Constrained Image Reconstruction from Incomplete and Noisy Data: A New Parametric Approach

by

Zhi-Pei Liang

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Thesis Advisor: E. Mark Haacke, Ph.D.

Department of Biomedical Engineering
Case Western Reserve University
May 14, 1989

CASE WESTERN RESERVE UNIVERSITY

GRADUATE STUDIES

We hereby approve the thesis of

Zhi-pei Liang

candidate for the Ph. D.

degree.*

Signed: Cecil W Thoma
(Chairman)

E. Mark Haacke

Steven Ng

A. Dennis Nelson

Date 13 April 89

*We also certify that written approval has been obtained for any proprietary material contained therein.

Constrained Image Reconstruction from Incomplete and Noisy Data: A New Parametric Approach

Abstract

by

Zhi-Pei Liang

The problem of reconstructing an image from limited Fourier transform data is addressed in this thesis. In order to reduce/remove the truncation artifacts (the Gibbs ringing and the loss of resolution) associated with the conventional FFT method, a new parametric image reconstruction approach is introduced. With this method, an image to be reconstructed is represented in terms of a generalized series model, and then the reconstruction problem is transformed to a parameter estimation problem. Particularly, instead of using the classical basis functions from the eigenfunctions of a Sturm-Liouville system, such as the set of complex exponentials, we proposed to use localized polynomials as the basis functions. This localized polynomial approximation model (LPA) has several desirable properties: First, it can parsimoniously characterize piecewise continuous behavior, therefore, the Gibbs ringing can be effectively removed or reduced. Second, the model parameters can be found efficiently using the linear-prediction formalism.

In this thesis, background material including a brief review of several well-known non-parametric methods are first presented; then the LPA modeling concept is introduced, which is followed by a detailed discussion of the theory and numerical procedures for finding the LPA model parameters; finally, reconstruction results from simulated and real magnetic resonance data are presented to demonstrate the performance of the proposed method.

To my wife and my parents

Acknowledgments

My sincerest appreciation to Dr. E. Mark Haacke for his support and encouragement throughout the course of this research; to Dr. Cecil W. Thomas for his help and guidance in all facets of my six years graduate studies at CWRU; to Dr. Steven H. Izen for his valuable comments about the thesis; and to Dr. Dennis Nelson for providing the very much needed computing facilities.

I would also like to express my gratitudes to my friends Sundar Amartur, Hubert Vesselle, Piotr Wielopolski, and Dr. Donald Kormos for creating a memorable research environment.

My special thanks go to my brother Liang, Fang-Pei for his insightful and constructive suggestions; and to Josph Molter for his tremendous help in making excellent hardcopies for images off the computers whenever I needed.

Finally, but most of all, my deepest thanks go to my wife, Bo, for her love, understanding, and support; without them this work may have never come to this stage.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

The limited data image reconstruction problem is addressed in this thesis. Particularly, we assume that the data are collected in the Fourier domain and perturbed by random noise and other sources of error; the object function (image) is piecewise continuous but of bounded variations. Mathematically, this problem can be formulated as solving the following integral equation:

$$\begin{aligned} \hat{s}(k) &= \int_{-\infty}^{+\infty} \rho(x) e^{-i2\pi kx} dx + \varepsilon(k) \\ &= s(k) + \varepsilon(k) \,, \end{aligned} \tag{1.1}$$

where the object function $\rho(x)$ is a spatial-support-limited complex function, i.e., $\rho(x) = 0$ for $|x| \notin \Omega = (-\sigma/2, +\sigma/2)$; the data $\hat{s}(k)$ is corrupted by noise $\varepsilon(k)$, and is uniformly sampled at $k = n\Delta k$, $n \in \mathcal{N} = \{-N/2, \cdots, N/2 - 1\}$, with the sampling interval Δk satisfying the Nyquist criterion.

In general, an infinite number of data points of $\hat{s}(k)$ are necessary to reconstruct $\rho(x)$ exactly (in the zero mean-squared error sense) in the noiseless case. When N is finite, the solution of Eq. (1.1) is not unique; traditionally, $\rho(x)$ is approximated by a truncated Fourier series (TFS) as follows

$$\hat{\rho}_N(x) = \Delta k \sum_{n=-N/2}^{N/2-1} \hat{s}(n\Delta k) e^{i2\pi n\Delta kx}. \tag{1.2}$$

Although this method finds wide application in various disciplines of applied science owing to the existence of fast Fourier transform (FFT) algorithms

for efficient evaluation of $\hat{\rho}_N(x)$, it suffers from several drawbacks. First, the spatial resolution of $\hat{\rho}_N(x)$ is limited to $1/(N\Delta k)$. Second, when $\rho(x)$ contains step discontinuities, the Gibbs ringing resulting from the data truncation may mask subtle but important features. Third, when noise is present, improving spatial resolution and reducing the Gibbs ringing by increasing the number of data points N will deteriorate the signal-to-noise ratio (S/N) of $\hat{\rho}_N(x)$. These limitations render this technique undesirable for imaging applications when N is small and high resolution and high S/N are desired.

Therefore, the current problem of limited data image reconstruction is not simply to find a solution for Eq. (1.1), but instead to find an "optimal" one which can simultaneously

- overcome the resolution limit,
- remove/reduce the Gibbs ringing, and
- improve the signal-to-noise ratio.

1.2 Motivation

The major impetus for this research came from magnetic resonance imaging (MRI). This section describes the basic principle underlying MRI and shows how the limited-data image reconstruction problem arises in practice.

Magnetic resonance imaging is based on the so-called *nuclear magnetic resonance* (NMR) phenomenon that when certain atomic nuclei with non-zero magnetic moment are placed in a static magnetic field and stimulated by radio-frequency (RF) pulses of a particular frequency (the Larmor frequency), they will re-emit some of the absorbed energy in the form of radio signals [53]. By properly manipulating the bulk magnetization so that it has a non-zero component orthogonal to the static field, free induction decay (FID) signals can be measured following the RF pulse switch-off. Since its discovery in 1946

independently by the groups headed by Block and Purcell, NMR has evolved into one of the most powerful techniques for the elucidation of structures and dynamics of molecules in matter. Medical applications of NMR began in the seventies with Lauterbur's development of the spatial information encoding concept [49]. Spurred by key developments in magnet design, RF technology, and advanced pulse sequences in the last decade, MRI has now become a premier technique for producing anatomical images from humans.

The primary task for MR imaging is to identify individually the FID signals emitted from nuclei at each location in an object being imaged [53, 68]. This requires a method to encode the spatial information into the FID signals. The most widely used technique nowadays is to superimpose magnetic field gradients onto a main homogeneous excitation field. In two-dimensional Fourier transform MR imaging, for example, orthogonal x and y gradients, G_x and G_y, are applied to frequency- and phase-encode the spatial information. When G_x, the frequency-encoding gradient, is applied it causes the Larmor frequency (which is directly proportional to the local field) to vary linearly with position along the x axis. The phase-encoding gradient, G_y, on the other hand, is to encode the orthogonal coordinate, y, into the FID signals by governing the starting phase of the signals from different y locations. Therefore, the essence of MR imaging experiments is to encode the spatial information into the Fourier Domain; namely, it is the Fourier transform of the final image that is actually devised to be measured. This data acquisition strategy is often coined *Fourier imaging*. Ignoring the T_1 and T_2 relaxation time effects, the FID signal acquired is proportional to

$$s(k_x, k_y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(x,y) e^{-i2\pi(k_x x + k_y y)} dx dy \tag{1.3}$$

where $\rho(x,y)$ is the local spin density to be imaged. If square-wave gradients are used, k_x and k_y, the wave numbers in the x and y directions, are related

to G_x and G_y by

$$k_x = \gamma G_x t_x, \qquad \text{and} \qquad k_y = \gamma G_y t_y$$

where γ is the gyromagnetic ratio, t_x is the data acquisition time for each FID, and t_y is the phase-encoding (preparation) time.

To obtain an image of spatial resolution $1/(N_x \Delta k_x)$ and $1/(N_y \Delta k_y)$ in the x and y directions, respectively, with the conventional FFT reconstruction method, we need to collect N_x points in an interval of Δk_x in the k_x direction and N_y points in an interval of Δk_y in the k_y direction. This is usually done by sampling each FID signal N_x times in the data acquisition time t_x and repeating the experiment N_y times with G_y stepping through $-N/2\Delta G_y, \cdots, (N/2-1)\Delta G_y$ to get N_y FIDs. Consequently, the total imaging time for acquiring such an image is $T = N_y T_R$ (where T_R is the repeat time), which can amount to several minutes for a typical 256×256 image. For acquiring a three-dimensional data set, the scanning time can take up to 15 minutes even with rapid scanning techniques. Therefore, it is of great importance to reduce this scanning time for either imaging moving objects such as the heart or enhancing patient throughput. Since the reduction of the repeat time T_R is inherently limited by the relaxation time T_1, we are often forced to sacrifice the high-frequency phase-encoding measurements for fast MR imaging. However, this can cause serious truncation artifacts [51, 99]. Hence, how to reconstruct high-quality MR images without these high-frequency measurements is one of the challenges in current MRI research [62, 84, 99].

The problem of reconstructing a support-limited function from finitely many samples of its Fourier transform also arises in a number of other applications. For example, in synthetic-aperture radio astronomy, the primary data collected are the Fourier transform of the physical quantity (radio brightness) to be imaged. However, physical constraints on the setting of the radio

antennas put a fundamental limit on the number of Fourier components that can be measured. If the FFT method is used for image reconstruction, the lack of high-frequency data will introduce spurious features to the images [27].

Therefore, the limited Fourier transform data inversion problem is a very practical problem in medical and physical imaging. It is very much desirable to develop a new technique which can reliably extract the maximum amount of information from the limited data and take advantage of *a priori* information as well. It goes without saying that such a technique might well prove useful in areas other than medical and physical imaging.

1.3 Main Results

Since *a priori* information is necessary to handle incomplete data with noise, in this research, we focused on the issue of how to use appropriate *a priori* constraints to overcome the data truncation artifacts. The main results thus far obtained are as follows.

First, in order to reduce the Gibbs ringing, we need some mechanism(s) to handle discontinuities in an object function. Considering that non-parametric methods lack this capability, we introduced a localized modeling approach to accomplish this. Particularly, an adaptive localized polynomial approximation (LPA) model is established to constrain reconstructions. This model constraint is expected to be valid whenever an object function is characterized by piecewise continuous behavior, and can also potentially help to reduce/eliminate the Gibbs ringing, to improve signal-to-noise ratio, and to enhance image resolution in general. As compared with other existing parametric models, the LPA model is more appropriate for characterizing piecewise continuous behavior and, thus, will be more effective for reducing the Gibbs ringing and reconstructing sharp edges.

Secondly, an algorithm for solving for the LPA model parameters has been found. By making use of the linear prediction formalism, the nonlinearity of the model can be nicely absorbed into a polynomial. Specifically, when the model is exact and the data S/N is high, the nonlinear model parameters can be found very efficiently with the Levinson algorithm and a polynomial rooting procedure; when this fails, the truncated SVD-based methods can be used to combat noise and modeling offset; the classical nonlinear least squares fitting procedure has also been shown to be applicable to further lower the noise threshold if one of the linear prediction methods is used to provide an initial estimate and the variable projection method is utilized to perform searching.

Finally, the performance of the LPA model being applied to reconstructing object functions from simulated and real experimental data has been analyzed. It is found that the LPA model is very effective for reconstruction of piecewise continuous object functions. For example, in reconstructing box-car type of object functions, the LPA model is capable of recovering $N-1$ boxes exactly from $2N$ data points; For boxes larger than a pixel, their edges can usually be determined to an accuracy much smaller than the pixel size; Boxes of widths smaller than a pixel can also be picked up with a reasonable level of data S/N and, thus, superresolution is possible. For reconstruction of general object functions, the LPA model can fairly accurately reconstruct sharp edges if any, and track and rebuild smooth features with localized lower-order polynomials. Furthermore, since the LPA model uses localized basis functions, local image properties or constraints can be easily incorporated into the reconstruction to improve image resolution and signal-to-noise ratio; more importantly, local approximation (modeling) error resulting from noise and possible systematic modeling offsets will be localized in the reconstruction rather than propagat throughout the entire image as in case of using non-local basis functions (e.g., the basis functions used in the Fourier series).

1.4 Organization of the Dissertation

This dissertation is organized as follows:

In Chapter 2, we review some of the important concepts fundamental to understanding the material in this thesis. Specifically, the concept of ill-posedness and regularization is covered. Some relevant properties of the Fourier transform are also discussed.

Chapter 3 is devoted to a literature review. Several popular non-parametric type of constrained reconstruction methods, including the support-limited extrapolation methods, the POCS-based methods, and the maximum entropy methods, are briefly reviewed. Their major strengths and weaknesses are also discussed.

In Chapter 4, the parametric image reconstruction concept is introduced. Particularly, the LPA model is defined and its characteristics are described. A brief discussion on the principle of parametric modeling and a short review of some of the existing parametric models are also included for completeness.

In Chapter 5, theory and numerical procedures for finding the LPA model parameters are detailed.

In chapter 6, the performance of the LPA model is analyzed based on reconstruction results from simulated data and real magnetic resonance experimental data.

In Chapter 7, the areas of future research are outlined and briefly discussed with the emphasis being placed upon improving the LPA model to extend its utility.

Chapter 8 contains the conclusions.

Chapter 2

Background

2.1 Ill-Posed Problems and Regularization Techniques

It is believed that physical situations can only lead to well-posed problems. However, due to inexact modeling or inaccurate approximations or insufficient measurement information, many problems in mathematical physics are ill-posed. Unfortunately, most inverse problems in applied physics and engineering fall into this category.

The first formal definition of ill-posedness is due to Hadamard [89]. Under his definition, the inverse problem of an operator equation $y = Kx$ is well-posed provided the following conditions are satisfied:

(1) *Existence*: a solution x exists for every element y in the range space;

(2) *Uniqueness*: the solution x is unique;

(3) *Stability*: small perturbations in y result in small perturbations in the solution x without the need to impose additional constraints.

If any of the above conditions are violated, the problem is said to be ill-posed. Specifically, when the operator K does not satisfy the stability condition, it is termed *ill-conditioned*. For the problem addressed, K corresponds to the Fourier integral operator. The ill-conditioned nature of this operator is evidenced by the following *Riemann-Lebesgue* lemma [14].

Theorem 2-1 (Riemann-Lebesgue Lemma) *If a kernel H is Riemann or Lebesgue integrable in the interval (a, b), then it follows that*

$$\lim_{\omega \to \infty} \int_a^b H(k, x) \sin(\omega x) dx = 0$$

Let $H(k,x) = e^{-i2\pi kx}$, and $(a,b) = (-\infty, +\infty)$, then the theorem implies that $s(k)$ in Eq. (1.1) is insensitive to the addition of arbitrary high-frequency components to the solution $\rho(x)$. In other words, small differences in $s(k)$ can map into large differences in $\rho(x)$. Therefore, when an operator K is ill-conditioned, the solution of the associated inverse problem $x = K^{-1}y$ can be highly sensitive to perturbations in y. This noise amplification aspect of ill-conditioned operators is the main source of difficulties in solving ill-posed inverse problems.

The general approach for solving an ill-posed problem is to convert it to an associated well-posed problem [58, 72]. This approach is usually referred to as regularization of the ill-posed problem. Consider the linear system of equations $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} \in \mathcal{C}^{m\times n}$, $\mathbf{x} \in \mathcal{C}^n$, and $\mathbf{b} \in \mathcal{C}^m$. Suppose $\text{rank}(\mathbf{A}) = \min\{m,n\}$. Then, if $m > n$, the system is over-determined and thus has no solution in the classical sense. On the other hand, if $m < n$, the system is under-determined, and there exist infinite number of solutions which form a $(n-m)$-dimensional vector space. Therefore, according to Hadamard's definition, the problem of solving this linear system of equations is ill-posed in both cases.

The most common approach to alleviating the ill-posedness is to redefine the solution in the least-squares sense. Under this new definition, the solution exists and is uniquely determined by [50]

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} \tag{2.1}$$

where $\mathbf{A}^{+}$ is the so-called pseudoinverse matix of $\mathbf{A}$ defined as

$$\mathbf{A}^{+} = \begin{cases} (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H, & \text{if } m > n; \\ \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}, & \text{if } m < n. \end{cases} \tag{2.2}$$

However, when the matrix $\mathbf{A}^H\mathbf{A}$ or $\mathbf{A}\mathbf{A}^H$ is ill-conditioned or singular, the least squares solution (2.1) will be either very sensitive to noise or not uniquely

determined and the problem becomes ill-posed again. In this case, further regularization is necessary to obtain a unique stable solution. The most popular way to handle this ill-posedness is to use the so-called *Tikhnov regularization method* which redefines the solution $\mathbf{x}$ as the one minimizing the following functional [58, 72]:

$$F(\mu, \mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \mu\|\mathbf{C}\mathbf{x}\|^2 \tag{2.3}$$

where μ is a Lagarangian multiplier, here called *regularization parameter*, and $\mathbf{C}$ is a constraint operator to restrict the solution to a special class. One can find that the $\mathbf{x}$ minimizing $F(\mu, \mathbf{x})$ is given by

$$\mathbf{x} = \begin{cases} (\mathbf{A}^H\mathbf{A} + \mu\mathbf{C}^H\mathbf{C})^{-1}\mathbf{A}^H\mathbf{b}, & \text{if } m > n; \\ \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \mu\mathbf{C}\mathbf{C}^H)^{-1}\mathbf{b}, & \text{if } m < n. \end{cases} \tag{2.4}$$

Clearly, by properly choosing the constraint operator $\mathbf{C}$ and the regularization parameter μ, $(\mathbf{A}^H\mathbf{A} + \mu\mathbf{C}^H\mathbf{C})$ and $(\mathbf{A}\mathbf{A}^H + \mu\mathbf{C}\mathbf{C}^H)$ can be made to be well-conditioned so that the solution given in (2.4) is stable. The objective in general in choosing $\mathbf{C}$ is, therefore, on one hand to describe some known properties of the solution and, on the other hand to make the matrix $(\mathbf{A}^H\mathbf{A} + \mu\mathbf{C}^H\mathbf{C})$ or $(\mathbf{A}\mathbf{A}^H + \mu\mathbf{C}\mathbf{C}^H)$ well-conditioned. For image processing applications, such as in image restoration [1], $\mathbf{C}$ is usually chosen to be a smoothing operator. For example, $\mathbf{C}$ is frequently chosen to be a simple diagonal matrix to impose a minimum-norm constraint on $\mathbf{x}$, or a banded finite difference matrix, such as

$$\mathbf{C} = \begin{bmatrix} 1 & -2 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & -2 & 1 \end{bmatrix}$$

to constrain the "smoothness" of $\mathbf{x}$ by minimizing the energy of the second-order difference of $\mathbf{x}$.

For a given $\mathbf{C}$, the regularization parameter μ should be chosen to minimize $F(\mu, \mathbf{x})$. When the following bounded norm constraints are also used

$$\|\mathbf{Ax} - \mathbf{b}\| \leq \epsilon, \qquad \text{and} \qquad \|\mathbf{Cx}\| \leq E.$$

Miller [56] found that the regularization parameter μ should be set to $(\epsilon/E)^2$. This method is sometimes referred to as the *Miller regularization method.*

2.2 Basic Perturbation Theory for Linear Systems

In practice, many problems are formulated and solved as a linear system of equations. Even though it is often easy to find a solution to a linear system, the solution may not be always reliable. Suppose that $\mathbf{A}$ and $\mathbf{b}$ are subject to small perturbations $\delta\mathbf{A}$, the modeling error, and δb, the measurement noise, respectively, and that as a consequence the true solution, $\mathbf{x}$, is in error by the vector $\delta\mathbf{x}$. We have

$$(\mathbf{A} + \delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}.$$

Subtracting $\mathbf{Ax} = \mathbf{b}$ from this, we obtain that

$$\mathbf{A}\delta\mathbf{x} + \delta\mathbf{A}\mathbf{x} + \delta\mathbf{A}\delta\mathbf{x} = \delta\mathbf{b}.$$

Assume that $\mathbf{A}^{-1}$ exists, then

$$\delta\mathbf{x} = \mathbf{A}^{-1}(\delta\mathbf{b} - \delta\mathbf{A}\mathbf{x} - \delta\mathbf{A}\delta\mathbf{x}).$$

It follows that, for any choice of norm,

$$\|\delta\mathbf{x}\| \leq \|\mathbf{A}^{-1}\|\big(\|\delta\mathbf{b}\| + \|\delta\mathbf{A}\|\|\mathbf{x}\| + \|\delta\mathbf{A}\|\|\delta\mathbf{x}\|\big).$$

Since

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|$$

we have

$$(1-\|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|)\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}^{-1}\|\left\{\frac{\|\delta\mathbf{b}\|}{\|\mathbf{x}\|}+\|\delta\mathbf{A}\|\right\} \leq \|\mathbf{A}^{-1}\|\left\{\frac{\|\mathbf{A}\|\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}+\|\delta\mathbf{A}\|\right\}.$$

When $\|\mathbf{A}^{-1}\|\|\delta\mathbf{A}\|$ is less than unity, we have the upper bound for $\|\delta\mathbf{x}\|/\|\mathbf{x}\|$ as

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa}{1-\kappa\|\delta\mathbf{A}\|/\|\mathbf{A}\|}\left\{\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}+\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}\right\} \tag{2.5}$$

where $\kappa = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$ is called the *condition number* of $\mathbf{A}$, which provides a bound to the magnification of the perturbation in $\mathbf{b}$, and approximately to that of the perturbation in $\mathbf{A}$ also. Under the spectral norm [26], which is defined as for any matrix $\mathbf{A}$ (square or rectangular)

$$\|\mathbf{A}\| = \max_{\mathbf{u}\neq\mathbf{0}}\frac{\|\mathbf{A}\mathbf{u}\|}{\|\mathbf{u}\|}, \tag{2.6}$$

the condition number κ can be expressed as

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{2.7}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest singular values of $\mathbf{A}$, respectively. Therefore, when the singular values of the coefficient matrix $\mathbf{A}$ of a linear system are distributed in a wide range, regularization techniques as described in the previous section should be used to obtain a *reliable* solution.

When $\mathbf{A}$ is rectangular matrix, the above derivations hold true for the least squares solution except that $\mathbf{A}^{-1}$ should be replaced with $\mathbf{A}^{+}$. Correspondingly, the condition number becomes $\kappa = \|\mathbf{A}\|\|\mathbf{A}^{+}\|$, and is again given by Eq. (2.7) under the spectral norm if $\lambda_{\min}$ is taken to be the smallest non-zero singular value.

2.3 Asymptotic Behavior of the Fourier Transform

In practice, high frequency data are frequently missing due to limited bandwidth of real systems or other physical constraints. Naturally, we wish these data are insignificant. However, this may not be always so. In general, the behavior of the Fourier transform $s(k)$ of a function $\rho(x)$ is governed by the continuity property of $\rho(x)$ as indicated by the following theorem [61].

Theorem 2.2 *If a spatial support-limited function $\rho(x)$ and all its derivatives of order up to p exist and are of bounded variations, then its Fourier transform $s(k)$ (if exists) tends to zero at least as fast as $1/k^{p+1}$ as $|k| \to \infty$:*

$$s(k) = O\left(\frac{1}{k^{p+1}}\right) \qquad |k| \to \infty$$

This theorem indicates that when $\rho(x)$ has jump discontinuities, $s(k)$ can decay as slowly as $1/k$. Consequently, large amounts of high frequency data are often necessary to recover a discontinuous function. It is the main subject of this thesis to discuss how to reconstruct this type of functions when significant high frequency data are unavailable.

2.4 Truncation Artifact and the Gibbs Phenomenon

There are several consequences of losing high frequency data, such as the loss of sharp edges and the presence of spurious ringing. The purpose of this section is to formally derive and characterize these artifacts.

Consider an arbitrary Ω-support-limited function $\rho(x)$ defined as in Eq. (1.1). Let $\Omega = (-\sigma/2, +\sigma/2)$ and $\rho_\sigma(x)$ be its periodic extension, i.e., $\rho_\sigma(x) = \rho(x)$ for $|x| < \sigma/2$ and $\rho_\sigma(x) = \rho_\sigma(x + n\sigma)$ for $n \in \mathcal{I}$. Then, if the sampling interval Δk equals the fundamental frequency of $\rho_\sigma(x)$, i.e., $\Delta k = 1/\sigma$, the Fourier

series expansion of $\rho_\sigma(x)$ is given by

$$\hat{\rho}_\sigma(x) = \Delta k \sum_{n=-\infty}^{\infty} s(n\Delta k) e^{i2\pi n\Delta k x} \tag{2.8}$$

This series converges uniformly to $\rho_\sigma(x)$ when $\rho_\sigma(x)$ is a continuous function or pointwise to $[\rho_\sigma(x^+) + \rho_\sigma(x^-)]/2$ when $\rho_\sigma(x)$ is piecewise continuous but has only jump discontinuous points [14].

