Reconstruction of Irregularly Sampled Interferograms in Imaging Fourier Transform Spectrometry

A Thesis
Presented to
The Academic Faculty

by

Jialin Tian

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

School of Electrical and Computer Engineering
Georgia Institute of Technology
February 2004

Reconstruction of Irregularly Sampled Interferograms in Imaging Fourier Transform Spectrometry

Approved by:

Monson H. Hayes III, Chairman

Douglas B. Williams

Russell M. Mersereau

Rick P. Trebino

William T. Rhodes

Date Approved: 3/5/04

To my mother, Yabin,

and to the memory of my grandmother, Peggy.

ACKNOWLEDGEMENTS

To my mother, Yabin, I could not have done it without her emotional, professional, and financial assistance through my college years. I want to thank her for her love, support and everything she has done for me. Many thanks to my father, Richard, for always being there for me when I needed help, and for cheering me up whenever I was feeling down.

My deepest appreciation goes to my thesis advisor, Professor Hayes, I am grateful for his guidance throughout my graduate school years. I want to thank him for believing in me when I was discouraged; showing me the right direction when I was lost; being patient with me when I was difficult and stubborn, and most of all, I want to thank him for his trust, understanding, and friendship. My sincere gratitude goes to Professors Rhodes and Mersereau who have taking their valuable time to review and comment on my work. Thank you to Professors Williams and Trebino for serving on my thesis committee. In addition, my thanks to Dave Johnson, Jim Dempsey, and Jack Stoughton at NASA Langley Research Center for their help on this project.

Heartfelt thanks to Dr. Mac Wisler at Computalog Ltd. who introduced me to real-world engineering problems when I was a sophomore student. To Janice Van Derbur, my piano teacher and my friend who taught me the importance of music. I always admire her boundless energy and enthusiasm for life. My appreciation to Professor Ozturk at North Carolina State University; I thank him for showing me the art of science, and the science in art.

Very special thanks to all my friends, and to those who have inspired me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

A/D	Analog to Digital
DFT	Discrete Fourier Transform
DN	Digital Number
EA	Evolutionary Algorithm
EM	Electromagnetic
ESS	Equal Space Sampling
ETS	Equal Time Sampling
FFT	Fast Fourier Transform
FPA	Focal Plane Array
FTS	Fourier Transform Spectrometer or Spectrometry
FWHM	Full Width at Half Maximum
GIFTS	Geostationary Imaging Fourier Transform Spectrometer
IFOV	Instantaneous Field of View
IFTS	Imaging Fourier Transform Spectrometer
ILS	Instrumental Line Shape
IR	Infrared
LW	Long-wave Infrared
MCT	Mercury Cadmium Telluride
MI	Michelson Interferometer
MW	Medium-wave Infrared
NASA	National Aeronautics and Space Administration
OOB	Out-Of-Band
OPD	Optical Path Difference
SNR	Signal-to-Noise Ratio
SW	Short-wave Infrared

TOA	Top of the Atmosphere
UV	Ultraviolet
ZPD	Zero Path Difference

LIST OF SYMBOLS

α_λ	Absorptance
B_C	Calculated real cold blackbody spectrum
B_H	Calculated real hot blackbody spectrum
B_S	Calculated real blackbody spectrum from a space scene
C_C	Uncalibrated observed complex cold blackbody spectrum
C_H	Uncalibrated observed complex hot blackbody spectrum
C_σ	Uncalibrated observed complex spectrum from an earth scene
C_S	Uncalibrated observed complex spectrum from a space scene
δ_λ	Optical depth
ϵ	Complex noise
$E(\sigma)$	Intensity spectrum as a function of wavenumber
E	Irradiance (in W m^{-2})
h	Planck's constant ($6.6260755 \times 10^{-34}$J s)
$I(t)$	Interferogram as a function of time
$I(x)$	Interferogram as a function of OPD
k	Boltzmann's constant (1.380658×10^{-23}J K^{-1})
λ	Wavelength (in cm or m)
$n\Delta t$	Uniform temporal sampling location
$n\Delta x$	Uniform spatial sampling location
ν	Frequency (in Hz)
L	Radiance (in W $m^{-2}sr^{-1}$)
Ω	Solid angle (in steradian, sr)
O_σ	Instrument emission
Φ	Radiant Flux (in J/s or W)
Q	Radiant Energy (in J)

ρ_λ	Reflectance
R_σ	Instrument responsivity
σ_b	Stefan-Boltzmann constant (5.67051×10^{-8}W m^{-2}K^{-4})
σ	Wavenumber (in cm^{-1})
τ_λ	Transmittance
t_n	Nonuniform temporal sampling location
ε_λ	Emittance
x	Optical path difference (in cm)
x_n	Nonuniform spatial sampling location

SUMMARY

A common problem that exists in FTS is concerned with how to compensate for sampling errors when an interferogram is sampled at nonuniform instants in the path-difference domain. These errors, due to various mechanical irregularities, are generally associated with a continuous scanning system, which samples the interferogram at either equal space or equal time intervals. In both systems, the accuracy of the reconstructed signal can be significantly compromised if no error correction is performed. For an IFTS, the off-axis effects also affect the accuracy of the sampling locations. In addition, if the nonuniform sampling locations are unknown, which is the case when a laser reference is not present, the reconstruction algorithm must be able to correct the sampling errors "blindly." The current technique for solving this problem in the FTS industry involves a low-pass interpolation/resampling process, which only has been applied to a single detector problem, and it does not offer a solution when sampling locations are unknown. Furthermore, this method does not allow any control over the quality of the outputs. Based on these facts, a reconstruction algorithm that provides optional controls over the quality of the reconstructed signals is needed when sampling locations are obtainable. On the other hand, a recovery method that is capable of correcting the irregular sampling errors without the knowledge of sample locations must be developed when no positional information is present. In both cases, these algorithms must be simple and efficient enough for applying to a multi-dimensional problem.

When sampling positions are available, two alternatives are presented in this thesis. The first method recovers the data using a truncated "sinc" interpolation, whereas

the second method solves the problem using a linear interpolation based on the interferogram's symmetry property. Each algorithm has its own unique strength: the linear interpolation method is easy to implement, highly efficient, and is able to produce exceptionally accurate results under low-noise conditions; the "sinc" interpolation method is more robust to noise, and is capable of controlling the output quality through an adjustable truncation window length.

In the case where sampling locations are unknown, an optimization problem with multiple objective and constraint functions is designed based on the spatial and spectral characteristics of the data measurement. This problem is solved using an evolutionary approach, in which potential solutions are competing to be the fittest individual in a simulated natural environment. Evolutionary algorithms are stochastic global optimization methods that were founded on the principles of natural selection and adaptation. They can work on a variety of function types and require the least amount of mathematic resources. Because of their simplicity, flexibility, and effectiveness, two evolutionary algorithms are developed and compared to obtain the most desirable solution in the reconstruction without reference case. One of which emphasizes the estimation of sampling offsets while the other attempts to recover the actual values of all correct samples. The choice between the two is made according to each interferogram measurement's spectral characteristics. The extensions of these algorithms are made to solve the multi-dimensional array problem.

CHAPTER 1

INTRODUCTION

The study of the Earth's atmospheric radiation has been one of the most important scientific subjects for years. Various remote sensing instruments were designed for the purpose of observing the energy exchange mechanisms between the atmosphere and radiation. One particular optical instrument called a Fourier transform spectrometer (FTS) has played a significant role in the development of remote sensing spectrometry. A FTS is designed to record the Fourier transform of a broadband spectrum using a photo-sensitive detector, and the actual spectrum can be obtained by numerical transformation afterward. In recent years, the advancement of solid-state detector technology has made it possible for the development of an Imaging FTS (IFTS). Instead of using a single detector element, an IFTS utilizes a detector array and is capable of collecting high-resolution spectral images across a large ground area. The NASA's New Millennium Program GIFTS (Geostationary Imaging Fourier Transform Spectrometer) is an example of a high-resolution IFTS.

A common problem that exists in FTS is concerned with how to compensate for sampling errors when an interferogram is sampled at nonuniform instants in the path-difference domain. These errors, resulting from various mechanical irregularities, are generally associated with a continuous scanning system, which samples the interferogram at either equal space or equal time intervals. In both systems, the accuracy of the reconstructed signal can be significantly compromised if no error correction is performed. For an IFTS, the off-axis effects also affect the accuracy of the sampling locations. Furthermore, if the nonuniform sampling locations are unknown, which is the case when a laser reference is not present, the reconstruction algorithm must be

able to correct the sampling errors "blindly." The current technique for solving this problem in the FTS industry involves a low-pass interpolation/resampling process, which has been applied only to a single detector problem, and it does not offer a solution when sampling locations are unknown. Thus, a reconstruction algorithm that is capable of correcting the irregular sampling errors without the knowledge of sample positions must be developed. This is possible only because additional information on the data itself allows us to impose a set of constraints to the solution.

These two cases, i.e., with-references and without-references, are treated individually for both the single-pixel and array problems. When sampling positions are available, two alternatives are presented. The first method recovers the data using a truncated "sinc" interpolation, whereas the second method solves the problem using a linear interpolation based on the interferogram's symmetry property. In the case where sampling locations are unknown, an optimization problem with multiple objective and constraint functions is designed based on the spatial and spectral characteristics of the data measurement. This problem is solved using an evolutionary approach, in which potential solutions are competing to be the fittest individual in a simulated natural environment. Two evolutionary algorithms are developed and compared to obtain the most desirable solution. One emphasizes the estimation of sampling offsets while the other attempts to recover the actual values of all correct samples. The extensions of these algorithms are made to solve the multi-dimensional array problem.

The organization of this thesis is as follows: Chapter 2 provides the scientific and technological background of the atmospheric radiation and remote sensing. Chapter 3 explores the theoretical background, underlining problems, and the nature of the problem. Chapter 4 presents a survey of some of the current techniques for nonuniform signal reconstruction. In Chapter 5, potential application-based solutions for solving the problem are presented. Chapter 6 discusses the techniques developed for

solving the "with-reference" case. Two of the interpolative methods along with their performance are introduced in this chapter. In Chapter 7, the designs of the two evolutionary algorithms for reconstructing interferograms in the "without-reference" case are presented in full details. And finally, Chapter 8 concludes the thesis.

CHAPTER 2

ATMOSPHERIC RADIATION AND REMOTE SENSING SPECTROMETRY

To develop an application-based signal reconstruction algorithm that compensates for a particular type of instrumental error, the nature of the measurement itself and methods for obtaining it must be studied and understood. The observation of energy exchange mechanisms between the Earth's atmosphere and radiation is of no exception. The goal of this chapter is to glance through several major aspects of the remote sensing Fourier transform spectrometry as well as to provide some fundamental background information in the study of atmospheric radiation.

This chapter is organized as if it were answering a series of questions, and relevant topics are arranged and introduced in a general-to-specific fashion. Starting off with the most fundamental question, what is electromagnetic radiation? How does it behave when it is in contact with matter, and specifically, how does it interact with the Earth's atmosphere? We examine the basic concept of observing and recording atmospheric radiance via remote sensing methods, and then focus our interests on one particular type of remote sensing optical instrument called a Fourier transform spectrometer. The operational principle of the instrument along with its measurement characteristics are reviewed. Then, we ask the question, what is the current development trend in Fourier transform spectrometry? The answer is an imaging Fourier transform spectrometer, which offers the latest remote sensing techniques. Finally, details on the procedure and methods of processing the measured data are explored in the last section.

2.1 Electromagnetic Radiation

What is the nature of light? For centuries, scientists have been searching for an answer; inventing and re-inventing the properties of light. Today, we understand light as traveling electromagnetic radiation by its dual nature – possessing the characteristics of both waves and particles. In this section, some of the basic concepts of electromagnetic radiation are introduced. We will also investigate the interactions between electromagnetic radiation and matter.

2.1.1 Electromagnetic Waves and Spectrum

Electromagnetic (EM) radiation can be described as energy transmitted by alternating electric and magnetic waves, propagating through a vacuum at the speed of light, $c = 2.9979 \times 10^8$ m/s. The distance between adjacent peaks of an oscillating EM field observed at a given time is known as its *wavelength*, λ. Alternatively, EM radiation can be specified by its *frequency*, ν, which is related to the wavelength by

$$\nu = \frac{c}{\lambda}, \tag{2.1}$$

in units of Hz or cycles per second. Because λ is inversely proportional to the energy, $E = \frac{hc}{\lambda}$, radiation is often preferably expressed in wavenumber, $\sigma = 1/\lambda$ (in cm^{-1}). The conversion between these two systems is as follows: 1 cm^{-1} in frequency is approximately equal to 30 GHz [9].

A *spectrum* can be defined as the distribution of all EM radiant energies according to frequency. Figure 2.1 shows regions of the EM spectrum grouped by their radiation types, extending from the very energetic short-wave γ-rays to a narrow range of the visible region, at the other end of the spectrum lie the long-wave microwave and radio-frequency regions [19, 28]. Different types of radiation are produced and detected based on their unique properties. For instance, at the lower frequency end of the spectrum, EM energy fields can be produced and detected by electronic circuits, whereas in the visible region, human eyes are sensitive to radiation ranging

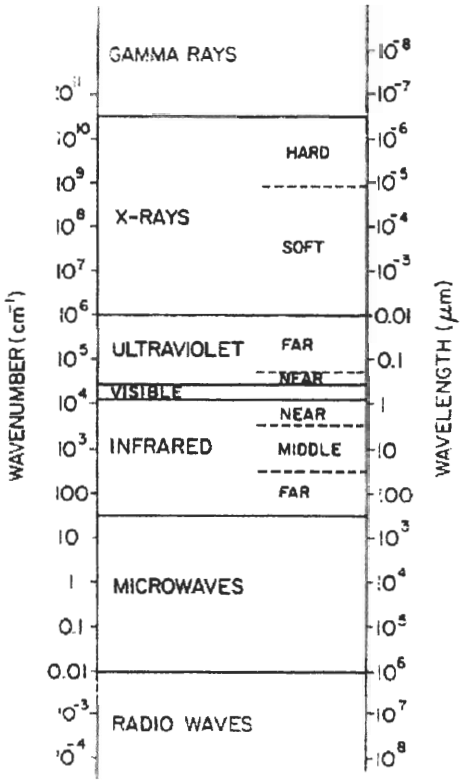


Figure 2.1: The electromagnetic spectrum (from [39]).

from 0.4 - 0.7 μm. Because most interactions between radiation and the Earth's atmosphere occur in the so-called "optical" window extending over wavelengths from the far ultraviolet (UV) to the far infrared (IR), approximately between 0.3-14 μm, most discussions in the following sections will focus on atmospheric radiative phenomena occurring in this particular region of the spectrum. Based on the properties of the atmosphere, the optical window can be further subdivided into a reflective region ($\sim$ 0.3-3 μm) and a thermal infrared region ($\sim$ 3-14 μm) [76]. The reflective region is mostly dominated by interactions between reflected solar radiation and atmospheric constituents, whereas the absorption and emission of thermal radiation by the atmosphere is mainly of interest in the thermal infrared region.

The fundamental quantity for measuring transported EM radiation is called *radiant energy* (Q) in units of Joules. The rate of radiant energy being transferred per unit time can be expressed in *radiant flux* (Φ) of units J/s or Watts, W. *Irradiance* (E) is the radiant flux incident upon a unit area, denoted by W m^{-2}. The monochromatic irradiance specifies the irradiance centered at a particular wavelength and is written as E_λ. When radiation is no longer coming from a single direction, irradiance can be determined only by integrating the incoming energy over all directions. The quantity of *radiance* (or *intensity*) L is defined as the irradiance per unit solid angle (in steradian) [39]. The solid angle Ω can be interpreted as the area of the projection

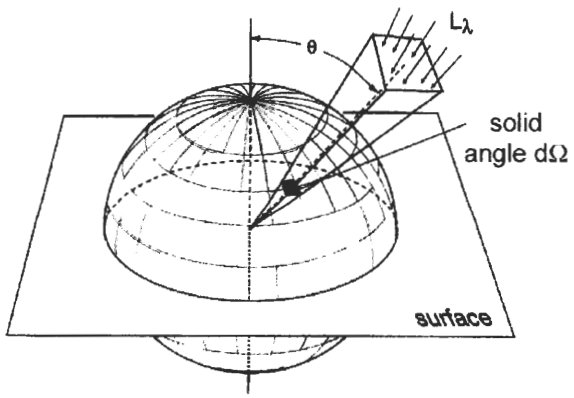


Figure 2.2: The concept of a solid angle is expressed as the relationship between radiance L_λ and irradiance E_λ (from [35]).

onto a unit spherical surface from an object at a given distance. A differential element of the solid angle is written as $d\Omega = \sin\theta d\theta d\phi = -d\cos\theta d\phi$, where θ and ϕ are zenith and azimuth angles, respectively. The solid angle surrounded by a point is

$$\Omega = \int_0^{2\pi} \int_0^{\pi} \sin\theta d\theta d\phi = 4\pi \text{ sr}. \tag{2.2}$$

For a spherical segment, the solid angle is

$$\Omega = \int_0^{2\pi} \int_{\theta_1}^{\theta_2} \sin\theta d\theta d\phi = 2\pi[\cos\theta_1 - \cos\theta_2]. \tag{2.3}$$

Using the small angle approximation for θ, the solid angle Ω subtended by a spherical cap is

$$\Omega = \int_0^{2\pi} \int_0^{\theta} \sin\theta' d\theta' d\phi = 2\pi[1 - \cos\theta] \approx \pi\theta^2. \tag{2.4}$$

The monochromatic irradiance E_λ can be obtained by integrating the radiance L_λ over a hemisphere (see Figure 2.2). Because L_λ represents radiation leaving or incident on an area normal to the beam, at an arbitrary angle θ, L_λ is multiplied by $\cos\theta$. Hence,

$$E_\lambda = \int_0^{2\pi} \int_0^{\pi/2} L_\lambda \cos\theta \sin\theta \, d\theta d\phi = \int_0^{2\pi} L_\lambda \cos\theta \, d\Omega \tag{2.5}$$

is the monochromatic irradiance [35, 41, 70].

2.1.2 Blackbody Radiation

EM radiation over a wide range of spectral regions is emitted from the surface of a heated body; every object that has a temperature above absolute zero emits EM energy. The spectral distribution of emitted energy depends on 1) the temperature of the emitting object and 2) the ability of a given substance to absorb and emit radiation at particular wavelengths. An ideal absorber, called a *blackbody*, is capable of absorbing and emitting energy at all wavelengths with 100% efficiency. For such an object, the spectral distribution is only temperature dependent, as shown in Figure 2.3. In 1900, Max Planck derived the theoretical solution that quantifies the blackbody radiation distribution; the *Planck function* expresses the radiance emission of a blackbody according to its wavelength as

$$B_\lambda(T) = \frac{2hc^2\lambda^{-5}}{e^{hc/\lambda kT} - 1}, \tag{2.6}$$

where h is *Planck's constant* and k is *Boltzmann's constant.* As $\lambda \to \infty$, $B_\sigma = 2kTc\sigma^2$; this is known as the *Rayleigh-Jeans distribution* [19, 39]. Alternatively, Equation (2.6) can be written as a function of wavenumber σ. This conversion would require the energy integrated over both systems to be equivalent, which is satisfied

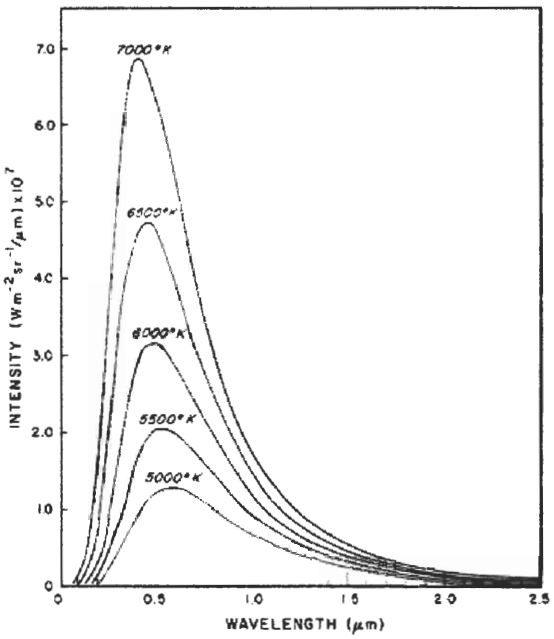


Figure 2.3: Blackbody radiation as a function as wavelength for a number of temperatures (from [41]).

by the following relationship:

$$B_\lambda(T)\, d\lambda = -B_\sigma(T)\, d\sigma. \tag{2.7}$$

Since $\lambda = -1/\sigma$, then $d\lambda/d\sigma = -\sigma^{-2} = -\lambda^2$, and this would give us the equation of $B_\sigma(T) = \lambda^2 B_\lambda(T)$. Using Equation (2.6), we can get

$$B_\sigma(T) = \frac{2hc^2\sigma^3}{e^{hc\sigma/kT} - 1}. \tag{2.8}$$

Blackbody emission has a single peak at λ_{max}, which can be determined by solving the equation $\partial B_\lambda(T)/\partial\lambda = 0$, the result is

$$\lambda_{max} T = 2897.9\ \mu\text{m K}, \tag{2.9}$$

commonly known as *Wein's displacement law.* Equation (2.6) represents the blackbody radiation at a single wavelength; the total radiant flux can be derived by integrating $B_\lambda(T)$ over all wavelengths as $B(T) = \int_0^\infty B_\lambda(T)\, d\lambda$. Its final solution has

the form of

$$B(T) = \sigma_b T^4, \tag{2.10}$$

and is called the *Stefan-Boltzmann law*, where σ_b is the *Stefan-Boltzmann constant.* Radiation intensity sometimes is expressed in units of temperature commonly known as the *brightness temperature.* At the given wavelength, this is the temperature required to match the measured intensity, I, to the Planck function. Therefore, the brightness temperature, T_b, can be obtained by inverting Equation (2.8), and

$$T_b = \frac{hc\sigma}{k}\left[\ln\left(\frac{2hc^2\sigma^3}{I} + 1\right)\right]^{-1}. \tag{2.11}$$

In the microwave portion of the spectrum, the brightness temperature can be simply approximated using the Rayleigh-Jeans distribution [70].

Unlike a blackbody, energy incident on real materials (graybodies) will experience a combination of interaction mechanisms, most of which can be categorized into transmitted, reflected, and absorbed energies. If they were normalized by the total incident energy to yield three related quantities, absorptance (α_λ), transmittance (τ_λ), and reflectance (ρ_λ), under the assumption that no scattering is present, then the law of thermodynamic equilibrium prevails, and

$$\alpha_\lambda + \rho_\lambda + \tau_\lambda = 1. \tag{2.12}$$

Similarly, the emittance of a substance is the total emitted energy normalized by the blackbody radiation at the given wavelength and is defined as $\varepsilon_\lambda = E_\lambda / B_\lambda$, where B_λ is the radiation emitted by a blackbody. The relation between the emittance and absorptance is given by *Kirchhoff's law* as $\alpha_\lambda = \varepsilon_\lambda$, which simply states that objects are as good absorbers as they are emitters [14, 19, 21, 35, 36, 39, 41].

2.1.3 Photons and Spectroscopy

Like most scientific theories that were discovered when well established principles could no longer explain certain phenomena, the quantum standpoint of view was

proposed by Max Planck to solve the problem of blackbody radiation, which could not be adequately explained by the classical theory of radiation. Based on the particle nature of light, EM radiation interacts with matter, absorbing or emitting energy only in quantized small bundles, known as *quanta*. Quanta that have the lowest amount of discrete energy can be viewed as the most elementary discrete units of the transported energy are called *photons* [14, 19, 21]. The energy being carried by a single photon is related to the frequency of the radiation by Planck's equation:

$$E = h\nu. \tag{2.13}$$

A transition is made between discrete energy states when atoms or molecules under excitation absorb or emit a photon. Alternatively, Equation (2.13) can be rearranged to express the relation between the wavelength and photon energy as

$$E = \frac{hc}{\lambda}, \tag{2.14}$$

where $hc = 1.986 \times 10^{-25}$ J m.

Spectroscopy is the study of spectral distributions of photon energy when radiation interacts with matter. Since various materials respond to radiation differently, depending on their structural properties, the spectral distribution shows the energy being exchanged between radiation and matter that can be used to determine the material in question. The interaction between molecules and radiation can be classified into five categories: ionization-dissociation, electronic, vibrational, rotational, and forbidden transitions. When a molecule absorbs certain amount of radiation at a given wavelength, its internal energy level makes a transition to a higher state [14, 21]. The type of the transition is frequency dependent. For example, in the IR portion of the spectrum, transitions are mostly associated with changes of vibrational and rotational states of a molecule; and electronic transitions occur in regions of UV and visible portions of the spectrum. Vibrational transitions in the IR spectral region are particularly interesting in the study of atmospheric spectroscopy; two of

the major absorbers in this section of the spectrum are carbon dioxide and water vapor [35, 36, 41].

Apart from the wave nature of radiation, in which the spectrum of radiation from a body is continuous, the emission spectrum of excited molecules is characterized by discrete lines. This, of course, is idealized for motionless molecules experiencing no external influences from other molecules. In non-ideal cases, spectral lines of finite widths are broadened by collisions and Doppler effect in the upper atmosphere and are mainly dominated by collisions in the lower atmosphere. The collision-broadened spectral lines resulting from molecular interactions generally take on the form of *Lorentz* lineshape

$$f_L(\nu - \nu_o) = \frac{\alpha_L/\pi}{(\nu - \nu_o)^2 + \alpha_L^2}, \tag{2.15}$$

where f_L represents the shape factor of a spectral line, α_L is the half width at the half maximum of the line and is a function of atmospheric pressure and temperature, ν is the frequency, and ν_o is the idea center frequency line [35, 41, 70]. Doppler broadening due to Doppler shifts resulted from molecular thermal motions has a lineshape function of

$$f_D(\nu - \nu_o) = \frac{1}{\alpha_D\sqrt{\pi}} \exp\Big[- \Big(\frac{\nu - \nu_o}{\alpha_D}\Big)^2\Big], \tag{2.16}$$

where α_D is a measure of the Doppler width of the line [30, 41, 70]. These two line shapes are plotted in Figure 2.4.

