use of singular value decomposition and decimation in discrete-time band-limited signal extrapolation. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(6):1201–1212, Dec. 1984.

[48] V. T. Tom, T. F. Quatieri, M. H. Hayes, and J. H. McClellan. Convergence of iterative nonexpansive signal reconstruction algorithms. *IEEE Trans. Acoust., Speech, Signal Processing*, 29(5):1052–1058, Oct. 1981.

[49] R. S. Varga. *Matrix Iterative Analysis.* Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[50] C. K. W. Wong, F. Marvasti, and W. G. Chambers. Implementation of recovery of speech with impulsive noise on a DSP chip. *Electron. Letters*, 31(17):1412–1413, Aug. 1995.

[51] W. Y. Xu and C. Chamzas. On the extrapolation of band-limited functions with energy constraints. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(5):1222–1234, Oct. 1983.

[52] D. C. Youla. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Trans. Circuits Syst.*, 25(9):694–702, Sept. 1978.

[53] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1 — theory. *IEEE Trans. Med. Imag.*, 1(2):81–94, Oct. 1982.

[54] D. M. Young. *Iterative Solution of Large Linear Systems.* Academic Press, Orlando, third edition, 1971.