Let $\rho_N(x)$ represent the corresponding truncated series:

$$\rho_N(x) = \Delta k \sum_{n=-N/2}^{N/2-1} s(n\Delta k) e^{i2\pi n\Delta k x} \tag{2.9}$$

using the following trigonometric sum formula:

$$\begin{aligned} H(x) &= \Delta k \sum_{n=-N/2}^{N/2-1} e^{i2\pi n\Delta k x} \\ &= \Delta k \frac{\sin(\pi N \Delta k x)}{\sin(\pi \Delta k x)} e^{-i\pi\Delta k x}, \end{aligned}$$

it is easy to derive that

$$\begin{aligned} \rho_N(x) &= \int_\Omega \hat{\rho}_\sigma(\tau) H(x-\tau) d\tau \\ &= \int_{-\infty}^{+\infty} \rho(\tau) H(x-\tau) d\tau, \end{aligned}$$

or

$$\rho_N(x) = \rho(x) * H(x). \tag{2.10}$$

Therefore, representing a support-limited function $\rho(x)$ by a truncated Fourier series, $\rho_N(x)$, is equivalent to convolving the function with the convolution kernel $H(x)$. Due to the oscillatory and non-local nature of $H(x)$ (Fig. 2.1b), this convolution contributes spurious ringing to $\rho_N(x)$ and merges closely spaced

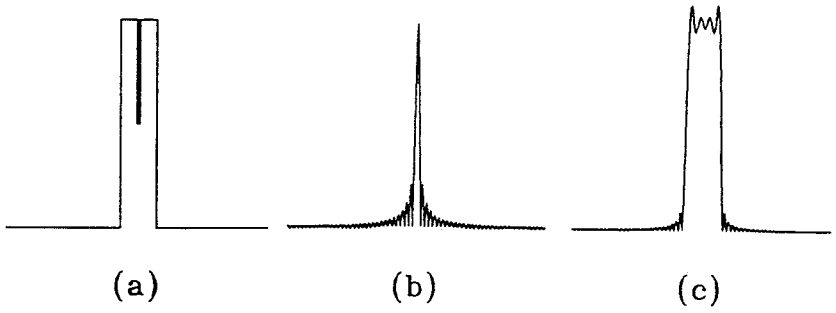


Figure 2.1: The truncation artifact.

features together as illustrated in Fig. 2.1c. This convolution effect is usually referred to as the *truncation artifact.*

The truncation artifact manifests itself differently for functions with different continuity properties. When $\rho(x)$ is a smooth function, $\rho_N(x)$ uniformly converges to $\rho(x)$ as $N \to \infty$ for $x \in \Omega$. However, if $\rho(x)$ contains step discontinuities, the convergence behavior of this limit in the neighborhood of a point of discontinuity will be "anomalous". Without loss of generality, let us assume that $\rho(x)$ has a step discontinuity at $x = x_0$ such that $\rho(x_0^+) \geq \rho(x_0^-)$. Then the maximum overshoot location x_N in $\rho_N(x)$ approaches x_0^+ as $N \to \infty$. It can be proved that [14]

$$\lim_{N\to\infty} \left[\rho_N(x_N) - \rho(x_0^+)\right] \approx 0.08949 \left[\rho(x_0^+) - \rho(x_0^-)\right]$$

Similarly, for the maximum undershoot, we have

$$\lim_{N\to\infty} \left[\rho_N(x_N) - \rho(x_0^-)\right] \approx -0.08949 \left[\rho(x_0^+) - \rho(x_0^-)\right]$$

Therefore, as $N \to \infty$ the value of the maximum overshoot (undershoot) does not tend to zero, but instead tends to a finite limit. The existence of this

finite, non-zero, limiting value of the overshoot (undershoot) is due to the non-uniform convergence of $\rho_N(x)$ to $\rho(x)$ in the vicinity of discontinuous points of $\rho(x)$. This non-uniform behavior of the limit $\rho_N(x) \rightarrow \rho(x)$ as $N \rightarrow \infty$ for $x \in \Omega$ is called the *Gibbs phenomenon.*

In some engineering literature, the terms truncation artifact and Gibbs phenomenon are often used interchangeably. However, there exists a fundamental difference between these two concepts. The former refers to the error (resolution limit and spurious ringing) resulting from the data truncation and the latter characterizes the limiting behavior of this error as $N \rightarrow \infty$. In this thesis, we attempt to distinguish these two, but keep using the "eclectic" term Gibbs ringing for the spurious oscillations

2.5 Resolution, the Uncertainty Principle and Noise

An imaging parameter of considerable interest in virtually all imaging systems is spatial resolution. Qualitatively, spatial resolution can be defined as the smallest separation, δ, of two point-sources necessary for them to remain resolvable in the image. Equivalently, spatial resolution of an imaging system characterizes how accurately the dimension, typically the edge locations, of an object can be determined from the image. In Fourier transform imaging systems, clearly, δ depends on the "width" of the convolution kernel $H(x)$. Usually the width, W_H, is taken to be the duration of the equivalent rectangular pulse approximating $H(x)$ in a cycle, i.e.,

$$W_H = \frac{1}{H(0)} \int_{-\frac{1}{2\Delta k}}^{+\frac{1}{2\Delta k}} H(x)dx = \frac{1}{N\Delta k}.$$

In general, if the data $s(k)$ is weighted with a window function $w(k)$, the corresponding convolution kernel becomes

$$H_w(x) = \Delta k \sum_{n=-N/2}^{N/2-1} w(n\Delta k) e^{i2\pi n\Delta k x}.$$

In this case, we have

$$\begin{aligned} W_H &= \frac{1}{H_w(0)} \int_{-\frac{1}{2\Delta k}}^{+\frac{1}{2\Delta k}} H_w(x)dx \\ &= \frac{\int_{-\frac{1}{2\Delta k}}^{+\frac{1}{2\Delta k}} \sum_{(n=-N/2)}^{(N/2-1)} w(n\Delta k) e^{i2\pi n\Delta k x}\, dx}{\sum_{n=-N/2}^{N/2-1} w(n\Delta k)} \\ &= \frac{1}{\sum_{-N/2 \le n \le N/2-1} \left[w(n\Delta k)/w(0)\right] \Delta k}. \end{aligned}$$

Assuming that $w(n\Delta k)$ are real and $w(n\Delta k) \le w(0)$ [32], we will have

$$W_H \ge \frac{1}{N\Delta k}.$$

Since $\delta \le W_H$, then for any reconstruction algorithms which do not lead to a recovery of some of the missing data, the best spatial resolution is

$$\delta = \frac{1}{N\Delta k}. \tag{2.11}$$

The right-hand side of Eq. (2.11) is often called the Fourier pixel size Δx. Clearly, in order for two point-sources to be distinguishable, their separation has to be larger than the pixel size. Another implication of Eq. (2.11) is that if we are to determine the object edge locations from the image thus obtained, an accuracy of Δx is the best we can attain. Eq. (2.11) also indicates that δ and N can not be reduced simultaneously, namely, we can not improve image resolution and reduce the number of measurements at the same time. This assertion is known as the *uncertainty principle* of the Fourier transform [81].

Measurement noise is also a limiting factor for obtaining high spatial resolution in practical imaging systems. If we assume that the noise term, $\varepsilon(k)$, in Eq. (1.1) is not zero, its contribution to the m^{th} pixel of the image will be $e_m = e(m\Delta x)$ given by

$$e_m = \Delta k \sum_{n=-N/2}^{N/2-1} \varepsilon(n\Delta k) e^{i2\pi nm/N} \tag{2.12}$$

and Eq. (1.2) becomes

$$\hat{\rho}_N(m\Delta x) = \rho_N(m\Delta x) + e_m \tag{2.13}$$

Furthermore, if we assume that $\varepsilon(k)$ is an ergodic, zero-mean, Gaussian white noise with all the sample points $\varepsilon_n = \varepsilon(n\Delta k)$, $\forall n$, being independent, i.e., $E\{\varepsilon_i \varepsilon_j^*\} = 0$ for $i \neq j$, then for the new random variable e_m the mean value is

$$\begin{aligned} \mu_e &= E\{e_m\} \\ &= E\left\{ \Delta k \sum_{n=-N/2}^{N/2-1} \varepsilon(n\Delta k) e^{i2\pi nm/N} \right\} \\ &= N\Delta k \mu_\varepsilon = 0, \end{aligned}$$

and the variance is

$$\begin{aligned} \sigma_e^2 &= E\left\{(e_m - \mu_e)(e_m - \mu_e)^*\right\} \\ &= \Delta k^2 \sum_{n=-N/2}^{N/2-1} \sum_{\ell=-N/2}^{N/2-1} E\{\varepsilon_n \varepsilon_\ell^*\} e^{i2\pi(n-l)m/N} \\ &= \Delta k^2 \sum_{n=-N/2}^{N/2-1} E\{\varepsilon_n \varepsilon_n^*\} \\ &= N\Delta k^2 \sigma_\varepsilon^2. \end{aligned}$$

The averaged signal energy of each pixel is

$$\begin{aligned}\bar{A}^2 &= \frac{1}{N}\sum_{m=-N/2}^{N/2-1}\rho_N(m\Delta x)\rho_N^*(m\Delta x)\\ &= \frac{1}{N}\sum_{m=-N/2}^{N/2-1}\sum_{n=-N/2}^{N/2-1}\sum_{\ell=-N/2}^{N/2-1}s_n s_\ell^* e^{i2\pi(n-\ell)m/N}\\ &= \sum_{n=-N/2}^{N/2-1}|s_n|^2,\end{aligned}$$

where $s_n = s(n\Delta k)$. Hence, the image signal-to-noise ratio per pixel is

$$SNR = \frac{\bar{A}}{\sigma_e} = \frac{\sum_{n=-N/2}^{N/2-1}|s_n|^2}{\sqrt{N}\Delta k\sigma_\varepsilon} = \sqrt{N}\frac{\Delta x}{\sigma_\varepsilon}\sum_{n=-N/2}^{N/2-1}|s_n|^2 \tag{2.14}$$

Since $s_n \sim O(1/n)$ (Theorem 1.1), $\bar{A}$ is a very slowly growing function of N. Thus, if we attempt to improve image resolution by a factor of 2 by increasing N to $2N$, the image signal-to-noise ratio will suffer a loss by approximately a factor of $\sqrt{2}$. Therefore, in a practical implementation, we need to compromise between image resolution and the signal-to-noise ratio. This property is characteristic of TFS (or FFT) reconstruction methods.

Chapter 3

Constrained Reconstruction: A Review

3.1 Introduction

The Fourier series expansion theory was well-established in the nineteenth century. However, its widespread use in engineering was not started until the re-discovery of the FFT algorithm by Cooley and Turkey in 1965. Since then, an enormous amount of research has been devoted to the improvement of the TFS technique to alleviate the truncation artifact — to overcome the resolution limit and to reduce the Gibbs ringing. In the sixties and the early seventies, computational efficiency was a major concern; people had to embrace the FFT algorithm while searching for other remedies. As a result, many filtering (windowing) schemes were proposed bearing this compromise [32]. Common to all these methods is that they treat the unknown data as zero and attempt to modify the shape of the convolution kernel $H(x)$ to reduce the width of its main-lobe and the amplitudes of its side-lobes by properly weighting the available data. However, as implied by the uncertainty principle, no weighting functions of this kind are possible to give resolution better than a pixel. Therefore, these windowing methods are currently used only for suppressing the spurious ringing but at the expense of the already limited resolution.

With the computational constraints being lifted in the eighties, researchers started looking into the possibility of overcoming the resolution limit by utilizing *a priori* knowledge to compensate the missing experimental data. This shift of research emphasis represents a major step forward in attaining super-resolution (resolution better than a pixel, $1/(N\Delta k)$). So far, several promising

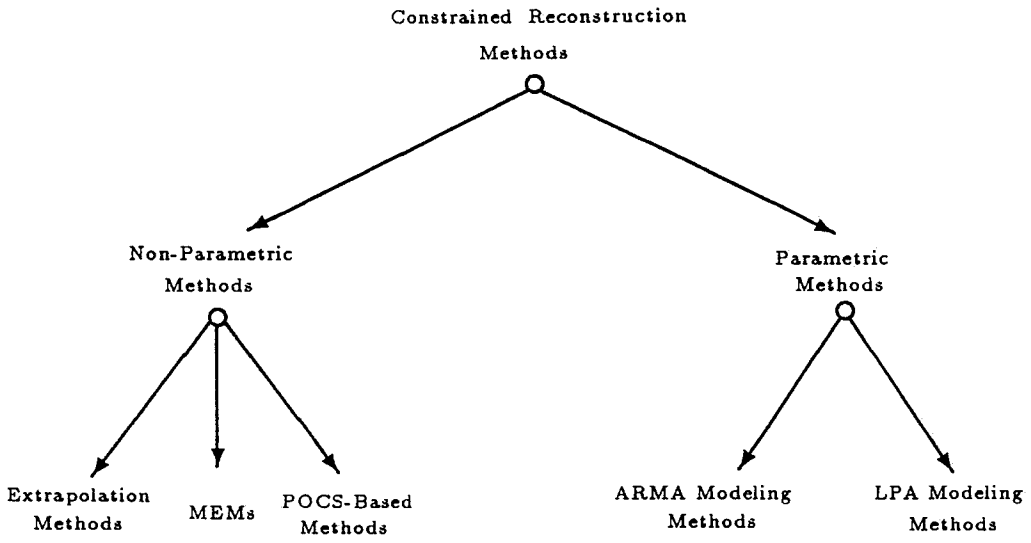


Figure 3.1: Constrained reconstruction methods.

constrained reconstruction methods have resulted and new methods continue to appear (e.g., [18]). Based on what *a priori* constraints are used and how they are used, these techniques can be roughly classified as in Fig. 3.1.

In this chapter, we attempt to give an overview of the non-parametric methods. This class of methods employ the non-parametric type of constraints such as *positivity, smoothness, support-limitedness*, or *noise statistics*. While retaining the Fourier series representation for $\rho(x)$, these methods attempt to alleviate the truncation artifact by recovering some or all of the unmeasured high-frequency Fourier components (terms) in the series. Typical in this class are the support-limited extrapolation methods, the projection-onto-convex-set-based (POCS) methods, and the maximum entropy methods (MEMs). In order to keep our discussion as brief as possible, we will concentrate on their basic ideas, major strengths and weaknesses but ignore their implementation, an important and rich part of these methods.

3.2 Support-Limited Extrapolation Methods

It is well-known that for a support-limited function $\rho(x)$, its Fourier transform $s(k)$ is analytical over the entire real axis and, therefore, can be recovered in principle from any finite segment of it. Based on this fact several celebrated continuous extrapolation algorithms have been developed, such as the prolate-spheroidal-wave-function-decomposition-based methods by Slepian *et al.* [82], and the iterative support-limited extrapolation method developed independently by Papoulis [59] and Gerchberg [23]. These methods have enjoyed great success in restoring optical images from diffraction-limited systems [4, 23, 70]. For the problem we are considering here, however, there are only N sampled data points available. In this discrete case, the analyticity property of $s(k)$ becomes meaningless; nevertheless, the basic idea of making use of the support-bound information is still appropriate.

Let's assume that $\rho(x)$ vanishes off $\Omega = (-\sigma/2, +\sigma/2)$, then we have

$$\rho(x) = \chi_\sigma(x) \sum_{m=-\infty}^{\infty} a(m\Delta k) e^{i2\pi m\Delta k x} \tag{3.1}$$

where

$$\chi_\sigma(x) = \begin{cases} 1, & |x| \le \sigma/2; \\ 0, & \text{otherwise.} \end{cases}$$

Accordingly, its Fourier transform is

$$s(k) = \sum_{m=-\infty}^{\infty} a(m\Delta k) \frac{\sin[\pi\sigma(k - m\Delta k)]}{\pi(k - m\Delta k)} \tag{3.2}$$

This is the well-known *Shannon interpolation formula.* Since $s(k)$ is only known for $k = n\Delta k, n \in \mathcal{N}, \mathcal{N} = \{-N/2, -N/2+1, \cdots, N/2-1\}$, the non-uniqueness of the solution for the a's is evident. Therefore, unlike continuous extrapolation, discrete extrapolation is inherently under-determined. This seemingly

small difference makes the latter much less successful.

If the minimum-norm solution is sought, namely, $\sum_m |a(m\Delta k)|^2$ is minimized, one will find that $a(m\Delta k)$ must satisfy

$$\begin{cases} a(m\Delta k) = 0, & m \notin \mathcal{N}; \\ s(n\Delta k) = \sum_{m=-N/2}^{N/2-1} a(m\Delta k) \dfrac{\sin[\pi\sigma(m-n)\Delta k]}{\pi(m-n)\Delta k}, & n \in \mathcal{N}, \end{cases}$$

which can be put in matrix form as

$$\mathbf{Ea} = \mathbf{s} \tag{3.3}$$

where $\mathbf{E} \in \mathcal{R}^{N\times N}$, $\mathbf{a} \in \mathcal{C}^N$ and $\mathbf{s} \in \mathcal{C}^N$ with

$$\mathbf{a} = \left[a(-N/2\Delta k), \cdots, a((N/2-1)\Delta k)\right]^T$$

$$\mathbf{s} = \left[s(-N/2\Delta k), \cdots, s((N/2-1)\Delta k)\right]^T$$

$$E_{mn} = \frac{\sin[\pi\sigma(m-n)\Delta k]}{\pi(m-n)\Delta k}, \qquad m, n = 1, 2, \cdots, N.$$

$\mathbf{E}$ is usually called the extrapolation matrix. It has N positive eigenvalues such that $1 > \lambda_1 > \lambda_2 > \cdots > \lambda_N > 0$. The first $N\sigma\Delta k$ eigenvalues are very close to 1, and the others gradually decay to zero [83]. Since $\mathbf{E}$ is positive-definite, $\mathbf{a}$ can be uniquely determined from Eq. (3.3). Therefore, the minimum-norm solution for the discrete extrapolation problem can be obtained in the following two steps [37, 71]:

(1) solve Eq. (3.3) for $\mathbf{a}$; and

(2) substitute $\mathbf{a}$ into Eq. (3.1) to get $\rho(x)$ or into Eq. (3.2) to get the extrapolated signal $s(k)$.

An iterative implementation of this two-step procedure is also possible based on the following successive approximation method for solving the linear

system (3.3) due to Van Citter [73]:

$$\begin{cases} \mathbf{a}_0 = 0; \\ \mathbf{a}_{\ell+1} = \mathbf{a}_\ell + \mu(\mathbf{s} - \mathbf{E}\mathbf{a}_\ell), \qquad \ell \geq 0, \end{cases}$$

where $0 < \mu < 2$ controls the convergence rate. Substituting it into Eq. (3.2), we have

$$\begin{cases} s_0(n\Delta k) = 0; \\ s_{\ell+1}(n\Delta k) = s_\ell(n\Delta k) + \mu \sum\limits_{m=-N/2}^{N/2-1} \dfrac{\sin[\pi\sigma(m-n)\Delta k]}{\pi(m-n)\Delta k}[s(m\Delta k) - s_\ell(m\Delta k)]. \end{cases} \tag{3.4}$$

When $\mu = 1$, this procedure corresponds to the discrete Papoulis-Gerchberg iterative scheme, which can also be attained by directly discretizing its continuous counterpart [59].

One point worth noting is that when the data is sampled at the Nyquist rate, i.e., $\Delta k = 1/\sigma$, we have that $\mathbf{E} = \sigma\mathbf{I}$ and $\mathbf{a} = \Delta k\mathbf{s}$ in Eq. (3.3), and thus Eq. (3.1) becomes identical to Eq. (1.2). This indicates that in order for the support-bound to contain useful information for extrapolation, the data have to be sampled at a rate higher than its Nyquist rate, i.e., $\Delta k < 1/\sigma$.

The major difficulty in solving the support-limited extrapolation problem (i.e., reducing the truncation artifact by extrapolating the data under the object-support-bound constraint) lies in the fact that the solution is very sensitive to noise, particularly to broadband noise because the extrapolation matrix $\mathbf{E}$ is usually ill-conditioned. Paradoxically, the smaller the sampling rate Δk with respect to the Nyquist rate $1/\sigma$, the more information the support-bound contains, but the more ill-conditioned the extrapolation matrix $\mathbf{E}$ will be. The usual way to cope with this ill-conditioned problem is to terminate recursion (3.4) before the noise gets magnified or to regularize Eq. (3.3) by

perturbing $\mathbf{E}$ with a diagonal matrix $\mu\mathbf{I}$. In this case, Eq. (3.3) becomes

$$(\mathbf{E} + \mu\mathbf{I})\mathbf{a} = \mathbf{s} \tag{3.5}$$

However, in practice, the determination of when to terminate the iterative process or how to select the regularization parameter μ is by and large non-trivial and often demands the knowledge of the data noise statistics.

In summary, when the object function support-bound is known and the data happen to be over-sampled, it is possible to recover some of the unmeasured high-frequency data. One of the main advantages of the above extrapolation methods is their computational efficiency. For example, if the iterative scheme is used, the discrete convolution (which is the major computation step) can be evaluated using an FFT algorithm; on the other hand, if the non-iterative method is preferred, the Toeplitz system (3.3) can be solved efficiently using the Levinson algorithm [8, 11, 44]. Unfortunately, these methods suffer a serious drawback: They require extremely high data signal-to-noise ratio to assure a reasonable gain in resolution, which makes them unpractical.

One may notice that the same subject has also been discussed in the context of constrained deconvolution/restoration [6, 7, 64, 73]. It is not surprising that the same conclusion has been reached in terms of either extrapolation or deconvolution.

3.3 POCS-Based Methods

The super-sensitivity of the support-limited extrapolation methods to noise reveals that the support-bound constraint alone is too weak to compensate the missing data, especially in the discrete case. With more valid constraints being used, we can expect that the reconstruction will accordingly improve. The

mechanism of projection onto convex sets (POCS) was introduced to the area of image reconstruction to meet the need of handling additional constraints. The principle of POCS has been discussed in great detail in the early literature, for example, in [74, 75, 90, 100, 101]. We review here only the central ideas of POCS. First, let's introduce the following concepts regarding convex sets and their projection operators.

Definition 3-1. A subset ω in the Hilbert space $\mathcal{H}$ is said to be convex if for any elements x_1 and x_2 in ω, it also contains $\mu x_1 + (1-\mu)x_2$ for all μ, $0 \leq \mu \leq 1$.

Definition 3-2. For any $x \in \mathcal{H}$, the projection $P_\omega x$ of x onto ω is the element in ω closest to x. If ω is closed and convex, P_ω exists and is uniquely determined by x and ω from the following minimality criterion

$$\|x - P_\omega x\| = \min_{y \in \omega} \|x - y\|.$$

In brief, the method of POCS is simply an iterative algorithm that finds a solution subject to a number of convex-type constraints by alternating projections. The following theorem is central to this technique which assures the convergence of such an iterative process [90, 101].

Theorem 3-1. *Given m closed convex sets $\mathcal{C}_i$, $i = 1, 2, \cdots, m$, in $\mathcal{H}$ and their corresponding projection operators P_i, if $\mathcal{C}_0 = \bigcap_{i=1}^{m} \mathcal{C}_i$ is non-empty, the sequence generated by*

$$f_{\ell+1} = P_m P_{m-1} \ldots P_1 f_\ell, \qquad \ell = 0, 1, \cdots$$

converges (weakly) to an element $f \in \mathcal{C}_0$ for any initial value $f_0 \in \mathcal{H}$.

Therefore, for a problem with m pieces of *a priori* constraints of which each restricts the solution to a convex set in the Hilbert space $\mathcal{H}$, POCS is an ideal

formalism for finding a solution. Various type of convex constraints exist for the reconstruction problem addressed. For example, the following constraints are of convex-type.