2.2 Interactions of Radiation with the Atmosphere

The atmospheric constituents interact with solar and terrestrial radiation to create the environment we are living in. Surrounded by nature, we often marvel at the beauty of a rainbow and aurora, and wonder why the sky is so blue and the grass so green. The answers can be found by exploring the radiative behaviors occurring in the atmosphere.

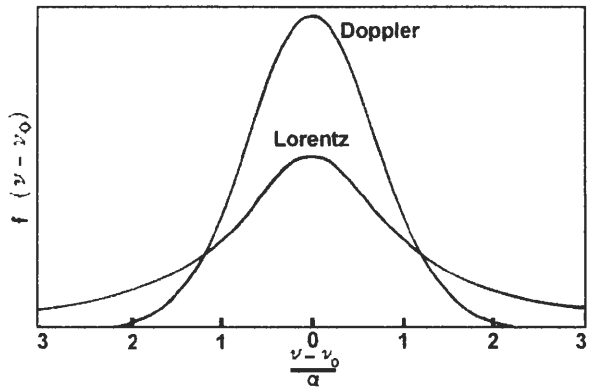


Figure 2.4: The Doppler and Lorentz lineshapes (modified from [70]).

2.2.1 Solar and Terrestrial Spectra

The radiation that reaches the Earth is originated at the surface of the Sun, referred to as the *photosphere*. Most solar radiation reaching the top of the Earth's atmosphere (TOA) can be considered as direct incident parallel beams, the rest includes lunar reflected and particle-scattered solar radiation. In general, solar radiation is quantified in irradiance because the solid angle subtended from the Sun at the Earth is considerably small because of the distance traveled; therefore, it is reasonable to assume the radiation is coming from a single direction. The Sun can be approximated as a blackbody with a surface irradiance of $E = 6.33 \times 10^7 \text{W m}^{-2}$. Applying the Stefan-Boltzmann law in Equation (2.10) yields the effective blackbody temperature of 5780 K. From Wein's displacement law, the solar energy has a maximum at about $\lambda_{max} = 0.5\mu$m (green in the visible region) [35, 39, 41]. For this reason, solar radiation is often referred to as the *short-wave* radiation because its energy is concentrated in the UV, visible, and near IR regions with shorter wavelengths. In practice, it is often convenient to convert this energy into photon units. By applying Equation (2.14), the photon energy E is obtained at 0.5μm ($E = 3.972 \times 10^{-19}$J); the equivalent irradiance in photon units can be then computed

as $E_{photon} = 6.33 \times 10^7 \text{W m}^{-2} / 3.972 \times 10^{-19} \text{J} \approx 1.6 \times 10^{26}$ photons $\text{s}^{-1}\text{m}^{-2}$ [19, 63].

Only a small portion of the Sun's surface radiation reaches the TOA (at about 1368 W m^{-2}), among which 40% of this incoming irradiance is absorbed by the Earth's surface. Based on the law of thermodynamic equilibrium, these energies must be balanced by the emission of outgoing terrestrial radiation. Thus, the terrestrial radiation spectrum can also be modeled as if it were emitted from a blackbody, but at a much lower effective mean temperature of 255 K, which peaks around 10μm. Similarly, terrestrial radiation is also known as the *long-wave*, IR, or thermal radiation because the wavelength range of maximum terrestrial emission is located in the IR range. Unlike solar radiation, terrestrial radiation does not emit energy in visible wavelengths because the Earth's temperature is too low. This is why it is dark during the nighttime. Thermal radiation is typically measured in radiance because in this case, radiation can come from all directions. Figure 2.5(a) shows normalized blackbody spectra plotted at solar and terrestrial effective temperatures. These two spectra do not experience any significant overlaps, which makes it possible to study their properties separately [41].

Among all incoming solar radiation at the TOA, about 20% is reflected back into space and the rest is absorbed by the Earth's atmosphere and surface. To maintain thermal equilibrium within both systems (the Earth and its atmosphere), this would require the absorbed energy to be reflected, scattered, or emitted outward [35].

2.2.2 Absorption, Scattering, and Emission of Radiation

The attenuation of solar radiation by the atmosphere is primarily dominated by absorption and scattering, whereas interactions between the atmosphere and thermal radiation are usually associated with emission and absorption [16, 30, 35]. The attenuated solar and thermal radiation spectra are shown in Figures 2.6 and 2.7. In

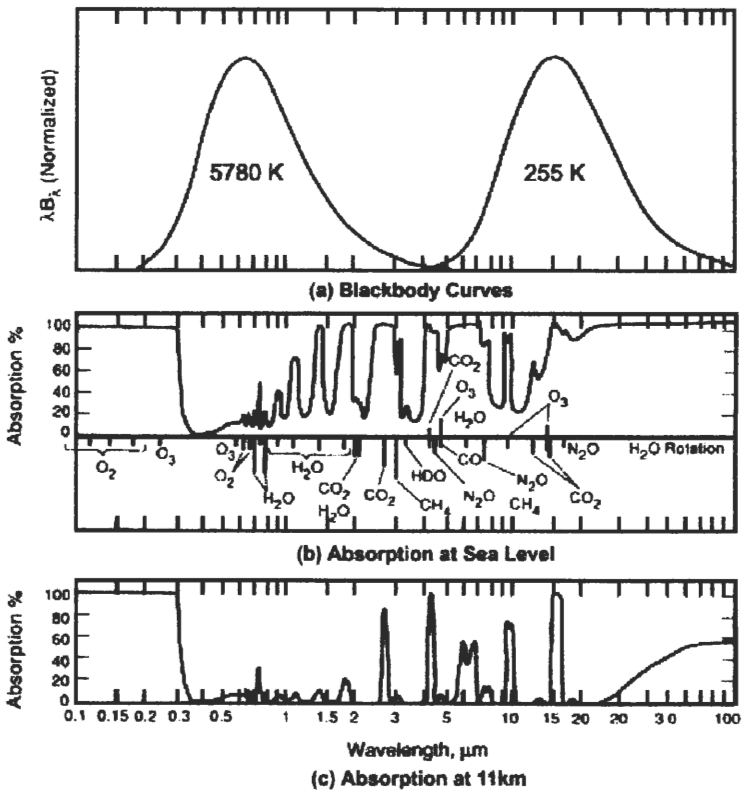


Figure 2.5: The attenuation of solar radiation. (a) Normalized blackbody spectra at solar and terrestrial effective temperatures. (b) Solar and thermal absorbance at sea level. (c) solar and thermal absorbance at 11 km, near the middle-latitude tropopause (adapted from [78]).

Figure 2.6, the top solid curve indicates solar irradiance at the TOA, and the bottom solid curve is estimated solar radiation at sealevel. The shaded area represents radiation absorbed by the atmosphere. In Figure 2.7, terrestrial radiance at sealevel is plotted against blackbody spectra at several temperatures (in dashed curves) since the envelope of the thermal radiation spectrum can be viewed as a combination of these curves [36, 39].

The Earth's atmosphere is composed of groups of gases, of both permanent and variable concentration, along with various solid and liquid particles, such as clouds,

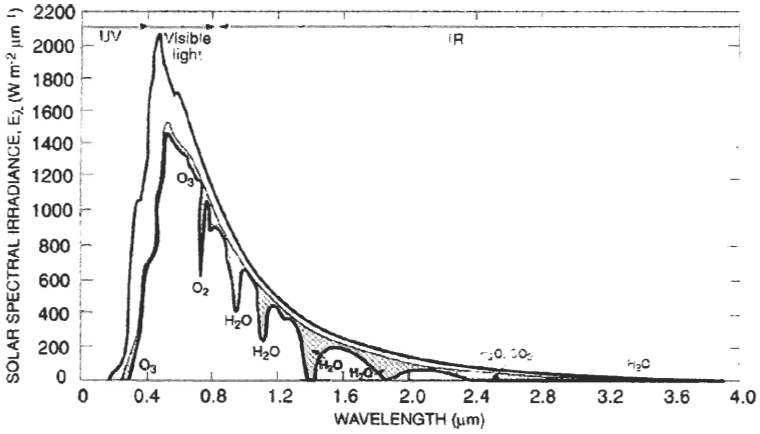


Figure 2.6: The solar irradiance and absorption with spectra at the TOA and sea-level. The shaded area indicates the absorption by the atmosphere (from [35]).

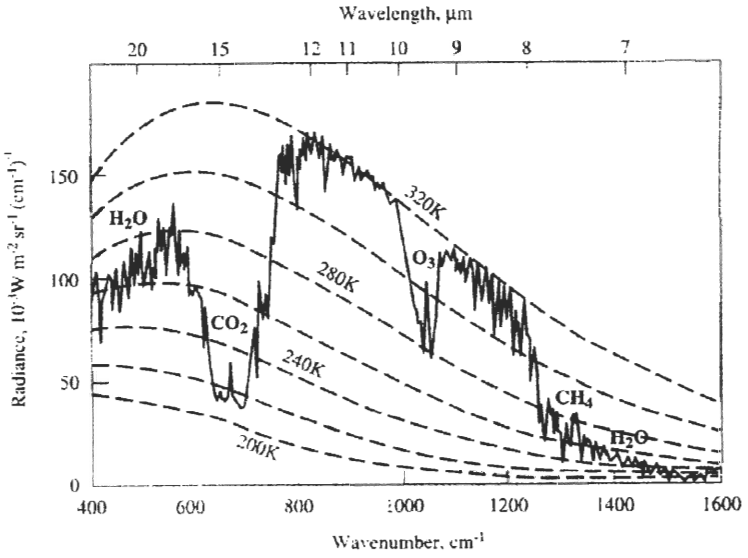


Figure 2.7: Terrestrial radiance with blackbody curves at different temperatures plotted in dashed lines (from [36]).

aerosols, water drops, etc. These constituents interact with radiation at different spectral bands according to their molecular structural properties. Figures 2.6 and 2.5(b) have most of the major absorbers identified. In the UV region, solar radiation is mostly absorbed by molecular and atomic oxygen and nitrogen species in the upper atmosphere; O_2 and N_2 absorb for wavelengths less than 0.2μm. In the near UV region ($\simeq 0.2-0.3\mu$m), ozone (O_3) occurs in the stratosphere as a result of electronic transitions. Radiation in the range of 0.3 and 0.4 μm reaches the surface. There is little absorption of solar radiation in the visible portion of the spectrum because the corresponding energies are too lower for electronic transitions and too high for vibrational transitions. Although ozone exhibits several weak absorption bands in the visible and near IR portions at about 0.44 to 1.18 μm, so does molecular oxygen, which has two weak bands at red wavelengths. Chief absorbers lie in the IR region are mainly composed of triatomic molecules such as H_2O, O_3, and CO_2. Water vapor absorbs solar radiation in bands centered at 0.94, 1.1, 1.38, and 1.87 μm due to the vibrational and rotational transitions. Carbon dioxide also has a number of weak absorption bands at 1.4, 1.6, and 2.0 μm. Absorption is often accompanied by the phenomenon of scattering, which occurs at all wavelengths in the spectrum. Atmospheric scattering is caused by the presence of gas molecules, aerosols, clouds, etc. When molecules have dimensions less than the wavelengths they affect, the amount of scattering is proportional to λ^{-4} and is referred to as *Rayleigh scattering*. As a consequence, the blue color of the sky is due to a relatively large amount of scattered blue light. Atmospheric constituents, such as water vapor and dust with dimensions that are of the order of the radiation wavelengths, will experience a different type of scattering, which is called *Mie scattering*. A third type of scattering, *non-selective scattering*, deals with particles that have even larger dimensions such as the cloud droplets; in this case, the scattering is wavelength independent [30, 35, 36].

In the IR region, thermal radiation is mostly absorbed by carbon dioxide, water

vapor, and ozone (Figures 2.7 and 2.5(b)). Carbon dioxide has a strong absorption band centered at 15 μm. Water vapor absorbs thermal radiation at 6.3 and 20 μm bands. From 8-12 μm, the atmosphere is relatively transparent, except for the O_3 feature at 9.6 μm. Most of the Earth's atmosphere is opaque due to the absorption of gases. However, in several regions, the atmosphere is relatively transparent, allowing radiation to propagate through; these regions are called *atmospheric windows*. In the visible and IR regions, these windows are located at 3.7 μm, 8.5-12.5 μm (with the exception of ozone at 9.3 μm). In Figure 2.7, the absorption spectrum of the Earth's atmosphere is plotted as a function of the wavelength, with most absorbers labeled [35, 36, 41].

Radiation traversed through the atmosphere experiences three basic types of processes. It can be absorbed, scattered, or emitted by gases and aerosols. In the scenario where a beam of solar radiation propagates through the atmosphere vertically downward, let the irradiance entering the atmosphere be E_λ, the thickness of the atmosphere be dz, and the existing irradiance be $E_\lambda + dE_\lambda$; then,

$$dE_\lambda = b_{e\lambda} E_\lambda dz, \tag{2.17}$$

with

$$b_{e\lambda} = b_{s\lambda} + b_{a\lambda}, \tag{2.18}$$

$b_{e\lambda}, b_{s\lambda}$, and $b_{a\lambda}$ are the *scattering*, *absorption* and *extinction coefficients* in units of m^{-1}. Note that both scattering and absorption can remove energy from the light beam; this attenuation is called *extinction*. The irradiance from altitude z to the TOA is called *optical depth* δ_λ, which can be obtained from the following integration:

$$\delta_\lambda \equiv \int_z^\infty b_{e\lambda} dz = \int_z^\infty \frac{dE_\lambda}{E_\lambda}. \tag{2.19}$$

This would give us the solution in the form of

$$E_{\lambda z} = E_{\lambda\infty} e^{-\delta_\lambda}, \tag{2.20}$$

and it is known as *Beer's law*. The transmittance, defined earlier, at between altitude z and the TOA can be expressed as

$$\tau_\lambda = \frac{E_{\lambda z}}{E_{\lambda \infty}} = e^{-\delta_\lambda}. \tag{2.21}$$

Naturally, the absorptance is $\alpha_\lambda = 1 - \tau_\lambda = 1 - e^{-\delta_\lambda}$. This unique nonlinear relationship gives rise to the absorption bands in an absorption spectrum. As δ_λ increases, discrete spectrum lines are broadened and merged together [30, 35].

Thermal radiation can be absorbed by the atmosphere, following an upward direction; the atmosphere itself can also emit long-wave radiation. In contrast to solar radiation, radiance L_λ is used for describing thermal radiation because of its diffused nature, instead of irradiance. Here, the upward absorption by the atmosphere is $-dL_\lambda = L_\lambda \alpha_{a\lambda} \rho dz$, where $\alpha_{a\lambda}$ is the *mass absorption coefficient* (equals to $b_{a\lambda}/\rho$) with ρ representing the density of the absorber. The emission from the atmosphere is $dL_\lambda = B_\lambda \alpha_{a\lambda} \rho dz$. The net contribution due to a layer of the atmosphere of thickness dz is the sum of these two terms, which produces the final solution

$$L_\lambda - B_\lambda = (L_{\lambda 0} - B_\lambda) e^{-\delta_\lambda \sec\theta}. \tag{2.22}$$

Notice that as δ_λ increases, the radiance L_λ approaches the blackbody radiation [35].

2.3 Remote Sensing of the Atmosphere

The term *remote sensing* refers to the technique of gathering information about certain objects at a distance. Atmospheric remote sensing emphasizes the measurement of EM radiation emitted or reflected from the Earth and its atmosphere in one or several spectral regions.

Distant observations of terrestrial radiative activities are often collected from deployed airborne or spaceborne remote sensing instruments. Although airborne remote sensing systems are relatively easy to launch, they can be difficult to control and maintain compared to spaceborne platforms. A satellite is an example of a spaceborne

vehicle that records EM radiation data while it circles the Earth like a man-made moon, along a predetermined orbit. The orbiting system is defined by two key parameters: the orbital altitude (the height above the Earth's surface) and inclination (the angle between the circle of the orbit and the equator). Two types of commonly used weather satellite orbiting systems include the low earth orbit system, which has a typical altitude of about 700 to 1500 km, and geosynchronous orbits at around 36,000 km with a fixed position relative to the Earth's surface. A geostationary satellite is similar to a geosynchronous satellite, but without an inclination angle [50, 76].

In addition to its navigational methods, another major characteristic of a remote sensing satellite is the spectral range it represents. Not all visible/IR frequencies are suitable for remote sensing applications because, at certain wavelengths, the spectrum is mostly opaque due to gaseous absorption. As mentioned in Section 2.1.1, the spectral region between 0.3-14 μm is mostly of interest for monitoring modulated EM radiation induced by atmospheric constituents. Several windows are available for measurements located at between 0.3-3 μm and 3-14 μm. The former (visible/near-IR) is appropriate for collecting reflected solar energy, whereas in the latter case (mid- and far-IRs), thermal radiation and re-emitted energy are measured. The remote sensing devices described so far are called *passive* because they measure the energy provided by external sources such as the Sun and the Earth [50, 76]. Instruments designed to operate at microwave frequencies are usually *active* because they collect the returned signals that have been transmitted by themselves. Radar is an example of an active remote sensing instrument.

As described in [63], atmospheric radiation can be interpreted as a continuous "rain" of photons. The purpose of a remote sensing instrument is to collect these photons. The data acquired by a digital remote sensing system usually takes the form of digital numbers (DN) that represent the radiance measurement, or equivalently, the number of photons being collected. The quality of these measurements is typically

characterized by four types of system resolutions: *spatial*, *radiometric*, *spectral*, and *temporal* [15, 50, 62, 76]. Spatial resolution defines the smallest object that can be resolved by a sensor as it integrates the radiation received across the instantaneous field of view (IFOV). The IFOV is the angle subtended by a single detector element, called a *pixel*, along the axis of the optical system, and is determined by IFOV $\cong D/f$, where D is the detector size and f is the focal length. The ground resolution element GIFOV is equal to $H \cdot (D/f)$, with H indicating the altitude of the sensor. The initial measurements made by the sensor are continuous electrical signals such as voltages. These voltages are then converted into DNs using an analog-to-digital (A/D) converter. The maximum number of different values that a sensor can distinguish is resolved by the dynamic range of a quantizer. Radiometric resolution is proportional to the number of bits Q used in this quantizer. For example, if the total number of available discrete DNs is 2^Q, then the dynamic range is between $[0, 2^Q - 1]$. Similarly, spectral resolution represents a sensor's ability to resolve spectral features [60, 62]. Besides the bandwidth of each spectral band and the total number of measured bands, spectral resolution is often expressed as the full width at half maximum (FWHM) of the instrument response function. Finally, temporal resolution is usually associated with the time period of repeated path converges.

2.4 Fourier Transform Spectrometry

In the previous section, remote sensing concepts and basic satellite operational principles were introduced. In this section, we turn our attention to the design and operational characteristics of an actual optical sensing system onboard a satellite. Various types of remote sensing devices exist. This thesis is mainly concerned with one particular optical instrument called a *Fourier transform spectrometer* (FTS). A spectrometer is a device that measures the spectral distribution of the EM energy radiated from a source. A FTS is a spectrometer designed to record the Fourier

transformation of a broadband spectrum, and the actual spectrum can be obtained via numerical transformation afterward.

2.4.1 The Michelson Interferometer

At the heart of a FTS is the Michelson interferometer (MI), originally invented for studying the speed of the light by A. A. Michelson, the Michelson interferometer has been adapted as a simple yet powerful tool in the study of Fourier transform spectroscopy.

The basic optical arrangement consists of four arms (Figure 2.8). The first arm contains a light source. The second has a fixed mirror, while the third arm has a moving mirror, which is traveling at a constant velocity. The fourth arm is open. At the intersection is a beam splitter, which transmits half of the incoming radiation and reflects the other half when an IR source is present. One of the separated beams strikes the moving mirror while the other strikes the stationary mirror; after reflecting off their respective mirrors, these light beams are recombined at the fourth arm. An *optical path difference* (OPD), x, is introduced as the difference in path lengths traveled by these two light beams. When the OPD is zero, which occurs as both mirrors are equally apart from the beam splitter, this condition is known as *zero path difference* (ZPD) [66].

Suppose a monochromatic source of wavelength λ and intensity E_0 is traveling through the MI. The interference of these two beams recombined at the beam splitter has an intensity of

$$I_0 = 2E_0\,[1 + \cos(\frac{2\pi x}{\lambda})] = 2E_0\,[1 + \cos(2\pi\sigma x)]. \tag{2.23}$$

The OPD varies as the mirror is moving at a constant velocity. As a result, an intensity interference pattern with dark and bright fringes is formed. The intensity reaches its maxima when $x = n\lambda$, with $n = 0, 1, 2, \ldots$, and this is known as the *constructive interference*, which describes the condition when all wavelengths of light

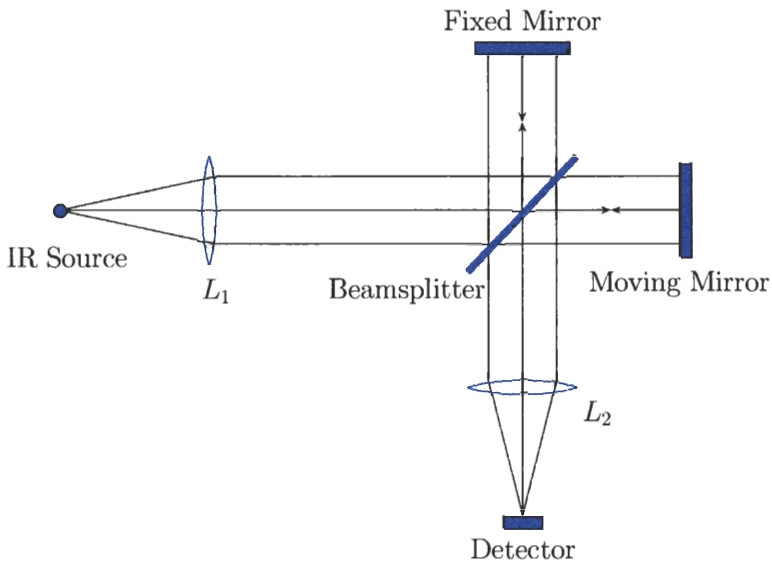


Figure 2.8: The Michelson interferometer as an IR spectrometer (modified from [9]).

are in-phase. In a similar fashion, the *destructive interference* occurs when $x = (n + \frac{1}{2})\lambda$; in this case, the intensity reaches its minima and all wavelengths of light are out-of-phase [28, 38, 63].

Now suppose the monochromatic source is replaced by a polychromatic light, which is composed of many frequencies, and has a spectrum of $E(\sigma)$. Then, at a given OPD x, the differential element of the intensity according to Equation (2.23) can be expressed as

$$dI_0(x, \sigma) = 2E(\sigma)[1 + \cos(2\pi\sigma x)]\, d\sigma. \tag{2.24}$$

By integrating both sides, the total intensity over the entire spectral band is

$$I_0(x) = 2\int_0^\infty E(\sigma)[1 + \cos(2\pi\sigma x)]\, d\sigma. \tag{2.25}$$

If we rearrange Equation (2.25) to obtain

$$I_0(x) = 2\int_0^\infty E(\sigma)\,d\sigma + 2\int_0^\infty E(\sigma)\cos(2\pi\sigma x)\,d\sigma, \tag{2.26}$$

then we could subtract the constant term of $I(0) = 2\int_0^\infty E(\sigma)\,d\sigma$ from $I_0(x)$, and derive an expression of

$$I(x) = I_0(x) - I(0) = I_0(x) - 2\int_0^\infty E(\sigma)d\sigma = 2\int_0^\infty E(\sigma)\cos(2\pi\sigma x)d\sigma. \tag{2.27}$$

The expression $I(x)$ is called an *interferogram*, which represents a plot of light intensity versus the optical path difference. To generate a complete interferogram, the moving mirror is moved back and forth once, which is known as a *scan*. The interferogram has its maximum centered at ZPD because all wavelengths of light constructively interfere at ZPD; as the mirror moves away from ZPD, the OPD increases. Consequently, rays at different frequencies are growing more and more out-of-phase or destructively interfering, causing the interferogram's intensity to decay rapidly. An illustrative diagram that explains the concept of broadband interfering can be seen in Figure 2.9. Equation (2.27) relates the interferogram and its spectrum by a cosine transform, which is suitable for an ideal interferogram that is symmetric about the ZPD. When complex noises or phase offset errors are present, the relationship is more appropriately expressed by a Fourier transform. Hence, we have the following transform pair:

$$\begin{aligned} I(x) &= \int_{-\infty}^{\infty} E(\sigma)e^{j2\pi\sigma x}d\sigma = \mathcal{F}^{-1}\{E(\sigma)\}, \\ E(\sigma) &= \int_{-\infty}^{\infty} I(x)e^{-j2\pi\sigma x}dx = \mathcal{F}\{I(x)\}. \end{aligned} \tag{2.28}$$

The FT pair reveals the fundamental principle of a FTS, which simply states that the desired spectrum is the Fourier transform of the measured interferogram. In addition, the FTS can also be viewed as a *frequency modulator* that translates optical

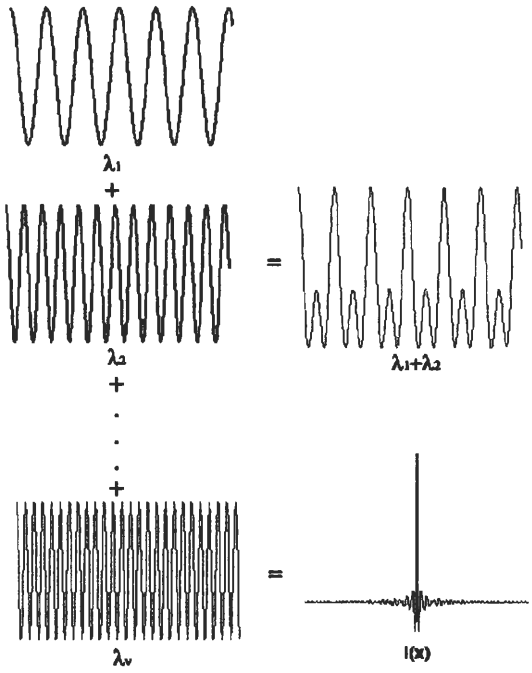


Figure 2.9: The concept of broadband interfering.

frequencies into electrical ones [9, 20]. To see the logic behind this operation, assume the mirror is traveling at a constant velocity of Vm/s, then the displaced OPD x during each scan can be written as $x = 2Vt$. Substitute this term into the frequency variable in Equation (2.28), and get $2\pi\sigma x = 2\pi\sigma(2Vt) = 2\pi\nu t$, where $\nu = 2\sigma V$ is the newly converted frequency variable in Hz within electrical frequency range, and σ is the original optical frequency variable in wavenumber.