1. *Data consistency constraint*:
   $C_1 = \left\{\rho(x)\text{: } \mathcal{F}\{\rho(x)\} = s(k), \text{ for } k = n\Delta k, \forall n\right\}$

2. *Limited-support constraint*:
   $C_2 = \left\{\rho(x)\text{: } \rho(x) = 0, \text{ for } x \notin \Omega\right\}$

3. *Bounded-magnitude constraint*:
   $C_3 = \left\{\rho(x)\text{: } |\rho(x)| \leq B\right\}$

4. *Phase constraint*:
   $C_4 = \left\{\rho(x)\text{: } \arg\{\rho(x)\} = \theta(x)\right\}$

5. *Bounded-noise variance constraint*:
   $C_5 = \left\{\rho(x)\text{: } \|s - \mathcal{F}\{\rho(x)\}\|^2 \leq \sigma_\varepsilon^2\right\}$

Some of the above constraints have been applied to limited-data image reconstruction problems with some success [74, 75]. With regard to constrained extrapolation, in particular, POCS offers two important advantages over the support-limited extrapolation methods (1) it enables any number of convex-type constraints to be incorporated in the algorithm; and (2) it guarantees convergence: weak in general but strong in finite-dimensional spaces. These desirable properties have encouraged much research into this method in recent years. It can be expected that POCS will play an important role in eventually solving the limited-data reconstruction problem as well as many others. However, POCS is not without limitations. Three of the most serious limitations are: First, it can handle only convex-type constraints, which prevents some effective but non-convex-type constraints to be used; Second, POCS can

be computationally expensive, since convex set projection operators are non-linear in nature and sometimes involve constrained non-linear optimization; for example, P_{C_5} for the convex set C_5 defined above requires solving a quadratic programming step; Third, POCS may converge very slowly and also *not necessarily to a unique solution* when the projection operators are not a contractive mapping; therefore, POCS can also converge to a "bad" solution. Thus far, POCS has not yet been very effective for overcoming the truncation artifact because practical, effective convex-type constraints have not been "identified".

3.4 Maximum Entropy Methods

One of the major concerns in superresolution reconstruction is image fidelity. Since the superresolution reconstruction is inherently under-determined, it heavily relies on the use of *a priori* constraints to alleviate this ill-posedness. However, it is more often than not that the data and the *a priori* constraints are still not sufficient to ensure a unique solution. In this case, the minimum energy constraint (the least squares principle) is frequently used in practice to select a solution from many "feasible" ones. Unfortunately, the minimum energy constraint suffers seriously from being "blind" to false image structures, especially to spurious ringing. In recent years, the maximum entropy principle has been invoked and advocated by many noted researchers for incomplete data image reconstruction to overcome this shortcoming [12, 27, 28, 38, 80].

The maximum entropy principle has its origin in information and communication theory largely due to Shannon. The background to this principle is well covered in the literature [22, 38]. Although there has been some dispute about the correct definition of entropy when the maximum entropy principle is applied to other areas, it appears that in imaging applications it has been finally come to agree that the correct form of "configurational" entropy of a

positive image is in the Shannon form [80]:

$$H_S = -\sum_{i\in\mathcal{M}} \rho_i \log \rho_i$$

where $\mathcal{M} = \{-M/2, -M/2+1, \cdots, M/2-1\}$ with M being the number of elements (pixels) necessary to characterize $\rho(x)$. Accordingly, the data constraint becomes

$$s_n = \Delta x \sum_{m\in\mathcal{M}} \rho_m e^{-i2\pi mn/N} + \varepsilon_n, \qquad n \in \mathcal{N}$$

Assume that the ε_n's are Gaussian random error with zero mean and standard deviation σ_ε, then the following chi-squared statistic test can be used to measure the misfit to the data

$$\chi^2 = \frac{1}{\sigma_\varepsilon^2} \sum_{n\in\mathcal{N}} \left| s_n - \Delta x \sum_{m\in\mathcal{M}} \rho_m e^{-i2\pi mn/M} \right|^2.$$

Therefore, the maximum entropy estimation of ρ_m is obtained by maximizing H_S subject to the constraint $\chi^2 = N$. Using the Lagrange multiplier method, we maximize the functional

$$f(\lambda, P) = -\sum_{i\in\mathcal{M}} P_i \log P_i - \lambda\chi^2.$$

In the absence of any data constraints, the maximization of H_S results in all the ρ_i being equal to e^{-1}. This fact seems to suggest (and many of us believe) that the maximum entropy principle selects the most uniform (or the least structural) reconstruction consistent with the data. This point has been demonstrated by examples from the application of the maximum entropy methods to NMR spectroscopic data analysis [35, 76, 77] and radio astronomy image reconstruction [27] as well as many others [20].

When an *a priori* guess (estimate) ρ_i^0 of ρ_i is available, this constraint can be incorporated into the entropy expression as

$$H_C = -\sum_{i \in \mathcal{M}} \rho_i \log \frac{\rho_i}{\rho_i^0}$$

This is the so-called cross-entropy method [85, 86]. Intuitively, maximizing H_C enforces the structure of ρ_i^0 on ρ_i while keeping ρ_i as smooth as the data permit. Therefore, in case the rough structure of ρ_i is known, this entropy measure will be more effective.

Another form of entropy definition due to Burg has also received great attention in the last decade, which expresses the entropy as [38, 60]

$$H_B = \int_\Omega \log \rho(x) dx$$

Maximizing this entropy under the data constraint that

$$s(n\Delta k) = \int_\Omega \rho(x) e^{-i2\pi n \Delta k x} dx, \qquad |n| \leq N/2.$$

leads to a closed-form solution for $\rho(x)$ as

$$\rho(x) = c_0 \Big/ \left| 1 + \sum_{m=1}^{N/2} c_m e^{i2\pi m \Delta k x} \right|^2,$$

where the coefficients are determined from the following Toeplitz system

$$\sum_{m=0}^{N/2} s[(n-m)\Delta k] c_m = \begin{cases} 1, & n = 0; \\ 0, & n = 1, \ldots, N/2 - 1. \end{cases}$$

The discovery of this elegant solution has amazed many researchers and led to its widespread uses in various applications, particularly in high-resolution spectrum estimation [42, 41, 54, 95]. However, as have been pointed out by several researchers [80], H_B is not an appropriate configurational entropy measure for images, and therefore, the "smoothest" argument does not apply to

H_B. In fact, it has been observed that this method can introduce spurious features of its own, such as spectral line splitting [42, 54].

Recently, the maximum entropy principle has also applied to reconstruct complex object functions. In this case, the following modified Shannon entropy measures are used:

$$H_S = -\sum_{m\in\mathcal{M}} (\rho_m^R \log \rho_m^R + \rho_m^I \log \rho_m^I)$$

when the real ρ_m^R and imaginary ρ_m^I parts of the object function ρ_m can be considered as two positive and independent qualities, such as the case in microwave tomography [57], or

$$H_S = -\sum_{m\in\mathcal{M}} |\rho_m| \log |\rho_m|$$

when we are only interested in the modulus of ρ_m, such as in magnetic resonance imaging [34, 78].

In summary, the main strength of the maximum entropy principle lies in its ability to select the smoothest image consistent with the data and other *a priori* constraints, and therefore, false features (resulting from data truncation and measurement noise) usually associated with the minimum energy constraint can be reduced or eliminated. However, the maximum entropy constraint itself does not necessarily contain any "built-in" *a priori* information about a problem (unless the positivity constraint is valid), therefore, the plain maximum entropy methods do not promise superresolution in general. The other weakness of the maximum entropy methods seems due to their heavy computational burden. Even though several attempts have made to relieve this [20, 79, 104], efficient and practical algorithms remain to be found.

Chapter 4

Parametric Image Reconstruction

4.1 Parametric Image Reconstruction: Rationale

In the previous chapter, we have pointed out that the non-parametric methods are common in that they all express the object function in terms of truncated Fourier series even though with a considerably larger number of terms than the number of data points available; and then *a priori* constraints are utilized to determine these newly introduced terms. This approach has the advantage of being straightforward and in some cases being computationally efficient. However, there are a number of disadvantages with this approach. One of them is the inefficiency in extracting information from the limited amount of available experimental data; with the Fourier series representation, usually a large number of terms are required to reconstruct even a very simple object function (e.g., a box-car function); therefore, many high frequency terms need to be recovered to attain some noticeable improvement in image quality. The second disadvantage can be attributed to the ineffectiveness in utilizing *a priori* information; in order to attain supersolution, we naturally hope to be able to use as much valid *a priori* information as possible. However, with the Fourier series framework, some *a priori* information are very difficult to incorporate into the reconstruction scheme.

For example, if an object function $\rho(x)$ is known to be a δ-function, i.e., $\rho(x) = A\delta(x-x_0)$, with an unknown amplitude A and location x_0. In principle, two parameters are sufficient to characterize the object function if a δ-function model is used to represent the object function; correspondingly, two data points

are enough to determine the unknowns. However, with the Fourier series representation, we need the entire series to represent the δ-function exactly and, therefore, an infinite number of data points are necessary to reconstruct the object function. Worse than this is that there is no obvious and easy way for the Fourier series framework to make use of the known δ-function structure of the object function to help to recover the missing terms in the series.

Therefore, when some physical properties of an object function or the signal are known *a priori*, it usually becomes more effective to constrain the reconstruction by directly incorporating these known properties into a parametric model for the object function or the signal. Clearly, if an appropriate model can be built with the number of model parameters less than the number of data points, the parameters may be exactly determined and thus the data truncation artifact can be eliminated. Hence, if we consider the non-parametric methods as a means to alleviate the truncation artifact by recovering some high-frequency components, the parametric methods attempt to completely remove this artifact by recovering the whole series; namely, the parametric methods implicitly extrapolate the known data to the infinite range. Another desirable advantage of the parametric methods is their ability for noise removal. Particularly, if the number of parameters characterizing the model is less than the number of data points, a least squares parameter estimation procedure can effectively "smooth" out some of the random noise in the measured data.

4.2 Principle of Parametric Modeling

There are a variety of ways to parametrize an object function or the corresponding signal depending on the assumptions made. One usual and particularly advantageous approach is to represent the object function $\rho(x)$ in terms

of a generalized series as

$$\rho(x) = \sum_{\ell} c_\ell \varphi_\ell(\theta, x), \tag{4.1}$$

where $\{\varphi_\ell(\theta, x)\}$ is the parametrized basis function set which are selected to incorporate and emphasize the physical properties of the class of object functions to be found; the parameter vector θ is included to make the basis functions adaptive so that the model is optimized for representing each individual object function. For instance, in the above described example, $\varphi_\ell(\theta, x)$ can be chosen to be a δ-function with θ parametrizing the location of the δ-functions so that any δ-function with arbitrary location can be accurately modeled.

Similar to model (4.1), one can also build a parametric model for the signal if *a priori* information is available in the Fourier domain or if it leads to a "simpler" model

$$s(k) = \sum_{\ell} c_\ell \Phi_\ell(\theta, k), \tag{4.2}$$

where

$$\Phi_\ell(\theta, k) = \int_{-\infty}^{+\infty} \varphi_\ell(\theta, x) e^{-i2\pi\Delta kx} dx, \qquad \forall \ell.$$

When the basis function $\varphi_\ell(\theta, x)$ is a linear function of the θ_ℓ's, model (4.1) is called a linear model; otherwise, it is referred to as a nonlinear model. The same definition applies to model (4.2).

The key step in setting up model (4.1) or model (4.2) lies in selecting the basis functions $\varphi_\ell(\theta, x)$ or $\Phi_\ell(\theta, k)$. Traditionally, for representing smooth functions, $\varphi_\ell(\theta, x)$ is chosen to be the eigenfunctions of a *Sturm-Liouville* system. For example, the Fourier series selects $\varphi_\ell(\theta, x) = e^{i2\pi\ell\Delta kx}$. This basis function set has the widely enjoyed advantage of being general for representing periodic smooth functions (or periodically-extended spatial-support-limited functions), however suffers from the same property (requirement of periodicity and smoothness) for limited data image reconstruction since we need a set of

specialized basis functions to handle *a priori* constraints as well as discontinuities. Considering that the general principles governing the selection of $\varphi_\ell(\theta, x)$ or the selection of $\Phi_\ell(\theta, k)$ is the same, in the following, we will concentrate our discussions on only the first case, in which the general principles can be summaried as

- *Sufficiency*: the set of basis functions should be complete in the class of all acceptable object functions;

- *Efficiency*: a linear combination of the basis functions can accurately approximate the object functions with only a finite number of terms;

- *Robustness*: the model parameters should be stable with respect to perturbations of reasonable levels;

- *Computability*: the model parameters can be found in a reasonable amount of time and cost with the available computing technology.

It is quite possible that no single set of basis functions are optimal under all these criteria, but they do provide a guideline for establishing an optimal model under some reasonable tradeoffs. Usually, the sufficiency is not a problem since there exits many known basis-function sets which are complete in representing spatial-support-limited object functions, such as the set of prolate spheroidal wave functions [82]. The problem arises in practice because we are only given a finite number of freedoms in selecting the basis functions due to the finiteness of the available experimental data, namely, we are limited to constructing model (4.1) in terms of only a finite number of basis functions. Since the object function space is of infinite dimension, with a finite number of basis functions the best that we can ask for such a model is that any object function be accurately approximated. Given a fixed number of basis functions, some sets provide better approximations than the others. For

example, we expect a piecewise-linear function to furnish a better approximation to a smooth function than a piecewise-constant one. Therefore, the art of parametric modeling lies in selecting the basis functions which can represent the object functions parsimoniously with respect to the number of terms (basis functions) used so that the approximation error is minimized. In this regard, *a priori* knowledge of a particular problem as well as users' ingenuity is important.

Robustness of a parametric model is important since in practice we always have to deal with measurement random noise and systematic modeling errors. Ideally, the model should be stable so that the model parameters are insensitive to perturbations in the assumed working environments. However, when building a model, particularly, a nonlinear one, it is extremely difficult to design the model with a pre-determined parameter sensitivity. Usually, we may want to select a set of orthogonal basis functions with certain desired properties; but orthogonality alone does not ensure a robust model. Therefore, building an appropriate, robust, and useful model often requires a heuristic approach and lots of testing.

The computational complexity of a model is also of great importance in practice since we eventually need to find all the model parameters. It is desirable that a model is computationally as simple as possible so that the model parameters can be found efficiently with the available computing technology. Since effective *a priori* constraints often require highly non-linear models, we frequently have to make some compromise in performance to gain computational efficiency. For example, for a given model, we usually prefer the L_2-norm to the other norms, such as the L_1-norm, to measure the goodness of fit in estimating the model parameters. Also, linearization techniques are often used in practice to simplify highly non-linear models.

4.3 Existing Parametric Methods: A Brief Review

Even though parametric tomographic image reconstruction is still in its infancy, the parametric modeling concept is not new in solving inverse problems. As a matter of fact, parametric methods have been successfully applied to various areas of data processing, such as time-series analysis, direction-of-arrival estimation of radar echo signals, and geophysical data processing. And recently, great effort has been made to apply some of these concepts to the area of image reconstruction. For example, in [31], Hanson *et al.* recognized that by properly choosing the basis functions $\varphi_\ell(\theta, x)$ in (4.1), many *a priori* features in the image domain can be incorporated into the model and, thus, some degree of improvement over the image quality can be achieved. Unfortunately, their models have been limited to the cases of known θ, i.e., the linear models, and consequently lack the desirable capability for adaptive modeling.

Independently, Smith *el al.* [84] have investigated another class of models, called the *autoregressive moving average* models (ARMA), for magnetic resonance image reconstruction. These models, having their origin in high-resolution spectrum estimation from short data records [13, 54], attempt to approximate an object function with a rational function as

$$\rho(x) = \frac{\sum_{m=0}^{p} a_m e^{i2\pi m \Delta k x}}{\sum_{n=0}^{q} b_n e^{i2\pi n \Delta k x}}, \tag{4.3}$$

or in terms of the generalized series as

$$\rho(x) = \sum_{m=0}^{p} a_m \varphi_m(\theta, x) \tag{4.4}$$

where the basis function is given by

$$\varphi_m(\theta, x) = e^{i2\pi m \Delta k x} \Big/ \sum_{n=0}^{q} b_n e^{i2\pi n \Delta k x}, \tag{4.5}$$

with $\theta = [b_0, b_1, \ldots, b_q]^T$. This set of basis functions has several appealing feartures. First, the nonlinearity of the basis functions can be nicely handled in the linear prediction framework [13, 54, 84]. Second, when an object function contains δ-function like behavior, it will take a lot of high frequency terms to represent it accurately with the Fourier series. But with the ARMA model this feature can be easily modeled by the behavior of the basis functions at their singular points [the zeros of the denominator in the right hand side of Eq. (4.5)]. Because of this capability, this model has proved very effective for reconstructing spiky type of object functions. The major weakness of the ARMA model lies in modeling a general piecewise continuous object function since frequently a large value of q is required for the basis functions to produce sharp transitions. In addition, a large value of q can mean instability of the model, which can result in spurious spikes in the reconstructions as observed in [84] when the model order was over-determined.

4.4 An Improved Parametric Method: The LPA Model

In the following, we describe a new parametric model, called the *localized polynomial approximation* model (LPA) to overcome the ineffectiveness with the linear models and the ARMA model for reconstructing piecewise continuous object functions.

4.4.1 Basic Idea and Definition

The basic idea underlying the LPA model stems from the following considerations: 1) *necessity of localized modeling*: discontinuities of an object function, $\rho(x)$, are the source of the Gibbs ringing; it is, therefore, crucial for a model to be able to "efficiently" characterize such behavior so that when the model is fit to the data, information embedded in the data will be best utilized for its

recovery; and 2) *optimality of localized polynomial approximation*: for a piecewise continuous function which is sufficiently smooth in each local region, it can be expanded in terms of local Taylor series. More importantly, such expansion can lead to a smaller truncation error than a localized Fourier series expansion does, since it does not impose periodicity on the function which can in turn create discontinuities at the ends of the local intervals and thus Gibbs ringing as the Fourier series expansion does.

Based on these considerations, an LPA model adopts an adaptive localized modeling approach to approximate an object function. Specifically, it first assumes that $\rho(x)$ is Ω-spatial-support-limited, i.e., $\rho(x) = 0$, for $x \notin \Omega$. Then it adaptively divides the support into M local regions, $(\alpha_\ell, \alpha_{\ell+1})$, for $\ell = 1, 2, \ldots, M$, and in each of these regions it approximates $\rho(x)$ with a polynomial (truncated Taylor series) of a certain order.

Equivalently, the LPA model expresses $\rho(x)$ as

$$\rho(x) = \sum_{\ell=1}^{M} \left[\sum_{r=0}^{P_\ell} c_{\ell r} (x - \beta_\ell)^r \right] W_\ell(x). \tag{4.6}$$

where

$$W_\ell(x) = \begin{cases} 1, & \text{for } \alpha_\ell < x < \alpha_{\ell+1} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\beta_\ell = (\alpha_\ell + \alpha_{\ell+1})/2, \qquad 1 \le \ell \le M$$

which is the center of the ℓ^{th} interval. Thus, the independent parameters of an LPA model include

M:	the number of local regions;
α_ℓ's:	the edge locations determining each local region;
P_ℓ's:	the local polynomial orders; and
$c_{\ell r}$'s:	the local polynomial coefficients.

If the maximum local polynomial order allowed in the model is L, namely, $\max\{P_1, P_2, \ldots, P_M\} = L$, we may then call the model the L^{th}-order LPA model.

4.4.2 Characteristics

Comparing Eq. (4.6) with Eq. (4.1), one can readily recognize that an LPA model chooses the following basis functions

$$\varphi_{\ell r}(\theta, x) = (x - \beta_\ell)^r W_\ell(x), \qquad \forall \ell \forall r$$

such that

$$\rho(x) = \sum_{\ell=1}^{M} \sum_{r=1}^{P_\ell} c_{\ell r} \varphi_{\ell r}(\theta, x).$$

From the definition of $W_\ell(x)$, it is clear that each of the basis functions, $\varphi_{\ell r}(\theta, x)$, is a local function, i.e., the support of $\varphi_{\ell r}(\theta, x)$, $(\alpha_\ell, \alpha_{\ell+1})$, is a local interval of the support, Ω, of the object function $\rho(x)$. The locality of this set of basis functions have several important implications for modeling an object function given a limited number of data points.

(1) *Capability for Handling Piecewise Continuous Behavior*: The capability to model piecewise continuous behavior of an object function is an immediate result of using localized basis functions. Hence, the difficulty in modeling discontinuities with non-local basis functions, such as the set of complex exponentials used by the Fourier series, can be easily overcome. Especially, the LPA model parametrizes the edge locations into the basis function which enables the model to pick up discontinuities at arbitrary locations and, thus, to model the piecewise continuous behavior adaptively.

(2) *Localized Modeling Error*: Another advantage of using localized basis functions comes from the fact that when the model is fit to the data, the

object function will be approximated separately in different intervals by the corresponding localized basis functions; therefore, approximation error from one interval will not propagate to the other intervals through the basis functions. This is a powerful advantage over the non-local basis function approach, since it makes it more straightforward to interpret the reconstructions.

(3) *Optimized Modeling*: As pointed out in the last section, given a limited number of data points, it is paramountly important for a basis function set to be able to model an object function using only a small number of terms. The use of localized basis functions is an important step for achieving this since it is usually very easy to model the local behavior of an object function with a few parameters provided the local intervals are properly chosen. For example, assume that the object function $\rho(x)$ contains jump discontinuities, the convergence rate of the Fourier series can be as slow as $1/N$ with N being the number of terms used. However, using the LPA modeling approach with the local regions being chosen based on the distributions of the discontinuities, it is possible to achieve a local convergence of $1/(1 \times 2 \times \cdots \times P_\ell)$ with P_ℓ being the order of the local polynomial fitted. Therefore, it is potentially more efficient to use an LPA model to characterize a piecewise continuous object function than the Fourier series.

Chapter 5

Parameter Estimation of The LPA Model

5.1 Introduction

Assume that the signal $s(k)$ is generated by an L^{th}-order LPA model, it can then be expressed as

$$s(k) = \sum_{\ell=1}^{M} \sum_{r=0}^{P_\ell} c_{\ell r} \Phi_{\ell r}(k), \tag{5.1}$$

where $\max\{P_1, P_2, \ldots, P_M\} = L$, and the basis function $\Phi_{\ell r}(k)$ is given by

$$\Phi_{\ell r}(k) = \int_{\alpha_\ell}^{\alpha_{\ell+1}} (x - \beta_\ell)^r e^{-i2\pi kx} dx, \tag{5.2}$$

which can be evaluated by the following recursion

$$\begin{cases} \Phi_{\ell 0}(k) = \dfrac{\sin(2\pi\Delta_\ell k)}{\pi k} e^{-i2\pi\beta_\ell k}, \\ \Phi_{\ell r}(k) = \dfrac{\Delta_\ell^r}{\pi k} \sin\left[(2\Delta_\ell k + r/2)\pi\right] e^{-i\pi(2\beta_\ell k - r/2)} + \dfrac{r}{i2\pi k} \Phi_{\ell(r-1)}(k) \end{cases} \tag{5.3}$$

where $\Delta_\ell = (\alpha_{\ell+1} - \alpha_\ell)/2$ and $\beta_\ell = (\alpha_\ell + \alpha_{\ell+1})/2$.

Given N sampled points of $s(k)$ at $k = n\Delta k$, for $n \in \mathcal{N}$, we need to find all the model parameters including P_ℓ, α_ℓ, and $c_{\ell r}$, for $\ell = 1, 2, \cdots, M$, and $r = 1, 2, \cdots, P_\ell$. If the α_ℓ's are known, the P_ℓ's can be determined accordingly from the α_ℓ's, and the $c_{\ell r}$'s can be found from Eq. (5.1) by solving the linear system. However, in practice, precise information of the α_ℓ's is seldom available, and it is the freedom to allow them to vary that gives the model the desirable adaptivity to pick up sharp edges at arbitrary locations and thus the capability

to reduce the Gibbs ringing resulting from them. In this case, solving the model parameters becomes a highly non-linear problem. Due to the nature of the model and the problem addressed, we can reasonably decouple the determination of the α_ℓ's, the P_ℓ's, and the $c_{\ell r}$'s. Therefore, the LPA model parameter estimation problem can be divided into the following three sub-problems:

(1) Finding the α_ℓ's in a weighted least squares sense, i.e.,

$$\min_{\alpha'_\ell s} \sum_{n\in\mathcal{N}} \left| \left[s(n\Delta k) - \sum_{\ell=1}^{M} \sum_{r=0}^{P_\ell} c_{\ell r} \Phi_{\ell r}(n\Delta k) \right] w(n\Delta k) \right|^2, \tag{5.4}$$

where $w(k)$ is a weighting function used to emphasize the high frequency components so that the "smoothing effect" of the least squares norm can be reduced.

(2) Determination of the P_ℓ's based on the α_ℓ's.

(3) Solving the $c_{\ell r}$'s from Eq. (5.1) based on the α_ℓ's and the P_ℓ's.

5.2 Theoretical Developments

Clearly, the key step in estimating the LPA model parameters is to find the α_ℓ's. Since $s(k)$ is a highly non-linear function of the α_ℓ's, it is computationally very difficult to solve the weighted least squares problem (5.4) without a good initial estimate [19, 21]. In the following, we give several propositions to simplify the problem so that the linear-prediction framework [52] can be used. Implementation of the algorithm is discussed in the next section.

Proposition 5-1: *By taking the $(L+1)^{th}$ derivative an L^{th}-order LPA model can be converted to a generalized δ-function model consisting of a summation of the Dirac δ-functions and their derivatives.*

Proof. This proposition is obvious from the definitions of Dirac δ-function and its derivatives [14].

Even though this proposition appears to be trivial, its usefulness will be seen soon. For the LPA model given in Eq. (5.1), it is easy to see that

$$\rho^{(P)}(x) = \sum_{\ell=1}^{M+1} \sum_{r=0}^{R_\ell} a_{\ell r} \delta^{(r)}(x - \alpha_\ell), \tag{5.5}$$

where $P = L + 1$, the superscripts $^{(P)}$ and $^{(r)}$ denote the P^{th}- and r^{th}-order derivative operators, respectively, and the amplitude parameter $a_{\ell r}$ is a linear function of the $c_{\ell r}$'s. The value of R_ℓ is dependent on the continuity property of $\rho(x)$. If $x = \alpha_\ell$ is a step discontinuous point (the zero-order edge point), $R_\ell = P - 1$; otherwise, if $\rho(x)$ is continuous up to but not including the J^{th}-order derivative at this point (the J^{th}-order edge point), $R_\ell = P - J - 1$ (obviously in this case, $P \geq 1$ and $J \leq P - 1$ hold).

Based on Eq. (5.5) and the differentiation property of the Fourier transform, we have

$$(i2\pi k)^P s(k) = \sum_{\ell=1}^{M+1} \left[\sum_{r=0}^{R_\ell} a_{\ell r} (i2\pi k)^r \right] e^{-i2\pi\alpha_\ell k}. \tag{5.6}$$

For simplicity, let $\tilde{s}(k) = (i2\pi k)^P s(k)$, $\tilde{a}_{\ell r} = (i2\pi\Delta k)^r a_{\ell r}$, $\omega_\ell = 2\pi\alpha\Delta k$, $k_n = (n - N/2)\Delta k$, $M_1 = M + 1$ and use $\tilde{s}_n$ to denote $\tilde{s}(k_n)$, Eq. (5.6) can then be rewritten as

$$\tilde{s}_n = \sum_{\ell=1}^{M_1} \left[\sum_{r=0}^{R_\ell} \tilde{a}_{\ell r} n^r \right] e^{-i\omega_\ell n}, \qquad n = 0, 1, \cdots, N-1. \tag{5.7}$$

This results indicates that by simply filtering (weighting) the measured signal s_n with a high-pass filter $(i2\pi k_n)^P$, the new signal thus obtained $\tilde{s}_n$ is a summation of polynomial-weighted complex exponentials. This special structure of $\tilde{s}_n$ leads to the following useful proposition.

Proposition 5-2: *Let* $\mathbf{A}$ *be the following Hankel matrix, usually called the backward linear prediction data matrix* [45],

$$\mathbf{A} = \begin{bmatrix} \tilde{s}_0 & \tilde{s}_1 & \cdots & \tilde{s}_q \\ \tilde{s}_1 & \tilde{s}_2 & \cdots & \tilde{s}_{q+1} \\ \vdots & \vdots & & \vdots \\ \tilde{s}_{N-q-1} & \tilde{s}_{N-q} & \cdots & \tilde{s}_{N-1} \end{bmatrix}. \tag{5.8}$$

If $N - q \geq \sum_{\ell=1}^{M_1} (R_\ell + 1)$, *the row space of* $\mathbf{A}$, *denoted as* $\mathcal{R}(\mathbf{A})$, *is identical to the spanning space of the following vector set* Ψ, i.e., $\mathcal{R}(\mathbf{A}) = \text{span}(\Psi)$,

$$\Psi = \left\{ \begin{array}{llllll} (1, & e^{-i\omega_\ell}, & e^{-i\omega_\ell 2}, & \cdots, & e^{-i\omega_\ell q}), \\ (0, & e^{-i\omega_\ell}, & 2e^{-i\omega_\ell 2}, & \cdots, & qe^{-i\omega_\ell q}), \\ & & \cdots\cdots\cdots & & \\ (0, & e^{-i\omega_\ell}, & 2^{R_\ell} e^{-i\omega_\ell 2}, & \cdots, & q^{R_\ell} e^{-i\omega_\ell q}) \end{array} \middle| \ \ell = 1, 2, \cdots, M_1 \right\}.$$

Proof. We will prove an illustrative special case. Let's assume that $M = 1$ and $R_\ell = 1$, $\forall \ell$. To prove $\mathcal{R}(\mathbf{A}) = \text{span}(\Psi)$ it is sufficient to show $\mathcal{R}(\mathbf{A}) \subseteq \text{span}(\Psi)$ and $\text{span}(\Psi) \subseteq \mathcal{R}(\mathbf{A})$. Observe the following decomposition of $\mathbf{A}$

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Psi}$$

where

$$\mathbf{T} = \begin{bmatrix} \tilde{a}_{10} & \tilde{a}_{11} & \tilde{a}_{20} & \tilde{a}_{21} \\ (\tilde{a}_{10} + \tilde{a}_{11})e^{-i\omega_1} & \tilde{a}_{11}e^{-i\omega_1} & (\tilde{a}_{20} + \tilde{a}_{21})e^{-i\omega_2} & \tilde{a}_{21}e^{-i\omega_2} \\ \vdots & \vdots & \vdots & \vdots \\ (\tilde{a}_{10} + p\tilde{a}_{11})e^{-i\omega_1 p} & \tilde{a}_{11}e^{-i\omega_1 p} & (\tilde{a}_{20} + p\tilde{a}_{21})e^{-i\omega_2 p} & \tilde{a}_{21}e^{-i\omega_2 p} \end{bmatrix}$$

with $p = N - q - 1$, and $\boldsymbol{\Psi}$ is the corresponding matrix of the vector set Ψ

$$\mathbf{\Psi} = \begin{bmatrix} 1 & e^{-i\omega_1} & e^{-i\omega_1 2} & \cdots & e^{-i\omega_1 q} \\ 0 & e^{-i\omega_1} & 2e^{-i\omega_1 2} & \cdots & qe^{-i\omega_1 q} \\ 1 & e^{-i\omega_2} & e^{-i\omega_2 2} & \cdots & e^{-i\omega_2 q} \\ 0 & e^{-i\omega_2} & 2e^{-i\omega_2 2} & \cdots & qe^{-i\omega_2 q} \end{bmatrix}.$$

$\mathcal{R}(\mathbf{A}) \subseteq \text{span}(\Psi)$ is obvious from this decomposition, since it implies that each row vector of $\mathbf{A}$ can be expressed as a linear combination of the row vectors of $\mathbf{\Psi}$, i.e., the vectors in Ψ. The relation $\text{span}(\Psi) \subseteq \mathcal{R}(\mathbf{A})$ can be proved similarly by the fact that $\mathbf{T}$ has a full column rank if $N - q \geq 4$ (that $\tilde{a}_{11}$ and $\tilde{a}_{21}$ are not zero is assured by the assumption that $R_\ell = 1$), therefore, the matrix $\mathbf{T}^H\mathbf{T}$ is non-singular, and $\mathbf{\Psi}$ can be expressed as

$$\mathbf{\Psi} = (\mathbf{T}^H\mathbf{T})^{-1}\mathbf{T}^H\mathbf{A} = \tilde{\mathbf{T}}\mathbf{A}.$$

The general case of arbitrary M and R_ℓ can be shown following the same procedure.

Remark 5-1: Matrix $\mathbf{A}$ is at most of rank $R = \sum_{\ell=1}^{M_1} (R_\ell + 1)$.

Proposition 5-3: *If matrix* $\mathbf{A}$ *satisfies the dimension condition:* $q \geq R$ *and* $N - q \geq R$, *then (1) the homogeneous equation* $\mathbf{Ag} = 0$ *has non-trivial solutions; and (2) for any non-zero solution* $\mathbf{g} = [g_0, g_1, \ldots, g_q]^T$ *the polynomial defined by*

$$\mathcal{P}(Z) = g_q Z^q + g_{q-1} Z^{q-1} + \cdots + g_0 \tag{5.9}$$

has zeros at $Z_\ell = e^{-i\omega_\ell}$ *with multiplicity at least of* $R_\ell + 1$, *for* $\ell = 1, 2, \cdots, M_1$.

Proof. The first part is obvious from the fact that if $\mathbf{A}$ satisfies the specified dimension conditions, $\text{rank}(\mathbf{A}) = R < q + 1$, the number of unknowns. The

second part can be proved as follows. Since any non-zero vector $\mathbf{g}$ satisfying $\mathbf{Ag} = 0$ is orthogonal to $\mathcal{R}(\mathbf{A})$, it is also orthogonal to span(Ψ) from Proposition 5-2. Let $\mathbf{v}_{\ell r}$, $\ell = 1, 2, \ldots, M_1$ and $r = 0, 1, \ldots, R_\ell$, denote the vectors in Ψ. Clearly, $\mathbf{g} \perp \mathbf{v}_{\ell r}$, $\forall \ell \forall r$. Then from the definition of Ψ, we have

$$\begin{cases} (\mathbf{g}, \mathbf{v}_{\ell 0}) = 0 & \Rightarrow \quad \mathcal{P}(Z)\Big|_{Z=e^{-i\omega_\ell}} = 0, \\ (\mathbf{g}, \mathbf{v}_{\ell 1}) = 0 & \Rightarrow \quad (Z\mathcal{P}'(Z))\Big|_{Z=e^{-i\omega_\ell}} = 0, \\ (\mathbf{g}, \mathbf{v}_{\ell 2}) = 0 & \Rightarrow \quad (Z(Z\mathcal{P}'(Z))')\Big|_{Z=e^{-i\omega_\ell}} = 0, \\ \quad \cdots & \cdots \quad \cdots \\ (\mathbf{g}, \mathbf{v}_{\ell R_\ell}) = 0 & \Rightarrow \quad (Z(\cdots(Z(Z\overbrace{\mathcal{P}'(Z))')'\cdots)'}^{R_\ell})\Big|_{Z=e^{-i\omega_\ell}} = 0, \end{cases}$$

which is equivalent to

$$\mathcal{P}(Z)\Big|_{Z=e^{-i\omega_\ell}} = (Z\mathcal{P}'(Z))\Big|_{Z=e^{-i\omega_\ell}} = \cdots = (Z^{R_\ell}\mathcal{P}^{(R_\ell)}(Z))\Big|_{Z=e^{-i\omega_\ell}} = 0.$$

Since $e^{-i\omega_\ell} \neq 0$, it readily follows that

$$\mathcal{P}(e^{-i\omega_\ell}) = \mathcal{P}'(e^{-i\omega_\ell}) = \cdots = \mathcal{P}^{(R_\ell)}(e^{-i\omega_\ell}) = 0, \quad \forall \ell.$$

Thus, $\mathcal{P}(Z)$ has zeros at $Z_\ell = e^{-i\omega_\ell}$ with multiplicity at least of $R_\ell + 1$, $\forall \ell$.

Proposition 5-4: *The sequence $\{\tilde{s}_n\}$ defined in Eq. (5.7) is R point linearly predictable (or R^{th}-order autoregressive).*

Proof. According to Proposition 5-3, the homogeneous equation $\mathbf{Ag} = 0$ has non-trivial solutions if $q \geq R$. Setting $q = R$, $\mathbf{Ag} = 0$ implies that

$$\tilde{s}_n = -\sum_{\ell=1}^{R} (g_{R-\ell}/g_R)\tilde{s}_{n-\ell}, \qquad \forall n, \tag{5.10}$$

when $g_R \neq 0$; otherwise, there exists a non-negative integer $m < R$ such that

$$\begin{cases} g_R = g_{R-1} = \cdots = g_{R-m+1} = 0, \\ g_{R-m} \neq 0, \end{cases}$$

and then

$$\tilde{s}_n = -\sum_{\ell=1}^{R-m} (g_{R-m-l}/g_{R-m})\tilde{s}_{n-l}, \qquad \forall n,$$

which implies that any point in the sequence $\{\tilde{s}_n\}$ is linearly predictable from its previous $R-m$ ($m \geq 0$) points. Therefore, $\{\tilde{s}_n\}$ is at least R point linearly predictable.

Remark 5-2: Polynomial $\mathcal{P}(Z)$ is the characteristic polynomial of the linear prediction equation (5.10). If $\mathcal{P}(Z)$ is treated as a discrete filter in the Z-domain (i.e., Z^{-1} is taken to be the unit delay operator), it is also called the prediction-error filter since $\{\mathcal{P}(Z)\tilde{s}_n\}$ gives the prediction error sequence in the presence of noise [47].

Remark 5-3: An important special case results when $R_\ell = 0$, $\forall l$ (the zero-order LPA model). In this case, $\tilde{s}_n$ in Eq. (5.7) becomes

$$\tilde{s}_n = \sum_{\ell=1}^{M_1} \tilde{a}_\ell e^{-i\omega_\ell n}, \qquad n = 0, 1, \cdots, N-1. \tag{5.11}$$

The following statements can be made about this signal:

(1) $\mathbf{A} = \mathbf{A}_L \mathbf{D} \mathbf{A}_R$

where

$$\mathbf{A}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ Z_1 & Z_2 & \cdots & Z_{M_1} \\ \vdots & \vdots & & \vdots \\ Z_1^p & Z_2^p & \cdots & Z_{M_1}^p \end{bmatrix}, \qquad \mathbf{A}_R = \begin{bmatrix} 1 & Z_1 & \cdots & Z_1^q \\ 1 & Z_2^2 & \cdots & Z_2^q \\ \vdots & \vdots & & \vdots \\ 1 & Z_{M_1} & \cdots & Z_{M_1}^q \end{bmatrix},$$

and

$$\mathbf{D} = \text{diag}(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{M_1})$$

with $p = N - q - 1$ and $Z_\ell = e^{-i\omega_\ell}$, $\forall \ell$.

(2) $\mathbf{A}^+ = \mathbf{A}_R^+ \mathbf{D}^{-1} \mathbf{A}_L^+$.

(3) The condition number $\kappa = \|\mathbf{A}\| \|\mathbf{A}^+\|$ increases rapidly as M increases or $|\omega_i - \omega_j|$ decreases. For a fixed set of ω_ℓ's, M, and N, the condition number κ is minimized when $q = N/2$.

Statement (3) is intuitively clear but a rigorous proof is difficult. Assume that the ω_ℓ's are uniformly spaced such that $\omega_\ell = \Delta\omega(\ell - [M/2] - 1)$, the plots in Fig. 5.2 are obtained by calculating κ as a function of M, $\Delta\omega$, and q.

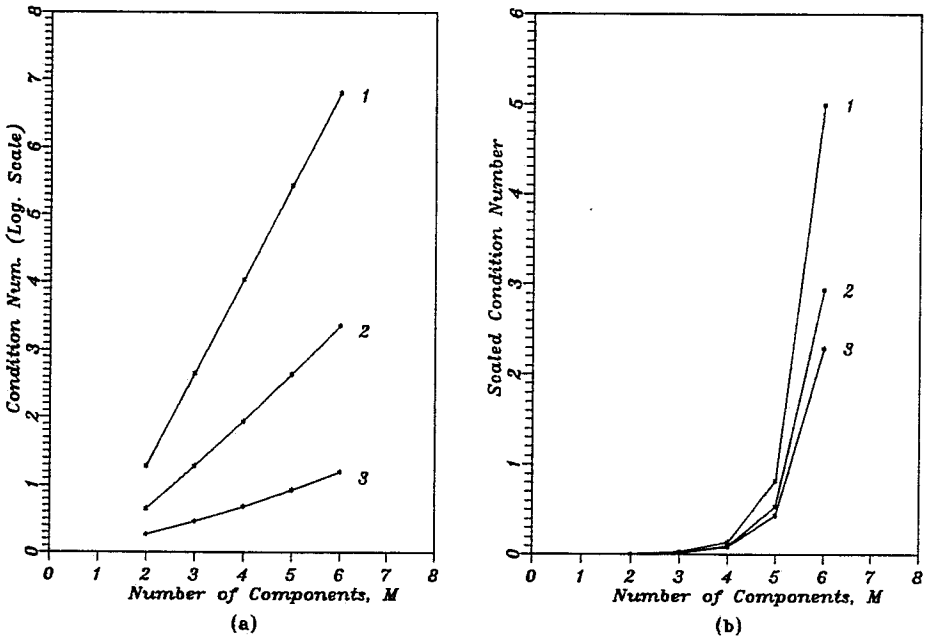


Figure 5.1: Plots of κ as a function of M, $\Delta\omega$, and q: (a) $\Delta\omega$ equals to π, 2π, and 3π, for curves 1, 2, and 3, respectively; (b) $\omega = 2\pi$ and the condition number has been scaled by 10^{-3}.

5.3 Algorithmic Implementation

Based on the above results, an algorithm for finding the LPA model parameters is given below. For the moment we assume that the model is exact and the signal is noise-free. Techniques to handle conditions deviating from these assumptions are discussed in the next section.

5.3.1 Determination of the Edge Locations, α_ℓ's

Solution of this set of highly non-linear parameters is the most difficult step. The tactic we use is to transform the original problem into an easy-to-solve one so that the linear prediction framework can be used. This is achieved by the following three-step procedure.

Step 1: High-pass filtering, $\tilde{s}_n = (i2\pi k_n)^P s_n$

From the previous section, we know that when P is chosen properly, the sequence $\{s_n\}$ can be converted to an R^{th}-order autoregressive sequence $\{\tilde{s}_n\}$ such that the α_ℓ's can be extracted from the roots of the corresponding characteristic polynomial equation. In theory, P should be chosen to be $L+1$, if the signal is known to be generated by an L^{th}-order LPA model. However, since this filter emphasizes the high spatial frequency components, noise in the high frequency range will correspondingly get amplified. Therefore, in practice, P should be chosen as small as possible under the trade-off between the amplification of measurement random noise and the reduction of systematic modeling error.

Another consideration concerning the selection of P is parameter sensitivity, that is, the stability of the polynomial roots. Let the roots of $\mathcal{P}(Z)$ be denoted by $Z_1, Z_2, \ldots, Z_q$, then differentiating both sizes of $\mathcal{P}(Z) = 0$ yields

$$|\Delta Z_\ell| \leq \frac{|\mathbf{z}_\ell \cdot \Delta \mathbf{g}|}{\prod_{\mu \neq \ell} |Z_\ell - Z_\mu|}, \qquad \forall \ell \tag{5.12}$$

where $\mathbf{z}_\ell = [1, Z_\ell, Z_\ell^2, \ldots, Z_\ell^q]^T$ and $\Delta \mathbf{g} = [\Delta g_0, \Delta g_1, \ldots, \Delta g_q]^T$.

Therefore, the potential error in a multiple root Z_ℓ ($\exists \mu,\ Z_\mu = Z_\ell$) due to perturbations in the coefficients $\mathbf{g}$ can go unbounded. This suggests that when the underlying signal s_n being modeled is not far deviating from the zero-order LPA model, it is beneficial to choose P to be one so that the characteristic polynomial has only simple roots. On the other hand, when the signal is dominated by higher-order behavior (such as ramps and parabolas) containing few zero-order edge points, it is appropriate to use P larger than one so that the systematic modeling offset is minimized for the estimation of the zero-order edges.

Step 2: Solving the system of linear equations, $\mathbf{Ag} = 0$

From Proposition 5-2, if $q \geq R$, $\mathbf{Ag} = 0$ has non-trivial solutions. Setting $g_0 = 1$ and rearranging the order of the equations, the following Toeplitz system results

$$\tilde{\mathbf{A}}\tilde{\mathbf{g}} = \mathbf{h}, \tag{5.13}$$

where

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{s}_p & \tilde{s}_{p+1} & \cdots & \tilde{s}_{N-1} \\ \tilde{s}_{p-1} & \tilde{s}_p & \cdots & \tilde{s}_{N-2} \\ \vdots & \vdots & & \vdots \\ \tilde{s}_1 & \tilde{s}_2 & \cdots & \tilde{s}_q \end{bmatrix}, \quad \tilde{\mathbf{g}} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_q \end{bmatrix}, \quad \text{and} \quad \mathbf{h} = - \begin{bmatrix} \tilde{s}_{p-1} \\ \tilde{s}_{p-2} \\ \vdots \\ \tilde{s}_0 \end{bmatrix}$$

with $p = N - q$. According to *Remark 5-3*, q should be chosen to $N/2$ to make $\mathbf{A}$ a square matrix. Therefore, when $R = N/2$, matrix $\tilde{\mathbf{A}}$ has full rank and

Eq. (5.13) can then be solved efficiently with the Levinson algorithm [44] or other fast algorithms [8]. When $R < N/2$, $g_\ell = 0$ for $R+1 \leq \ell \leq q$ since the sequence $\{\tilde{s}_n\}$ is R^{th}-order autoregressive (Proposition 5-4). In principle $\tilde{\mathbf{g}}_R = [g_1, g_2, \ldots, g_R]^T$ can be directly determined from the truncated version of Eq. (5.13) with $p = q = R$ using any linear system solver. However, it requires the *a priori* knowledge of the exact value of R. The Levinson algorithm is an ideal solution for this problem since it solves Eq. (5.13) inherently by iterating on each truncated problem with p and q stepping through 1 to $N/2$. The iterative process will terminate at step R returning the desired solution for $\tilde{\mathbf{g}}$. Hence, in the noiseless case, the Levinson algorithm can not only offer a fast solution to Eq. (5.13) but also help to determine the order, R, of the autoregressive sequence $\{\tilde{s}_n\}$.

Step 3: Polynomial rooting of $\mathcal{P}(Z)$

This step can be accomplished using a modified Newton's method [21] or the Jenkins-Traub method [39] to handle the multiple root case. Once all the roots Z_ℓ, $\ell = 1, 2 \cdots, q$, are found, α_ℓ can be easily extracted from them through the following relationship given in Proposition 5-3,

$$\alpha_\ell = \frac{1}{2\pi\Delta k}\left\{\mathrm{Im}[\ln(Z_\ell)]\,\Big|_{\,\mathrm{mod}[-\pi,\pi]}\right\} \tag{5.14}$$

where operator Im denotes taking the imaginary part of a complex number.

An efficient method for finding the signal roots is also possible by making use of the fact that all the signal roots lie on the unit circle (Proposition 5-3). Consider the following all-pole filter defined by the prediction-error filter $\mathcal{P}(Z)$

$$H(e^{i\omega}) = \frac{1}{\mathcal{P}(e^{i\omega})} = \frac{1}{\sum_{\ell=1}^{q} g_\ell e^{i\omega\ell}}.$$

The denominator can be evaluated using an FFT algorithm. Since $|H(e^{i\omega})|$ exhibits spikes at the signal poles, $e^{-i\omega_\ell}$, $\forall \ell$, and also at the extraneous poles (when $q > R$) but with lower amplitudes, these poles can be identified with a peak detector and differentiated based on the amplitudes. Clearly, high numerical accuracy in estimating the signal poles can be easily obtained by simply increasing the number of points evaluated.

After all the roots are found, an extra step is necessary to remove the extraneous roots due to using $q > R$, and more importantly, to handle the tightly clustered roots caused by using P larger than one or by noise. Clearly, this "root management" step is to ensure a stable model and to eliminate possible noisy spikes from false narrow intervals. Some of the extraneous roots can be easily removed in practice based on the fact that they usually wander away from the unit circle [47]. The others close to the unit circle can be eliminated by using the following test if they get clustered around signal roots, or otherwise, are retained without causing model instability.

To remove clustered roots (edges), we can sort the α_ℓ's in a non-decreasing order. Then, in the m-multiple root case, $\alpha_{\ell+m-1} - \alpha_\ell$ will be smaller than a predetermined threshold δ (usually δ can be chosen to be a fraction of the Fourier pixel size $\Delta x = 1/(N\Delta k)$); correspondingly the edges at $\alpha_\ell, \alpha_{\ell+1}, \cdots, \alpha_{\ell+m-1}$ should be replaced by a new single edge $\tilde{\alpha}_\ell$ at their mean location, i.e., $\tilde{\alpha}_\ell = \sum_{\mu=\ell}^{\ell+m-1} \alpha_\mu/m$. However, in order to prevent true zero-order edges from narrow image features (boxes) from being lumped together, some safeguard is necessary. One way to accomplish this is to perform an additional check on the amplitudes of the boxes corresponding to the small intervals, which can be estimated from the windowed-FFT reconstruction $\hat{\rho}_N(x)$ from this purpose. If their variations are within the noise level, the edges should be lumped together as described. Otherwise, the μ^{th} local interval defined by α_μ and $\alpha_{\mu+1}$ should

be retained if the corresponding box amplitude is significantly different from those of its neighboring boxes.