2.4.2 FTS Spectral Properties

In practice, the spectrum obtained from the collected interferogram data often experiences some form of distortion due to various sources. As a result, the true spectrum is transformed and modified by a distortion function. Several major contributors responsible for this effect include 1) the natural spectral broadening due to molecular interactions described in Section 2.1.3, 2) the error caused by the truncation of an

ideal infinitely long interferogram, and 3) the self-apodization errors resulting from the off-axis effects of an extended source and a finite detector size. The first type of line contribution is a physical effect that reflects the gas properties of the measurement; therefore, it should not be removed or manipulated. The second and third types of contribution are due to instrumental errors and limitations; together, they produce the *instrumental line shape* (ILS) function [38]. This section focuses on the derivations of these two types of distortion functions.

Because it is impossible to acquire an interferogram of infinite length, the measurement is truncated to the total scan length of the OPD, L. In the space domain, this effect can be written as

$$I_L(x) = \Pi_{2L}(x)I(x), \tag{2.29}$$

where Π_{2L} is a rectangular window

$$\Pi_{2L}(x) = \begin{cases} 1, & |x| \leq L, \\ 0, & |x| > L. \end{cases} \tag{2.30}$$

And L is the total length of the OPD traveled by the mirror. In the frequency domain, the equivalent relationship is expressed as a convolution operation:

$$E_L(\nu) = W_L(\nu) * E(\nu), \tag{2.31}$$

with

$$W_L(\nu) = 2L \ \mathrm{sinc}(2\pi\nu L). \tag{2.32}$$

Therefore, the new spectrum line shape has a resolution of $\delta\nu \approx 1.207/(2L)$, and is defined by the FWHM (Full Width Half Maximum) of the sinc function. To reduce the relatively large sidelobes caused by a sinc function, a different kind of windowing function can be applied. These distortion functions are also known as the *apodization functions* [9, 38].

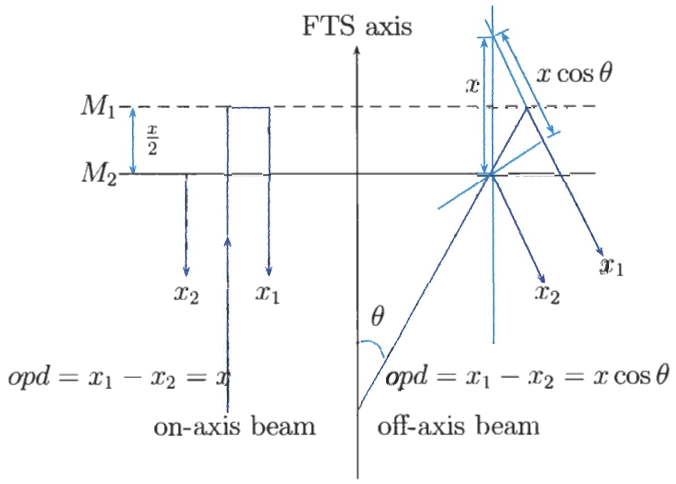


Figure 2.10: The OPDs for an on-axis beam and an off-axis beam with an off-axis angle θ (modified from [20]).

A practical Michelson interferometer often uses a collimator (for example, a convex lens) for collecting the light from the radiation source (Figure 2.8), and a detector of finite size is placed at the fourth arm to measure the collected intensity. Real light sources are never just points, but instead, they are extended sources with finite sizes. As a consequence, light is entering the interferometer at different angles with respect to the optical axis. This effect is known as the *off-axis effect*. Suppose the incoming ray makes an angle θ with respect to the optical axis, as shown in Figure 2.10; then, the OPD traveled by this off-axis ray equals $x\cos(\theta)$, with x as the OPD path along the optical axis in an ideal case. Therefore, these nonparallel (off-axis) rays in the interferometer experience a slightly shorter OPD than the on-axis paths. To see how the off-axis effects are reflected in the spectral line shape, first we use the small angle approximation $\cos(\theta) \approx 1 - \frac{\theta^2}{2} + \ldots$ to get $x\cos(\theta) \approx x(1 - \frac{\theta^2}{2})$. In Figure 2.11, assume the area of a circular radiation source is a, and the focal length is f. Then, the radius of the circle $r \approx f\theta$. Substitute the value for r into a to obtain an expression of $a = \pi r^2 \approx \pi(f\theta)^2$. The solid angle Ω of the source as seen from the collimator is $\Omega \approx \frac{a}{f^2} = \pi\theta^2$. Hence, $x\cos(\theta) \approx x(1 - \frac{\Omega}{2\pi})$. According to Equation (2.28), the

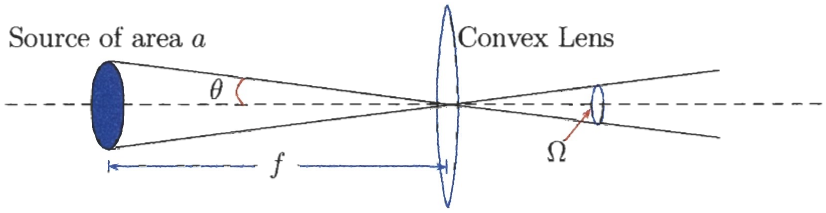


Figure 2.11: The solid angle for a circular radiation source (modified from [38]).

intensity at the detector for a point source is

$$I(x) = \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x} d\sigma; \tag{2.33}$$

for a circular source at an angle θ, the intensity is

$$I_\theta(x) = \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x \cos\theta} d\sigma; \tag{2.34}$$

then a small increment in solid angle $d\Omega$ would result in

$$dI_\theta(x) = d\Omega \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x \cos\theta} d\sigma \tag{2.35}$$

$$= d\Omega \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x(1-\frac{\Omega}{2\pi})} d\sigma. \tag{2.36}$$

The total intensity in the detector is obtained by integrating over the whole solid angle of the radiation source, and it is written as

$$I_\Omega(x) = \int_0^{\Omega} I_\theta(x) d\Omega' \tag{2.37}$$

$$I_\Omega(x) = \int_0^{\Omega} \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x(1-\frac{\Omega'}{2\pi})} d\sigma \, d\Omega' \tag{2.38}$$

$$= \Omega \int_{-\infty}^{\infty} E(\sigma) \operatorname{sinc}\left(\frac{\sigma x \Omega}{2}\right) e^{j2\pi\sigma x(1-\frac{\Omega}{4\pi})} d\sigma. \tag{2.39}$$

The off-axis effects are characterized by a shift of the measured wavenumber, $\sigma_\Omega = \sigma(1 - \Omega/4\pi)$ with σ_Ω as the measured wavenumber, and by the sinc function broadening of the spectral lines [38]. The former is due to the off-axis angle, whereas the

latter is caused by the integration over the solid angle associated with the finite detector size. Since the distortion is non-local in both space and wavenumber domains, and it depends on both the OPD and wavenumber, it is often referred to as the *self-apodization effect.*

2.5 Imaging FTS – GIFTS

The advancement of solid-state detector technology has made it possible for the development of an Imaging FTS (IFTS). Instead of using a single detector element, an IFTS utilizes an array detector consisting of multiple photosensitive elements. As a result, the IFTS is capable of collecting high resolution spectral images across multiple ground pixels.

2.5.1 IFTS – GIFTS Instrumentation

The elements in an array detector are usually arranged as rectangular or square, and they are placed in the focal plane of an IFTS. For this reason, it is often referred to as the *focal plane array* (FPA). The FPA contains two components: a detector array composed of photovoltaic (i.e., essentially no current is drawn) diodes, and a readout integrated circuit that is responsible for reading the photo current from each pixel of the detector array [6, 7].

The working principle behind an IR detector can be explained through the band theory of solids. For a semiconductor material, electrons under excitation can rise from the valence band to the conduction band if the photon energy during interactions is greater than the band gap, creating an obtainable current, voltage, or charge [54]. For an intrinsic semiconductor, electrons can be liberated from the valence band to the conduction band under sufficient thermal excitation. Because the photon energy is inversely proportional to the wavelength, it grows relatively weaker at longer wavelengths such that it becomes comparable to the thermal noise generated by thermal agitation of the electrons in the conductor. Since the thermal noise is temperature

dependent as $\sim kT$, the control of temperature is critical in order to reduce the thermal noise to a level where it does not affect the measurements significantly, i.e., the thermal noise is much less than the photon energy at the desired wavelength. In the IR wavelength range between 1 and 12 μm, two of the most commonly used detector semiconductor materials are indium antimonide (InSb) and mercury cadmium telluride (HgCdTe, or MCT). To reduce the thermal noise, a cooling method is required so that the detector has a typical temperature range of 50-80 K [9]. Once a voltage (or a current) signal is measured by each pixel, it is successively read through a cascaded circuitry; signals are then being sent out in a desired sequence to form a two-dimensional FPA image [7].

The NASA's New Millennium Program GIFTS (Geostationary Imaging Fourier Transform Spectrometer) is a high resolution sounding system designed to measure temperature, moisture, and wind profiles over the IR spectral bands 685 to 1130 cm^{-1} and 1650 to 2250 cm^{-1} (Figure 2.12) using advanced IFTS technologies. GIFTS

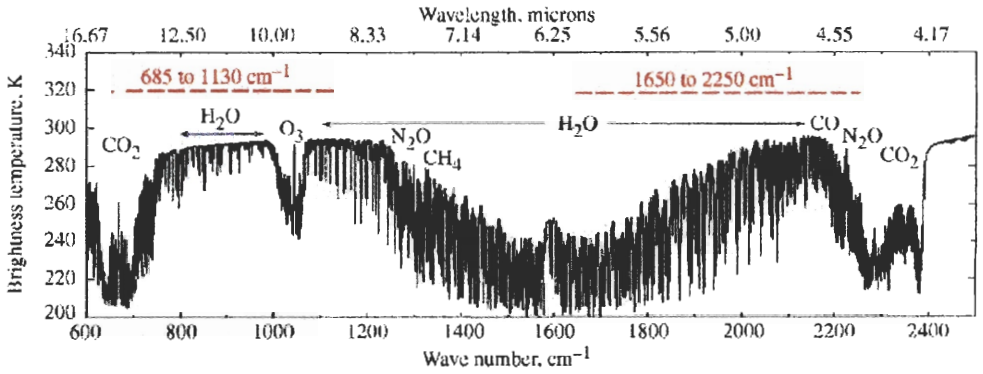


Figure 2.12: GIFTS spectral coverage by two detector arrays plotted as brightness temperature (from [49]).

gathers atmospheric spectral radiance measurements across 1724 spectral channels simultaneously with a resolution up to 0.3 cm^{-1} over a area of 512-km $\times$ 512-km for every 10 seconds. To achieve a wide range of spectral resolutions and a large area

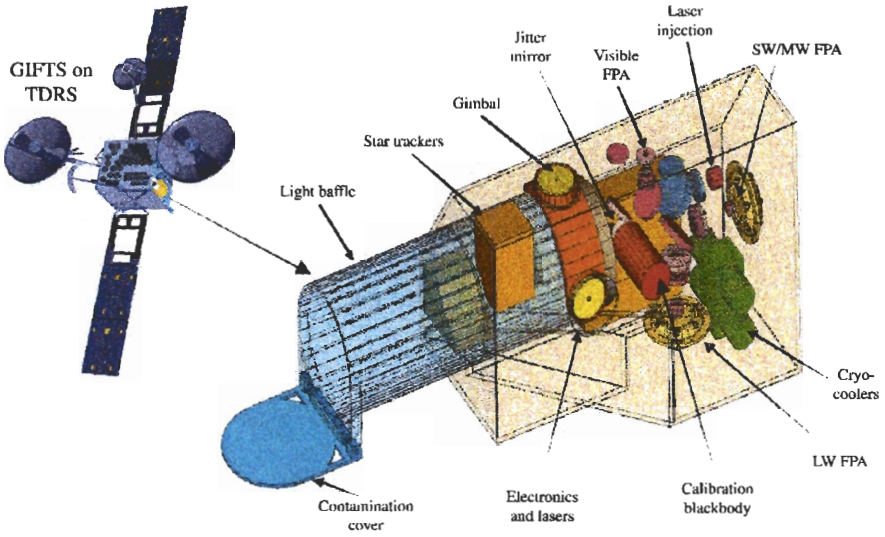


Figure 2.13: GIFTS electro-optical module (from [68]).

coverage, GIFTS uses an IFTS mounted on a geostationary satellite that employs two 128×128 MCT FPAs sensitive to LW emission of CO_2 and O_3 and SW/MW emission of H_2O, CO, N_2O and CO_2. These detector arrays are cooled to 60 K in order to reduce the thermal noise in the IR frequency regions. To realize high radiometric accuracy, GIFTS uses two on-board blackbody calibration sources, along with a deep space look. A layout of the GIFTS electro-optical module is shown in Figure 2.13. The collected radiance data is processed by the supporting electronics, which consist of required sampling mechanisms, a 14-bit A/D, and a high performance vector processor. A list of selected parameter specifications can be seen in Table 2.1 [49, 68, 69].

2.5.2 Hyperspectral Image Characteristics

The spectral data obtained from a high resolution imaging FTS is often classified as "hyperspectral." A *hyperspectral image* has successive frames corresponding to adjacent spectral channels, and each frame representing a two-dimensional spatial scene. Hyperspectral data sets are usually recorded over many narrow, contiguous

Table 2.1: Selected Specifications for GIFTS.

Configuration	Michelson-type; continuous scan
Spectral coverage	685–1130 cm^{-1} (LW) 1650–2250 cm^{-1} (SW/MW)
Scan interval	10 sec
Max. spectral resolution	0.3 cm^{-1}
Detector	MCT 128 × 128-pixel FPA (2) at 60 K
Path difference control	Stabilized 852-nm laser
IR footprint size	4 km
Sensitivity	0.2 $mW/cm^2sr\ cm^{-1}$ (LW) 0.006 $mW/cm^2sr\ cm^{-1}$ (SW/MW)
Calibration	Absolute ε (Tb) $< 0.2K$
IFOV	512 × 512-km
A/D	14-bit

bands across a wide range of the EM spectrum. In addition, hyperspectral images are generally oversampled such that the number of spectral bands is much higher than the required number of bands for identifying the materials in question. This would provide more detailed and accurate spectral images [8, 37]. Figure 2.14 illustrates the concept of a hyperspectral data set, which can be seen as a stack of spatial images over different spectral channels or as a stack of spectra arranged in an array of pixels. Because of the spatial-spatial-spectral dimensional nature of the hyperspectral data sets, they are often referred as image/data cubes [8]. In Figure 2.15(a), the interferogram for GIFTS center pixel is plotted for 2048 samples; in Figure 2.15(b), one frame of the intensity data from the GIFTS FPA is shown. It's known that imaging spectrometer remote sensing data are often correlated both spatially (intra-band) and spectrally (inter-band) with the tendency of a higher spectral correlation. The spatial correlations are caused by the homogeneities of ground features covered by neighboring pixels, while the unique spectral characteristics are in part due to the overlap of information across adjacent channels.

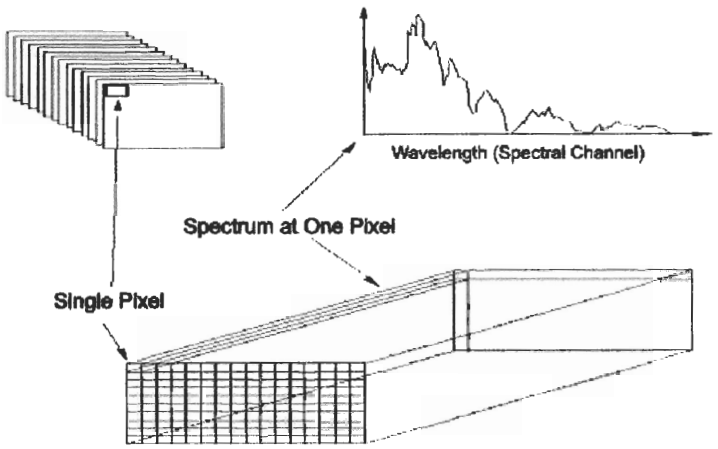
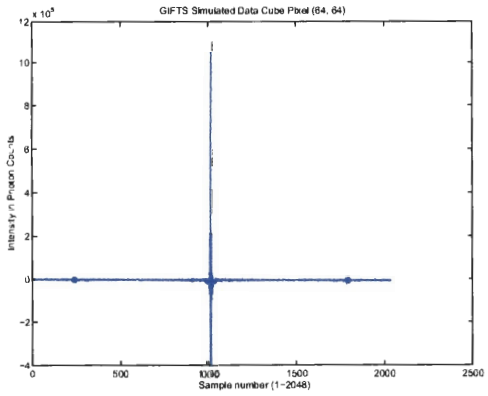
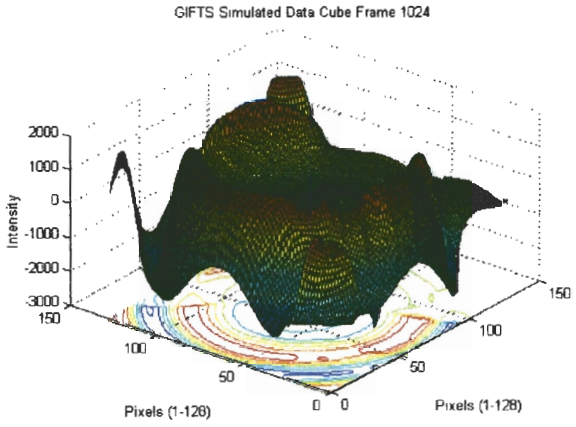


Figure 2.14: Hyperspectral data concept (modified from [8]).



(a) GIFTS simulated interferogram at pixel (64,64)



(b) GIFTS simulated FPA intensity at frame 1024

Figure 2.15: GIFTS simulated data cube with one pixel and one frame of data shown.

2.6 Processing of Measured Data

For a complicated remote sensing system, desired atmospheric quantities are not directly available from the measurements, instead, observed raw sensor data must

undergo a hierarchy of processing stages in order to extract useable environmental information required by users [1, 62]. One of the most important steps is the calibration of acquired measurements, i.e. raw data must be calibrated to correct instrumental offsets and errors. The calibrations for a FTS can be generally divided into three basic types: *radiometric*, *spectral*, and *geometric*. Radiometric calibration assigns absolute radiance values to the measured intensity; similarly, spectral calibration deals with the assignment of absolute wavenumbers to the spectrum. Finally, geometric calibration is typically concerned with the process of assigning latitude and longitude values to the measured spectrum. Since the geometric calibration is not closely related to FTS instrumentations, we will only discuss radiometric and spectral calibrations in the following sections.

2.6.1 From Raw Sensor Data to Environmental Profiles

The measured raw sensor data are usually received as DNs, which bear no physical meanings, therefore, incoming data packets must be reformatted before any further computation can be performed. This stage can be viewed as the pre-processing step because it is relatively independent from the measurement itself. The detected interferogram passes through the analog front-end electronics, then sampled by an A/D before being filtered, trimmed, and packed in the target processor, and downlink to the Earth. The received data (as DNs) must be loaded and sorted, and converted from interferograms to spectra. In the second stage, raw spectra are calibrated and assigned with absolute physical units according to calibration reference measurements. Details on FTS calibrations will be reviewed in the following section. Once the corrected spectra are obtained, the final stage involves the retrieval of desired environmental profiles such as temperature, moisture, and pressure. The study of atmospheric retrieval algorithms is a subject on its own, and it is well beyond the scope of this report. However, retrieval methods generally require the finding of an

inverse model and solving the inverse problem based on the assumed model [61].

2.6.2 Calibrations

Measurements made from any type of instrument will suffer from various distortions, for example, the data could be contaminated by random noises, distorted by instrumental gain and offset factors, or transformed by a linear/nonlinear operator. For this reason, an instrument must be calibrated against some form of reference so that measurements can be corrected using these calibration information. For a FTS, the uncalibrated spectrum can be modeled as

$$C_\sigma = \left[(F \cdot L)_\sigma + O_\sigma\right] R_\sigma + \epsilon, \tag{2.40}$$

where C_σ is the observed uncalibrated complex spectrum from a scene, L_σ is the incident spectral radiance, R_σ is the spectral responsivity of the instrument, also referred to as *gain*, O_σ is the instrument emission, or the *offset* term, F is the instrument operator, which accounts for instrumental effects such as self-apodization, modulation, etc., and ϵ is the complex noise. The linear (gain slope/offset) intensity response is treated in *radiometric calibration*, whereas the instrumental operator is typically compensated for during *spectral calibration*.

The references in radiometric calibration are usually provided by on-board internal cavities that closely approximate blackbodies. Typically, the cavity has a conical shape and its internal surfaces are coated with highly emissive paint so that the absorptance of the cavity is ≈ 1 [70]. For the GIFTS radiometric calibration, two on-board blackbody targets (300 K and 265 K) along with deep space are viewed during regular intervals. The instrument responsivity (gain slope) is determined by the difference between the two blackbody views, while the self-emission (offset) of the instrument is corrected using the deep space view. The blackbody cavity has an emittance range of 0.993 to 0.996 [10]. In Equation (2.40), suppose we neglect the instrument operator F for now, and take the transmission of the telescope into

account, then $(F \cdot L)_\sigma \equiv L_\sigma = N\tau$, with τ given as the transmittance of the telescope, and N as the true incident radiance that we wish to recover. Assume no scattering is present, the self-emission of the telescope, based on Equation (2.12), becomes $B_t\varepsilon = B_t(1-\tau)$, where B_t is the Planck emission at the temperature of the telescope. Thus, Equation (2.40) can be rewritten as

$$C_\sigma = \big[N\tau + B_t(1-\tau)\big]R_\sigma + \epsilon. \tag{2.41}$$

Assume the blackbody targets have a complete emittance of 1 so that $\tau = 0$, the uncalibrated observed spectra C_H and C_C for the hot and cold blackbodies can be expressed as

$$C_H = B_H R_\sigma + \epsilon \quad \text{and} \quad C_C = B_C R_\sigma + \epsilon, \tag{2.42}$$

where B_H and B_C are known blackbody radiances at the hot and cold temperatures. Therefore, the gain slope (or the responsivity) can be obtained as

$$R_\sigma = \frac{C_H - C_C}{B_H - B_C}. \tag{2.43}$$

To determine the instrument self-emission offset term, a deep space view is used for solving the transmittance τ [10]. The uncalibrated deep space spectrum C_S for a space scene B_S is

$$C_S = \big[B_S\tau + B_t(1-\tau)\big]R_\sigma + \epsilon, \tag{2.44}$$

then τ can be found by differencing the space and the cold blackbody views:

$$C_S - C_C = \big[(1-\tau)B_t + \tau B_S - B_C\big]R_\sigma. \tag{2.45}$$

Note that by differencing multiple references during the calibration process, the unknown noise term has been eliminated from the general equation.

As described in Equation (2.40), spectral calibration accounts for the instrument operator F, which is frequency dependent. The calibration process typically can be divided into three basic stages. The first step involves the assignment of correct

wavenumbers to the raw spectrum. This is accomplished using the spectral references provided by a metrology laser. The operation of a metrology laser and its purpose as a sampling reference will be discussed further in the next chapter. The second step corrects the sinc apodization effect introduced in Section 2.4.2, whereas the self-apodization effect is compensated for during the last step of the calibration procedure [51, 59].

CHAPTER 3

FTS AND IMAGING FTS SAMPLING SYSTEMS

A common problem that exists in FTS is concerned with how to compensate for sampling errors when an interferogram is sampled at nonuniform instants in the path-difference domain. These errors, due to various mechanical irregularities, are generally associated with a continuous scanning system, which samples the interferogram at either equal space or equal time intervals. In an equal time sampling (ETS) scheme, the samples collected at equal time intervals do not necessarily represent uniform sample values taken as a function of the OPD if the scanning mirror drive has a varying speed. In an equal space sampling (ESS) scheme, the interferogram data is collected at uniform path difference positions, in which case the error arises when the mirror velocity is varying during the integration. In both systems, the accuracy of the reconstructed signal can be significantly compromised if no error correction is performed. For an IFTS, the off-axis effects also affect the accuracy of the sampling locations.

3.1 FTS/IFTS Sampling Techniques

There are two basic techniques for recording and sampling an interferogram. In the first approach, the movable mirror is translating at constant velocity so that a continuous interferogram is obtained at the detector; the interferogram is then sampled by a separate sampling circuitry. This is referred to as the *continuous* or *rapid scanning* scheme. In the second approach, the mirror steps through sample points by quickly moving to and stopping at a sampling location, and then holding the current position for the desired integration interval. This is known as the *step scan* method, in which discrete interferogram samples are produced directly without

additional sampling mechanism. The underlining goal of a sampling system is to collect samples at precisely known uniform locations. Based on this criterion, the step scan approach offers several advantages over the continuous scanning method in terms of accurate positioning controls. For example, the step scan system is static during the integration, which is desirable in order to guarantee that the temporal acquisition of data is consistent. For a continuous scanning system, this condition cannot be satisfied because the mirror is continuously moving, and this results in some relatively small position deviations. However, the step scan is not as widely used as the continuous scan for several reasons. First, the step scan requires a complex servo design that is too demanding for most FTS applications; second, to be able to accomplish precise accelerating and decelerating movements, the sample rate has to be kept relatively low [9, 64]. Therefore, only continuous scanning method will be considered.

The continuous scanning method can be further subdivided into two categories: the equal space sampling and equal time sampling systems. Both systems are guided by a stabilized monochromatic laser reference. The sinusoidal laser signal is passed through the interferometer, and generates light and dark fringes as the mirror scans back and forth. The fringes are then detected by a separate detector, and used as inputs to a zero-crossing circuit and a counter, which indicates the exact elapsed clock cycles between fringes. Because the laser wavelength is known, fringes of the detector laser signal are counted to accurately measure the displacement of the moving mirror [52, 67].

In the ESS method (see Figure 3.1), the detector sampling is guided by the laser reference; in some cases, the mirror movement is also monitored and controlled by the same laser reference. The detected IR signals are sampled at equal spaced intervals by an A/D, which is clocked by the laser fringes. At selected zero or fractional zero crossings, a sample-and-hold circuit is triggered and held momentarily while the A/D

measures the magnitude of the signal and outputs it to the DSP. In the ETS method (Figure 3.2), the laser signal only serves as a reference of the mirror movement because in this case, the IR signals are sampled at equal time intervals by the A/D that is triggered by an external clock [9, 67].