5.3.2 Determination of the Local Polynomial Order, P_ℓ's

Once all the α_ℓ's are estimated, the P_ℓ's can be determined based on the size of the local intervals and the image amplitude variations. The primary concern in selecting the local polynomial orders is, of course, the model stability. Usually, the higher the orders, the easier the data consistency can be obtained, but the less stable the model will be. Since data consistency does not usually promise good estimates, especially for short data records, we need to have enough information to define each local polynomial to ensure a stable model. We recommend the following procedure for determining the P_ℓ's:

(1) Get the windowed-FFT reconstruction $\hat{\rho}_N(x)$ (this step is actually carried out at the "root management" stage described above) and from it calculate the image mean value m_ℓ and standard deviation σ_ℓ for each local interval;

(2) At the ℓ^{th} interval, if $\alpha_{\ell+1} - \alpha_\ell$ is not larger than two pixels or σ_ℓ is on the noise level, set $P_\ell = 0$; otherwise,

(3) Iteratively fit $\hat{\rho}_N(x)$ in this interval with increasingly higher order polynomials until the residual is on the noise level or the order reaches L (the maximum local polynomial order allowed), and then set P_ℓ to this order. Clearly, the P_ℓ thus determined satisfies that $P_\ell \leq \min\{L, \max\{0, [(\alpha_{\ell+1} - \alpha_\ell)/\Delta x - 1]\}\}$.

5.3.3 Determination of the Local Polynomial Coefficients, $c_{\ell r}$'s

After the P_ℓ's and the α_ℓ's are determined, the basis function $\Phi_{\ell r}(k)$ can be calculated from recursion (5.3). Let $\mathbf{c} = [c_{10}, c_{11}, \ldots, c_{\ell r}, c_{\ell(r+1)}, \ldots, c_{MP_M}]^T$, Eq. (5.1) can then be rewritten in matrix form as

$$\mathbf{\Phi c} = \mathbf{d} \tag{5.15}$$

where $\mathbf{d}$ is the data vector, i.e., $\mathbf{d} = [s_0, s_1, \cdots, s_{N-1}]^T$ and the model matrix $\mathbf{\Phi}$ is given by

$$\mathbf{\Phi} = \begin{bmatrix} \Phi_{10}(0) & \cdots & \Phi_{\ell r}(0) & \Phi_{\ell(r+1)}(0) & \cdots & \Phi_{MP_M}(0) \\ \Phi_{10}(1) & \cdots & \Phi_{\ell r}(1) & \Phi_{\ell(r+1)}(1) & \cdots & \Phi_{MP_M}(1) \\ \vdots & & \vdots & \vdots & & \vdots \\ \Phi_{10}(\tilde{N}) & \cdots & \Phi_{\ell r}(\tilde{N}) & \Phi_{\ell(r+1)}(\tilde{N}) & \cdots & \Phi_{MP_M}(\tilde{N}) \end{bmatrix},$$

with $\tilde{N} = N - 1$. Linear system (5.15) is over-determined and can be solved in a least squares sense such that

$$\mathbf{c} = (\mathbf{\Phi}^H \mathbf{\Phi})^{-1} \mathbf{\Phi}^H \mathbf{d}. \tag{5.16}$$

If the "root management" step is activated and the local polynomial orders are properly chosen, $\mathbf{\Phi}$ is usually well-conditioned. Therefore, the least squares solution given in Eq. (5.16) can be obtained numerically by a Cholesky decomposition based procedure or a Householder transform based method.

5.4 Practical Consideration and Improvements

In the above discussions, we have assumed that the data are noise-free and the model is exact. In practice, this is usually not true, and measurement noise and modeling error may limit the performance of the proposed method, especially for the estimation of the nonlinear parameters, α_ℓ's. In this section, techniques to cope with these non-ideal conditions are presented. Particularly, we will focus on discussing how to use the truncated SVD regularization concept to handle the ill-conditionedness that may arise with the linear prediction formalism, and to remove noise. Toward this end, we will also discuss the variable projection method for nonlinear least squares fitting.

5.4.1 The Use of Truncated SVD

When noise is present, Eq. (5.13) becomes

$$\tilde{\mathbf{A}}_e(\tilde{\mathbf{g}} + \Delta\tilde{\mathbf{g}}) = \mathbf{h}_e \tag{5.17}$$

where $\tilde{\mathbf{A}}_e = \tilde{\mathbf{A}} + \mathbf{E}$ and $\mathbf{h}_e = \mathbf{h} + \mathbf{e}$. Since the polynomial root locations are sensitive to $\Delta\tilde{\mathbf{g}}$, it is desirable that $\|\Delta\tilde{\mathbf{g}}\|$ is as small as possible. When $R = N/2$ and $\|\mathbf{E}\|$ is small such that $\text{rank}(\tilde{\mathbf{A}}_e) = \text{rank}(\tilde{\mathbf{A}})$, $\|\Delta\tilde{\mathbf{g}}\|$ satisfies

$$\frac{\|\Delta\tilde{\mathbf{g}}\|}{\|\tilde{\mathbf{g}}\|} \leq \frac{\kappa}{1 - \kappa\|\mathbf{E}\|/\|\tilde{\mathbf{A}}\|} \left\{ \frac{\|\mathbf{e}\|}{\|\mathbf{h}\|} + \frac{\|\mathbf{E}\|}{\|\tilde{\mathbf{A}}\|} \right\},$$

where κ is the condition number of $\tilde{\mathbf{A}}$. Since $\tilde{\mathbf{A}}$ is usually ill-conditioned, especially when the edges are closely spaced (Fig. 5.2), the relative error $\|\Delta\tilde{\mathbf{g}}\|/\|\tilde{\mathbf{g}}\|$ for $\tilde{\mathbf{g}}$ can be several orders of magitude bigger than that for the coefficient matrix $\tilde{\mathbf{A}}$ or that for the right hand side vector $\mathbf{h}$. Therefore, the linear prediction formalism is ineffective for extracting a large number of closely spaced edges. Fortunately, in practice, the number of sharp edges is usually significantly smaller than $N/2$, which makes it possible through the use of a truncated SVD to estimate sharp but closely spaced edges fairly accurately even in the presence of noise (superresolution).

Rationale for Using the Truncated SVD

When $R < q \leq N/2$, $\tilde{\mathbf{A}}$ is rank-deficient. If the minimum-norm solution is sought, one obtains from Eq. (5.17) that

$$\tilde{\mathbf{g}} + \Delta\tilde{\mathbf{g}} = \tilde{\mathbf{A}}_e^+(\mathbf{h} + \mathbf{e}).$$

Substitution of $\tilde{\mathbf{g}}$ with $\tilde{\mathbf{A}}^+\mathbf{h}$ yields

$$\Delta\tilde{\mathbf{g}} = (\tilde{\mathbf{A}}_e^+ - \tilde{\mathbf{A}}^+)\mathbf{h} + \tilde{\mathbf{A}}_e^+\mathbf{e}. \tag{5.18}$$

Clearly, to minimize $\|\Delta\tilde{\mathbf{g}}\|$, it is desirable that $\|\tilde{\mathbf{A}}_e^+ - \tilde{\mathbf{A}}^+\|$ is as small as possible. It is well-known, however, that the pseudoinverse of a matrix is not a continuous function of its elements. Consequently, small perturbations in the elements can lead to a huge perturbation in the pseudoinverse; particularly, if $\text{rank}(\tilde{\mathbf{A}}_e) \neq \text{rank}(\tilde{\mathbf{A}})$, $\|\tilde{\mathbf{A}}_e^+ - \tilde{\mathbf{A}}^+\| \geq 1/\|\mathbf{E}\|$ [5, 50, 87, 98]. Therefore, when the perturbation $\mathbf{E}$ causes a rank change (increase) in the matrix $\tilde{\mathbf{A}}$ (as is always the case in practice), it is beneficial to replace $\tilde{\mathbf{A}}_e$ with an approximation $\hat{\mathbf{A}}_e$ of rank $\hat{R} = R$. Naturally, $\hat{\mathbf{A}}_e$ should be chosen in such a way so that $\|\hat{\mathbf{A}}_e - \tilde{\mathbf{A}}_e\|$ is minimized (under the Fobenius norm [26]). An effective tool for this lower rank approximation is provided by the SVD.

The SVD of $\tilde{\mathbf{A}}_e$ is given by

$$\begin{aligned}\tilde{\mathbf{A}}_e &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \\ &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix},\end{aligned}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices with their column vectors being the left and right singular vectors of $\tilde{\mathbf{A}}_e$, respectively, and $\boldsymbol{\Sigma}$ is a diagonal matrix containing the singular values in non-increasing order, i.e., $\lambda_1, \lambda_2, \ldots, \lambda_q$. With this factorization, $\hat{\mathbf{A}}_e$ can be obtained by retaining the first $\hat{R}$ principal singular values ($\boldsymbol{\Sigma}_1$) with their corresponding singular vectors ($\mathbf{U}_1$ and $\mathbf{V}_1$), namely,

$$\hat{\mathbf{A}}_e = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^H$$

Accordingly, the minimum-norm solution for $\tilde{\mathbf{g}}$ is given by

$$\begin{aligned}\hat{\mathbf{g}}_e &= \mathbf{V}_1\boldsymbol{\Sigma}_1^{-1}\mathbf{U}_1^H\mathbf{h}_e \\ &= \sum_{\ell=1}^{\hat{R}} \lambda_\ell^{-1} \left(\mathbf{u}_\ell^H\mathbf{h}_e\right) \mathbf{v}_\ell,\end{aligned}$$

where $\mathbf{u}_\ell$ and $\mathbf{v}_\ell$ are the ℓ^{th} column vectors of $\mathbf{U}_1$ and $\mathbf{V}_1$, respectively.

The above truncated SVD minimum-norm solution was first proposed by Kumaresan and Tufts. Discussions of its properties are well-published [42, 45, 54, 91, 92]. Two of the most important ones are:

(1) *Noise Removal*: In the noiseless case, $\lambda_\ell = 0, \ell > R$, then $\hat{\mathbf{A}}_e = \tilde{\mathbf{A}}_e = \tilde{\mathbf{A}}$. When noise is present, these singular values are not zero, but will be relatively smaller than the principal (signal) singular values if the data S/N is not too low. Suppose that e_n is of zero-mean and uncorrelated with $\tilde{s}_n$ such that

$$\sum_n e_n e^*_{n+\ell} = \begin{cases} \sigma_e^2, & \text{for } \ell = 0, \\ 0, & \text{otherwise,} \end{cases}$$

then

$$\lambda_\ell^2 \approx \sum_{n=1}^{q} e_n e_n^* \approx \sigma_e^2, \qquad \ell > R.$$

Hence, by setting the "noise" singular values (λ_ℓ, $\ell > R$) to zero, the noise in the noise-subspace spanned by their corresponding "noise" singular vectors ($\mathbf{u}_\ell$, $\mathbf{v}_\ell$, $\ell \geq R$) will be removed; correspondingly, $\hat{\mathbf{A}}_e$ becomes less perturbed than $\tilde{\mathbf{A}}$, and thus the same for $\hat{\mathbf{g}}_e$. This noise removal effect of the truncated SVD is sometimes called *numerical filtering*.

(2) *Reduced Noise Sensitivity*: It is evident from Eq. (5.12) that the sensitivity of the signal roots Z_ℓ's to the perturbations in the coefficient vector, $\Delta\tilde{\mathbf{g}}$, strongly depends on the locations of all the roots. Since we have no control over the locations of the signal roots, it is desirable that the extraneous roots are well-behaved. Clearly, the optimal distribution of the extraneous roots is such that all the roots are uniformly distributed along the unit circle and, therefore, the denominator in the right hand side of inequality (5.12) reaches the maximum. It has been demonstrated that this optimality is met by the truncated SVD minimum-norm solution for the coefficient vector in the noiseless case [47]. When noise is present, the extraneous roots will be driven

off these positions, but this "uniform" root distribution property will provide great resistance to the extraneous roots clustering around the signal roots.

Determination of the Truncation Level $\hat{R}$

Let $\tilde{\mathbf{E}}$ be the residual matrix between $\hat{\mathbf{A}}_e$ and $\tilde{\mathbf{A}}$, i.e., $\tilde{\mathbf{E}} = \hat{\mathbf{A}}_e - \tilde{\mathbf{A}}$, under the above conditions for $\hat{\mathbf{A}}_e$, the following inequality holds [98]:

$$\|\hat{\mathbf{A}}_e^+ - \tilde{\mathbf{A}}^+\| \leq \mu \|\hat{\mathbf{A}}_e^+\| \|\tilde{\mathbf{A}}^+\| \|\tilde{\mathbf{E}}\|$$

where $\mu = (1+\sqrt{5})/2$. Substituting it into Eq. (5.18) leads to

$$\|\Delta\tilde{\mathbf{g}}\| \leq \mu \|\hat{\mathbf{A}}_e^+\| \|\tilde{\mathbf{A}}^+\| \|\tilde{\mathbf{E}}\| \|\mathbf{h}\| + \|\hat{\mathbf{A}}_e^+\| \|\mathbf{e}\|.$$

Furthermore, $\|\tilde{\mathbf{A}}\| \|\tilde{\mathbf{g}}\| \geq \|\mathbf{h}\|$, therefore,

$$\begin{aligned} \|\Delta\tilde{\mathbf{g}}\| &\leq \mu \|\tilde{\mathbf{A}}^+\| \|\tilde{\mathbf{A}}\| \|\hat{\mathbf{A}}_e^+\| \|\tilde{\mathbf{E}}\| \|\tilde{\mathbf{g}}\| + \|\hat{\mathbf{A}}_e^+\| \|\mathbf{e}\| \\ &\approx \mu \frac{\lambda_1}{\lambda_{\hat{R}}} \frac{\|\hat{\mathbf{E}}\|}{\lambda_{\bar{R}}} \|\tilde{\mathbf{g}}\| + \frac{\|\mathbf{e}\|}{\lambda_{\hat{R}}} \\ &\approx \mu \frac{\lambda_1}{\lambda_{\hat{R}}} \|\tilde{\mathbf{g}}\| + \frac{\|\mathbf{e}\|}{\lambda_{\hat{R}}}. \end{aligned} \tag{5.19}$$

Eq. (5.19) indicates that in order for $\|\Delta\mathbf{g}\|$ to be small, it is necessary for both terms in the right hand side to be small. When $\mathbf{e}$ merely comes from random measurement noise (in the case that the model is appropriate), the first term is usually the dominating error and, thus, it is benificial to select a small $\hat{R}$ so that the effective condition number, $\lambda_1/\lambda_{\hat{R}}$, is small. However, when $\hat{R}$ gets too small ($\hat{R} < R$), over-truncation will result and, consequently, $\mathbf{e}$ will become large and $\Delta\mathbf{g}$ will increase again as $\hat{R}$ decreases. Therefore, ideally $\hat{R}$ should be chosen to be R when λ_1/λ_R is small and prior knowledge of R is available or R can be estimated via some order-determination procedures [102, 103, 97]. However, if λ_1/λ_R is too large, it is necessary to sacrifice some less significant edges to protect the significant ones by "slightly" over-truncating the matrix

($\hat{R} < R$). Hence, in practice, whether R is known or not, the truncation level $\hat{R}$ should be chosen in such a way so that the effective condition number of $\hat{\mathbf{A}}_e$, $\lambda_1/\lambda_{\hat{R}}$, is under a threshold pre-determined based on the data S/N. However, the problem of under-truncation and over-truncation may also arise in practical application with this simple condition number threshold criterion, since the λ_ℓ's usually decay gradually in the presence of noise and, therefore, it is difficult to decide a strict threshold. In order to avoid under-truncation or too much over-truncation resulting from a poorly-selected threshold, the following test can be used to complement this procedure.

Recall that for any acceptable value of $\hat{R}$, $\hat{\mathbf{A}}_e = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H$ is supposed to contain mainly the signal part ($\tilde{s}_n$) of the data ($\tilde{d}_n = \tilde{s}_n + e_n$). Therefore, an approximated signal sequence $\hat{s}_n$ can be extracted from $\hat{\mathbf{A}}_e$ by averaging the elements of $\hat{\mathbf{A}}_e$ along the diagonals to restore its Toeplitz structure. If the $\hat{R}$ is properly chosen, $\mathrm{Var}\{\tilde{d}_n - \hat{s}_n\}$ should be at the noise level. Therefore, this value can provide a good check on under- or over-truncation.

Noise Compensation

Removing/reducing the noise in the coefficient matrix $\tilde{\mathbf{A}}_e$ in Eq. (5.17) is crucial for controlling the error in $\tilde{\mathbf{g}}$. From Eq. (5.19), it is also desirable to reduce the noise $\mathbf{e}$ in the right hand side vector $\mathbf{h}_e$. This can be accomplished by replacing $\mathbf{h}_e$ with a new vector $\hat{\mathbf{h}}_e$ formed from the approximated signal sequence $\{\hat{s}_n\}$ obtained as described above.

Other SVD-based methods

Since Kumaresan and Tufts introduced the truncated SVD regularization concept to the linear prediction problem (sometimes referred to as the LPSVD method [2]), many other related SVD-based methods have also been proposed.

Two of the best recognized ones are the *noise eigenvector method* and the *signal subspace method.* As we will see from the followings, the former tend to use only the noise eigenvectors to represent the polynomial coefficient vector (an idea initially from Pisarenko [63]), and on the other hand, the latter attempts to fit the whole signal subspace directly with the known signal structure.

Noise Eigenvector Method (NESVD): Consider applying SVD directly to $\mathbf{A}_e$ (not to $\tilde{\mathbf{A}}_e$), namely,

$$\begin{aligned} \mathbf{A}_e &= \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^H + \mathbf{U}_2\boldsymbol{\Sigma}_2\mathbf{V}_2^H \\ &= \hat{\mathbf{A}}_e + \hat{\mathbf{E}}. \end{aligned} \tag{5.20}$$

Then the coefficeint vector $\mathbf{g}$ should be estimated from

$$\hat{\mathbf{A}}_e\mathbf{g} = 0,$$

or equivalently from

$$\mathbf{V}_1^H\mathbf{g} = 0. \tag{5.21}$$

Based on the orthonormality of the singular vectors, $\mathbf{v}_\ell$'s, Eq. (5.21) implies that $\mathbf{g}$ can be any vector in the noise subspace spanned by the noise singular vectors $\mathbf{v}_\ell$, $\ell > R$. As with the LPSVD, if g_0 is constrained to 1 and let $\tilde{\mathbf{g}} = [g_1, g_2, \ldots, g_R]^T$, we have from Eq. (5.21) that

$$\mathbf{B}\tilde{\mathbf{g}} = -\mathbf{b}$$

where

$$\mathbf{B} = \begin{bmatrix} v_{11}^* & v_{12}^* & \cdots & v_{1q}^* \\ v_{21}^* & v_{22}^* & \cdots & v_{2q}^* \\ \vdots & \vdots & & \vdots \\ v_{R1}^* & v_{R2}^* & \cdots & v_{Rq}^* \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} v_{10}^* \\ v_{20}^* \\ \vdots \\ v_{R0}^* \end{bmatrix}.$$

The minimum-norm solution of $\tilde{\mathbf{g}}$ is

$$\tilde{\mathbf{g}} = -\mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{b}.$$

Since

$$(\mathbf{B}\mathbf{B}^H)^{-1} = (\mathbf{I}_R - \mathbf{b}\mathbf{b}^H)^{-1} = \mathbf{I}_R + \frac{\mathbf{b}\mathbf{b}^H}{1 - \|\mathbf{b}\|^2}$$

we have that

$$\tilde{\mathbf{g}} = \mathbf{B}^H \left(\mathbf{I}_R + \frac{\mathbf{b}\mathbf{b}^H}{1 - \|\mathbf{b}\|^2} \right) \mathbf{b}$$

or

$$\tilde{\mathbf{g}} = \frac{\mathbf{B}^H\mathbf{b}}{1 - \|\mathbf{b}\|^2}. \tag{5.22}$$

An equivalent expression can also be derived in terms of the noise singular vectors as

$$\tilde{\mathbf{g}} = \frac{\mathbf{C}^H\mathbf{c}}{\|\mathbf{c}\|^2} \tag{5.23}$$

where

$$\mathbf{C} = \begin{bmatrix} v^*_{(R+1)1} & v^*_{(R+1)2} & \cdots & v^*_{(R+1)q} \\ v^*_{(R+2)1} & v^*_{(R+2)2} & \cdots & v^*_{(R+2)q} \\ \vdots & \vdots & & \vdots \\ v^*_{(q+1)1} & v^*_{(q+1)2} & \cdots & v^*_{(q+1)q} \end{bmatrix}, \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} v^*_{(R+1)0} \\ v^*_{(R+2)0} \\ \vdots \\ v^*_{(q+1)0} \end{bmatrix}.$$

One of the desirable properties of this noise eigenvector is that the error in the right hand side vector $\mathbf{h}_e$ in Eq. (5.17) is automatically taken care of by the noise removal capability offered by the truncated SVD formalism. For this reason, this method has been identified as the *total least squares method* and claimed to be of superior performance [65]. There is also a potential computational advantage of this method over the LPSVD since it requires only

the right singular vectors or the left singular vectors and thus eliminates the need for computing a full decomposition. Another subtle point worth noting is that this method places equal importance on the singular vectors used. However, it is well-known that the singular vectors corresponding to the large singular values are usually less perturbed than those associated with the small singular values. Therefore, this noise eigenvector method is potentially more vulnerable to noise than the LPSVD, which contradicts the claimed superior performance of [65]. In fact, by careful examination of the results in [65], one can find that the extraneous roots from this method are not as well-behaved as the LPSVD. Weighting the singular vectors with their corresponding singular values may be helpful for reducing the noise effect but will be at the expense of the computational efficiency gained from expression (5.22) or (5.23).

Signal Subspace Method (SGSVD): When $R_\ell = 0$ (then, $R = M_1$), a substantially different way to extract the ω_ℓ's from the linear prediction matrix $\mathbf{A}_e$ is possible [48]. Comparing the factorization in *remark 5-3* and the SVD in Eq. (5.20), one immediately gets

$$\mathbf{V}_1^H = \mathbf{T}^H \mathbf{A}_R$$

or

$$\mathbf{V}_1 = \mathbf{A}_R^H \mathbf{T}$$

where $\mathbf{T}$ is a non-singular linear transformation. Let $\mathbf{\Omega} = \mathrm{diag}(e^{i\omega_1}, e^{i\omega_2}, \ldots, e^{i\omega_R})$ and $\theta = (\overbrace{1, 1, \ldots, 1}^{R})$, then

$$\mathbf{V}_1 = \begin{bmatrix} \theta \\ \theta\mathbf{\Omega} \\ \vdots \\ \theta\mathbf{\Omega}^q \end{bmatrix} \mathbf{T} = \begin{bmatrix} \theta\mathbf{T} \\ \theta\mathbf{T}\mathbf{T}^{-1}\mathbf{\Omega}\mathbf{T} \\ \vdots \\ \theta\mathbf{T}\mathbf{T}^{-1}\mathbf{\Omega}^q\mathbf{T} \end{bmatrix}.$$

Further, let $\mathbf{M} = \mathbf{T}^{-1}\mathbf{\Omega}\mathbf{T}$, then the R eigenvalues of $\mathbf{M}$ are exactly equal to $e^{i\omega_\ell}$, $\ell = 1, 2, \ldots, R$, in the noiseless case. Therefore, under this formulation the problem of finding the ω_ℓ's reduces to finding the $\mathbf{M}$. Note that

$$\mathbf{V}_b\mathbf{M} = \mathbf{V}_t \tag{5.24}$$

where $\mathbf{V}_b$ and $\mathbf{V}_t$ are the matrices obtained from $\mathbf{V}_1$ with the bottom row and the top row being removed, respectively. $\mathbf{M}$ can be determined from Eq. (5.24) in a least squares sense as

$$\mathbf{M} = (\mathbf{V}_b^H\mathbf{V}_b)^{-1}\mathbf{V}_b^H\mathbf{V}_t. \tag{5.25}$$

Making use of the following identity:

$$(\mathbf{V}_b^H\mathbf{V}_b)^{-1} = (\mathbf{I}_R - \mathbf{v}_{q+1}^H\mathbf{v}_{q+1})^{-1} = \mathbf{I}_R + \frac{\mathbf{v}_{q+1}^H\mathbf{v}_{q+1}}{1 - \|\mathbf{v}_{q+1}\|^2},$$

Eq. (5.25) reduces to

$$\mathbf{M} = \left(\mathbf{I}_R + \frac{\mathbf{v}_{q+1}^H\mathbf{v}_{q+1}}{1 - \|\mathbf{v}_{q+1}\|^2}\right)\mathbf{V}_b^H\mathbf{V}_t, \tag{5.26}$$

where $\mathbf{v}_{q+1}$ is the $(q+1)^{th}$ row vector of $\mathbf{V}_1$.

A similar expression is also available for the matrix $\mathbf{M}$ in terms of the right singular vectors as

$$\mathbf{M} = \left(\mathbf{I}_R + \frac{\mathbf{u}_{q+1}^T\mathbf{u}_{q+1}^*}{1 - \|\mathbf{u}_{q+1}\|^2}\right)\mathbf{U}_b^T\mathbf{U}_t^*, \tag{5.27}$$

where $\mathbf{u}_{q+1}$ is the $(q+1)^{th}$ row vector of $\mathbf{U}_1$, and similarly, $\mathbf{U}_b$ and $\mathbf{U}_t$ are the matrices obtained from $\mathbf{U}_1$ with the bottom row and the top row being removed, respectively.

One appealing feature of this signal subspace method lies in the fact that it directly gives the signal roots and thus further measures to deal with the extraneous roots are not necessary. For NMR spectroscopy application, this is a big advantage since the signal roots are always off the unit circle and can get mixed in with the extraneous roots if the LPSVD or NESVD is used. For this reason, Barkuijsen *et al.* strongly advocate this method [3]. For imaging applications, however, this may be a week point since the introduction of extraneous roots (i.e., higher-order linear predictor) can help to absorb part of the noise and protect signal roots [45]. In addition, this method may suffer the same the problem as the NESVD does for handling noise since the singular vectors are treated equally.

5.4.2 Combined Forward-Backward Linear Prediction

The fact that the Eq. (5.9) has all its roots on the unit circle can also be exploited to further improve the estimate of $\tilde{\mathbf{g}}$. This can be achieved using the so-called combined forward-backward linear prediction formalism [91], in which $\tilde{\mathbf{g}}$ is estimated from the following modified linear prediction equations

$$\left[\begin{array}{cccc} \tilde{s}_{N-q} & \tilde{s}_{N-q-1} & \cdots & \tilde{s}_{N-1} \\ \tilde{s}_{N-q-1} & \tilde{s}_{N-q} & \cdots & \tilde{s}_{N-2} \\ \vdots & \vdots & & \vdots \\ \tilde{s}_{1} & \tilde{s}_{2} & \cdots & \tilde{s}_{q} \\ \hline \tilde{s}^{*}_{N-2} & \tilde{s}^{*}_{N-3} & \cdots & \tilde{s}^{*}_{N-q-1} \\ \tilde{s}^{*}_{N-3} & \tilde{s}^{*}_{N-2} & \cdots & \tilde{s}^{*}_{N-q-2} \\ \vdots & \vdots & & \vdots \\ \tilde{s}^{*}_{q-1} & \tilde{s}^{*}_{q-2} & \cdots & \tilde{s}^{*}_{0} \end{array}\right] \left[\begin{array}{c} g_1 \\ g_2 \\ \vdots \\ g_q \end{array}\right] = - \left[\begin{array}{c} \tilde{s}_{N-q-1} \\ \tilde{s}_{N-q-2} \\ \vdots \\ \tilde{s}_{0} \\ \hline \tilde{s}^{*}_{N-1} \\ \tilde{s}^{*}_{N-2} \\ \vdots \\ \tilde{s}^{*}_{q} \end{array}\right]. \tag{5.28}$$

As one can see, the upper part of Eq. (5.28) is identical to Eq. (5.13), of which the corresponding polynomial has all its roots located inside and/or on

the unit circle. The lower part is the forward prediction conterpart, of which the polynomial roots reside outside and/or on the unit circle. Therefore, as a consequence of Eq. (5.28), the roots of the polynomial defined by $\tilde{\mathbf{g}}$ will be forced onto the unit circle. The truncated SVD concept can also be applied to Eq. (5.28) as described above to gain robustness.

5.4.3 Nonlinear Least Squares Fitting

When the data signal-to-noise ratio is low and the number of data points, N, is small, the bias for the estimates of the non-linear parameters, α_ℓ's, with the linear prediction formalism may be big even when the truncated SVD and the forward-backward linear prediction formalism are used. In this case, the nonlinear least squares (NLS) fitting approach as formulated in (5.4) can be used to refine the estimates of the α_ℓ's obtained from using any of the above described procedures.

There are two major reasons that the NLS method can do better than the LP-based methods. First, the LP-based methods are only asymptotically equivalent to the NLS method for the extraction of the α_ℓ's from the sequence $\{\tilde{d}_n\} = \{\tilde{s}_n + \tilde{e}_n\}$ at high data S/N [46]. If the signal-to-noise of the modified data sequence $\{\tilde{d}_n\}$ is low, the LP-based methods is known to be inferior to the NLS method for estimating a moderate number of closely spaced components. Secondly, the high-pass filter $(i2\pi k_n)^P$ used by the LP-based methods can drastically amplify the noise in the high frequency range even for $P = 1$. Whereas, with the NLS method, it is possible to employ a different high-pass filter (weighting function) to emphasize the edge information with less noise amplification.

The nonlinear least squares problem (5.4) is usually solved by an iterative procedure that calculates an approximation to the solution starting from an

initial estimate θ_0 and obtaining a sequence θ_n by the iteration

$$\theta_{n+1} = \theta_n + \gamma_n \mathbf{p}_n$$

where θ is the parameter vector containing the α_ℓ's and the c_ℓ's, $\mathbf{p}_n$ is the search direction vector and γ_n is the steplength at the n^{th} step. In the following, we describe how to determine the $\mathbf{p}$ and γ at each step. Particularly, the special structure of the linear parameters, c_ℓ's, being separable from the nonlinear parameters, α_ℓ's, is exploited to reduce the dimension of the search space such that $\theta = [\alpha_1, \alpha_2, \ldots, \alpha_{M_1}]^T$. Improvement of convergence rate offered by this specialized algorithm over other general methods is usually significant since the number of nonlinear parameters is only a small fraction of the total number of parameters.

The Variable Projection Method

Following the notations in Section 5.3.3, the residual vector $\mathbf{r} = [r_1, r_2, \ldots, r_N]^T$ can be expressed as

$$\mathbf{r} = \mathbf{\Phi c} - \mathbf{d} \tag{5.29}$$

where $\mathbf{\Phi}$, $\mathbf{c}$ and $\mathbf{d}$ are as described in Eq. (5.15) except that $\Phi_{\ell r}(k)$ and $s(k)$ are multiplied by a weighting $w(k)$ defined in (5.4). And the α_ℓ's are to be determined by minimizing the functional $\|\mathbf{r}\|^2$. Replacing the linear parameter vector $\mathbf{c}$ by $\mathbf{\Phi}^+\mathbf{d}$, Guttman *et al.* [29] proved that one can equivalently minimizes the *variable projection functional* $\|\hat{\mathbf{r}}\|^2$, with $\hat{\mathbf{r}}$ being defined as

$$\hat{\mathbf{r}} = \mathbf{\Phi}\mathbf{\Phi}^+\mathbf{d} - \mathbf{d} = (\mathbf{\Phi}\mathbf{\Phi}^+ - \mathbf{I}_N)\mathbf{d}. \tag{5.30}$$

When the residual $\|\hat{\mathbf{r}}\|^2$ is small and the initial estimate is close to the solution, $\mathbf{p}$ can be chosen as the Gauss-Newton search direction given by

$$\mathbf{p} = -(\mathbf{J}^H\mathbf{J})^{-1}\mathbf{J}^H\hat{\mathbf{r}}, \tag{5.31}$$

or the Levenberg-Marquardt direction given by

$$\mathbf{p} = -\left(\mathbf{J}^H\mathbf{J} + \mu\mathbf{I}\right)^{-1}\mathbf{J}^H\hat{\mathbf{r}}, \tag{5.32}$$

where $\mathbf{J}$ is the Jacobian matrix defined by

$$\mathbf{J} = \left[\frac{\partial\hat{\mathbf{r}}}{\partial\alpha_1}, \frac{\partial\hat{\mathbf{r}}}{\partial\alpha_2}, \cdots, \frac{\partial\hat{\mathbf{r}}}{\partial\alpha_{M_1}}\right],$$

and μ is the regularization parameter usually chosen to make the modified matrix $\mathbf{J}^H\mathbf{J} + \mu\mathbf{I}$ positive definite. The other frequently used trust region criterion for the selection of μ is discussed in [21].

Superimposing the realness constraint on $\mathbf{p}$, Eq. (5.31) and Eq. (5.32) become

$$\mathbf{p} = -\left[\mathbf{J}^H\mathbf{J} + \left(\mathbf{J}^H\mathbf{J}\right)^*\right]^{-1}\left[\mathbf{J}^H\hat{\mathbf{r}} + \left(\mathbf{J}^H\hat{\mathbf{r}}\right)^*\right]$$

and

$$\mathbf{p} = -\left[\mathbf{J}^H\mathbf{J} + \left(\mathbf{J}^H\mathbf{J}\right)^* + \mu\mathbf{I}\right]^{-1}\left[\mathbf{J}^H\hat{\mathbf{r}} + \left(\mathbf{J}^H\hat{\mathbf{r}}\right)^*\right].$$

Clearly, the vector $\mathbf{p}$ given above will not be affected by replacing $\mathbf{J}$ and $\hat{\mathbf{r}}$ with $\mathbf{J}_Q = \mathbf{Q}\mathbf{J}$ and $\hat{\mathbf{r}}_Q = \mathbf{Q}\hat{\mathbf{r}}$ provided $\mathbf{Q}$ is an orthonormal matrix. Therefore, the key step in calculating $\mathbf{p}$ is to find $\dfrac{\partial\hat{\mathbf{r}}}{\partial\alpha_\ell}$ or $\mathbf{Q}\dfrac{\partial\hat{\mathbf{r}}}{\partial\alpha_\ell}$, which involves the differentiation of the pseudoinverse matrix $\mathbf{\Phi}^+$ if one proceeds from Eq. (5.30). This subject is treated in great detail in [25, 40] but with a prolific use of mathematical symbols. In the following, we give a simpler derivation.

Comparing Eq. (5.29) and Eq. (5.30), one can readily see that $\hat{\mathbf{r}}$ is obtained from Eq. (5.29) by treating $\mathbf{c}$ as a function of the α_ℓ's governed by the normal equation

$$\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{c} = \mathbf{\Phi}^H\mathbf{d}. \tag{5.33}$$

Then, taking the partial derivative in both sides of Eq. (5.29) with respect to α_ℓ under this constraint, we have

$$\frac{\partial \hat{\mathbf{r}}}{\partial \alpha_\ell} = \mathbf{\Phi}\frac{\partial \mathbf{c}}{\partial \alpha_\ell} + \frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c}. \tag{5.34}$$

Now, we need to find $\dfrac{\partial \mathbf{c}}{\partial \alpha_\ell}$ in terms of $\dfrac{\partial \mathbf{r}}{\partial \alpha_\ell}$ and $\mathbf{r}$. From Eq. (5.33), we obtain

$$\mathbf{\Phi}^H\left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c} + \mathbf{\Phi}\frac{\partial \mathbf{c}}{\partial \alpha_\ell}\right) = \left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\right)^H(\mathbf{d} - \mathbf{\Phi}\mathbf{c})$$

Rearrangements lead where to

$$\mathbf{\Phi}^H\mathbf{\Phi}\frac{\partial \mathbf{c}}{\partial \alpha_\ell} = -\left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\right)^H\mathbf{r} - \mathbf{\Phi}^H\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c} \tag{5.35}$$

Applying the QR decomposition to the matrix $\mathbf{\Phi}$, i.e.,

$$\mathbf{\Phi} = \mathbf{Q}^H\begin{bmatrix}\mathbf{U}\\ \mathbf{0}\end{bmatrix}$$

where $\mathbf{Q}$ is an $N \times N$ orthonormal matrix and $\mathbf{U}$ is an $R \times R$ upper triangular matrix, Eq. (5.35) becomes

$$\mathbf{U}^H\mathbf{U}\frac{\partial \mathbf{c}}{\partial \alpha_\ell} = -\left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\right)^H\mathbf{r} - \mathbf{\Phi}^H\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c}$$

After some simplifications and rearrangements, one gets

$$\begin{bmatrix}\mathbf{U}\\ \mathbf{0}\end{bmatrix}\frac{\partial \mathbf{c}}{\partial \alpha_\ell} = -\begin{bmatrix}\mathbf{U}^{-1} & 0\end{bmatrix}^H\left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\right)^H\hat{\mathbf{r}} - \begin{bmatrix}\mathbf{I}_R & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c}. \tag{5.36}$$

Multiplying both sides of Eq. (5.34) with $\mathbf{Q}$ gives

$$\mathbf{Q}\frac{\partial \hat{\mathbf{r}}}{\partial \alpha_\ell} = \begin{bmatrix}\mathbf{U}\\ \mathbf{0}\end{bmatrix}\frac{\partial \mathbf{c}}{\partial \alpha_\ell} + \mathbf{Q}\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c}. \tag{5.37}$$

Finally, substituting Eq. (5.36) into Eq. (5.37), we obtain

$$\mathbf{Q}\frac{\partial \hat{\mathbf{r}}}{\partial \alpha_\ell} = -\begin{bmatrix} \mathbf{U}^{-1} & \mathbf{0} \end{bmatrix}^H \left(\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\right)^H \hat{\mathbf{r}} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-R} \end{bmatrix} \mathbf{Q}\frac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}\mathbf{c}. \tag{5.38}$$

Because of the structure of $\mathbf{\Phi}$, the matrix $\dfrac{\partial \mathbf{\Phi}}{\partial \alpha_\ell}$ is sparse with at most $P_\ell + P_{\ell+1} + 2$ non-zero columns. Therefore, its computation and the subsequent multiplication with the vector $\mathbf{r}$ and $\mathbf{c}$ do not usually consititute a big computational burden. And also, the inverse matrix $\mathbf{U}^{-1}$ in the first term of Eq. (5.38) need not be found explicitly because $\mathbf{U}^H$ is lower triangular matrix and therefore, the whole term can be computed fairly accurately and efficiently with the forward substitution method [26].

Line Search Strategy

When each residual function $\hat{r}_\ell(\theta)$ can be reasonably approximated by a linear function in the search region, the steplength γ in the Newton direction can be satisfactorily chosen to be one. Otherwise, an "optimal" value of γ should be found in each step to speed up the convergence or to prevent divergence when $\|\hat{\mathbf{r}}(\theta_n + \mathbf{p}_n)\|^2 > \|\hat{\mathbf{r}}(\theta_n)\|^2$ is encountered.

Given a search direction $\mathbf{p}_n$, minimizing the residual $\|\hat{\mathbf{r}}(\theta_n + \gamma \mathbf{p}_n)\|^2$ with respect to γ is a one-dimensional problem, and can be solved with various search methods [21]. One particularly advantageous method is to make polynomial approximations to each individual function $\hat{r}_\ell(\theta_n + \gamma \mathbf{p}_n)$. For example, if $\hat{r}_\ell(\theta_n + \gamma \mathbf{p}_n)$ is approximated by a local quadratic model as

$$\hat{r}_\ell(\theta_n + \gamma \mathbf{p}_n) \approx a_n \gamma^2 + b_n \gamma + c_n \tag{5.39}$$

then we have

$$c_n = \hat{r}_\ell(\theta_n), \quad b_n = \hat{r}'_\ell(\theta_n) = [\nabla \hat{r}_\ell(\theta_n)]^T \mathbf{p}_n,$$

and

$$a_n = \hat{r}_\ell(\theta_n + \mathbf{p}_n) - \hat{r}_\ell(\theta_n) - [\nabla \hat{r}_\ell(\theta_n)]^T \mathbf{p}_n.$$

Correspondingly, the residual vector is

$$\begin{aligned} \hat{\mathbf{r}}(\theta_n + \gamma \mathbf{p}_n) &\approx \hat{\mathbf{r}}(\theta_n) + \gamma \mathbf{J}\mathbf{p}_n + \gamma^2 \left[\hat{\mathbf{r}}(\theta_n + \mathbf{p}_n) - \mathbf{J}\mathbf{p}_n - \hat{\mathbf{r}}(\theta_n)\right] \\ &= \mathbf{c}_n + \mathbf{b}_n \gamma + \mathbf{a}_n \gamma^2, \end{aligned} \tag{5.40}$$

and $\|\hat{\mathbf{r}}(\theta_n + \gamma \mathbf{p}_n)\|^2$ is a quartic in γ. One can easily find that the γ minimizing $\|\hat{\mathbf{r}}(\theta_n + \gamma \mathbf{p}_n)\|^2$ satisfies the following cubic equation

$$2\mathbf{a}_n^T \mathbf{a}_n \gamma^3 + 3\mathbf{a}_n^T \mathbf{b}_n \gamma^2 + (2\mathbf{a}_n^T \mathbf{c}_n + \mathbf{b}_n^T \mathbf{b}_n)\gamma + \mathbf{b}_n^T \mathbf{c}_n = 0.$$

Therefore, the optimal γ in each iteration can be found from the real roots of the above polynomial equation.

If $\mathbf{J}_Q$ and $\hat{\mathbf{r}}_Q$, instead of $\mathbf{J}$ and $\hat{\mathbf{r}}$, have been evaluated in calculating $\mathbf{p}_n$, then, $\mathbf{J}$ and $\hat{\mathbf{r}}$ can be replaced by $\mathbf{J}_Q$ and $\hat{\mathbf{r}}_Q$ in Eq. (5.40), respectively, and the rest remains the same.

If the γ thus found leads to $\|\hat{\mathbf{r}}(\theta_n + \gamma \mathbf{p}_n)\|^2 \geq \|\hat{\mathbf{r}}(\theta_n)\|^2$, one may consider using higher-order model for $\hat{r}_\ell(\theta_n + \gamma \mathbf{p}_n)$, or terminating the iteration since an extrema may have been reached.

5.5 Summary

Thus far, we have discussed in great detail how to estimate the LPA model parameters, particularly, the α_ℓ's. The basic ideas can be briefly summaried as: if the LPA model is appropriate and the data S/N is high, the Levinson algorithm is appropriate; otherwise, the various truncated-SVD-based methods should be used; if the estimate bias is still big, the non-linear least squares fitting procedure can further refine the estimates. The detailed procedure for estimating the LPA model parameters is given in the following diagram.

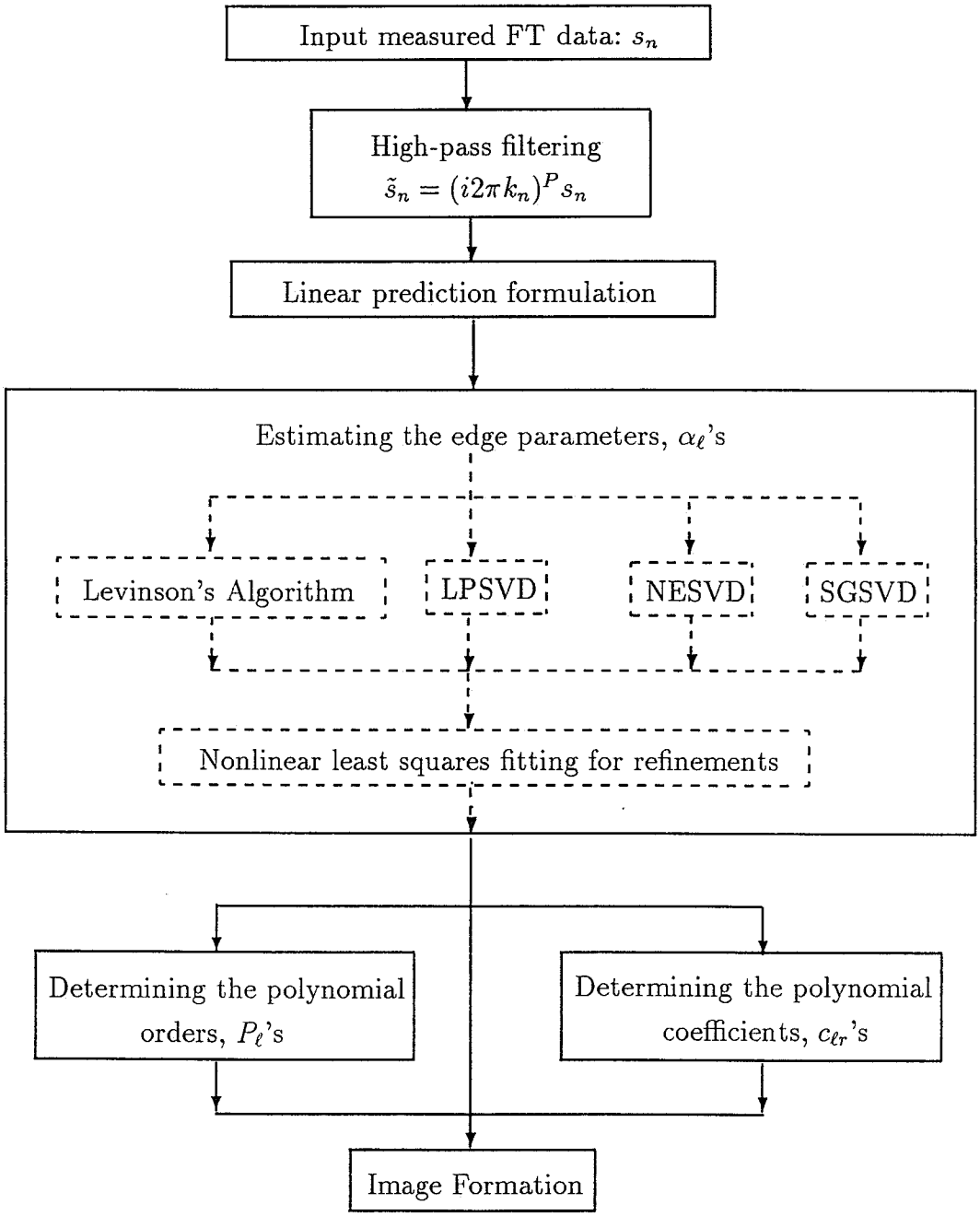


Figure 5.2: The procedure for LPA model-based image reconstruction.

Chapter 6

Results and Discussions

6.1 Introduction

The performance of the LPA model being applied to reconstructing object functions from a limited number of data points is analyzed in this chapter. Since theoretical analysis is very difficult due to the non-linear nature of the model, our discussions will be mainly based on results from simulated data and real magnetic resonance experimental data.

Some theoretical dissusions on the parameter sensitivity of the linear prediction based methods are available in the literature [36, 43, 67]; however, one may be disappointed to find out that the results are still too preliminary and some even full of errors [36]. In the following, instead of introducing many abstract formulas of little meaning we will demonstrate various points through concrete figures.

6.2 Reconstruction of Box-Car Object Functions

Box-car type of behavior is very common in medical and physical images representing sharp transitions from one object to another with different physical or chemical properties being imaged. However, it has been very difficult to accurately reconstruct such behavior with the FFT method due to the discontinuities involved. The LPA modeling approach offers an ideal solution to this problem since the box-car behavior can be easily described within the LPA model framework. Additionally, the LPA modeling approach has the following desirable features: 1) the capability to reconstruct boxes with dimensions less

than the Fourier pixel size, i.e., superresolution is possible; 2) the number of data points needed for reconstructing box-car type of object functions depends on the number of boxes present. For an object containing $N-1$ consecutive boxes (N edges) with unknown widths, amplitudes and locations, $2N$ data points may be sufficient to reconstruct it in the ideal noiseless case. Therefore, when an object to be imaged is structurally simple, such as a blood vessel, it is possible to reconstruct its sectional images accurately with only a few data points without causing any truncation artifacts; and 3) when the number of data points is larger than the number of parameters in the LPA model describing the object, an improvement of image S/N is possible by the smoothing effect of the least squares parameter estimation procedure.

In Fig. 6.1, the reconstruction results of a box-car object model (Fig. 6.1a) using the FFT and the zero-order LPA model with 64 data points are illustrated. In order to demonstrate the above mentioned capabilities, the model has been built with several selected features. First, the superresolution capability is tested by building into the model boxes with width smaller than a pixel, (for example, the left two boxes are a half pixel wide), as well as boxes with widths of non-integer multiple of the pixel size, (for example, boxes 4 and 6 are 2.75 pixel wide, and box 5 is 5.5 pixel wide). Second, the Gibbs ringing problem and hence the loss of resolution problem inherent to the FFT method are demonstrated by placing two boxes close together. Third, parameter sensitivity is tested by adding Gaussian noise to the data, and its dependence on amplitudes is revealed by building the boxes with different amplitudes.