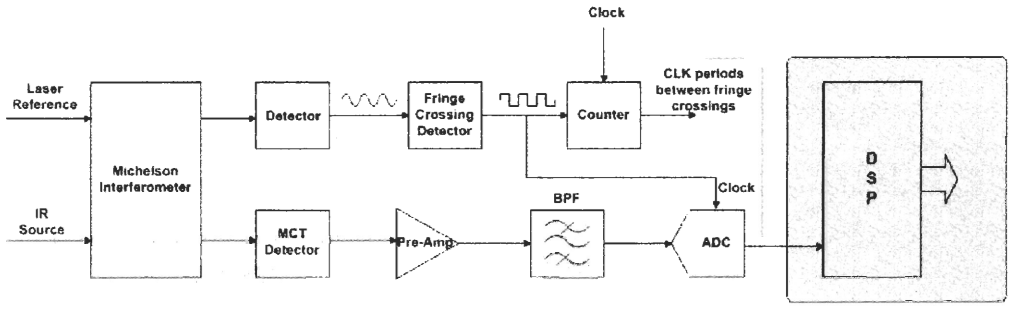


Figure 3.1: The equal space sampling system.

Suppose the mirror velocity cannot be controlled precisely due to mechanical irregularities, surface roughness, external disturbances, etc., this would result in undesirable sampling errors in both the ETS and ESS systems. For an ETS, the fluctuations of mirror speed cause the interferogram to be sampled at irregular sampling locations in the space domain, even though samples are taken at equal time instants. The laser reference signal monitors exactly how much deviations are present in the final measurement. For an ESS, the sampling position error is less severe when compared to the ETS because samples are collected at selected laser fringes. However, errors are still present since the mirror velocity may vary during the integration [67, 77]. For instance, the collection of samples is triggered at equal space instants, but during this period of time, a varying mirror velocity could cause an incorrect measurement.

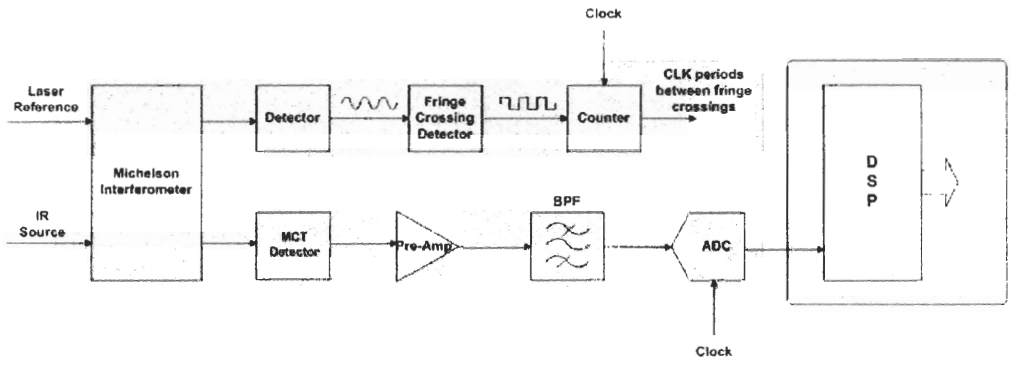


Figure 3.2: The equal time sampling system.

3.2 Nonuniform Interferogram Reconstruction – Problem Defined

Previously we have reviewed basic sampling methodology for a FTS. Now we will investigate the sampling positioning error in detail and its impact on sampling reconstruction algorithms. Throughout the rest of the discussion, the words "nonuniform" and "irregular" are used interchangeably to describe samples taken at nonuniform locations with positioning uncertainties due to instrumental imperfections that require to be corrected, which should be distinguished from samples that are purposely taken at nonuniform instants.

The first section is devoted to sampling theorems, which explain the fundamental algorithm applied to reconstruct uniformly sampled data. Then the nonuniform sampling problem is addressed for a one-dimensional FTS and for a multi-dimensional IFTS.

3.2.1 Sampling Theorems

This section serves as a theoretical background of the sampling theorem, in addition, the reconstruction algorithm for uniformly sampled interferograms will be presented. In Equation (2.28), a continuous doubled-sided interferogram $I(x)$ and its spectrum $E(\sigma)$ form a Fourier transform pair. Suppose $I(x)$ is sampled at a set of equally spaced

discrete points x_n to produce a sample set $I(x_n)$, then the original interferogram can be uniquely determined by this collection of values when certain constraints are superimposed on both the signal and its samples. The interferogram can be modeled in the form of

$$I(x) = \sum_{n\in\mathbb{Z}} I(x_n)S_n(x), \tag{3.1}$$

where S_n represents a set of expansion functions that interpolates over the sample values. In other words, the interferogram can be completely recovered from its samples when a suitable set of interpolation functions is selected. The Shannon (WKS) sampling theorem states that if $I(x)$ contains no frequencies higher than W, the bandwidth of $I(x)$, then $I(x)$ can be uniquely determined by its equidistant points spaced $1/2W$ apart. To derive a kernel function for the interpolation, let $\{e^{j\sigma x_n}\}$ be a basis function in the frequency domain, where $x_n = 2\pi n/2W$. Then the spectrum $E(\sigma)$ can be written as a linear combination of the basis function as

$$E(\sigma) = \sum_{n\in\mathbb{Z}} c_n e^{-j2\pi\sigma n/2W}, \qquad |\sigma| \le W. \tag{3.2}$$

The Fourier coefficients c_n are:

$$c_n = \frac{1}{2W}\int_{-W}^{W} E(\sigma)e^{j2\pi\sigma n/2W}d\sigma = \frac{1}{2W}I(\frac{n}{2W}).$$

Note that c_n is essentially composed of $I(n/2W)$, which are sample points of $I(x)$ spaced $1/2W$ apart. Multiply both sides of Equation (3.2) by $e^{j2\pi\alpha x}$ and integrate over $[-\pi, \pi]$, it becomes

$$\int_{-W}^{W} E(\sigma)e^{j2\pi\sigma x}d\sigma = \sum_{n\in\mathbb{Z}} \frac{1}{2W}X(\frac{n}{2W})\int_{-W}^{W} e^{j2\pi\sigma(x-n/2W)}d\sigma.$$

Notice that the left-hand side of the equation above is $I(x)$ by Fourier analysis. Finally, the reconstruction formula is

$$I(x) = \sum_{n\in\mathbb{Z}} I(\frac{n}{2W})\ \mathrm{sinc}(2Wx - n), \tag{3.3}$$

where $\mathrm{sinc}(x)$ is defined as

$$\mathrm{sinc}(x) := \begin{cases} \frac{\sin \pi x}{\pi x}, & x \neq 0 \\ 1, & x = 0. \end{cases} \tag{3.4}$$

The above equation has the general form of Equation (3.1), and offers an explicit solution for recovering a bandlimited signal from its samples when at least two samples per cycle of the highest frequency component are present. The sinc kernel forms an orthogonal set that spans the signal space of $I(x)$. Note that it takes on the value of one at available sample points and zero at the others.

The reconstruction formula can be seen as a low-pass process in the frequency domain. It is known that the spectrum of a periodic impulse train is another periodic impulse train

$$\sum_{n\in\mathbb{Z}} \delta(x - n\Delta x) \leftrightarrow \sum_{n\in\mathbb{Z}} e^{j2\pi\sigma n\Delta x} = \frac{1}{\Delta x} \sum_{n\in\mathbb{Z}} \delta(\sigma - \frac{n}{\Delta x}), \tag{3.5}$$

in which $n\Delta x$ are uniformly spaced sample points with Δx as its sampling interval. The sampled version of a continuous interferogram can be expressed as follows:

$$I_s(x) = I(x) \sum_{n\in\mathbb{Z}} \delta(x - n\Delta x), \tag{3.6}$$

and

$$I_s(x) = I(n\Delta x). \tag{3.7}$$

Using the relation in Equation (3.5), the spectrum of a sampled interferogram can be represented as

$$E^{\Delta x}(\sigma) = \frac{1}{\Delta x} \sum_{n\in\mathbb{Z}} E(\sigma - \frac{n}{\Delta x}), \tag{3.8}$$

where $E(\cdot)$ is the Fourier transform of the original interferogram, and the right-hand side of the above equation contains multiple replicas of this spectrum. Therefore, a suitable reconstruction function must be chosen so that only one copy of the spectrum is obtained. A low-pass filter in the frequency domain would be sufficient for signal recovery, and furthermore its inverse Fourier transform is equivalent to a sinc function.

3.2.2 Nonuniformly Sampled Interferograms

When the sampling locations are nonuniformly distributed, the reconstruction method described previously no longer applies. This situation arises in FTS interferogram sampling if the mirror velocity is varying. For an ETS system with a single detector element, samples are collected at equal time intervals, and their corresponding sampling locations in space can be derived according to laser fringe crossings. For example, in Figure 3.3, a typical clock cycle signal between fringes as a function of the OPD is shown. Each data point indicates the number of 25ns clock periods that has passed since the last laser fringe crossing. This elapsed clock periods during each crossing would be constant if the mirror is moving at a constant velocity. Using this measurement, we can obtain accurate estimates of the sampling location in space. Similarly, for an ESS system, uncertainties in the path difference domain during integration or collection time periods can be obtained based on the laser reference signal. For both the ETS and ESS cases, the varying mirror velocity could result in an incorrect measurement. For the ETS system, the problem is defined as how to reconstruct the interferogram from samples collected at known irregular locations. However, the laser reference only has a finite lifetime. If this reference information is unavailable, which means samples are taken at unknown locations, the reconstruction problem becomes less tangible. For the ESS system, the problem can be stated as how to fix the errors caused by incorrect measurements taking during the collection periods. Under these conditions, the more rational approach would be to take all available factors into account, which include the characteristics of the interferogram and its spectrum, and develop an application based reconstruction algorithm.

An IFTS with an array of detectors experiences the same problems that were defined for the 1-D sampling systems. In addition, the distortion in sampling locations caused by off-axis effects becomes too large to be neglected for a FPA. Recall in Section 2.4.2, the fundamental formulation of a differential element for the off-axis

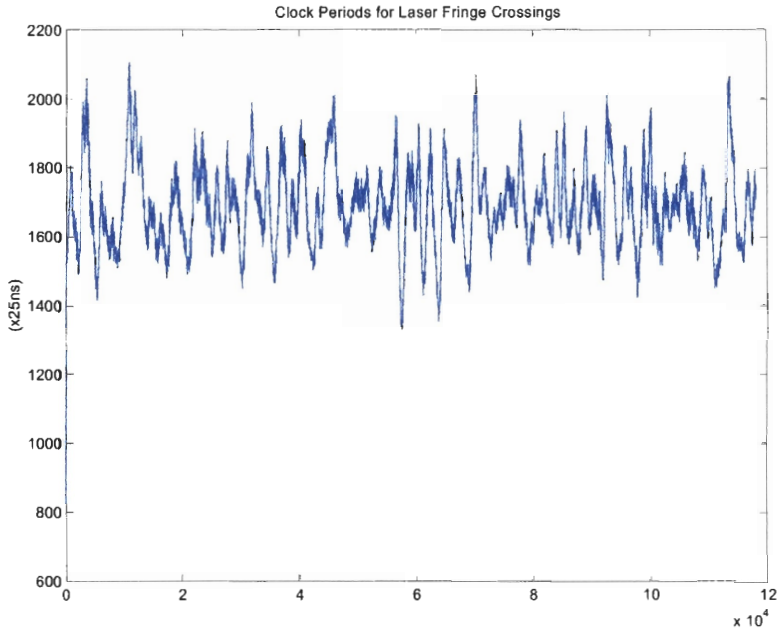


Figure 3.3: Typical fringe crossing clock periods.

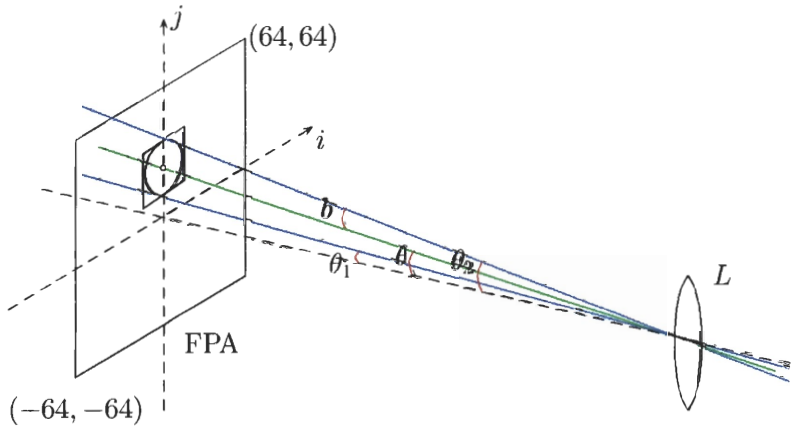


Figure 3.4: GIFTS FPA off-axis geometry for an array of detectors with one pixel enlarged to show details.

interferogram is written as

$$dI_\theta(x) = d\Omega \int_{-\infty}^{\infty} E(\sigma) e^{j2\pi\sigma x \cos\theta} d\sigma. \tag{3.9}$$

For the GIFTS geometry, the maximum angle subtended by the FPA is $\sim$ 97.39 milliradians. Then the half angle subtended by a single pixel is $b \approx ((97.39/2)/128 =$ 0.38 milliradians (see Figure 3.4). If θ is the off-axis angle to the center of a given pixel, θ can be viewed as its mean off-axis angle. Then the measured interferogram at each pixel as a function of θ can be derived from the following integration:

$$I_\theta(x) = \frac{\int_\Omega d\Omega' \int E(\sigma) e^{j2\pi\sigma x \cos\theta} d\sigma}{\int_\Omega d\Omega'}, \tag{3.10}$$

where $I_\theta(x)$ is normalized by the total solid angle subtended by a given pixel. Replacing $\int_\Omega d\Omega'$ with $\int d\phi \int_{\theta_1}^{\theta_2} \sin\theta d\theta'$ to incorporate the contribution of the off-axis angles for an individual detector, which can be achieved by integrating θ between $\theta_1 = \theta - b$ and $\theta_2 = \theta + b$, the resulting intensity for each off-axis pixel element is

$$I_\theta(x) = \int_{-\infty}^{\infty} E(\sigma)\ \mathrm{sinc}(2\pi\sigma x b\theta) e^{j2\pi\sigma x\left(1-\frac{\theta^2+b^2}{2}\right)} d\sigma. \tag{3.11}$$

Note that the shape of each square pixel is approximated as small segments of a circular ring with angular radius θ and width $2b$. If the pixels are labelled with a coordinate system (i, j), with corresponding edge pixels expressed as $(64, 64)$ and $(-64, -64)$, θ is related to b by

$$\theta = b\sqrt{(2|i| - 1)^2 + (2|j| - 1)^2}. \tag{3.12}$$

Figure 3.5 maps the circular rings for all detector pixels. Each ring has a width of $2b$; its corresponding off-axis angle θ is also labelled. According to Equation (3.12), the off-axis angle is constant for a ring at a given radius where pixel (i, j) is located on that ring. As a result, the off-axis interferograms are sampled at slightly shorter OPDs by a factor of $(1 - (\theta^2 + b^2)/2)$; using the inverse relationship, the off-axis

spectra are expanded to slightly higher wavenumbers. In Figure 3.6 three on-axis OPD sample points are shown, with their off-axis sample locations ranging from the center to a corner pixel [72, 74, 75].

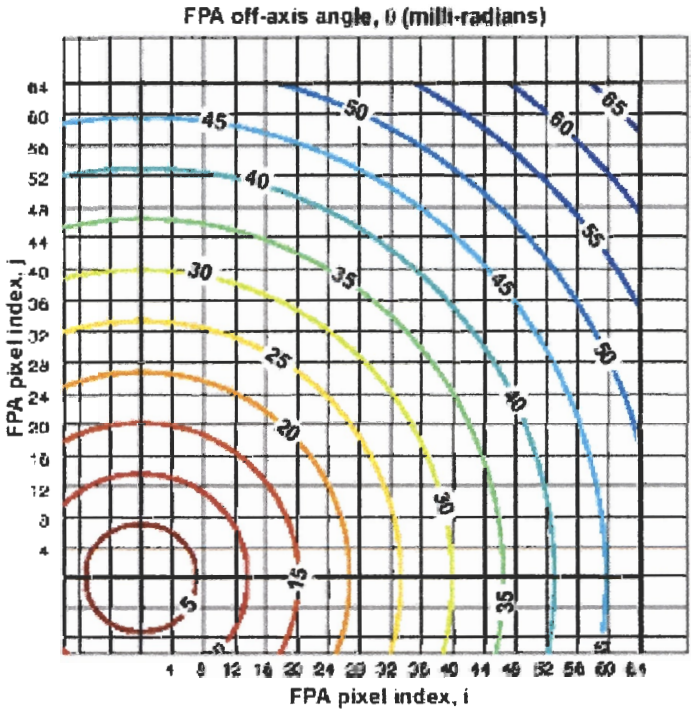


Figure 3.5: GIFTS off-axis angle approximation (from [74]).

In summary, for a single array FTS, the interferogram requires to be reconstructed from irregularly sampled data due to mirror velocity errors. For an IFTS, besides the random velocity uncertainty, off-axis effects will also cause deterministic distortions in the sampling position. These will apply to both the ETS and ESS scanning systems, and when the laser reference is unavailable, the reconstruction is performed without specifically knowing the exact sampling locations.

Now let's consider the error that occurs when an interferogram is sampled at irregular intervals where samples are no longer equidistantly spaced. If we represent the time jitter error by a set of small perturbation values, is it possible to recover

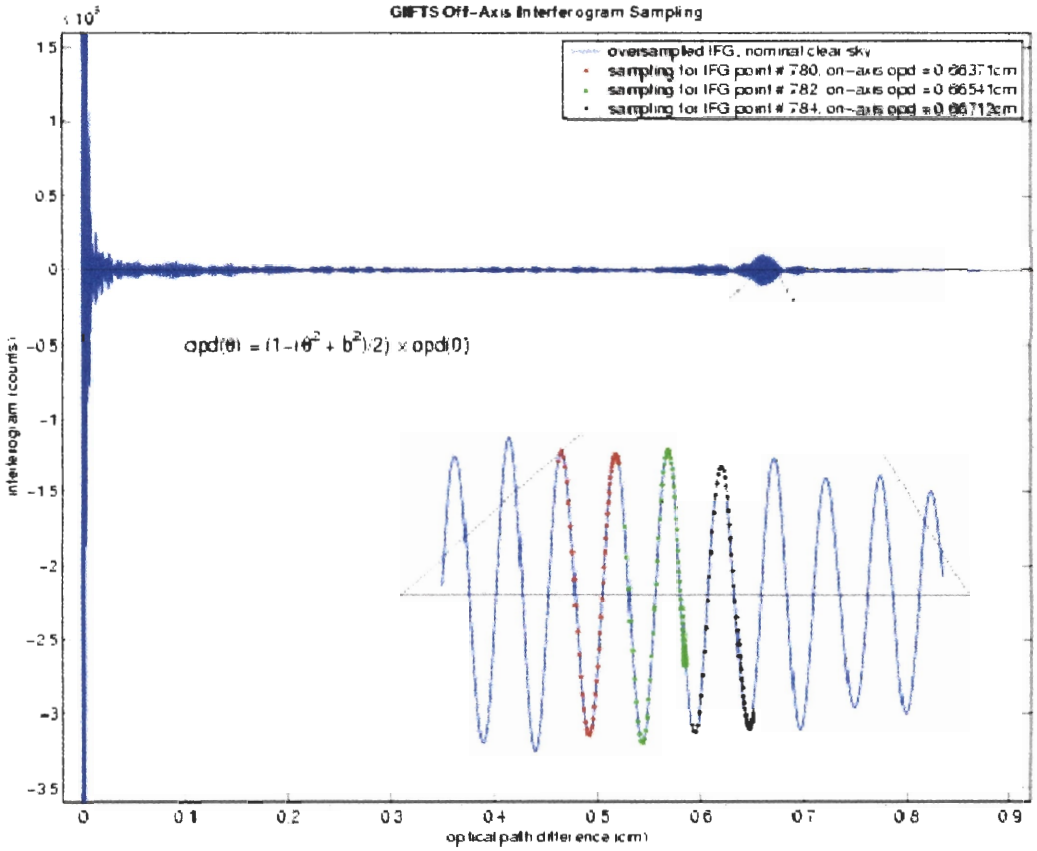


Figure 3.6: Off-axis sampling of three interferogram sample points. Pixels are selected at (0.6637, 0.6654, 0.6671 cm); the off-axis variations are plotted for every pixel along the diagonal from the center to a corner pixel of the detector array (from [10]).

the original signal accurately from these nonuniform sample values? In theory, it has been proven that as long as the density of the sample points $\{x_n\}$ is higher than the Nyquist rate, then a bandlimited signal is uniquely defined by its irregular sample values [11, 34, 40, 43, 44, 53, 73]. Having established that, the next question would be what type of reconstruction technique can be employed to accomplish this task. Observations made in the frequency domain suggest that the spectrum of a nonuniformly sampled signal is heavily aliased, and that the spectrum repetition property is lost. Consequently, a linear time-invariant filter is no longer sufficient for signal reconstruction. The reconstruction error using the ordinary sinc kernel function can be written as

$$e(x) = I(x) - \sum_{n\in\mathbb{Z}} I(\frac{n}{2W} + \delta_n)\ \mathrm{sinc}(2Wx - n), \tag{3.13}$$

where δ_n denotes offset values from the uniform sample points. Many attempts have been made to minimize this error, including techniques that are connected to different variations of the sampling theorem and those that are based on other interpolation or estimation methods [34, 43, 44]. The simplest method for reducing the time jitter error is to over-sample the signal at a much higher rate [12, 20, 52]. Although this does not necessarily minimize the error, it does recover the original signal with a negligible error [43, 44]. Details on this technique are discussed in the following chapter.

CHAPTER 4

RECONSTRUCTION OF IRREGULARLY SAMPLED DATA

This chapter presents a survey of various techniques for signal reconstruction from nonuniformly sampled data. An extensive amount of work has been developed in this particular area, where many of these techniques are closely connected to deviations of the original Shannon's sampling theorem; other branches such as linear modeling, splines, Lagrange interpolation, etc. have also been studied extensively. For finding an optimal reconstruction solution in a practical application, besides the accuracy of the algorithm one must also take its feasibility and complexity into consideration. Algorithms that produce slightly superior results may not be the best solution if they require large computational overhead. In addition, prior knowledge about the data characteristics can be taken into consideration in the formulation of such a problem.

First, we review some of the currently applied algorithms in the FTS industry, which can be classified as interpolative methods. Then we explore the option of using iterative methods to reconstruct nonuniform samples.

4.1 Reconstruction by Interpolative Methods

In the previous chapter, we have seen that even though a unique solution cannot be obtained from irregular samples via the conventional reconstruction method, when the signal is oversampled at a much higher rate compared to the Nyquist rate the low-pass method can produce a relatively reasonable reconstruction with negligible errors. The equal time sampling (ETS) with digital filter and numerical resampling method introduced in [12, 20, 52] was developed based on this useful property of oversampling.

In this particular reconstruction strategy, known as the Brault algorithm [12, 20, 52, 57], the interferogram $I(x)$ is acquired as a voltage signal $V(t)$, and therefore, is a function of time. The signal $V(t)$ is sampled at equal time instants $n\Delta t$. The sample points t_n corresponding to equal space intervals, $n\Delta x$, can be determined according to the elapsed clock periods between fringe zeros crossings. Then the interferogram $I(t)$ as a function of time is reconstructed by a low pass filter, and consequently, $I(t_n)$ can be found from $I(t)$, which is ideally equivalent to the data obtained as equal space samples $I(n\Delta x)$. This was made possible with tolerable errors based on the fact that the original interferogram is oversampled.

When an interferogram is sampled at near critical frequency, the digital filter with resampling strategy becomes less reliable since a low-pass filter cannot compensate for the smearing error of spectra caused by time jitters. Many approaches based on the sampling theorem have been developed and can be roughly categorized into non-iterative and iterative methods. For example, in [53] Papoulis has proposed a non-iterative method that involves the mapping of irregular samples into regular samples using a one-to-one transformation between a set of nonuniform and uniform sample points. Another example would be the reconstruction of a band-limited signal by interpolating the sample sequence and its derivatives, details can be found in [58]. However, more promising strategies for recovering the original signal sampled at the critical frequency would be an algorithm that is iterative in nature. In the following section, most discussions will be focused on iterative reconstruction schemes.

4.2 Reconstruction by Iterative Methods

In [45, 46], Marvasti has proposed an iterative method for recovering a band-limited signal from its nonuniform samples. The algorithm adapted the concept of frames, which can be found in problems related to non-harmonic Fourier series analysis. Beginning with an initial estimate set to zero, at each successive iteration, the difference

between known sample values and reconstructed sequence from the previous update convolved with the sinc kernel function, together multiply by a convergence constant λ to form an adjustment term, where λ has a range between 0.5 and 1. Then a new set of estimates is found by summing up this adjustment with previous estimates. The algorithm can be outlined as follows: Let $\{x_n\}_{n=0}^{N-1}$ be a sampling set with a maximum gap $\delta = sup\ (x_{n+1} - x_n) < \frac{\Delta x}{2B}$, then $I(x)$ can be reconstructed from its nonuniform samples $I(x_n)$ by the iterative algorithm:

$$I_0(x) = 0 \tag{4.1}$$

$$I_{j+1}(x) = I_j(x) + \lambda \sum_{n=0}^{N-1} \big(I(x_n) - I_j(x_n)\big)\ \mathrm{sinc}(x - x_n) \tag{4.2}$$

where $\mathrm{sinc}(x) = \frac{\sin(\pi x/\Delta x)}{\pi x/\Delta x}$. The bandwidth of the signal is B, and Δx is the (uniform) sampling period. To avoid the aliasing effect, the bandwidth B must satisfy the condition $0 < B < \frac{1}{2\Delta x}$, thus, the *Nyquist* interval is defined as $\frac{\Delta x}{2B}$ in this case.