In the noiseless case, the object is reconstructed perfectly, as expected, with the zero-order LPA model (Fig. 6.1b), while the FFT reconstruction (Fig. 6.1c) suffers significant ringing which renders the reconstruction useless for fine details (the small box). Therefore, when the zero-order LPA model is adequate to

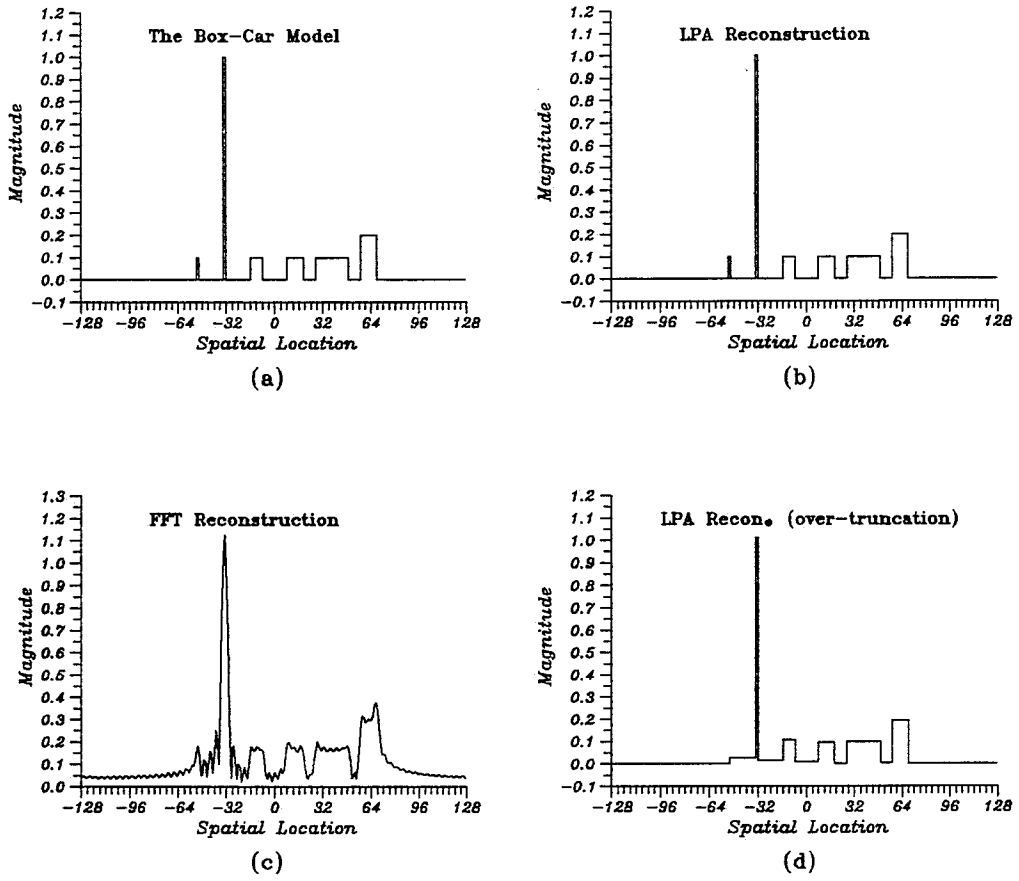


Figure 6.1: Reconstructions of a box car object model in (a) with 64 data points using (b) the zero-order LPA model with a correct SVD truncation order; (c) the FFT method; and (d) the zero-order LPA model but with over-truncation of the SVD.

describe an object, it can tremendously improve the image quality over the FFT method. In practice, noise may limit its performance, particularly its superresolution capability. Based on this model, a Monte-Carlo study on the model parameter sensitivity (for amplitude, width, and area) to Gaussian noise perturbation has been conducted. The exact values (in terms of the number of pixels) of the model parameters are shown in part I of Table 6.1, while the

Table 6.1:

Monte-Carlo study results of the box-car model in Fig 6.2a: I) the true parameters, and II) the mean (μ) and standard deviation (σ) obtained from 100 trials of noise with S/N=50:1.

Part I: the true parameters

BOX	AMPLITUDE	WIDTH	AREA
1	0.1	2.0	0.2
2	1.0	2.0	2.0
3	0.1	8.0	0.8
4	0.1	11.0	1.1
5	0.1	22.0	2.2
6	0.2	11.0	2.2

Part II: mean and standard deviations

BOX	AMPLITUDE ($\mu \pm \sigma$)	WIDTH ($\mu \pm \sigma$)	AREA ($\mu \pm \sigma$)
1	0.1108 ± 0.7921	2.1518 ± 0.8468	0.2021 ± 0.0074
2	1.0043 ± 0.0673	2.0018 ± 0.1403	2.0010 ± 0.0110
3	0.1000 ± 0.0010	7.9926 ± 0.1073	0.7992 ± 0.0089
4	0.1000 ± 0.0007	10.9965 ± 0.0765	1.0997 ± 0.0095
5	0.1001 ± 0.0006	21.9976 ± 0.0916	2.2023 ± 0.0131
6	0.2000 ± 0.0009	11.0033 ± 0.0421	2.1998 ± 0.0090

results obtained from the noisy data set with a S/N of 50:1 and 100 independent trials of noise are given in part II of Table 6.1. One can see from these results that the LPA model is rather robust with respect to noise of reasonable levels. This robust behavior is obtained through the utilization of the SVD in estimating the α_ℓ's and the least squares fitting procedure in finding the $c_{\ell r}$'s. One may also notice that the higher energy box components are less vulnerable to noise perturbation, since they are associated with the principal singular vectors.

Besides random noise, systematic errors may also arise in practical application of the LPA model even for reconstructing box-car object functions. One is due to a false selection of the truncation level $\hat{R}$ in the SVD step. For the object in Fig. 6.1a, there are 12 edges leading to 12 "non-zero" singular values in the SVD of the LP data matrix $\tilde{\mathbf{A}}_e$. If only 11 principal singular values and their associated singular vectors are kept, Fig. 6.1d results. This systematic error leads to the loss of the box component with the lowest energy, and a false extended box containing the missing energy appears due to conservation of energy. However, the nature of these deviations tends to be local because of the localized nature of basis functions used in the model, which is, of course, a powerful advantage over the FFT method.

6.3 Reconstruction of Higher-Order Behavior

Reconstruction of higher-order behavior is also possible with a higher-order LPA model. Presented in Fig. 6.2a is a model built of a combination of box-car, ramp, and parabolic functions. Assume that only 96 data points are available, the FFT reconstruction in Fig. 6.2b exhibits a considerable amount of Gibbs ringing. Using the second-order LPA model, an exact reconstruction, Fig. 6.2e, is obtained, as expected. When the zero-order LPA model (Fig. 6.2c) and the

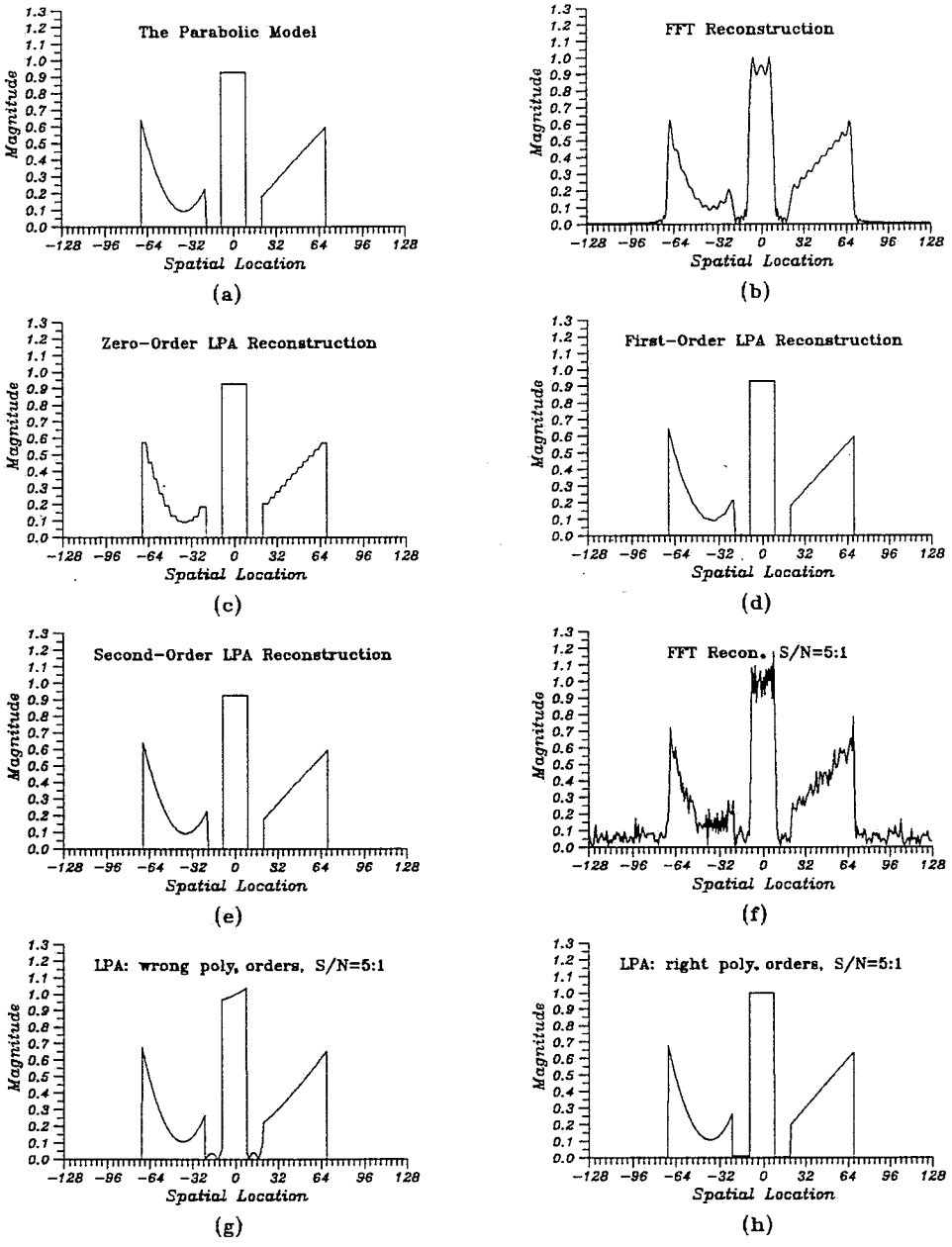


Figure 6.2: Reconstructions of a parabolic object model in (a) with 96 data points using (b) the FFT method; (c) the zero-order LPA model; (d) the first-order LPA model; (e) the second-order LPA model; (f) the FFT method with S/N=5:1; (g) the second-order LPA model with over-determined local polynomial orders and S/N=5:1; and (h) the second-order LPA model with correct local polynomial orders and S/N=5:1.

first-order LPA model (Fig. 6.2d) are used, systematic modeling error exists. However, the effect of this error has been minimized by the capability of these models to adaptively rebuild higher-order behavior using localized lower-order behavior. As can be seen from Fig 6.3c, the zero-order LPA model automatically accumulates a number of boxes to reproduce the ramp and parabolic type of changes fairly well. In Fig 6.3d, the first-order LPA model improves the reconstruction of the parabolic profile further by using localized ramp functions.

To test the robustness of the higher-order LPA model, Gaussian noise of moderate levels was added to the data to obtain the results shown in Figs. 6.3f-h and Table 6.2. In Fig. 6.2f, the FFT reconstruction with 256 data points and S/N=5:1 is presented. The noisy appearance of the FFT reconstruction is due to its incapability to handle noise, since the FFT method makes no discrimination between the true data and noise and, thus, the noise in the data domain is "faithfully" translated into the image noise. Using the LPA modeling approach with 96 data points and the same S/N, the noise is significantly reduced. In Fig. 6.2g, the image in each local region was purposely fit with a parabola and, therefore, the polynomial orders for the four local intervals in the right hand side were higher than those in the model. This over-determination of the local polynomial orders allowed some noise to be picked up in the corresponding intervals. When the orders are correctly chosen, the LPA reconstruction in Fig. 6.2h improves accordingly.

The Monte-Carlo study results in Table 6.2 were obtained with 100 trails of noise. The edge parameters, $\alpha_1 \sim \alpha_6$, were estimated with the high-pass filter order (P) being set to 1. Therefore, the estimation of the edge parameters is perturbed by both systematic error and Gaussian noise. From part I of Table 6.2, we can see that the estimates of the edges are fairly robust to these

Table 6.2:

Monte-Carlo study results of the parabolic model in Fig. 6.2a for I) the edge parameter, α_ℓ's, and II) the polynomial coefficients, $c_{\ell r}$'s.

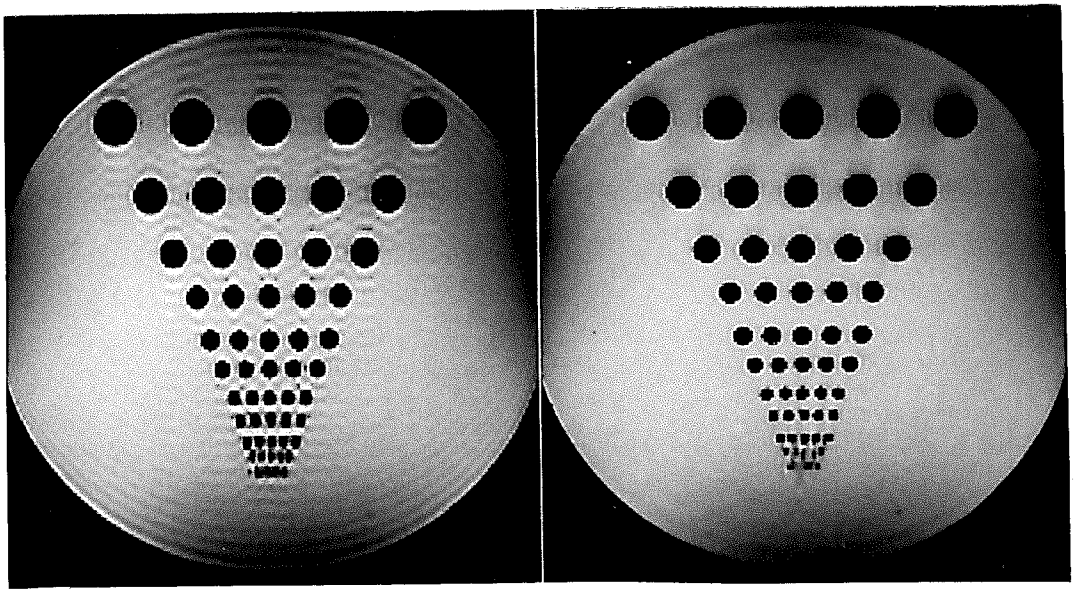
Part I: Edge Locations

true parameters (α_1 - α_6)	mean ± std (SNR=5:1)	mean ± std (SNR=10:1)	mean ± std (SNR=50:1)
-69.0	-69.24 ± 0.18	-69.22 ± 0.09	-69.22 ± 0.02
-21.0	-21.09 ± 0.51	-20.98 ± 0.24	-20.98 ± 0.05
-9.0	-8.98 ± 0.11	-8.98 ± 0.06	-8.98 ± 0.01
9.0	8.99 ± 0.11	8.99 ± 0.06	-8.99 ± 0.01
21.0	21.56 ± 0.62	21.34 ± 0.30	21.32 ± 0.06
69.0	69.08 ± 0.17	69.08 ± 0.09	69.08 ± 0.02

Part II: Polynomial Coefficients

true parameters ($c_{\ell r}$'s)	mean ± std (SNR=5:1)	mean ± std (SNR=10:1)	mean ± std (SNR=50:1)
c_{12}: 0.0375	0.0361 ± 0.003	0.0360 ± 0.001	0.0360 ± 0.0003
c_{11}: -0.0750	-0.0744 ± 0.005	-0.0744 ± 0.002	-0.0744 ± 0.0005
c_{10}: 0.1375	0.1419 ± 0.010	0.1406 ± 0.005	0.1405 ± 0.0011
c_{20}: 1.0000	1.0000 ± 0.012	1.0000 ± 0.006	1.0000 ± 0.0012
c_{31}: 0.0750	0.0725 ± 0.004	0.0734 ± 0.002	0.0734 ± 0.0004
c_{30}: 0.4167	0.4202 ± 0.007	0.4187 ± 0.004	0.4182 ± 0.0008

128 data points are not sufficient for MRI applications with the FFT reconstruction method, as observed by others [51, 62, 99]. Using the zero-order LPA model for reconstruction, the Gibbs ringing is substantially reduced without loss of image resolution. This dramatic image quality improvement of the LPA Modeling approach over the FFT method is because the phantom can be well-described by the LPA model; therefore, superimposing the model onto the reconstruction brought in much valid high-frequency information which has been assumed to be zero by the FFT method.



(a) (b)

Figure 6.3: Reconstructions of an MR phantom. The horizontal directions are reconstructed using the FFT with 256 data points, while the vertical directions are reconstructed with 128 data points using (a) the FFT and (b) the zero-order LPA model.

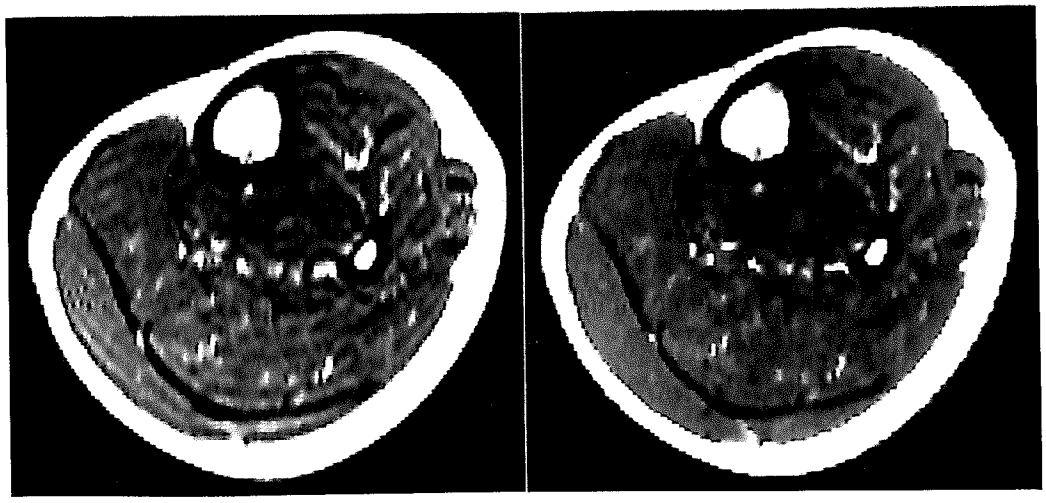
As another more practical example, data from a cut through a human leg is used. The images are first reconstructed in the horizontal direction with the FFT using 256 data points (in practice, acquiring 256 data points in this direction is often possible). Then, a 96 point FFT is applied in the vertical direction

errors. The largest bias (occurred for edge α_5) is less than a quarter of a pixel in terms of 96 data points used. The standard deviations are also smaller than a quarter of a pixel under all the values of S/N studied. As the signal-to-noise ratio increases, the standard deviations for these estimates reduced approximately linearly. However, the errors of these parameters between the mean values and the true values improve little due to the existence of the systematic error. The same observations also apply to the polynomial coefficients. From part II of Table 6.2, one can see that the relative errors for these polynomial coefficients, $|\Delta c_{\ell r}/c_{\ell r}|$, are well below the noise level. This is because the edges have been fairly accurately determined and the number of polynomial coefficients is much smaller than the number of data points available which allows room for the least squares fitting step to average out some of the data noise. One may also notice that the higher-amplitude edges, α_3 and α_4, are less perturbed than the other lower-amplitude edges as in the box-car model case. Correspondingly, the polynomial coefficients for the regions defined by these edges have smaller error and standard deviation, namely, the reconstruction in these region are more accurate. This is in the exact contrary to the FFT reconstruction where spurious ringing near the higher-amplitude edges causes bigger systematic errors.

6.4 Some Practical Examples

To test the general applicability of the LPA modeling approach, real magnetic resonance experimental data are used. In Fig. 6.3, the reconstructions of a high resolution phantom using 256 data points in the horizontal direction and 128 data points in the vertical direction are presented. The Gibbs ringing resulting from the sharp edges is clearly seen in the FFT reconstruction (Fig. 6.3a), which propagates throughout the entire image in the vertical direction and renders the image useless. This indicates that when sharp edges are present,

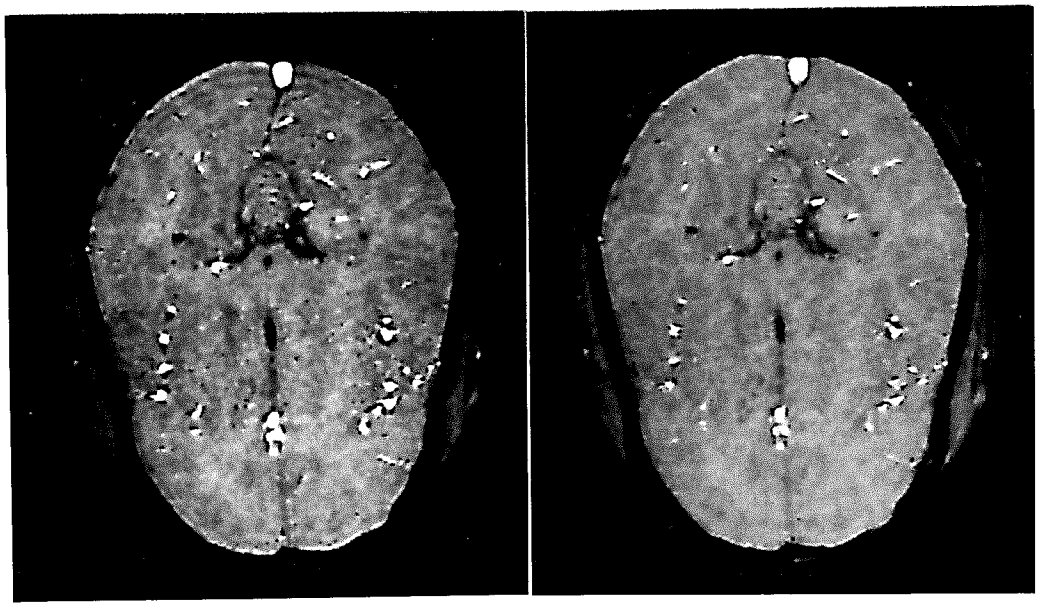
to get Fig. 6.4a. From this image, one can again appreciate the magnitude of the spurious ringing from the high intensity signal in the fat at the edge of the leg, and understand why people are so reluctant to reduce the imaging time by eliminating the high frequency phase encoding measurements. Using the LPA model to constrain the reconstruction, a relatively high quality image, Fig. 6.4b, is obtained. Since this image contains many higher-order features, the zero-order LPA model is not adequate. However, with the sharp edges being accurately recovered by the zero-order LPA model, reconstructing the higher-order smooth features is less of a problem. By properly re-introducing these higher-order features back to the model under the data consistency constraint, the image in Fig. 6.4b is reasonably well reconstructed for only 96 data points. Hence, in practice, we can always use lower-order LPA model to do the first reconstruction without worrying about lossing higher-order features. This property bodes well for the application of the LPA model to approximate general object functions.



(a) (b)

Figure 6.4: Reconstructions of a cut through a human leg. The horizontal directions are reconstructed using the FFT with 256 data points, while the vertical directions are reconstructed with 96 data points using (a) the FFT and (b) the LPA model.

With the FFT reconstruction method, the Gibbs ringing can also appear in the read direction (here, the vertical direction in Fig. 6.5) even with 256 data points when asymmetric echoes are acquired. Fig. 6.5a is a typical example where the Gibbs ringing along the sharp edges can be clearly seen. Since the image contains many blood vessels standing out of the background, the LPA model is ideal for representing these features. Using it to constrain the reconstruction, the image in Fig. 6.5b is obtained. As compared with the FFT reconstruction, it has significantly remove the Gibbs ringing along the tissue boundaries and sharpened up the blood vessels also.



(a) (b)

Figure 6.5: Reconstructions of a transversal cut through a human head. The horizontal directions (phase-encoding direction) are reconstructed using the FFT with 256 data points, while the vertical directions (read direction) are reconstructed with 256 data points from asymmetric echoes using (a) the FFT and (b) the LPA model.

Chapter 7

Future Research

7.1 Introduction

Even though the LPA modeling approach has shown superior performance over the classical FFT method for reconstructing piecewise continuous object functions from limited data, further research is necessary to make it practically useful as well as to make full use of its potential. In this chapter, we outline and briefly discuss some of the areas for future research.