A modified version called adaptive weights method was first introduced by Feichtinger and Gröchenig [23, 24, 25, 71]. The motivation for incorporating a weight vector into the interpolation process is to compensate for the irregularities among sample points. The weight vector is selected to be proportional to the local variations of the point densities in order to improve the convergence rate. Unlike the Marvasti method, the initial estimate is realized as the convolution between a sinc interpolation function and the data sequence at irregular intervals. The algorithm can be summarized as

$$I_0(x) = \sum_{n=0}^{N-1} I(x_n) w_n\ \mathrm{sinc}(x - x_n) \tag{4.3}$$

$$I_{j+1}(x) = I_j(x) + \sum_{n=0}^{N-1} \big(I(x_n) - I_j(x_n)\big) w_n\ \mathrm{sinc}(x - x_n). \tag{4.4}$$

The adaptive weights for Nyquist sampling sets are chosen to be the differences between midpoints of samples x_n and x_{n+1}, w_n is given by

$$\begin{aligned} w_n : &= m_n - m_{n-1} = \frac{x_{n+1} - x_{n-1}}{2}, \text{ and} \\ \sum_{n=0}^{N-1} w_n &= N. \end{aligned}$$

The convergence of both algorithms described above is established in [24] using the fact that the sinc kernel function can be seen as a frame operator. A system of functions $\{f_j\}_{j\in\Gamma}$ in a Hilbert space $\mathbb{H}$ is called a frame, if there exist two constants $0 < A \leq B$, such that

$$A\|I\|^2 \leq \sum_{n\in\Gamma} |\langle I, f_n\rangle|^2 \leq B\|I\|^2 \quad \forall I \in \mathbb{H}.$$

If $A = B$, the frame is said to be tight. For any frame, the frame operator S is defined as

$$SI = \frac{2}{A+B} \sum_{j\in\Gamma} \langle I, f_j\rangle f_j,$$

where $I \equiv I(x)$ is the continuous band-limited signal. Since S is a positive self-adjoint operator that is also invertible [42, 71], the non-orthogonal expansion of I is

$$I = SS^{-1}I = \sum_{j\in\Gamma} \langle I, f_j\rangle S^{-1} f_j.$$

This equation shows that I can be reconstructed via iteration from its coefficients $\langle I, f_n\rangle$ with a rate of convergence given by $\rho = \frac{B-A}{B+A}$ [71]. In the irregular sampling problem, if $\lim_{n\to\pm\infty} x_n = \pm\infty$, and the maximum sampling gap δ satisfies $\delta = sup\ (x_{n+1} - x_n) < \frac{\Delta x}{2B}$, then the frame operator is

$$SI = \sum_{n=0}^{N-1} I(x_n)\ \mathrm{sinc}(x - x_n),$$

and $\mathrm{sinc}(x - x_n)$ is a frame with bounds $A \geq (1 - \delta/\Delta x)^2$ and $B \leq (1 + \delta/\Delta x)^2$. [23, 24, 25, 42, 45, 46, 71].

Besides using an adaptive weight vector, the rate of convergence can be further improved by considering an acceleration technique. The method of steepest descent serves as a foundation for many other iterative algorithms. Consider the quadratic function

$$F(x) = \frac{1}{2}x^T Ax - b^T x \tag{4.5}$$

where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is assumed to be positive-definite and symmetric, because A is positive-definite, the surface defined by $F(x)$ is shaped like a parabolic bowl. The gradient of $F(x)$ is

$$\nabla F(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} F(x) \\ \frac{\partial}{\partial x_2} F(x) \\ \vdots \\ \frac{\partial}{\partial x_n} F(x) \end{bmatrix}. \tag{4.6}$$

To minimize $F(x)$, one can set $F'(x) = Ax - b$ to zero, therefore, the problem of minimizing $F(x)$ is equivalent to solving $Ax = b$, and the optimal solution point is $x = A^{-1}b$. Consider some arbitrary point p, and $p = x + e$, $e \neq 0$. From Equation 4.5,

$$\begin{aligned} F(p) &= F(x+e) \\ &= \frac{1}{2}(x+e)^T A(x+e) - b^T(x+e) \\ &= \frac{1}{2}x^T Ax + \frac{1}{2}x^T Ae + \frac{1}{2}e^T Ax + \frac{1}{2}e^T Ae - b^T x - b^T e \\ &= \frac{1}{2}x^T Ax - b^T x + \frac{1}{2}e^T Ae + e^T b - b^T e \ \text{ (by symmetry of } A) \\ &= F(x) + \frac{1}{2}e^T Ae \\ &= F(x) + \frac{1}{2}(p-x)^T A(p-x). \end{aligned}$$

Since A is positive-definite, then for all $p \neq x$, $F(p) > F(x)$. Hence, x is a global minimum of F. Let $r_k = b - Ax_k$ be the *residual* of x_k. If the residual is nonzero, then there exists a positive α_k such that $F(x_k + \alpha_k r_k) < F(x_k)$. Since

$$F(x_k + \alpha_k r_k) = \frac{1}{2}\langle x_k + \alpha_k r_k, Ax_k + \alpha_k Ar_k \rangle - \langle r_k + Ax_k, x_k + \alpha_k r_k \rangle,$$

and

$$\partial F/\partial \alpha_k = \alpha_k \langle r_k, Ar_k \rangle - \langle r_k, r_k \rangle = 0,$$

then

$$\alpha_k = \frac{\langle r_k, r_k \rangle}{\langle r_k, Ar_k \rangle}.$$

Table 4.1 outlines the adaptive weights steepest descent method, in this case, the residual r_k is defined as the difference between the initial and current reconstructed signals, and S represents the sinc operator.

Table 4.1: The adaptive weights steepest descent method.

Initialization:	$I_0(x)$ *is arbitrary*
Computation:	For $k = 0, 1, 2, \ldots$
	(a) $r_k = b - SI_k(x)$,
	$SI_k(x) = \sum_{n=0}^{N-1} I_k(x_n) w_n \ \mathrm{sinc}(x - x_n)$.
	$b = \sum_{n=0}^{N-1} I(x_n) w_n \ \mathrm{sinc}(x - x_n)$.
	(b) $\alpha_k = \frac{\langle r_k, r_k \rangle}{\langle r_k, Sr_k \rangle}$
	(c) $I_{k+1}(x) = I_k(x) + \alpha_k r_k$

The method of steepest descent uses residuals as its direction vectors, and very often the algorithm takes a direction similar to earlier steps. This leads to unnecessary repetitions in search directions. Conjugate gradient offers a superior alternative which involves the use of a separate set of search vectors. This method allows the search direction to be conjugate to all previous directions. The solution is to find a p_k so that it is A-orthogonal to p_{k-1}, i.e., $p_k^T A p_{k-1} = 0$. Let $I_k(x)$ be arbitrary and let successive approximation to the solution $I(x)$ be $I_{k+1}(x) = I_k(x) + \alpha_k p_k$, where p_k is a direction vector. The quantity α_k is found to minimize $F(I_{k+1}(x))$. Let $p_0 = r_0$ for $k = 0$, and $p_k = r_k + \beta_k p_{k-1}$ for $k \geq 1$. β_k is chosen such that p_k is the closest vector to r_{k-1} and is A-conjugate to p_{k-1} [13, 29]. Note that p_k and $I_k(x)$ are built up so that $I_{k+1}(x)$ minimize $F(x)$ over the whole vector space of all previous directions, $\{p_1, p_2, \ldots, p_k\}$. The algorithm is outlined in Table 4.2.

Table 4.2: The adaptive weights conjugate gradient method.

Initialization:	$r_0 = b - SI_0(x)$
	$p_0 = r_0$
Computation:	For $k = 1, 2, \ldots$
	(a) $\alpha_k = \frac{\langle r_{k-1}, r_{k-1} \rangle}{\langle p_{k-1}, Sp_{k-1} \rangle}$
	(b) $I_k(x) = I_{k-1}(x) + \alpha_k p_{k-1}$
	(c) $r_k = r_{k-1} - \alpha_k S p_{k-1}$
	(d) $\beta_k = \frac{\langle r_k, r_k \rangle}{\langle r_{k-1}, r_{k-1} \rangle}$
	(e) $p_k = r_k + \beta_k p_{k-1}$
	$SI_k(x) = \sum_{n=0}^{N-1} I_k(x_n) w_n \ \mathrm{sinc}(x - x_n).$
	$b = \sum_{n=0}^{N-1} I(x_n) w_n \ \mathrm{sinc}(x - x_n).$

CHAPTER 5

RECONSTRUCTION OF IRREGULARLY SAMPLED INTERFEROGRAMS

This chapter introduces methods for reconstructing nonuniformly sampled interferogram arrays. An interpolative algorithm is formulated in the case where sampling locations are available. When spatial location irregularities are unknown, a separate method that involves the recovery of a uniform sample spectrum is proposed based on interferometric characteristics of the data spectrum. In addition, the procedures of generating a nonuniform test data cube will be presented.

5.1 Nonuniform Interferogram Reconstruction – Revisited

Previously we have seen that the problem of reconstructing a nonuniformly sampled interferogram exists in both the ESS and ETS systems. Furthermore, the irregular sampling locations are available only when a laser signal is employed as a guiding reference. When this is no longer the case, the sampling error must be corrected without the specific knowledge on the irregularities involved in the mirror movement. The current correction strategy used in the area of FTS is described in Section 4.1, and is known as the Brault algorithm. In this method, the irregularly sampled interferogram is resampled at a regular grid, which is achieved by first applying a low-pass interpolation process to obtain the continuous-time interferogram, then resample this interferogram at regular locations in space. This is a straightforward, yet effective way of reconstructing a nonuniformly sampled interferogram especially when the interferogram is oversampled. However, the interpolation may only be done when the sample locations in space are known. These locations are provided by the laser fringe

crossing times. In Section 4.2, iterative techniques were introduced as general-purpose nonuniform reconstruction solutions. These algorithms are computationally more involved than the Brault algorithm. However, they generally produce more reliable results than the Brault algorithm near the Nyquist sampling rate. Again, however, the sampling location information is required when using these iterative reconstruction techniques. Based on these facts, a reconstruction algorithm that is capable of correcting the irregular sampling errors "blindly" is needed when the laser reference signal is not present. On the other hand, if the laser signal is obtainable and sampling locations are known, then an extension of the Brault algorithm that is suitable for solving the multi-dimensional (multi-pixel FPA) reconstruction problem is proposed.

Table 5.1: Interferogram reconstruction problems.

1-D with laser reference	M-D with laser reference
1-D without laser reference	M-D without laser reference

The interferogram reconstruction problems can be summarized as in Table 5.1. For a single-detector (1-D) problem, with the sampling location reference given to us, we could recover the correct interferogram easily using an interpolative method. For the multi-detector (M-D) array problem, the same 1-D interpolation technique can be applied to frames of data with the addition of an off-axis effect correction routine. Recall in Section 3.2.2, the off-axis effects cause the interferograms to be sampled at slightly shorter OPD, therefore, resulting a distortion in both the space and wavenumber domains. In the spectral domain, the off-axis spectra are expanded to slightly higher wavenumbers, causing them to have different spectral sampling

intervals. The wavenumber of the entire spectrum is shifted by an amount $\sigma(1-f)$ with $f = (1 - \frac{\theta^2+b^2}{2})$ (Section 3.2.2), and σ is the given wavenumber without the off-axis effects. Since b and θ are functions of pixel locations, the greatest distortion is experienced by the corner pixel elements as shown in Figure 5.1. Here the spectral

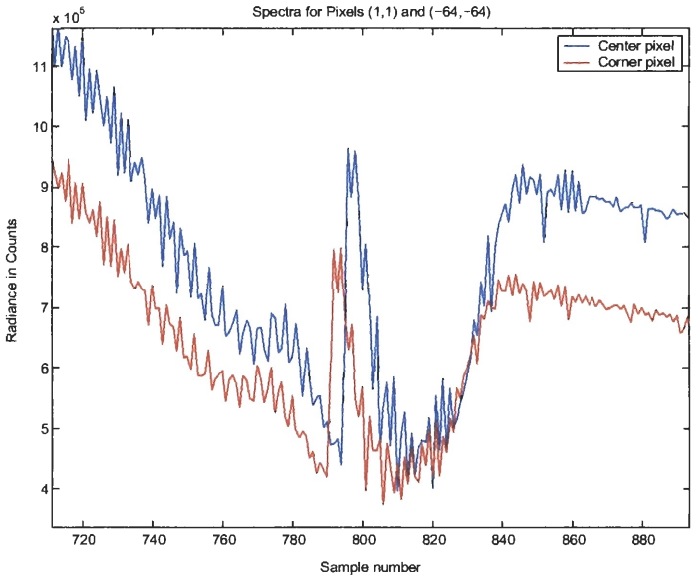


Figure 5.1: Wavenumber shifts between the center and corner pixels due to the off-axis effects. The center pixel (1,1) is plotted in blue, where as the corner pixel (-64,-64) is plotted in red. Both spectra are zoomed in near the ozone absorption band centered at 9.6 μm to show details.

shift can be seen by looking at shift in the position of each absorption band. Therefore, the off-axis correction algorithm must be able to re-map the spectra to their correct wavenumber grids.

If the nonuniform sampling locations are unknown, which is the case when a laser reference is not present, the reconstruction algorithm must be able to correct the sampling errors "blindly." To be able to correct the positional errors without actually knowing the positions can be a tricky task, unless we know certain properties of the data that we are trying to recover. In the following sections, we will discuss these known data characteristics and the possibilities of incorporating them into the

reconstruction algorithms. Similarly, the "without references" case applies to both single- and multi-detector problems.

Before we continue onto the next section, let's recall the notation that has been used so far. Table 5.2 shows the equivalency between an interferogram as a function of time and as a function of space.

Table 5.2: Interferogram notations.

Continuous interferogram as a function of space x or time t	$I(x)$	$I(t)$
Measured interferogram as a function of nonequal space x_n, or equivalent equal time $n\Delta t$.	$I(x_n)$	$I(n\Delta t)$
Desired interferogram as a function of equal space $n\Delta x$, or equivalent nonequal time t_n.	$I(n\Delta x)$	$I(t_n)$

5.2 The Problem with Known Sampling Locations

Start with the simpler case, in which the temporal sampling locations t_n corresponding to equal space intervals $n\Delta x$ can be derived from the collected laser fringe crossing information. If we have equal time samples, $I(n\Delta t)$, then the continuous interferogram as a function of time, $I(t)$, can be reconstructed via the following interpolation:

$$I(t) = \sum_{n=-\infty}^{\infty} I(n\Delta t)\ \mathrm{sinc}(t - n\Delta t) \tag{5.1}$$

$$\approx \sum_{n=0}^{N-1} I(n\Delta t)\ \mathrm{sinc}(t - n\Delta t), \tag{5.2}$$

where $I(n\Delta t)$ is the discrete interferogram measured at equal time instants triggered by an external clock signal, and $\mathrm{sinc}\{\cdot\}$ is the interpolation function defined as $\mathrm{sinc}(t) = \frac{\sin(\pi t/\Delta t)}{\pi t/\Delta t}$. Suppose $I(t)$ is resampled at t_n, then

$$I(t_n) = I(n\Delta t) * \mathrm{sinc}(t_n) \tag{5.3}$$

$$= \sum_{k=0}^{N-1} I(k\Delta t)\ \mathrm{sinc}(t_n - k\Delta t), \quad n = 0, 1, ..., N-1. \tag{5.4}$$

Since t_n is equivalent to $n\Delta x$ in space, then $I(t_n)$ is the desired interferogram sampled at equal space locations, $I(n\Delta x)$. For an array of detectors, the reconstructed interferogram data cube can be written as

$$I_{x,y}(n\Delta x) \equiv I_{x,y}(t_n) = \sum_{k=0}^{N-1} I_{x,y}(k\Delta t)\ \mathrm{sinc}(t_n - k\Delta t). \tag{5.5}$$

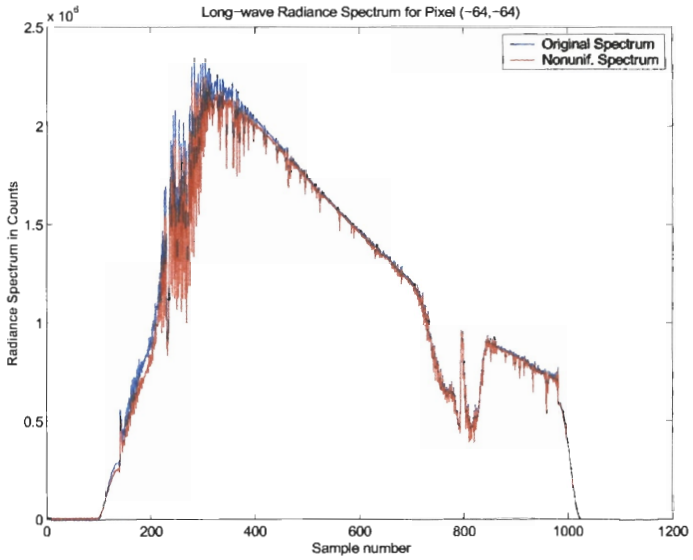
Where (x, y) indicates the pixel coordinates ranging from (-64, -64) to (64, 64) on a diagonal of the FPA, but excluding pixel (0, 0), that is pixels (0, 1), (0, -1), (-1, 0), and (-1, 0) do not exist. When applied to 16,384 pixels, the sinc interpolation can be computational demanding. Suppose we could have a solution that is slightly less accurate but requires significantly less computing power, would not that solution be a more desirable alternative? Or maybe we could offer a flexible solution that allows the user to decide which factor is the more important one for the application on hand: speed or accuracy. Two such alternatives are presented in the following chapter. The first involves the truncation of the sinc kernel function. The accuracy of the recovered data depends on the length of the truncation window being applied to the sinc kernel. A longer window length results in better data quality, but requires more computation, although this relationship is not necessarily linear. The second alternative takes advantage of a double-scanned interferogram. We can roughly view the two-sided scanning process as taking the data measurement twice, back-to-back, with a small time delay. That is the reason why an ideal interferogram is symmetrical about the ZPD. When random sampling offsets are present, each pair of symmetrical data measurements are in the neighborhood of the correct data value, and we know exactly how far they are from the desired point. Suppose that the interferogram does not experience any rapid changes within this neighborhood. Then we could easily apply a linear interpolation based on each symmetrical pair, therefore, approximating the correct value.

5.3 The Problem with Unknown Sampling Locations

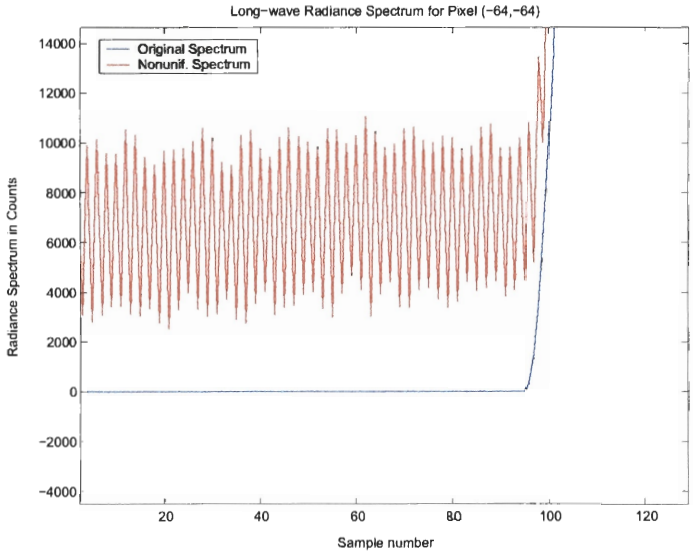
Now we consider the situation in which a laser reference is not available, and the nonuniform spatial sampling locations are unknown. In this case, the problem cannot be solved by any form of interpolative scheme because it is impossible to formulate and evaluate an interpolation function without actually knowing where the sample positions are. Therefore, additional information on the interferogram itself and/or its spectral properties must be taking into account.

First we note that the radiance spectrum of each interferogram is strictly bandlimited. The band coverage is pre-defined by the system specification, and is physically realized by selecting a particular type of detector (material) that is spectrally sensitive to the IR radiation within this band region. For example, the GIFTS spectral coverage for long-wave and short/mid-wave IR radiation is 685-1130 cm^{-1} and 1650-2250 cm^{-1}, respectively. Suppose nonuniform sampling position errors are encountered during data collection. What kind of effects will these errors have on the spectral band if we take the DCT of the measured interferogram? In Figure 5.2 (a), the spectrum of uniform spatial samples $E(m\Delta\sigma)$ is plotted along with the nonuniform spectrum $\hat{E}(k\Delta\sigma)$. From the close-up of these spectra in Figure 5.2 (b), we can see that the spectrum derived from the nonuniform samples exhibits oscillating distortions in the out-of-band portion of the spectrum, whereas the original spectrum has zero values in the same spectral region in an ideal case. Based on this observation, a bandlimited spectrum can be stated as a valid objective or constraint in the formulation of the reconstruction algorithm.

Another noticeable difference between these two spectra is in the phase angle. As described in Section 2.4.1, an ideal interferogram is symmetrical about ZPD and has a real spectrum; the imaginary part is present due to various phase errors and complex noises. Even though this is the case, the real part of the spectrum has a far greater amplitude than the imaginary portion of the spectrum. The amplitude



(a) Radiance spectra for pixel (-64,-64)



(b) Enlarged view of out-of-band region from the same plot

Figure 5.2: The original (in blue) and nonuniform (in red) long-wave radiance spectra for pixel (-64,64)

ratio between the two is on the order of 10^3 based on the simulation result obtained from a nonuniformly sampled interferogram with maximum sampling offsets within $\pm 10\%$ of the original locations. As a result, the phase angles are concentrated in the neighborhood of 0. This is, however, not the case when the interferogram is nonuniformly sampled. Here the real and imaginary parts of the spectrum are comparable in amplitudes. This effect can be seen in Figure 5.3, where unwrapped phase angles

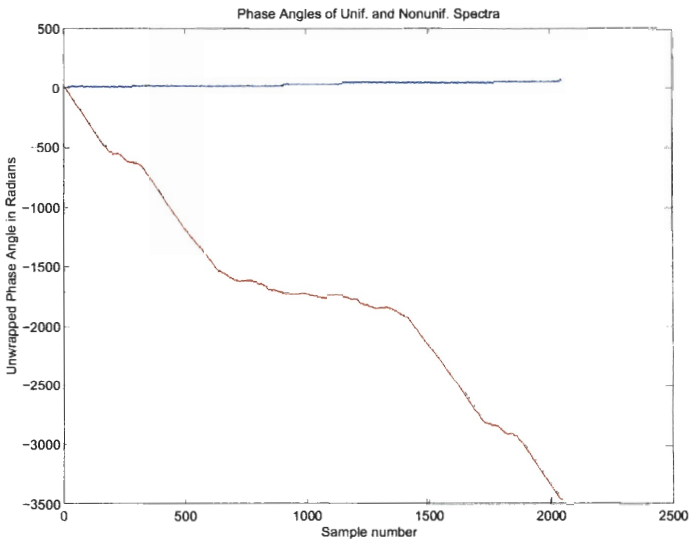


Figure 5.3: Phases of the original and nonuniform spectra for pixel (-64,-64).

are shown for both the original and nonuniform spectra. The blue curve indicates the unwrapped phase for the original spectrum, whereas the spectrum with nonuniform samples is plotted in red. Therefore, a second objective would cause the reconstructed spectrum to be real or "almost" real, or equivalently, a guaranteed symmetry in the space domain.

So far we have seen that the nonuniform sampling errors in an interferogram are reflected by its spectral changes, and it may be possible to correct these distortions more effectively by using a constrained iterative reconstruction algorithm that imposes known properties on the solution in both the space (interferogram) and

spectral domains, rather than working in only one domain as we did with the interpolative methods. Therefore, a reasonable tactic for solving the problem, without the reference provided by a laser signal, would be the one that attempts to solve the reconstruction problem in the frequency domain, i.e., to recover the correct spectrum instead of the interferogram samples. These iterative algorithms perform an update of the interferogram values in the space domain, along with an update of the spectrum in the frequency domain during each iteration. To ensure the corrected spectrum is consistent with the measurement, most reconstruction algorithms seek a cost function that combines the residual norm with a set of constraints. For example, the data consistency constraint can be expressed as the distance between the measurement and the estimate. The goal is to correct the errors in the measurement according to some known properties about the data, however, the estimate should still agree with the measurement to a certain degree. The optimal solution is obtained by finding a solution that minimizes this cost function, and satisfies the known constraints. A weighting parameter is usually used in conjunction with the constraints, which gives the solution the freedom to be "in favor" of the constraints or the measurement.

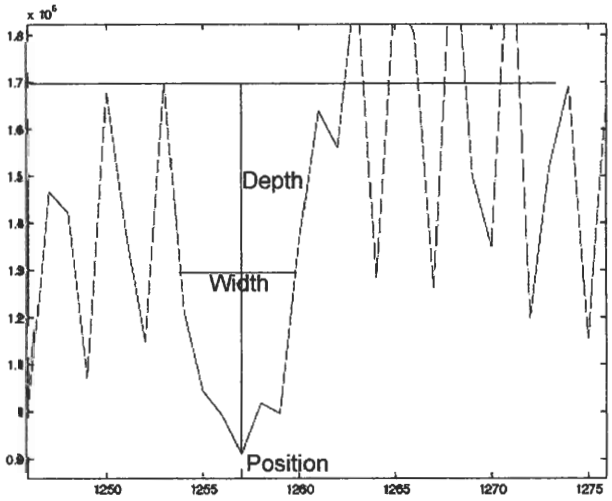


Figure 5.4: Three parameters that define an absorption band.

In our case, the magnitude response of the nonuniform spectrum differs from the original spectrum primarily in the amplitude, while its shape or envelop does not change significantly. Consequently, the residual norm would not be a useful measure of the data consistency because two correct spectra from different ground scenes usually only differ in amplitudes also, e.g., the ground area under the clear sky will have higher magnitude values than the ground scene covered by clouds. Therefore, verifying the data consistency by data values would not be a favorable choice because we have no way of knowing whether the amplitude change is caused by sampling errors or by nature. A more appropriate criterion can be found if we take a closer look at the absorption bands. Generally, each band can be characterized by three parameters: the position at band minimum, the width at FWHM, and the band depth (Figure 5.4). Since the band depth is an indication of the amplitude, any changes may result from either sampling errors or different ground scenes, it would not make a good measure for data consistency. This leaves us with band position and width, which are more reliable measures because they do not vary greatly due to atmospheric properties.

Based on the observations made above, we developed a reconstruction algorithm that has a top level structure as shown in Figure 5.5. First, we generate an initial estimate based on the nonuniform measurement. Then we compute the radiance spectrum by taking the FFT (Fast Fourier Transform) of the initial estimate. The "goodness" of this solution is then evaluated against a set of objective functions, which are defined according to the spectral properties as described previously. These objectives can be written as

$$\begin{aligned} \min_{\mathbf{I}\in\mathcal{F}\subseteq\mathbb{R}^n} \mathbf{f}: \quad & \mathbb{R}^n \rightarrow \mathbb{R} \\ \mathbf{f}(\mathbf{I}) \mapsto \quad & \sum_{i=1}^{k} w_i f_i(\mathbf{I}). \end{aligned} \tag{5.6}$$

where $\mathbb{R}^n$ denotes n-dimensional Euclidean space, the set $\mathcal{F} \subseteq \mathbb{R}^n$ is the feasible

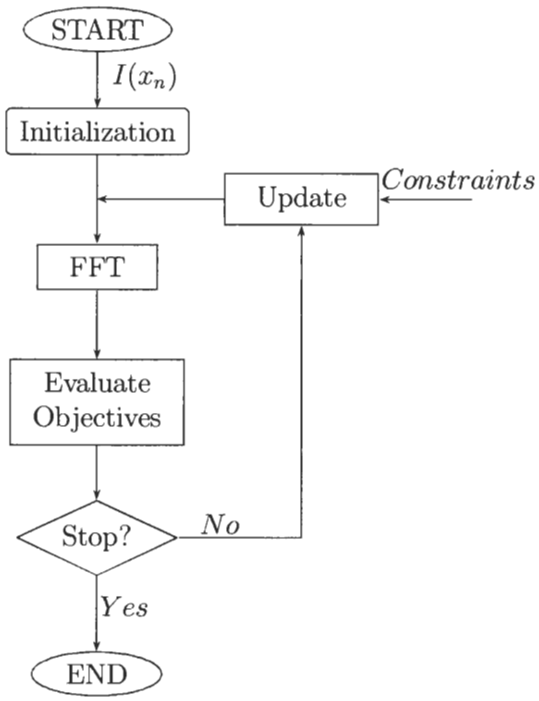


Figure 5.5: The basic structure of the reconstruction algorithm without sampling location references.

search space defined by a set of constraints, and $\mathbf{I} = (I_1, \ldots, I_n)$ is the interferogram vector that optimizes (minimizes) a set of objective functions $f_i(\cdot)$. Three objectives are being considered: the total OOB (out-of-band) spectral signal, the out-of-band spectral signal variation, and the entropy of the entire spectrum. The first objective restricts the OOB spectral noise, and is designed based on the fact that a uniform spectrum is bandlimited. Suppose a nonuniformly sampled spectrum has a high OOB signal that is due to an offset at ZPD. In this case, the OOB signal will not be a good indication of the quality of this solution since the error is the result of a sample DC offset. Therefore, the second objective is necessary to ensure that a good solution will have low out-of-band noise that is spectrally flat (white). This objective function

computes the total distances of each OOB spectral data value from the average OOB value. A lower distance measurement implies a lower data deviation, thus a flatter OOB signal. The third objective function computes the spectral entropy. As we have discussed previously, the location and width of each absorption band should not vary. However, detecting these features (location and width) over numerous bands for each spectrum at each evaluation would be a labor intense and rather difficult task. The minimum entropy criterion offers a measure for the overall shape of the absorption bands instead of monitoring each band individually. The entropy of a given spectrum is defined as the logarithm of the number of all possible distinct spectral configurations. The minimum entropy assumes the most spectrum information, i.e., known band locations and widths. As opposite to the well known maximum entropy method, which offers a uniform (flat) solution, the minimum entropy algorithm finds the "sharpest" spectrum, i.e., the one that has the spikiest or deepest absorption bands. A simple approach for dealing with multiple objectives is to assign each individual objective a positive coefficient w_i, $i = 1, ..., k$. The weighted cost functions are then summed up to produce a single scalar measure. In Equation (5.6), the vector to be optimized is represented by interferogram sample values, alternatively, a vector composed of nonuniform sampling time instants t_n can be used instead:

$$\begin{aligned} \min_{\mathbf{t}\in\mathcal{F}\subseteq\mathbb{R}^n} \mathbf{f}: \quad & \mathbb{R}^n \rightarrow \mathbb{R} \\ \mathbf{f}(I(\mathbf{t})) \mapsto \quad & \sum_{i=1}^{k} w_i f_i(I(\mathbf{t})), \end{aligned} \tag{5.7}$$

where $\mathbf{t} = (t_1, \ldots, t_n)$. If the optimal time instants, $\mathbf{t}^* \in \mathbb{R}$, are obtained, then the interferogram values at $\mathbf{t}^*$ will correspond to equal space sampling. After computing the objective function, the result of our current performance evaluation together with a set of constraints will lead to a new update. Three constraints are chosen to be considered: the symmetry of the interferogram, the lower and upper bounds of the estimation, and the consistency with the measured data. A symmetry constraint

will guarantee a real spectrum. The limits are designed to confine the estimate to a bounded region, which will ensure the interferogram sample values are within certain limits of the measured sample values. This region is assigned based on the assumption that the position offset errors are within 10% of the correct sampling locations. Finally, the estimate in the space domain also must be consistent with the measurement; this can be done by incorporating the measured interferogram into the estimation process, for example, it can be used as an initial estimate. Note that the consistency constraint is also monitored by the data limit constraint. After the update, the iteration continues until a stopping criterion is met. The implementation and derivation of these algorithms will be presented in the following chapters.

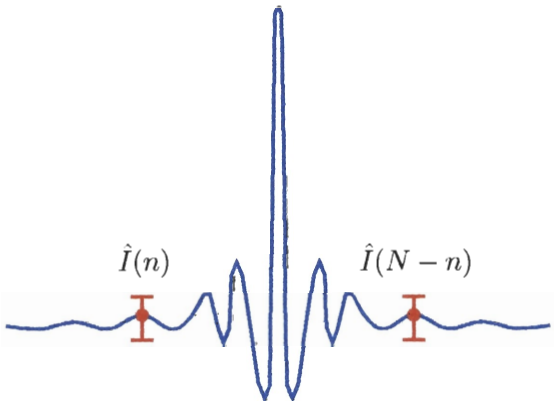


Figure 5.6: An illustration of the proposed constraints in the interferogram domain.

In summary, Figures 5.7 and 5.6 illustrate the concept of having multiple objectives and constraints in the reconstruction problem. In Figure 5.7, we use an arbitrary pair of samples as an example, here the estimated sample pair $\hat{I}(n)$ and $\hat{I}(N-n)$ must be equal in order to satisfy the symmetry constraint. Their possible solution values are confined within the indicated range, and the initial estimate is set to be the measured nonuniformly sampled interferogram (not shown in the Figure). These

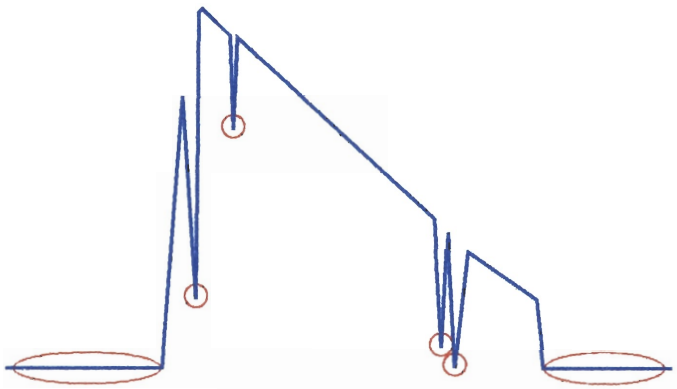


Figure 5.7: An illustration of the proposed objectives in the spectral domain.

two constraints will ensure the estimate is consistent with the measurement in the interferogram domain. Meanwhile, in Figure 5.7, the objective functions in the spectral domain are shown. The entropy objective ensures the "nulls" from the absorption bands are kept to be narrow, sharp, and as deep as possible. Whereas the out-of-band portion of the spectrum is monitored by the total OOB signal and the OOB signal deviation. All of these objectives and constraints are working in parallel as they were in a "tug of war." The desired solution is found to be the one that satisfies all of them the best.

5.4 Generating the Test Data Cube

Before any reconstruction algorithm can be developed, a set of nonuniformly sampled interferograms must be generated in the space domain, which requires the knowledge of irregular spatial sampling locations derived from the laser signal. In order to generate this set of nonuniform offset values, we adopt a clock signal from the test measurements that were obtained from a real FTS instrument. The experimental clock reference signal in Figure 3.3 is obtained with a HeNe laser of wavelength 633 nm, and the number of 25 ns clock periods c_n since the previous fringe crossing is counted, where $n = 0, ..., N-1$ crossings. To find a set of offset samples from the

clock signal, first, we compute the average clock periods, $\hat{c}$, over all fringe crossings:

$$\hat{c} = \frac{\sum_{n=0}^{N-1} c_n}{N}. \tag{5.8}$$

Then the offset samples normalized to an arbitrary sampling interval Δt is

$$e_n = \Delta t\left[\frac{c_n - \hat{c}}{\hat{c}}\right], \quad n = 0, \ldots, N-1. \tag{5.9}$$

The nonuniform time instants t_n, which is equivalent to $n\Delta x$, can be derived from $t_n = n\Delta t + e_n$; $n\Delta t$ represents equal time sampling locations. The parameter Δt is chosen so that when the interferogram $I(n\Delta t)$ is upsampled by an integer factor $M \equiv \Delta t$. The upsampling factor should be sufficiently large in order to create a smooth plot with a fine grid, from which irregular time samples can be taking with a relatively high resolution in terms of sampling positions. Figure 5.8 shows the steps of generating a nonuniformly sampled interferogram.

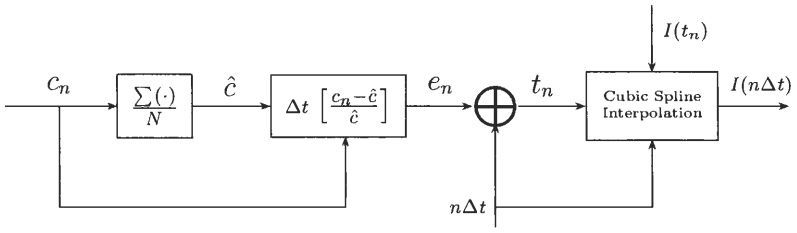


Figure 5.8: Generating a nonuniform interferogram.

To simulate the nonuniform sampling effect, an error-free discrete interferogram $I(t_n)$ of length 2048 is used as the original signal. With the uniformly sampled interferogram $I(t_n)$, t_n, and $n\Delta t$ available at hand, we need to determine the irregularly sampled interferogram $I(n\Delta t)$. We will do this by using a cubic spline interpolation of the upsampled interferogram. The cubic spline function is guaranteed to be smooth in the first derivative, and continuous through the second derivative [55]. Given the interferogram function $I_n = I(t_n)$, $n = 0, ..., N-1$, a linear interpolation function

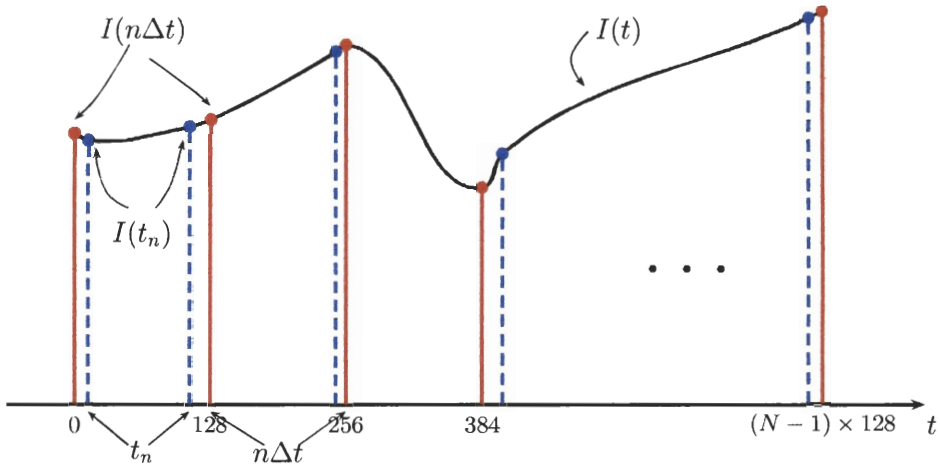


Figure 5.9: A sample plot of the generated nonuniform interferogram.

between a particular interval $t_m \equiv m\Delta t$ and $t_{m+1} \equiv (m+1)\Delta t$ is

$$I = AI_m + BI_{m+1}, \tag{5.10}$$

where

$$A = \frac{t_{m+1} - t}{t_{m+1} - t_m}, \text{ and } B = \frac{t - t_m}{t_{m+1} - t_m}. \tag{5.11}$$

A cubic spline formula will need some higher order terms in addition to Equation (5.10) to satisfy the 1st and 2nd derivative conditions. Thus, the complete cubic spline function [55] can be derived as

$$I = AI_m + BI_{m+1} + CI''_m + DI''_{m+1}, \tag{5.12}$$

where

$$C = \frac{1}{6}(A^3 - A)(t_{m+1} - t_m)^2, \text{ and } D = \frac{1}{6}(B^3 - B)(t_{m+1} - t_m)^2. \tag{5.13}$$

In Figure 5.9, a sample graph illustrates the relation between uniform and nonuniform time samples, along with their corresponding temporal sampling positions. For an array of detectors, the same data generation method has been applied to each single interferogram vector. In the following two chapters, we will discuss the algorithm details for correcting these nonuniform sampling errors.

CHAPTER 6

INTERFEROGRAM RECONSTRUCTION WITH KNOWN SAMPLING LOCATIONS

Continued from the previous chapter, this chapter presents the details of reconstructing nonuniformly sampled interferograms when their irregular sampling positions are known from a laser reference signal. In Section 6.1, we introduce the procedures for solving the one-dimensional reconstruction problem. Two algorithms will be discussed. The first algorithm is based on using truncated sinc interpolation kernels for reconstruction. The second method uses a simple linear interpolation and the symmetry property of the interferogram to solve the problem. In Section 6.2, these two methods are modified to recover an array of interferogram measurements. In addition, a routine for correcting the array off-axis effects is presented.

6.1 One-Dimensional Reconstruction

This section focuses on the correction of a single interferogram through a sinc or linear interpolation. Their formulation will be given in the following subsections. The advantages and disadvantages of each algorithm also will be discussed.

6.1.1 The sinc and Truncated sinc Interpolations

Recall in Section 5.2, the equal space samples at corresponding nonequal time instants is given by

$$I(t_n) = I(n\Delta t) * \text{sinc}(t_n) \tag{6.1}$$

$$\approx \sum_{k=0}^{N-1} I(k\Delta t)\ \text{sinc}(t_n - k\Delta t), \quad n = 0, 1, ..., N-1. \tag{6.2}$$

If the number of the sinc kernel functions is reduced to $L = 2W + 1 \ll N$, then we could estimate the interferogram as

$$\hat{I}(t_n) = \sum_{k=n-W}^{n+W} I(k\Delta t)\ \mathrm{sinc}(t_n - k\Delta t). \tag{6.3}$$

This formulation is equivalent to multiplying $\mathrm{sinc}(\cdot)$ by a rectangular window of length L, this is

$$\hat{I}(t_n) = \sum_{k=0}^{N-1} I(k\Delta t)\ \mathrm{sinc}(t_n - k\Delta t)\, w((n-k)\Delta t), \tag{6.4}$$

where

$$w(k\Delta t) = \begin{cases} 1, & |k| \leq W, \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

Notice that the truncation window function is centered at the sample of interest. For instance, if we wish to recover $I(t_j)$, then the windowed sinc kernels are centered at sample j, as the sample number is changing, so does the truncation window. If the sample is located at the beginning or the end of the signal where there are not enough samples on one side of it, zero values are used for these missing samples in the calculation. One of the advantages of applying a truncation window is that not all of the N sinc functions are required for calculating the interpolation outputs, therefore, saving a significant amount of computational power, especially when $L \ll N$. The second advantage would be the flexibility of the interpolation process. If a more accurate result is required, then we can set L equals to N, which is equivalent to the reconstruction algorithm without any truncation. On the other hand, if a less precise result is needed, then we can set the length of L to be much shorter than N, e.g., $L = 9$. In general, the quality of the reconstructed output improves as the window length L increases, however, this relationship is not necessarily linear. The rate of improvement decays as the window length grows. We will discuss these results further in Section 6.3.

6.1.2 Linear Interpolation

This section introduces an alternative reconstruction method based on the symmetry property of an interferogram. A double-sided interferogram contains data values that are symmetrical about the ZPD in an ideal situation. We can interpret the double scanning process as taking the same measurements twice during a short period of time. If these samples are irregularly spaced, then for a particular sample point, there are two measurements available in the neighborhood of the desired sample location after scanning. It is very unlikely that the sampling location offsets of these two measured values are equal during a double-scan. Since we have the values and locations of two distinct samples that are near the desired sample point, a linear interpolation will give us an estimate of the desired sample based on these available information. This algorithm is very simple to implement; it requires very little computational resources. The results are relatively accurate when compared to other reconstruction methods, especially when the original interferogram is oversampled.

For an arbitrary sample point j, where $1 \leq j \leq N/2 - 1$, the equal time symmetrical samples are $I(j\Delta t)$ and $I((N-j)\Delta t)$. Their absolute sampling location offsets from the desired point t_j and t_{N-j} can be written as

$$\delta_1 = \left| j\Delta t - t_j \right|, \text{ and} \tag{6.6}$$

$$\delta_2 = \left| (N-j)\Delta t - t_{N-j} \right|. \tag{6.7}$$

Note that $t_j = t_{N-j}$ for an ideal interferogram due to symmetry. The slope a and y-intersect factor b can be found from

$$\begin{cases} a\,\delta_1 + b &= I(j\Delta t) \\ a\,\delta_2 + b &= I\big((N-j)\Delta t\big). \end{cases} \tag{6.8}$$

Therefore, the estimated equal space sample value at t_j is

$$\hat{I}(t_j) = \begin{cases} a\,t_j + b, & \text{if } \delta_1 \neq \delta_2; \\ \frac{I(j\Delta t)+I\left((N-j)\Delta t\right)}{2}, & \text{if } \delta_1 = \delta_2, \end{cases} \tag{6.9}$$

and

$$\hat{I}(t_{N-j}) = \hat{I}(t_j). \tag{6.10}$$

If the two offsets are equal in some rare cases, then the average of these two measurements is used as the equal space estimate. Figure 6.1 shows how to obtain one equal space sample based on two symmetrical equal time sample points using the technique of linear interpolation.

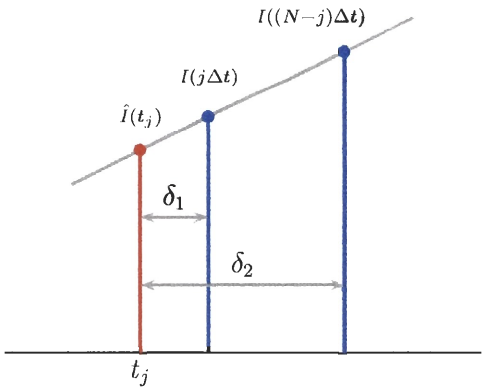


Figure 6.1: The reconstruction of sample j via linear interpolations.

The above reconstruction scheme applies to samples $n = 1, \ldots, N/2 - 1$. Additional steps must be considered when $n = 0$ and $N/2$ since there are no symmetrical counterparts available for these sample points at the ZPD and MPD. For the sample taken at MPD, we simply neglect the offset error because the value at MPD is very small compared with the others. A sampling positional error at ZPD will cause a DC offset in the frequency or wavenumber domain, thus, we can correct the ZPD sample

by subtracting this offset from its spectrum in the frequency domain. Please refer to Section 7.1.2.6 for details.

6.2 *Multi-Dimensional Reconstruction*

In this section, the interpolation methods described in the preceding section are modified to solve the multi-dimensional reconstruction problem. Furthermore, an off-axis effect correction algorithm is introduced for an array of detectors.

6.2.1 Linear and sinc Interpolations

For the sinc interpolation, the reconstructed equal space interferograms are

$$I_{x,y}(t_n) = I_{x,y}(n\Delta t) * \text{sinc}(t_n) \tag{6.11}$$

$$= \sum_{k=0}^{N-1} I_{x,y}(k\Delta t)\ \text{sinc}(t_n - k\Delta t),\ n = 0, 1, \ldots, N-1. \tag{6.12}$$

This equation applies to all pixels in the array, where $x, y = -64, \ldots, -1, 1, \ldots, 64$ indicate the coordinates of each detector element. Similar to the one-dimensional case, a moving truncation window can be applied to the detector array, the estimated interferograms can written as

$$\hat{I}_{x,y}(t_n) = \sum_{k=0}^{N-1} I_{x,y}(k\Delta t)\ \text{sinc}(t_n - k\Delta t)\, w((n-k)\Delta t). \tag{6.13}$$

The definition of the windowing function $w(k\Delta t)$ is identical to that given in Equations (6.5). Suppose we wish to reconstruct sample j for all pixel coordinates x and y such that a frame of data needs to be recovered, and the index of this frame is j, for $0 \leq j \leq N-1$. With a truncation window applied to the data cube, the desired equal space data frame is found by interpolating the equal time frames situated around frame j. A pictorial explanation of this process can be seen in Figure 6.2. The equal time frame j is shown in blue, along with its neighbors shown in gray. The length of the truncation window $L = 2W + 1$ determines the number of equal time

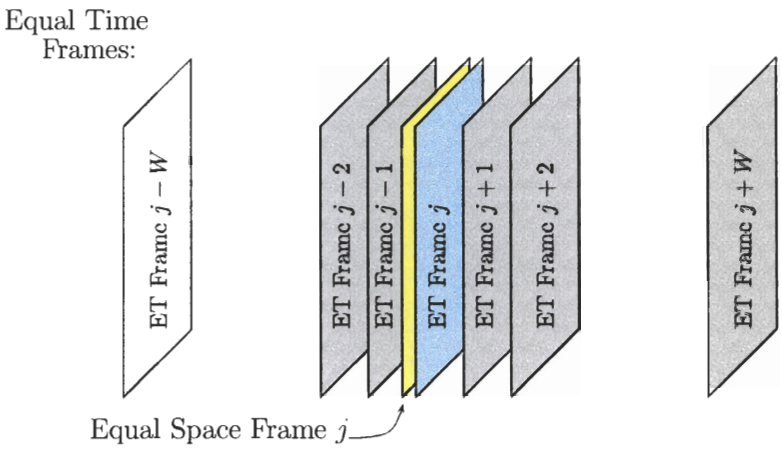


Figure 6.2: The sinc interpolation between interferogram frames.

frames being used in the interpolation. The desired equal space frame j is shown in yellow.

The concept of interpolating data between interferogram frames also applies to the linear method. Here the calculation of δ_1 and δ_2 remains the same since all array elements experience the same amount of nonuniform sampling offsets, thus

$$\delta_1 = \left| n\Delta t - t_n \right|, \text{ and} \tag{6.14}$$

$$\delta_2 = \left| (N-n)\Delta t - t_{N-n} \right|, \ \forall n. \tag{6.15}$$

The slope $a_{x,y}$ and y-intersect factor $b_{x,y}$ must be found independently for each coordinate, and they are

$$\begin{cases} a_{x,y}\,\delta_1 + b_{x,y} &= I_{x,y}(n\Delta t) \\ a_{x,y}\,\delta_2 + b_{x,y} &= I_{x,y}\big((N-n)\Delta t\big), \ \forall n, \ x \text{ and } y. \end{cases} \tag{6.16}$$

Therefore, the estimated equal space frames are

$$\hat{I}_{x,y}(t_n) = \begin{cases} a_{x,y}\,t_n + b_{x,y}, & \text{if } \delta_1 \neq \delta_2; \\ \frac{I_{x,y}(n\Delta t) + I_{x,y}\left((N-j)\Delta t\right)}{2}, & \text{if } \delta_1 = \delta_2, \end{cases} \tag{6.17}$$

and

$$\hat{I}_{x,y}(t_{N-n}) = \hat{I}_{x,y}(t_n), \ \forall n, \ x \text{ and } y. \tag{6.18}$$

The concept of linearly interpolating symmetrical interferogram frames can be seen in Figure 6.3. For the reconstruction of an arbitrary frame j, the desired equal space frames j and $N - j$ are shown in yellow whereas the equal time frames used for the interpolations are drawn in blue.

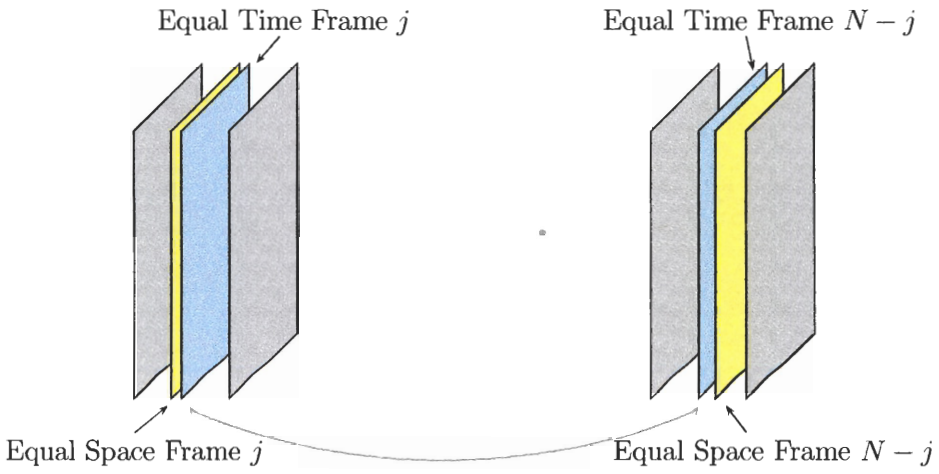


Figure 6.3: The linear interpolation between interferogram frames.