7.2 Fast Algorithms

One of the major and obvious obstacles in practical applications of the LPA modeling approach now is its computational intensity. Particularly, when SVD is used for alleviating the ill-conditionedness of the linear-prediction formalism, it takes several hours MicroVax CPU time to reconstruct a two-dimensional image of size 128 by 256 with our current implementation of the algorithm. There are several ways to overcome this problem. The easiest but not reasonably the best way is to use a supercomputer. By making use of the parallel computing capability of such a computer, each row/column of an image can be processed much faster and also at the same time; thus, a reduction of the reconstruction time by a 100-fold is much possible. Of course, it may also be possible to develop a piece of dedicated microchip-based hardware for the algorithm instead of implementing it in a general purpose computers. Such a decision depends on many considerations beyond the scope of pure "scientific" research.

Another and perhaps better way to reduce the computation time is by developing fast algorithms. Since the matrices we are dealing with have the nice Toeplitz structure, there may possibly exist efficient specialized algorithms to calculate their SVDs. This problem has been actively attacked by several investigators, and some very important and promising results have been obtained [9, 15, 16, 96]. It would not be surprising if some practical algorithms resulted from such research in the near future.

There are also many other technical strategies that can be used in a practical implementation to speed up the reconstruction. For example, instead of finding a full SVD of the linear prediction matrix $\mathbf{A}$ or $\tilde{\mathbf{A}}$, we can perform a partial decomposition to find only the significant singular values and their corresponding singular vectors. If the number of edges is considerably smaller than the number of data points available , the calculation of this partial decomposition via some specialized algorithms [30, 93] can save a significant amount of computation time compared with a full decomposition using the QR algorithm [26]. In some cases, we may also use fast asymptotic solutions to completely replace the SVD step. For example, we can perform only the QR decomposition on $\mathbf{A}$ using the Householder transform or the Given transform method [26]. By truncating the upper triangle matrix $\mathbf{R}$ to a proper level, some of the measurement noise in the data can be removed and the effective condition number of $\mathbf{A}$ will also be correspongingly improved. Since the QR decomposition is much less expensive than the SVD computationally, such an approximation will lead to a substantial gain in speed. It has been reported that this method performs equally well compared with the SVD-based methods in some cases in spectroscopic data analysis [88], however, further investigations are necessary to understand how it performs in general before it can be used to substitute for the SVD.

7.3 Higher Dimensional Modeling

Another line of research should be directed to further improving the performance and utility of the LPA model. Thus far, we have concentrated our discussions only on the one-dimensional modeling case. Extension of the LPA modeling concept to higher dimensions will be very useful if computationally feasible, since on one hand we practically deal with data sets of more than one-dimension and on the other hand, higher dimensional modeling can potentially better constrain the reconstruction and enable the data to be used much more effectively.

Some work has been done in the area of two-dimensional ARMA modeling for spectrum estimation [54], and its application to magnetic resonance image reconstruction is under investigation. Since the LPA model offers better approximation to piecewise continuous object functions, its extension to higher dimensions would be more effective than the ARMA model if computationally possible. Preliminary studies appear to confirm such a feasibility, but much work remains to be done to make it become reality.

7.4 Optimal Coverage of the Fourier Space

In many imaging modalities, such as magnetic resonance imaging, we often have great flexibility in designing the experimental procedure to acquire the data. Clearly, as long as the sampling rate satisfies the Nyquist criterion, we can sample the data in the Fourier space (or the K-space) in as many ways as human imagination allows. However, given a fixed amount of data acquisition time and a pre-selected image reconstruction scheme, an optimal sampling scheme exists. Such a data acquisition scheme coupled with an appropriate reconstruction method can obtain the maximum amount of information, or give the best image resolution and signal-to-noise ratio as an outcome.

Since the FFT method has been dominating in the field of image reconstruction, it has been unnecessary to develop sampling schemes other than the uniform symmetric coverage. However, with more sophisticated reconstruction methods being developed to make use of *a priori* constraints, such a uniform symmetric coverage becomes backward. For example, in MRI, it is quite often that the phase of a complex object function can be fairly accurately estimated from some low frequency data [55], and therefore, by simply acquiring these data and the data in one of the half-planes of the K-space, the data in the other half can be recreated using the phase constraint. This is the so-called *half-Fourier reconstruction* problem. Clearly, in this case, it will be a great waste of scanning time to perform the uniform symmetric coverage of the K-space. When other *a priori* constraints are utilized, non-uniform coverage may also be possible and useful. No doubt, research along this line will prove very important and would also be productive.

Another issue with data acquisition is how to reduce the random noise. Traditionally, experiments are repeated to get several sets of data and then the data are averaged. Other more intelligent ways may also be possible if *a priori* knowledge is available. This topic deserves further research as well.

7.5 The Cone-Beam Problem

Another actively investigated limited data image reconstruction problem is the cone-beam problem, so called because the data are available in the form of projections (line integrals) only in a small cone-shape region as depicted in Fig. 7.1. This problem usually arises in X-ray tomography where some of the projection data are not acquired to minimize exposure to X-rays. Like the limited Fourier transform data inversion problem, the cone-beam problem is ill-posed [17] and requires *a priori* knowledge to constrain the reconstruction to

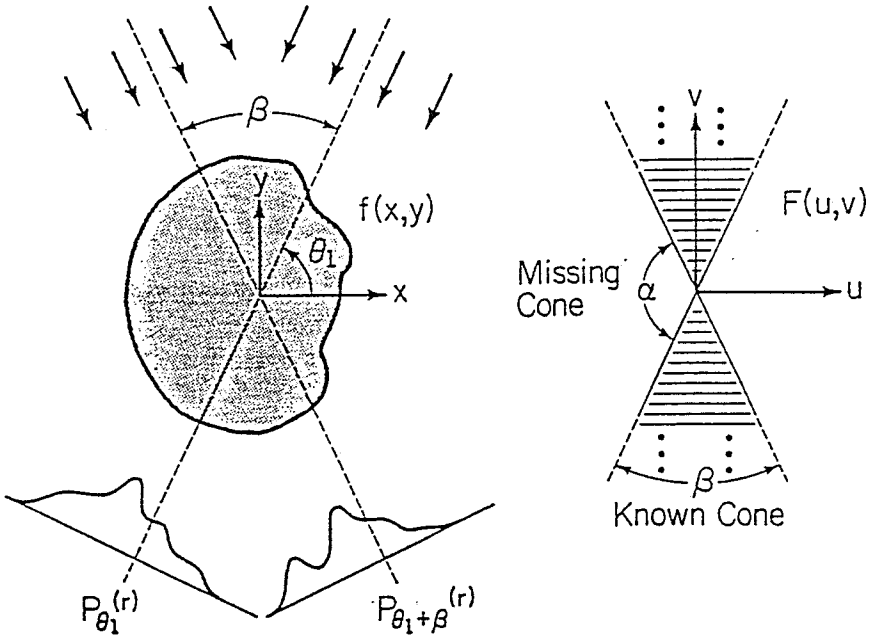


Figure 7.1: The cone-beam problem. Obtained from [33].

achieve better image quality. However, since the problem is inherently higher-dimensional, it has been much more computationally difficult to make use of *a priori* constraints. Thus far, extensive research has been focused on using non-parametric methods [24, 33, 66], particularly, the projection-onto-convex-set-based methods [74, 75], to incorporate *a priori* constraints, such as positivity and spatial-support-limitedness, but the success has been moderate due to the limitations of these methods as mentioned in Chapter 3.

Hanson *et al.* [31] broke out of the crowd and tried to use the parametric approach to regularize the problem, but only settled down to the linear models, possibly to avoid computational complications. With the rapid developments of computing technology, it might well be possible to solve this seemingly

awsome problem by using effective non-linear models. Bresler *et al.* [10] and Rossi *et al.* [69] have investigated this possibility and achieved some success with their statistical object models. The LPA modeling concept may also prove useful for recovery of sharp image features for this problem.

7.6 Utilization of Further Constraints

The LPA modeling approach proves effective for handling limited data when an object function can be *parsimoniously* represented with an LPA model. For some complex object functions, however, very high order LPA models may be required for a reasonable representation and, therefore, the efficiency of the model is lost. One consequence of using a higher-order LPA model is that many data points now are needed to uniquely determine the model parameters. More importantly, the use of a higher-order LPA model may cause instability. In these cases, other *a priori* constraints than the model itself are necessary to secure good reconstructions.

In practice, many other *a priori* constraints are indeed available, such as image magnitude lower and higher bounds, noise statistics, and phase/positivity. These constraints have been utilized in the non-parametric methods, and can also be incorporated into a model-based reconstruction scheme. How to modify the LPA model to make it flexible enough to handle these constraints deserves further investigations.

Chapter 8

Conclusions

The problem of reconstructing an image from limited Fourier transform data has been addressed in this thesis. Because of the ill-posedness of the problem, it has been recognized that *a priori* constraints are necessary to reduce/remove the truncation artifacts (the Gibbs ringing and the loss of resolution) associated with the conventional FFT method. In this thesis, a new parametric image reconstruction approach was introduced. With this method, an image to be reconstructed is represented in terms of a generalized series model, and then the reconstruction problem is transformed to a parameter estimation problem. Particularly, instead of using the classical basis functions from the eigenfunctions of a SL system, such as the set of complex exponentials, we proposed to use localized polynomials as the basis functions. This LPA model has several desirable properties. First, it can parsimoniously characterize piecewise continuous behavior and, therefore, the Gibbs ringing can be effectively removed or reduced. Second, the model parameters can be found efficiently using the linear-prediction formalism.

With these properties, the LPA model has been found very effective for reconstruction of piecewise continuous object functions. For example, in reconstructing box-car type of object functions, the LPA model is capable of recovering $N-1$ boxes exactly from $2N$ data points; For boxes larger than a pixel, their edges can usually be determined to an accuracy much smaller than the pixel size; Boxes of widths smaller than a pixel can also be picked up with a reasonable level of data S/N and, thus, superresolution is possible. For reconstruction of general object functions, the LPA model can fairly accu-

rately reconstruct sharp edges if any, and track and rebuild smooth features with localized lower-order polynomials. Furthermore, since the LPA model uses localized basis functions, local image properties or constraints can be easily incorporated into the reconstruction to improve image resolution and signal-to-noise ratio; more importantly, local approximation (modeling) error resulting from noise and possible systematic modeling offsets will be localized in the reconstruction rather than propagat throughout the entire image as in case of using non-local basis functions (e.g., the basis functions used in the Fourier series).

Although the LPA model is mainly developed for image reconstruction from limited Fourier transform data, other applications, such as image edge detection and general adaptive polynomial-based curve fitting, are also possible. One of the main drawbacks of the LPA model is its heavy computational burden which limits its application to small data sets. In addition, the LPA model is not very effective for characterizing higher-order behavior in general. Therefore, further research is necessary to improve these aspects of the model to enhance its performance and utility.

Bibliography

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1977.

[2] H. Barkhuijsen, R. DE Beer, W. M. M. J. Bovee, J. H. N. Creyghton, and D. VAN Ormondt, "Application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID," *Magn. Reson. Med.*, vol. 2, 86-89, 1985.

[3] H. Barkhuijsen, R. DE Beer, and D. VAN Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," *J. Magn. Reson.*, vol. 73, 553-557, 1987.

[4] C. W. Barnes, "Object restoration in a diffraction-limited imaging system," *J. Opt. Soc. Am. A*, vol. 56, 575-578, 1966.

[5] A. Ben-Israel, "On error bound for generalied inverses," *SIAM J. Numer. Anal.*, vol. 3, 585-592, 1966.

[6] M. Bertero, C. D. Mol, and E. R. Pike, "Linear inverse problems with discrete data. I: general formulation and singular system analysis", *Inverse problems*, vol. 4, 301-330, 1988.

[7] M. Bertero, C. D. Mol, and E. R. Pike, "Linear inverse problems with discrete data. II: stability and regularization," *Inverse problems*, vol. 4, 573-594, 1985.

[8] R. A. Blahut, *Fast Algorithms for Digital Signal Processing*, Addison-Wesley Publishing Company, Reading, Mass.,1985.

[9] A. W. Bojanczyk, R. P. Brent, and F. R. de Hoog, "QR factorization of Toeplitz matrices," *Numer. Math.*, vol. 49, 18-94, 1986.

[10] Y. Bresler and A. Macovski, "Three-dimensional reconstruction from projections with incomplete and noisy data by object estimation", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, No. 8, 1139-1152, 1986.

[11] J. R. Bunch, "Stability of methods for solving Toeplitz systems of equations," *SIAM J. Sci. Stat. Comput.*, vol. 6, No. 2, 349-364, 1985.

[12] S. F. Burch, S. F. Gull and J. Skilling, "Image restoration by a powerful maximum entropy method," *Comput. Vision Graphics Image Process.*, vol. 23, 113-128, 1983.

[13] J. A. Cadzow, "Spectral Estimation: an overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, 907-938, 1982.

[14] D. C. Champency, *A Handbook of Fourier Theorems*, Cambridge University Press, Cambridge, 1987.

[15] J. Chun, T. Kailath and H. Lev-Ari, "Fast parallel algorithms for QR factorization," *SIAM J. Sci. Stat. Comput.*, vol. 8, 899-913, 1987.

[16] G. Cybenko, "Fast Toeplitz orthogonalization using inner products," *SIAM J. Sci. Stat. Comput.*, vol. 8, 734-740, 1987.

[17] M. E. Davison, "The ill-conditioned nature of the limited angle tomography problem," *SIAM J. Appl. Math.*, vol. 43, 428-448, 1983.

[18] A. Dembo, "Signal reconstruction from noisy partial information of its transform," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-37, 1989.

[19] J. E. Dennis, Jr., "A user's guide to nonlinear optimization algorithms," *Proc. IEEE*, vol. 72, 1765-1776, 1984.

[20] J. Eriksson, "A note on solution of large sparse maximum entropy problems with linear equality constraints," *Math. Prog.*, vol. 18, 146-154, 1980.

[21] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, New York, 1987.

[22] B. R. Frieden, "Restoring maximum likelihood and maximum entropy," *J. Opt. Soc. Am. A*, vol. 62, 511-518, 1972.

[23] R. Gerchberg, "Superresolution through error energy reduction," *Opt. Acta*, vol. 21, 709-720, 1974.

[24] D. A. Girard, "Optimal regularized reconstruction in computerized tomography," *SIAM J. Sci. Stat. Comput.*, vol. 8, 934-950, 1987.

[25] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, 413-432, 1973.

[26] G. H. Golub and C. F. VanLoan, *Matrix Computations*, Baltimore, MD: John Hopkins University Press, 1983.

[27] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, 686-690, 1978.

[28] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *IEE proceedings*, vol. 131, Pt. F, 646-659, 1984.

[29] I. Guttman, V. Pereyra, and H. D. Scolnik, "Least squares estimation for a class of non-linear models," *Technometrics*, vol. 15, 209-218, 1973.

[30] R. Haimi-Cohen and A. Cohen, "Gradient-type algorithms for partial singular value decomposition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, 137-142, 1987.

[31] K. M. Hanson and G. W. Wecksung, "Local basis-function approach to computed tomography," *Appl. Opt.*, vol. 24, 4028-4039, 1985.

[32] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, 51-83, 1978.

[33] D. A. Hayner and W. K. Jenkins, "The missing cone problem in computer tomography," in *Advances in Computer Vision and Image Processing*, vol. 1, 83-144, edited by T. S. Huang, JAI Press Inc, 1984.

[34] J. C. Hoch, "Maximum entropy signal processing of two-dimensional NMR data," *J. Magn. Reson.*, vol. 64, 436-440, 1985.

[35] P. J. Hore, "NMR data processing using the maximum entropy method," *J. Magn. Reson.*, vol. 62, 561-567, 1985.

[36] Y. Hua and T. K. Sarkar, "Perturbation analysis of TK method for harmonic retrieval problems," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-36, 228-240, 1988.

[37] A. K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with application in spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-29, 830-845, 1981.

[38] E. T. Jaynes, "On the rational of maximum-entropy methods," *Proc. IEEE*, vol. 70, 939-952, 1982.

[39] M. A. Jenkins and J. F. Traub, "A three-stage variable shift iteration for polynomial zeros and its relation to generalized Rayleigh iteration," *Numer. Math.*, vol. 14, 252-263, 1970.

[40] L. Kaufman and V. Pereyra, "A method for separable nonlinear least squares problems with separable nonlinear equality constraints," *SIAM J. Numer. Anal.*, vol. 15, 12-20, 1978.

[41] S. Kawata, K. Minami, and S. Minami, "Superresolution of Fourier transform spectroscopy data by the maximum entropy method," *Appl. Opt.*, vol. 22, 3593-3598, 1983.

[42] S. M. Kay and S. L. Marple, Jr, "Spectrum analysis — A modern perspective," *Proc. IEEE*, vol. 69, 1380-1419, 1981.

[43] A. C. Kot, S. Parthasarathy, D. W. Tufts, and R. J. Vaccaro, "The statistical performance of the state-variable balancing and Prony's method in parameter estimation," *Proc. ICASSP*, 1549-1552, 1987.

[44] H. Krishna, and S. D. Morgera, "The Levinson recurrence and fast algorithms for solving Toeplitz systems of linear equations," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-35, 839-848, 1981.

[45] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-30, 833-840, 1982.

[46] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, 637-640, 1986.

[47] R. Kumaresan, "On the zeros of the linear prediction-error filter for deterministic signals," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-31, 217-220, 1983.

[48] S. Y. Kung, K. S. Akrun, and D. V. B. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Am. A*, vol. 73, 1799-1810, 1983.

[49] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, 190-191, 1973.

[50] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1974.

[51] R. B. Lufkin, E. Pusey, D. D. Stark, R. Brown, B. Leikind, and W. N. Hanafee, "Boundary artifact due to truncation errors in MR imaging," *AJR*, vol. 147, 1283-1287, 1986.

[52] J. Makhoul, "Linear prediction: a tutorial review," *Proc. IEEE*, vol. 63, 561-580, 1975.

[53] P. Mansfield and P. G. Morris, *NMR Imaging in Biomedicine*, Academic Press, New York, 1982.

[54] S. L. Marple, Jr., *Digital Spectral Analysis with Application*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1987.

[55] J. B. Mitchell, *Half-Fourier Imaging in Nuclear Magnetic Resonance*, M.S. thesis, Case Western Reserve University, Jan., 1989.

[56] K. Miller, "Least squares methods for ill-posed problems with a prescribed bound," *SIAM J. Math. Anal.*, vol. 1, 52-57, 1970.

[57] A. Mohammad-Djafari and G. Demoment, "Maximum entropy Fourier synthesis with application to diffraction tomography," *Appl. Opt.*, vol. 26, 1987.

[58] M. Z. Nashed, "Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory," *IEEE Trans. Antennas Propagat.*, vol. AP-29, 220-231, 1981.

[59] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits and Syst.*, vol. CAS-22, 735-742, 1975.

[60] A. Papoulis, "Maximum entropy and spectral estimation: a review," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-29, 1176-1186, 1981.

[61] A. Papoulis, *Signal Analysis*, McGraw-Hill Book Company, New York, 1977.

[62] D. L. Parker, G. T. Gullberg, and P. R. Frederick, "Gibbs artifact removal in magnetic resonance imaging," *Med. Phys.*, vol. 14, 640-645, 1987.

[63] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. astr. Soc.*, vol. 33, 347-366, 1973.

[64] R. Prost and R. Goutte, "Deconvolution when the convolution kernel has no inverse," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-25, 542-548, 1977.

[65] M. A. Rahman and K. B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-35, 1440-1454, 1987.

[66] R. Rangayyan, A. P. Dhawan, and R. Gordon, "Algorithms for limited-view computed tomography: an annotated bibliography and a challenge," *Appl. Opt.*, vol. 24, 4000-4012, 1985.

[67] D. V. B. Rao, "Perturbation analysis of a SVD based method for the harmonic retrieval problem," *Proc. ICASSP*, 624-627, 1985.

[68] S. J. Riederer, "Recent advances in magnetic resonance imaging," *Proc. IEEE*, vol. 76, 1095-1105, 1988.

[69] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects — Part I and II: Performance analysis and robustness analysis," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-32, 886-906, 1984.

[70] C. K. Rushforth, "Restoration, resolution, and noise," *J. Opt. Soc. Am. A*, vol. 58, 539-545, 1968.

[71] J. L. C. Sanz and T. S. Huang, "Support-limited signal and image extrapolation," in *Advances in Computer Vision and Image Processing*, vol. 1, 1-82, edited by T. S. Huang, JAI Press Inc, 1984.

[72] T. K. Sarkar, D. D. Weiner, and V. K. Jain, "Some mathematical considerations in dealing with the inverse problem," *IEEE Trans. Antennas and Propagat.*, vol. AP-29, 373-379, 1981.

[73] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative restoration algorithms," *Proc. IEEE*, vol. 69, 432-450, 1981.

[74] M. I. Sezan and H. Stark, "Image restoration by the method of convex projections: Part 2 — Applications and numerical results," *IEEE Trans. Med. Imaging*, vol. MI-1, 95-101, 1982.

[75] M. I. Sezan and H. Stark, "Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion," *IEEE Trans. Med. Imaging*, vol. MI-3, 91-97, 1984.

[76] S. Sibisi, "Two-dimensional reconstructions from one-dimensional data by maximum entropy," *Nature*, vol. 301, 134-136, 1983.

[77] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, and J. Staunton, "Maximum entropy signal processing in practical NMR spectroscopy," *Nature*, vol. 311; 446-447, 1984.

[78] B. C. De Simone, F. De Luca, and B. Maraviglia, " Maximum entropy method in phase-encoding NMR imaging," *Magn. Reson. Med.*, vol. 4, 78-82, 1987.

[79] J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: general algorithm," *Mon. Not. R. astr. Soc.*, vol. 211, 111-124, 1984.

[80] J. Skilling and S. F. Gull, "Algorithms and applications," in *Maximum Entropy and Bayesian Methods in Inverse Problems*, C. R. Smith and W. J. Grandy, Jr. (Eds.), 83-132, D. Reidel Publishing Company, 1985.

[81] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, 292-300, 1976.

[82] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — I ," *Bell Syst. Tech. J.*, vol. 40, 43-63, 1961.

[83] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — V: the discrete case ," *Bell Syst. Tech. J.*, vol. 57, 1371-1430, 1978.

[84] M. R. Smith, S. T. Nichols, R. M. Henkelman, and M. L. Wood, "Application of autoregressive moving average parametric modeling in magnetic resonance image reconstruction," *IEEE Trans. Med. Imaging*, vol. MI-5, 132-139, 1986.

[85] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, 26-36, 1980.

[86] J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-29, 230-237, 1981.

[87] G. W. Stewart, "On the perturbation of pseudo-inverses, projections and linear least squares problems," *SIAM Rev.*, vol. 19, 634-662, 1977.

[88] J. Tang, C. P. Lin, M. K. Bowman, and J. R. Norris, "An alternative to Fourier transform spectral analysis with improved resolution," *J. Magn. Reson.*, vol. 64, 343-346, 1985.

[89] A. N. Tikhnov and V. T. Arsenin, *Solutions of Ill-posed Problems*, V. H. Winston & Sons, Washington D. C., 1977.

[90] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-32, 201-212, 1984.

[91] D. W. Tufts, and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, 975-989, 1982.

[92] D. W. Tufts, R. Kumaresan, and I. Kirsteins, "Data adaptive signal estimation by singular value decomposition of a data matrix," *Proc. IEEE*, vol. 70, 684-685, 1982.

[93] D. W. Tufts and C. D. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high resolution estimation of frequencies," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-31, 1046-1053, 1986.

[94] R. J. Vaccaro and A. C. Kot, "A perturbation theory for the analysis of SVD-based algorithms," *Proc. ICASSP*, 1613-1616, 1987.

[95] V. Viti, E. Massaro, L. Guidoni, and P. Barone, "The use of the maximum entropy method in NMR spectroscopy," *J. Magn. Reson.*, vol. 70, 379-393, 1986.

[96] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-31, 1218-1221, 1983.

[97] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-33, 387-392, 1985.

[98] P. A. Wedin, "Perturbation theory for pseudo-inverse," *BIT*, vol. 13, 217-232, 1973.

[99] M. L. Wood and R. M. Henkelman, "Truncation artifacts in magnetic resonance imaging," *Magn. Reson. Med.*, vol. 2, 517-526, 1985.

[100] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.*, vol. CAS-25, 694-702, 1978.

[101] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 — Theory," *IEEE Trans. Med. Imaging*, vol. MI-1, 81-94, 1982.

[102] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, 1-25, 1986.

[103] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, 129-132, 1987.

[104] X. Zhuang, E. Ostevold, and R. M. Haralick, "A differential equation approach to maximum entropy image reconstruction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-35, 208-218, 1987.