6.2.2 Off-axis Effect Correction

In addition to the nonuniform sampling error caused by mechanical irregularities of the scanning system, for an array of detectors, we must also consider the off-axis effect that was formerly introduced in Section 3.2.2 as the off-axis effect. Unlike the random offset errors discussed previously, the off-axis effect errors are systematic. The value of the off-axis distortion factor is determined by the location of each pixel, which can be found from

$$f = 1 - \frac{\theta^2 + b^2}{2}, \tag{6.19}$$

where

$$\theta = b\sqrt{(2|x| - 1)^2 + (2|y| - 1)^2}. \tag{6.20}$$

Note that x and y indicate the pixel location, and b is the half angle subtended by a single pixel, which has the value of $b \approx 0.38$ milliradians. From these two equations, we know that the off-axis pixels are sampled at slightly shorter OPDs, therefore, the spectra of these pixels are expanded to higher wavenumbers. The largest distortion is experienced by the corner pixels. From [74], the off-axis distortions in the space and wavenumber domains can be summarized into Table 6.1.

Table 6.1: On-axis and off-axis interferogram sampling parameters.

Interferogram parameters	On-axis	Off-axis
Sampling interval in OPD:	dx	$dx' = dx \cdot f$
Maximum optical delay:	$X = (\frac{N}{2})\, dx$	$X' = X \cdot f$
Spectral sampling interval:	$d\sigma = \frac{1}{2X}$	$d\sigma' = \frac{d\sigma}{f}$
Maximum spectral range:	$\sigma = \frac{1}{2dx}$	$\sigma' = \frac{\sigma}{f}$

To correct the off-axis sampling errors, a frequency domain approach is implemented. The first step involves the correction of random sampling errors using one of the interpolation methods described in the previous section. The resulting interferograms can be written as $\hat{I}_{x,y}(n\Delta x)$, $\forall\, x, y$. Their corresponding spectra, $\hat{E}_{x,y}(k\Delta\sigma)$, can be obtained via FFT. Since the off-axis pixel spectra are expanded to higher wavenumbers, a frequency sample rate conversion is performed for each unique coordinate (x, y) so that the spectral values of the corresponding interferograms are restored onto a shorter wavenumber grid. These procedures are outlined in Figure 6.4. Note that factors M and D are chosen larger enough so that pixels with different distortion factors will be restored onto distinct wavenumber grids. For example, the corner

pixel $(64, 64)$ experiences the off-axis effect by an amount of $f = 0.9977$, whereas pixel $(32, 32)$ has a f value of 0.9994. If for both interferograms, M and D are set to have the values of 100 and 99, respectively, then these two pixel spectra are corrected by the same amount, which is undesirable if high-resolution results are needed. On the other hand, suppose M equals 1000 for both pixels; D is equal to 997 for pixel $(64, 64)$, and 999 for pixel $(32, 32)$, then we will be able to obtain two unique solutions for both of them.

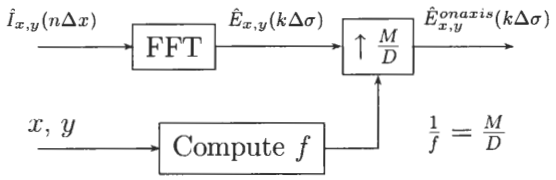


Figure 6.4: The off-axis effect correction algorithm outline.

6.3 Implementation and Results

This section presents the implementation and test results of the algorithms described previously. First, the complexity of each algorithm is discussed, which is followed by test results of these algorithms obtained from MATLAB simulations under noisy conditions and at different sampling rates. Finally, a comparison is made between these test results and those obtained from the existing methods.

6.3.1 Complexity, Noise, and Sampling Rate

The computational complexities of the algorithms introduced earlier are shown in Table 6.2. The required number of multiplications and additions are listed in columns. For the sinc interpolation, the variable L represents the truncation window length. In the multi-dimensional case, the numbers of operations are multiplied by the number of pixels to be recovered.

Table 6.2: An algorithm complexity analysis.

Method	Multiplications	Additions
sinc Interpolation	NL	$N(L-1)$
Linear Interpolation	$3(N/2-1)$	$5(N/2-1)$
ZPD Estimate	$2N\log_2(N)$	$2N\log_2(N)$

The next experiment examines the effects of an additive noise vector on the performance of each algorithm at different sampling rates. In this test, the signal is corrupted by the detector noise, which has the characteristic of a white (normally distributed) noise with an amplitude of $\sqrt{2N}$, where N is the number of samples. The detector noise is set to be real, asymmetrical about the ZPD, and assumed to be spectrally uncorrelated. The amplitude of the detector noise $\sqrt{2N}$ is determined such that its real part in the spectral count domain is NESR$\times R_I$, where NESR is the noise equivalent spectral radiance for the GIFTS instrument, and R_I is the instrument responsivity. The quantity NESR$\times R_I$ is equal to 1 count in the spectral count domain, which requires the noise amplitude of $\sqrt{2N}$ in the interferogram count domain. The factor of $\sqrt{2N}$ is generated when we take the inverse FFT of the detector noise in the spectral count domain. The derivation of the detector noise can be found in [74]. The errors of the reconstructed signals are computed for both the noise-corrupted and noise-free interferograms (Table 6.3). These errors are measured in 2-norm, which are defined by

$$\|\mathbf{e}\|_2 = \left\{ \sum_{n=0}^{N-1} e^2 \right\}^{1/2} = \left\{ \sum_{n=0}^{N-1} (I(n\Delta x) - \hat{I}(n\Delta x))^2 \right\}^{1/2}, \tag{6.21}$$

where $\hat{I}(n\Delta x)$ is the estimated uniformly spaced interferogram. For the sinc interpolation method, the results are obtained at three different window lengths, $L = 2W + 1 = 7$, 129, and 513. As discussed earlier in the chapter, the performance of a truncated sinc interpolation scheme improves as the window length increases, however, at a decaying rate, which can be seen from the errors of the reconstructed

outputs shown in Table 6.3. The linear interpolation method performed remarkably well in the test, especially when no detector noise is present. This is because each nonuniform sample pair is located near the desired equal space sample, even closer than the neighboring samples of the desired equal space sample. For example, the distances from $I(j\Delta t)$ to $I(t_j)$ and $I((N-j)\Delta t)$ to $I(t_j)$ are much closer than the distance from $I((j-1)\Delta t)$ to $I(t_j)$ because the maximum offset errors are within $\pm 10\%$ of the uniform sampling location. When an additive noise is included with the signal, the linear interpolation outputs experience relatively larger changes in the resulting errors. When the original nonuniform interferogram is oversampled at 4 times, as expected, all algorithms achieved better results compared to those obtained at near Nyquist rate.

Table 6.3: Reconstruction results (with references) from interferograms with the additive detector noise at different sampling rates .

Noise Condition	Method	Near Nyquist Rate	4× Oversampled
Without Noise (original error: $\|\mathbf{e}\| = 1.2107e5$)	sinc $L = 7$	5.9626e4	6.3840e3
	sinc $L = 129$	3.6740e4	2.5781e3
	sinc $L = 513$	3.6707e4	2.2691e3
	Linear	1.4442e4	1.6319e3
With Noise (original error: $\|\mathbf{e}\| = 1.2120e5$)	sinc $L = 7$	5.9669e4	8.6489e3
	sinc $L = 129$	3.6742e4	4.3588e3
	sinc $L = 513$	3.6709e4	4.1022e3
	Linear	1.7515e4	6.7941e3

6.3.2 Comparison of Reconstruction Test Results

Comparisons are made between the algorithms introduced in this chapter and the existing methods described in Chapter 4. Similarly, the tests are conducted under noisy and noise-free conditions at near Nyquist sampling rate. In general the iterative methods outperformed the Brault algorithm and any of the truncated sinc interpolation schemes, however, they do require more computational resources. In

Table 6.4: Comparison of test results (with references) between reconstruction methods.

Noise Condition	Method	Reconstructed Signals
Without Noise (original error: $\|\mathbf{e}\| = 1.2107e5$)	sinc $L = 7$	5.9626e4
	sinc $L = 129$	3.6740e4
	Linear	1.4442e4
	Brault	2.4793e4
	Marvasti (iter.=5)	2.3131e4
	Adaptive Weights (iter.=5)	2.4318e4
	Steepest Descent (iter.=5)	2.0237e4
	Conjugate Gradient (iter.=5)	1.9875e4
With Noise (original error: $\|\mathbf{e}\| = 1.2120e5$)	sinc $L = 7$	5.9669e4
	sinc $L = 129$	3.6742e4
	Linear	1.7415e4
	Brault	2.4830e4
	Marvasti (iter.=5)	2.3978e4
	Adaptive Weights (iter.=5)	2.4530e4
	Steepest Descent (iter.=5)	2.2767e4
	Conjugate Gradient (iter.=5)	2.0722e4

addition, the issues of stability and convergence rate must also be taking into consideration when using an iterative algorithm. Once again, the linear interpolation method produced the most favorable results when compared with outputs from any other techniques. For the GIFTS application, each interferogram data cube contains a huge amount of data. As a result, any one of the iterative methods that requires the repetitive computation of N^2 sinc kernels would be impractical. The digital filtering (Brault) method is a more suitable candidate, however, it does not offer any options for the output data quality. The truncated sinc interpolation method provides multi-resolution solutions, which is decided based on the desired speed vs. accuracy criterion. Finally, when there is no additive noise present, the linear interpolation method is simple to implement, highly efficient, and is capable of achieving exceptionally accurate results.

CHAPTER 7

INTERFEROGRAM RECONSTRUCTION WITH UNKNOWN SAMPLING LOCATIONS

This chapter presents the key elements of reconstructing the nonuniformly sampled interferogram array without the availability of the location reference. In Section 7.1, the minimization problem is briefly reintroduced; an evolutionary approach is chosen as the most suitable candidate for dealing with a complex model function. Also in this section, a short introduction of the evolutionary computational principle is given, followed by two algorithms that are designed for solving the nonuniform reconstruction problem. These two routines are compared side-by-side to determine the best scheme. In Section 7.2, the extension of the algorithm is made for the recovery of a multi-dimensional interferogram array. Finally, implementation details of these techniques and analyses of the test results are presented in the final section of the chapter.

7.1 One-Dimensional Reconstruction

This section focuses on solving the one-dimensional nonuniform interferogram reconstruction problem without knowledge of the sampling locations. The concept of optimizing a multiple-objective function with additional constraints was formulated in Section 5.3 based on the interferogram's spatial and spectral properties. Now the question remains: how do we find a suitable optimization method for solving such a problem defined in Section 5.3? Recall in Equation (5.6), the three objectives that were defined included the total OOB spectral signal, its variation, and the entropy of the spectrum. In addition, three constraints were presented as the data limits, the

symmetry of the interferogram and data consistency according to the measurement. Note that the vector to be optimized is in the spatial domain while all objective functions are in the wavenumber domain. Therefore, a Fourier transform is required at each function evaluation, and furthermore, both objectives and constraints include nonlinear functions. Based on these observations, an optimization algorithm that is too restrictive in terms of the required form and conditions of the evaluation function is undesirable; an algorithm that is too demanding in terms of computing power is also unfit for the problem. For example, traditional methods that involve the calculation or estimation of a gradient function is too costly, even impossible to compute in this case. An algorithm that requires the objective function to be convex, quadratic, or linear, etc. is also out of the question because our objective functions do not fit into any of these required forms. Non-derivative techniques that rely on the idea of transforming a current estimate into a better solution based on the merit of the evaluation function seem to be promising since they can be easily applied to our model. However, very often these methods search for a solution in a local neighborhood and, as a result, converge to a local optimum. This is especially a problem if the objective function has a complex "hilly" landscape. Therefore, a desirable algorithm should be able to find a global optimal solution; it should be relatively simple to implement, and it should be capable of dealing with a variety of model functions without too many restrictions. An evolutionary algorithm (EA) that simulates the natural selection process among a group of potential solutions fits the description well. In addition to its simplicity, this approach offers several advantages in terms of finding the optimal solution. First of all, EAs work on stochastic models, which generate and mutate solutions based on a set of probabilistic rules. Secondly, an EA evaluates a population of individual solutions simultaneously, they evolve and adapt according to the rule of *the survival of the fittest*. In the end, a population of the strongest solutions remains. Finally, local optima can be escaped since EAs are global search

methods. The actual algorithm structure is described in the following section.

7.1.1 Introduction to Evolutionary Algorithms

Evolutionary algorithms, as the name implies, capture the essence of a Darwinian biological evolutionary process. In the beginning, a randomly selected population composed of a number of individuals (potential solutions) is created. Each individual's fitness is judged based on its performance outcome at the evaluation function. Those individuals with the weakest results are quickly eliminated while the strongest are kept to generate offspring. This new population will acquire certain characteristics from their parents, yet, random mutations and variations are also taking place among them. Once again, these new individuals are competing to survive according to the evaluation function. As a result, a new generation emerges from the previous one. The cycle continues as each new generation becomes more fit to this simulated genetic environment than the last one until a termination condition is reached. In the end, an elite population of fittest individual solutions has been produced.

In Figure 7.1, the structure of an evolutionary optimization algorithm is shown. Before an initial population is selected, a representation of the data vector to be optimized is chosen. A representation is a mapping scheme that transforms the original state space of possible solutions into an encoded state space. For example, a fixed-length vector of symbols such as a string of binary bits, or integers can be used as valid representations [2, 27, 47, 48]. Once the data vector has been mapped into the encoded space, a population is created randomly. At this point, several parameters are required to be considered. For example, the number of individuals in the initial population, and the probability distribution of the population. There are no concrete principles for setting up these variables in the field of evolutionary computation, in general, so the best set of parameters are determined experimentally for a particular application. The next stage involves the evaluation of each individual's fitness. A

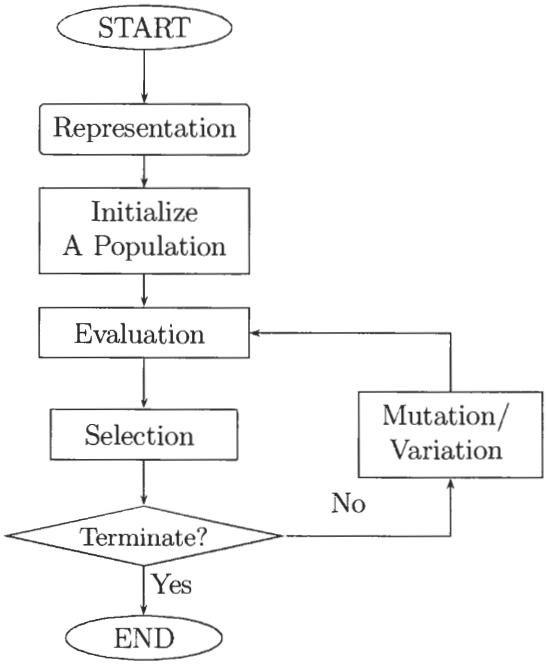


Figure 7.1: The evolutionary algorithm outline.

well-chosen evaluation function should satisfy some basic requirements. For example, the optimum solution should receive the most favorable evaluation. In addition, there should be a correlation between how fit the rest of the solutions are and their evaluation scores, i.e., the relative quality of each individual is directly reflected by its evaluation ranking [48]. Next, a selection operator is applied to the population. An example of a simple selection scheme eliminates the unfit individuals who have the lowest evaluation scores and promotes the higher ranking ones. The "survivors" remain and will be able to generate offspring. The new generation will experience some form of variation or mutation. A zero mean Gaussian random variable is a popular choice as a variation operator when the solution vector is represented as continuous or integer variables. Very often, the Gaussian mutation variable is added to the previous generation. On the other hand, a bit flip or crossing over (two or more solutions exchanging bits) would make a better choice as a variation operator

for a binary string since binary variation is usually done by changing one or several bit values. The iteration continues until a termination condition is reached.

There are several approaches for handling the constrained optimization problem using the evolutionary algorithms. The simplest solution is the "death penalty" method, in which individuals who do not satisfy the constraints are rejected permanently. Although it is simple to implement, this strategy seems to be too "crude" in certain cases. For example, suppose an individual has a high evaluation score, however, lies on the boundary between the feasible and infeasible regions of the search space. A total elimination of this particular solution may not be the most desirable choice. A better alternative considers penalizing or repairing these infeasible solutions, usually by incorporating a penalty factor in the evaluation function [3, 26, 48].

In summary, evolutionary algorithms are stochastic global optimization methods that were developed based on the principle of natural selection and adaptation. Unlike classic optimization techniques, EAs can work on a variety of function types and require the least amount of mathematic resources. Because of their flexibility, the design choices of each parameter are endless, ranging from simple decision rules to elaborate dynamic systems. There are no standard sets of rules of selecting them. Thus, the best evolutionary algorithm is the algorithm that works best for the application on hand. In the following sections, we will discuss how the nonuniform interferogram reconstruction problem fits into the evolutionary way of thinking.

7.1.2 Interferogram Reconstruction: Algorithm A

This section describes a recovery algorithm that emphasizes the estimation of sampling locations. Recall in Equation (5.7), the vector to be optimized is stated as the sample positions in time. This requires a mapping between the sample time instants and their corresponding interferogram function values. Both sides of the interferogram need to be corrected, the symmetry constraint is carried out by assigning a

penalty factor for the infeasible solutions. We name this routine *Algorithm A*. An alternative method that rejects all asymmetrical solutions, therefore, only estimates the values of a half interferogram, will be introduced later on. For now, we discuss the design details of *Algorithm A*.

7.1.2.1 Representation

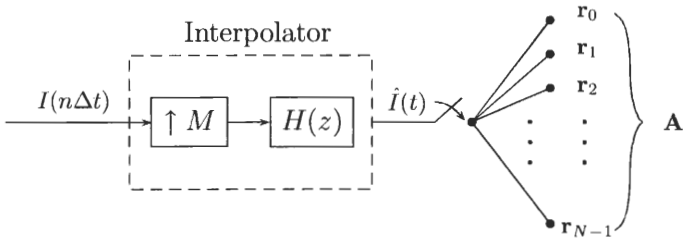


Figure 7.2: Reconstructing the continuous interferogram in time.

To find an appropriate representation for the optimization vector, consider a procedure similar to the case in which sample locations are known. In Chapter 6, the one-step reconstruction algorithm can be seen as a two-step process. That is the uniform time samples are reconstructed through an interpolator, followed by a resampling. The upsample rate is selected large enough so that the resulting interferogram has a relatively high resolution, thus, resembles a continuous signal. Similar steps are applied to the without-reference case, as shown in Figure 7.2, to generate a heavily over-sampled interferogram $\hat{I}(t)$ from equal time samples $I(n\Delta t)$. Here, the time instants t_n corresponding to equal space samples are unknown. However, it is assumed that we may place lower and upper bounds on the range of values where these samples may be located. This leads us to the limit constraint, which was described in Chapter 5. The nonuniform position offsets in time are within 10% of the original uniform locations, which are derived based on the characteristics of the

moving mirror. These offsets can be translated into equivalent offset sample numbers. For instance, if the upsample rate is chosen to be $M = 128$, then the desired data value is located within $\pm 128 \times 10\% \approx \pm 13$ samples of the original. If the limits are relaxed to 20% to account for the worst case scenario, then the desired data value lives in a neighborhood of ± 26 samples of the measured data value. In this case, each sample has $(2 \times 26) + 1 = 53$ possible candidates, which calls for a partition of $\hat{I}(t)$ into a matrix A (Figure 7.2), as follows,

$$\underset{N \times L}{\mathbf{A}} = \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0(L-1)} \\ a_{10} & a_{11} & \dots & a_{1(L-1)} \\ \vdots & \vdots & \dots & \vdots \\ a_{(N-1)0} & a_{(N-1)1} & \dots & a_{(N-1)(L-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_0^T \\ \mathbf{r}_1^T \\ \vdots \\ \mathbf{r}_{N-1}^T \end{bmatrix}, \tag{7.1}$$

where each row of A represents all possible data values (neighborhood) of a particular sample. The matrix A is of size $N \times L$, where N is the total number of samples to be corrected or the length of the interferogram to be recovered, and $L = 53$ is the length of the row vector $\mathbf{r}_n$. The vector can be written as

$$\mathbf{r}_n = \left[a_{n0},\ a_{n1},\ \dots,\ a_{n(L-1)}\right]^T, \text{ for } n = 0, \dots, N-1. \tag{7.2}$$

The relation between $\mathbf{r}_n$ and the estimated continuous interferogram $\hat{I}(t)$ can be expressed as

$$\mathbf{r}_n = \begin{cases} \left[\hat{I}(0),\ \hat{I}(1),\ \dots,\ \hat{I}(L-1)\right]^T & \text{if } n = 0 \\ \left[\hat{I}(M(n-1) - \frac{L-1}{2}),\ \dots,\ \hat{I}(M(n-1) + \frac{L-1}{2})\right]^T & \text{if } n = 1,\ \dots,\ N-1. \end{cases} \tag{7.3}$$

Note that when $n = 0$, all 53 samples of the vector $\mathbf{r}_0$ are taken on the right side of the measured value instead of from both sides. This is the case because if the first data point was sampled incorrectly, then it must be a delayed sample. In Figure 7.3, the concept of generating matrix A is illustrated. Each red line indicates an equal time sample value and its corresponding sample number, whereas the two blue neighboring

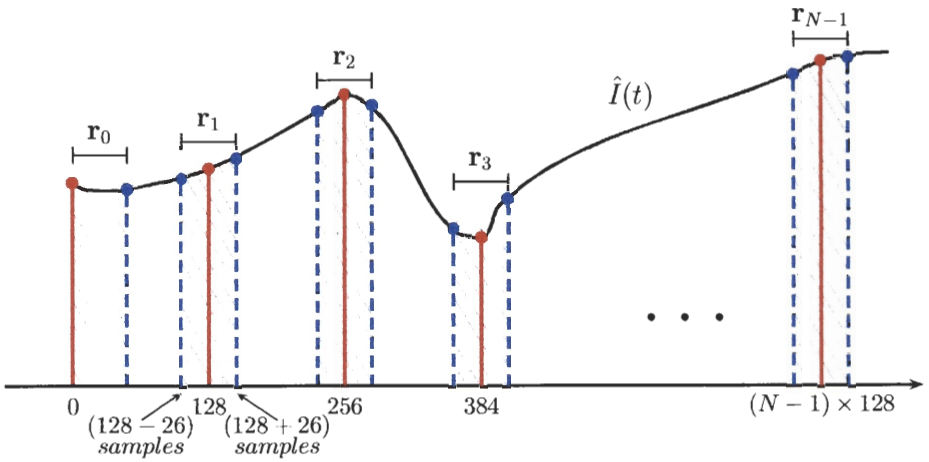


Figure 7.3: Generating the data matrix **A**.

dashed lines represent a region of the correct data value (and its location) for this particular sample. Each sample point n has a unique data vector $\mathbf{r}_n$ that is composed of all possible solution candidates as shown. Figure 7.4 shows an enlarged portion of Figure 7.3 that includes the vector $\mathbf{r}_1$.

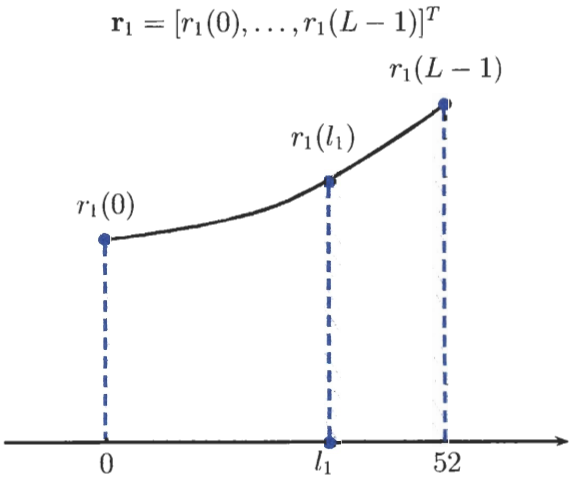


Figure 7.4: Generating the data matrix **A**.

Having established $\mathbf{r}_n$ for $n = 0, \ldots, N-1$, the next step is to find a suitable

representation so that the array elements in $\mathbf{r}_n$ can be retrieved easily through a simple mapping. Since the source for signal errors is in the sampling position, a representation that can reflect the positional variations would be a logical choice. Define a vector $\mathbf{l}$ as

$$\mathbf{l} = \left[l_0,\ l_1,\ \ldots,\ l_{N-1}\right]^T, \tag{7.4}$$

where $0 \leq l_n \leq L-1$, and $n = 0,\ \ldots,\ N-1$. The interferogram sample values for one possible solution out of all L^N solutions are found from

$$\begin{aligned} \mathbf{r}(\mathbf{l}) &= \left[r_0(l_0),\ r_1(l_1),\ \ldots,\ r_{N-1}(l_{N-1})\right]^T \\ &= \left[I_0,\ \ldots,\ I_{N-1}\right]^T \\ &= \mathbf{I}. \end{aligned} \tag{7.5}$$

For example, to recover the n^{th} sample, l_n is estimated to be any integer between 0 and $L-1=52$. Then the corresponding function value is found from $I_n = r_n(l_n)$ for this particular sample. Details on creating one data segment $\mathbf{r}_n$ are shown in Figure 7.5. Now the problem has reduced to estimating an integer vector of length N with the condition that each single element is bounded between 0 and 52. Notice that we have reduced the original task into a combinatorial optimization problem in which the solution of interest belongs to a finite and countable set.

7.1.2.2 Initialization

Two factors must be considered in order to initialize a population of potential solutions: the population size, and the way it is generated. A simple and popular approach is to randomly select individuals from all possible solutions using a uniform or Gaussian distribution. In our case, we could initialize an individual $\mathbf{l}^p$ as N truncated normally distributed random integers $\sim N(26, \sigma)$. The vector has a mean of 26 since it ranges from 0 to 52, and sample 26 is the default location; σ is the standard deviation of $\mathbf{l}^p$. The superscript p indicating that the vector is p^{th} member

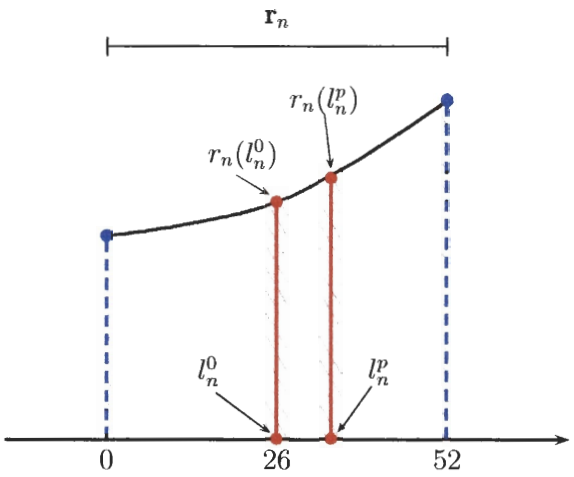
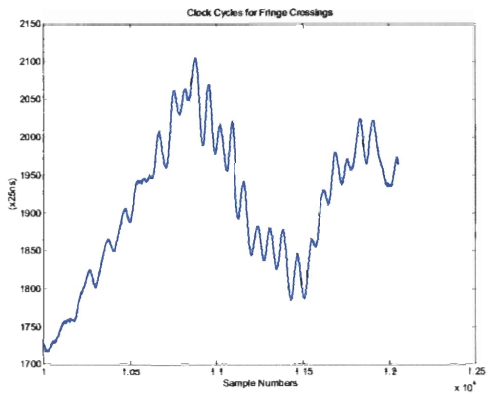


Figure 7.5: The representation of n^{th} data segment.

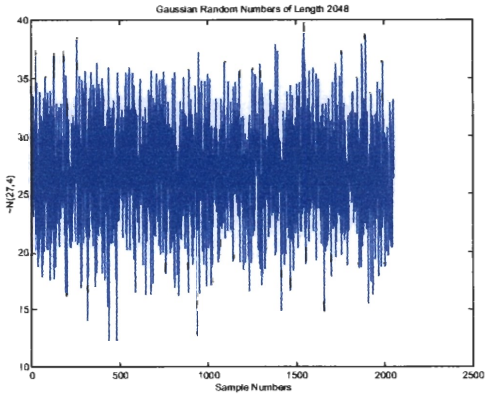
of the population, and $p = 0, \ldots, P-1$ with P representing the population size. This formulation seems to be an acceptable choice. However, if we take a closer look at a typical FTS clock cycle as shown in Figure 3.3, and zoom in on any data segment of length $N = 2048$ (Figure 7.6(a)), we see that the clock signal does not behave like random noises as shown in Figure 7.6(b), instead, it oscillates in a wave like motion with correlations among samples. How can we capture its characteristics and yet still preserve the randomness of the population? One potential solution is to use a simple one dimensional random walk to generate $\mathbf{l}^p$. First, we let $Z_1, Z_2, \ldots, Z_{N-1}$ be independent identically distributed random variables with $P(Z_n = 1) = P(Z_n = -1) = \frac{1}{2}$. The random walk process X_n is

$$X_n = \sum_{k=0}^{n-1} Z_k \quad n = 1, 2, \ldots, N-1, \tag{7.6}$$

and $X_0 = 0$. Next, we set $\mathbf{l}^p$ to be the sum of a constant vector and X_n so that $\mathbf{l}^p$ has a mean of 26. In addition, the elements of $\mathbf{l}^p$ are set to be bounded between 0 and 52. Figure 7.7 shows an example of an initialized population that is generated based on a random walk process.



(a) One segment of the clock signal with length of $N = 2048$.



(b) A normally distributed random vector $\sim N(26, 4)$ with length of $N = 2048$.

Figure 7.6: A comparison between the clock signal and a Gaussian random vector of the same length.

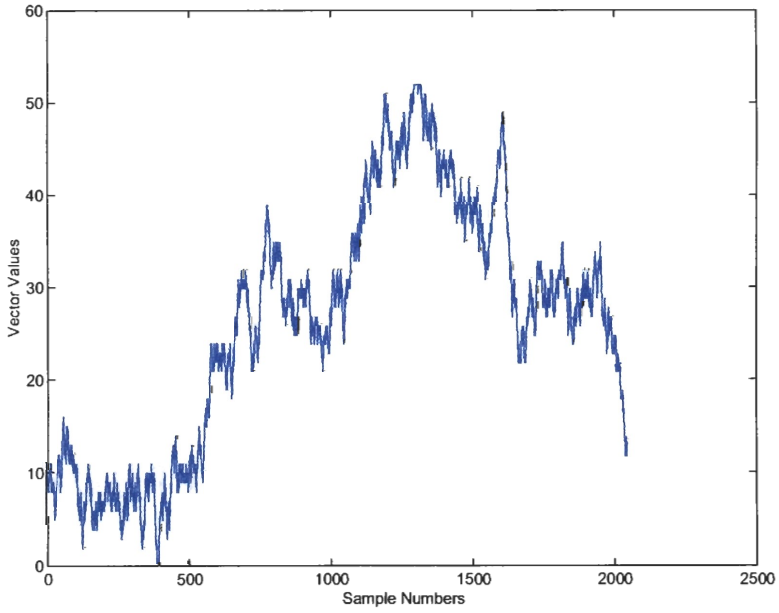


Figure 7.7: Generating an individual $\mathbf{l}^p$ based on a random walk process.

The next task in the initialization process is to determine a proper population size. This step is essentially experimental. A series of size choices are tested, ranging from 10 individuals to 100. The results (presented in the last section of this chapter) reveal that there are no significant differences in terms of the quality of restored interferograms. In general, a greater number of individuals has the advantage of introducing diversity to the population, but it requires more computing power. The reason that the size does not matter much in this case is because one individual of the population is set to be the default measurement in order to satisfy the data consistency constraint described in Section 5.3. Here we implant the equal time interferogram measurement in the population, so that other members can be influenced by it. We assign $\mathbf{l}^0$ to denote the measured equal time interferogram, and the rest of the members $\mathbf{l}^p$ are written as follows:

$$\mathbf{l}^0 = [l_0^0, l_1^0, \ \ldots, \ l_{N-1}^0]^T; \tag{7.7}$$

$$\mathbf{l}^p = [l_0^p, l_1^p, \ \ldots, \ l_{N-1}^p]^T, \tag{7.8}$$

for $p = 1, \ \ldots, \ P-1$. The actual interferogram population is expressed as:

$$\mathbf{I}^0 = \mathbf{r}(\mathbf{l}^0), \tag{7.9}$$

and

$$\begin{aligned} \mathbf{I}^p &= \mathbf{r}(\mathbf{l}^p) \\ &= \left[I_0^p, \ \ldots, \ I_{N-1}^p\right]^T \\ &= \left[r_0(l_0^p), \ r_1(l_1^p), \ \ldots, \ r_{N-1}(l_{N-1}^p)\right]^T, \ p = 1, \ \ldots, \ P-1. \end{aligned} \tag{7.10}$$

This scheme is illustrated in Figure 7.5 for an arbitrary sample n. Where $I_n^0 = r_n(l_n^0)$ is the measured function value for sample n; its corresponding index l_n^0 is centered at 26, which is true for all samples ($n = 0, \ldots, N-1$). The randomly generated p^{th} member of the population l_n^p is located anywhere between 0 and 52, and its function value is written as $I_n^p = r_n(l_n^p)$.

7.1.2.3 *Evaluation*

A properly selected evaluation function plays an important role in the success of an evolutionary algorithm. Based on the formulation in Equation (5.7), the evaluation function for the p^{th} individual is stated as

$$\begin{aligned} eval_{f_i}(\mathbf{I}^p) &= \sum_i w_i f_i(\mathbf{r}(\mathbf{I}^p)) \\ &= \sum_i w_i f_i(\mathbf{I}^p). \end{aligned} \tag{7.11}$$

This function is expressed as a weighted combination of three objects, namely, the entropy, the out-of-band signal, and the OOB signal deviation. To derive an expression for the minimum entropy criterion, first, we define the DFT of each estimate:

$$\begin{aligned} \mathbf{E}^p &= \mathcal{F}\{\mathbf{I}^p\} \\ &= \sum_{n=0}^{N-1} I_n^p e^{-j2\pi nk/N},\ k = 0,\ \ldots,\ N-1, \end{aligned} \tag{7.12}$$

which is composed of a real and a imaginary part

$$\mathbf{E}^p = \mathbf{E}_R^p + j\mathbf{E}_I^p. \tag{7.13}$$

Note that the spectrum is complex due to the fact that the estimated non-ideal interferogram may not be symmetrical about the ZPD. Now we can estimate the power spectrum of $\mathbf{I}^p$ as

$$\mathbf{P}^p = \frac{1}{N}\left|\mathbf{E}^p\right|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1} I_n^p e^{-j2\pi nk/N}\right|^2. \tag{7.14}$$

From the definition given in [33], the entropy H^p for a Gaussian random process with power spectrum $\mathbf{P}^p$ is

$$H^p = \sum_k \ln P_k^p. \tag{7.15}$$

In order to compare and rank the entropy values among all solution candidates, a normalization procedure is required so that all members in the population will have

an equal area under their spectra as written in Equation (7.16).

$$\sum_k \tilde{E}_k^0 = \sum_k \tilde{E}_k^1 = \cdots = \sum_k \tilde{E}_k^{P-1}. \tag{7.16}$$

Recall that the relationship between the area under an entire spectrum and the first sample value in space is

$$I_0^p = \frac{1}{N} \sum_k E_k^p, \tag{7.17}$$

then the requirement stated in Equation (7.16) can be satisfied by setting the ZPD values of all solutions to be equal. By selecting the anchor ZPD value to be the default ZPD measurement I_0^0, and setting the ZPDs of the rest individuals to have the same value as I_0^0, a modified population can be written as

$$\tilde{I}_n^p = \begin{cases} I_n^0 & n = 0 \\ I_n^p & n = 1, \ \ldots, \ N-1, \end{cases} \tag{7.18}$$

where

$$\tilde{I}_0^0 = \tilde{I}_0^1 = \cdots = \tilde{I}_0^{P-1}. \tag{7.19}$$

The updated spectrum and power spectrum of $\tilde{I}_n^p$ are

$$\tilde{\mathbf{E}}^p = \mathcal{F}\{\tilde{\mathbf{I}}^p\} \tag{7.20}$$

and

$$\tilde{\mathbf{P}}^p = \frac{1}{N} \left| \tilde{\mathbf{E}}^p \right|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} \tilde{I}_n^p e^{-j2\pi nk/N} \right|^2, \tag{7.21}$$

respectively. From this, we define the entropy measure as:

$$f_1'(\mathbf{I}^p) = \sum_k \ln \tilde{P}_k^p. \tag{7.22}$$

The second objective is the total out-of-band signal of the interferogram. Since the spectrum is confined to lie within the band between $w_1 < k < w_2$, any signal outside this region can be quantified using the objective function

$$f_2'(\mathbf{I}^p) = \sum_{k<w_1, \, k>w_2} \left| E_{R,\,k}^p \right|. \tag{7.23}$$

Another objective is imposed on the spectrum to monitor the out-of-band signal; this function measures the total deviation of the OOB spectral samples from their mean values. A high sample mean indicates that the spectrum has a "wavy" baseline, whereas a low average value implies a flat OOB signal. This objective is given by

$$f_3'(\mathbf{I}^p) = \sum_{k<w_1,\, k>w_2} \left| E^p_{R,\,k} - \frac{\sum\limits_{k<w_1,\, k>w_2} E^p_{R,\,k}}{N - (w_2 - w_1 + 1)} \right|. \tag{7.24}$$

Finally, to be able to combine the results from all three evaluation functions and generate a single output, a normalization step is necessary. This can be done using the equation

$$f_i(\mathbf{I}^p) = \frac{f_i'(\mathbf{I}^p)}{\sqrt{\sum_{p=0}^{P-1} |f_i'(\mathbf{I}^p)|^2}}, \text{ for } i = 1, 2, 3. \tag{7.25}$$

Table 7.1: A comparison between three evaluation functions.

Trial	Known Performance			Entropy		OOB Signal		OOB Sig. Dev.		Sum	Rank
	SNR_E	$\|\mathbf{e}\|$	Rank	Results	Rank	Results	Rank	Results	Rank		
$\hat{\mathbf{I}}^7$	54.88	0.71e5	1	0.3389	8	0.1338	3	0.3626	7	0.84	3
$\hat{\mathbf{I}}^6$	53.88	0.75e5	2	0.2459	1	0.3931	7	0.2541	3	0.89	4
$\hat{\mathbf{I}}^5$	51.72	0.83e5	3	0.3196	4	0.0331	1	0.2019	1	0.55	1
$\hat{\mathbf{I}}^1$	51.69	0.84e5	4	0.2856	2	0.1326	2	0.2456	2	0.66	2
$\hat{\mathbf{I}}^4$	50.54	0.89e5	5	0.3397	9	0.2427	4	0.3803	10	0.96	5
$\hat{\mathbf{I}}^0$	48.78	0.96e5	6	0.3250	6	0.4412	10	0.3562	6	1.12	10
$\hat{\mathbf{I}}^8$	47.46	1.03e5	7	0.3250	5	0.2742	5	0.3701	9	0.97	6
$\hat{\mathbf{I}}^9$	47.26	1.04e5	8	0.3386	7	0.3984	8	0.3658	8	1.10	9
$\hat{\mathbf{I}}^2$	46.17	1.10e5	9	0.3423	10	0.3593	6	0.2829	4	0.98	7
$\hat{\mathbf{I}}^3$	45.39	1.15e5	10	0.2873	3	0.4386	9	0.2840	5	1.01	8

How well does each of these objectives perform in terms of choosing and ranking a population of solutions? Suppose we randomly generate 10 individuals $\hat{\mathbf{I}}^0$ to $\hat{\mathbf{I}}^9$, and evaluate their true performance by the SNRs of the estimated spectra and $\mathcal{L}_2$ norm errors of the estimated interferograms. The error in the $\mathcal{L}_2$ norm is defined as

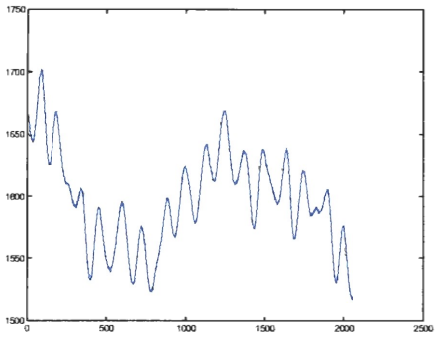
$$\|\mathbf{e}\| = \left\{ \sum_{n=0}^{N-1} e_n^2 \right\}^{1/2} = \left\{ \sum_{n=0}^{N-1} (I_n - \hat{I}_n)^2 \right\}^{1/2}, \tag{7.26}$$

and the spectral signal-to-noise (SNR) is

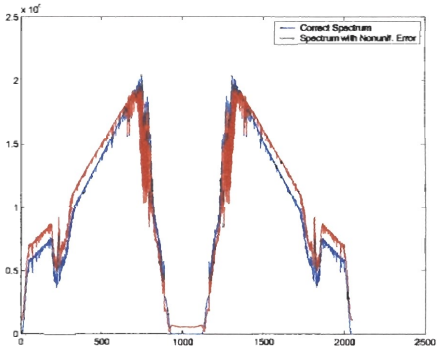
$$\mathrm{SNR}_E = 10 \log \left(\frac{\sum_{k=0}^{N-1} |E_k|^2}{\sum_{k=0}^{N-1} |E_k - \hat{E}_k|^2} \right). \tag{7.27}$$

The results along with their ranks are shown in Table 7.1. Note that the order of these test interferograms has been rearranged according to their ranks. Next, we test how accurate can each of these three objectives predict the correct ranking. The normalized outputs and their corresponding ranks for each evaluation function are shown in columns. The summed results from all three evaluation functions are shown in the last two columns. There are no obvious indications that one objective is significantly better than the others, however, a simple conclusion can be drawn from the summed results is that the objective functions can roughly identify an individual as in the top or bottom 50% of the population, which is a useful fact in the up-coming selection stage.

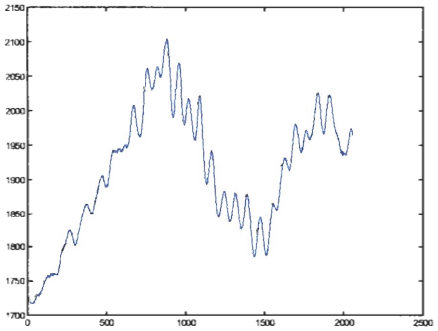
The next question is: how well does each objective perform in terms of recovering the uniform spectrum? Three nonuniform interferograms are generated for this test, their corresponding clock signals are plotted in Figures 7.8 (a), (c), and (e). The spectra of these irregular interferograms are plotted against uniform spectra in the background as seen in Figures 7.8 (b), (d), and (f). Every one of these signals has been reconstructed using the three evaluation functions separately, their individual performance in terms of SNRs and error norms are shown in Table 7.2. The last column in the table shows the results obtained using the equal-weighted sum of all three objectives. The test for each interferogram is repeated five times and the average results are shown in bold faced numbers. Again, no indication of major superiority of any one of the three functions, therefore, the weighting factors are simply set to be equal based on this observation. Note that the combined evaluation function



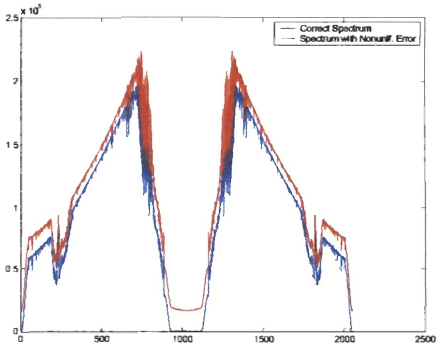
(a) The clock cycle of test signal 1.



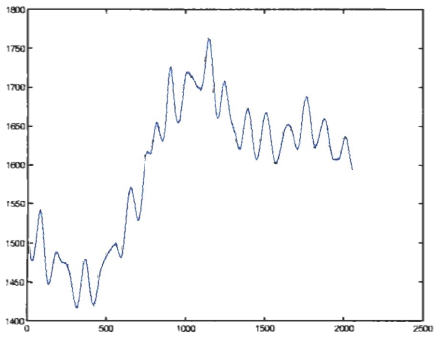
(b) The uniform (blue) and nonuniform (red) spectra of test signal 1.



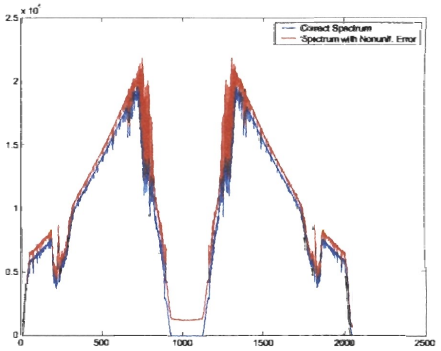
(c) The clock cycle of test signal 2.



(d) The uniform (blue) and nonuniform (red) spectra of test signal 2.



(e) The clock cycle of test signal 3.



(f) The uniform (blue) and nonuniform (red) spectra of test signal 3.

Figure 7.8: The clock cycle segments and the spectra of three corresponding test signals.

Table 7.2: Test results of each individual evaluation function from three nonuniform interferograms.

Signal	Entropy		OOB Signal		OOB Sig. Dev.		Combined	
	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
$I^1(n\Delta t)$	51.74	8.34e4	50.28	8.98e4	49.73	9.22e4	55.77	6.82e4
SNR_E=47.26	64.96	4.31e4	49.95	9.13e4	52.70	7.95e4	60.05	5.51e4
$\|\mathbf{e}\|$=1.04e5	50.33	8.95e4	66.68	3.95e4	48.70	9.71e4	55.73	6.83e4
	50.65	8.81e4	56.17	6.68e4	51.14	8.59e4	61.96	5.01e4
	52.82	7.90e4	60.53	5.38e4	52.68	7.96e4	61.21	5.19e4
Mean	**54.10**	**7.66e4**	**56.72**	**6.82e4**	**50.99**	**8.69e4**	**58.94**	**5.87e4**
	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
$I^2(n\Delta t)$	64.05	4.51e4	51.27	8.54e4	49.54	9.31e4	65.92	4.11e4
SNR_E=47.15	51.65	8.38e4	51.62	8.39e4	51.44	8.46e4	59.91	5.55e4
$\|\mathbf{e}\|$=1.73e5	47.65	1.02e5	67.39	3.81e4	48.29	9.91e4	59.41	5.69e4
	55.59	6.69e4	55.99	6.75e4	65.32	4.23e4	66.65	3.96e4
	48.67	9.73e4	52.30	8.11e4	48.68	9.72e4	59.12	5.76e4
Mean	**53.52**	**7.95e4**	**55.71**	**7.14e4**	**52.65**	**8.33e4**	**62.20**	**5.01e4**
	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
$I^3(n\Delta t)$	54.05	7.32e4	50.25	8.99e4	62.54	4.86e4	68.01	3.70e4
SNR_E=47.40	58.65	5.88e4	73.02	2.88e4	61.26	5.18e4	64.79	4.34e4
$\|\mathbf{e}\|$=1.04e5	49.65	9.47e5	56.17	6.69e4	56.86	6.46e4	76.75	2.39e4
	54.59	7.33e4	57.84	6.15e4	61.04	5.24e4	75.39	2.56e4
	56.67	6.69e4	50.51	8.87e4	60.67	5.34e4	59.97	5.53e4
Mean	**54.55**	**7.34e4**	**57.56**	**6.71e4**	**60.48**	**5.52e4**	**68.98**	**3.70e4**

out-performed any single function alone. In addition, when compared with single-objective outputs, the multi-objective test results show a more stable trait over the five runs.

So far, we have not yet considered the symmetry constraint. In Section 7.1.1, we discussed the advantage of having a soft constraint when solving optimization problems in general. This can be achieved by assigning a penalty factor to the infeasible solutions. For the interferogram reconstruction case, it is easy to impose a hard constraint that "forces" the interferogram to become symmetrical. However, this may not be a good choice in a non-ideal situation since real interferogram measurements experience varies noises, which make them become slightly asymmetrical. Therefore, we would not want to reject those solutions with high evaluation scores but failed

to satisfy the symmetry condition. If a symmetrical interferogram is required afterward, we can always correct the asymmetrical interferogram by setting the imaginary part of its spectrum to zero. There are many techniques for designing the constraint functions in evolutionary algorithms as described in [3, 48]. We select the ranking system because of its simplicity and effectiveness for this particular problem. To formulate the symmetry constraint, if we adopt the notation introduced in [48], then the evaluation function is

$$eval_u(\mathbf{I}^p) = eval_{f_i}(\mathbf{I}^p) + Q(\mathbf{I}^p), \tag{7.28}$$

where

$$Q(\mathbf{I}^p) = \omega\, q(\mathbf{I}^p) \tag{7.29}$$

represents the penalty function for asymmetrical solutions; similarly, ω is the weighting factor for this penalty function, which is defined according to the importance of the symmetry constraint. The penalty is quantified as follows

$$q'(\mathbf{I}^p) = \sum_{n=1}^{N/2-1} \left| I_n^p - I_{N-n}^p \right|, \tag{7.30}$$

which is normalized by

$$q(\mathbf{I}^p) = \frac{q'(\mathbf{I}^p)}{\sqrt{\sum_{p=0}^{P-1} |q'(\mathbf{I}^p)|^2}}. \tag{7.31}$$

7.1.2.4 Selection

Once a population is evaluated, the selection is done based on each member's fitness at the evaluation stage. Naturally, the rule of *the survival of the fittest* has come to shine in this step. From the previous observation, a selection is made where the top 50% fittest individuals from the last step are chosen. Define a function $\alpha(i)$ that arranges the order of the estimates according to each solution's relative fitness in the population. For example, $\alpha(0)$ stores the index of the strongest individual, whereas

$\alpha(P-1)$ stores the index of the weakest solution. Thus,

$$\alpha(i) = order\,\{eval_u(\mathbf{I}^p)\}, \tag{7.32}$$

for

$$\begin{aligned} i &= 0, \ldots, P-1 \\ 0 &\leq \alpha(i) \leq P-1. \end{aligned} \tag{7.33}$$

From the order in $\alpha(i)$, the corresponding evaluation function is

$$eval_u(\mathbf{I}^{\alpha(i)}) = sort\,\{eval_u(\mathbf{I}^p)\}, \tag{7.34}$$

such that

$$eval_u(\mathbf{I}^{\alpha(0)}) \leq eval_u(\mathbf{I}^{\alpha(1)}) \leq \cdots \leq eval_u(\mathbf{I}^{\alpha(P-1)}). \tag{7.35}$$

Notice that the highest ranking individual has the lowest evaluation result since this is a minimization problem. The selected set of candidates is

$$sel(\mathbf{I}^{\alpha(j)}) = \{\mathbf{I}^{\alpha(0)}, \mathbf{I}^{\alpha(1)}, \ldots, \mathbf{I}^{\alpha(J)}\}, \tag{7.36}$$

where

$$\begin{cases} J &= \frac{P-1}{2} - 1, \quad \text{odd } P \\ J &= \frac{P}{2} - 1, \quad \text{even } P. \end{cases} \tag{7.37}$$

An example of the selection rule is shown in Table 7.3, which borrows the results from the first few columns in Table 7.1. For a group of individuals $\hat{\mathbf{I}}^p$, where $p = 0, \ldots, 9$, we have their SNRs and corresponding ranks listed in the second and fourth rows, respectively. The indices of the top five fittest members are shown in row number five as function values of $\alpha(i)$. And finally, the "survivors" of this selection round are first five elements in the last row.

Table 7.3: An example of the selection rule.

Trial	$\hat{\mathbf{I}}^0$	$\hat{\mathbf{I}}^1$	$\hat{\mathbf{I}}^2$	$\hat{\mathbf{I}}^3$	$\hat{\mathbf{I}}^4$	$\hat{\mathbf{I}}^5$	$\hat{\mathbf{I}}^6$	$\hat{\mathbf{I}}^7$	$\hat{\mathbf{I}}^8$	$\hat{\mathbf{I}}^9$
SNR_E	48.78	51.69	46.17	45.39	50.54	51.72	53.88	54.88	47.46	47.26
p, i	0	1	2	3	4	5	6	7	8	9
rank	6	4	9	10	5	3	2	1	7	8
$\alpha(i)$	7	6	5	1	4	0	8	9	2	3
$\mathbf{I}^{\alpha(i)}$	$\mathbf{I}^7$	$\mathbf{I}^6$	$\mathbf{I}^5$	$\mathbf{I}^1$	$\mathbf{I}^4$	$\mathbf{I}^0$	$\mathbf{I}^8$	$\mathbf{I}^9$	$\mathbf{I}^2$	$\mathbf{I}^3$

7.1.2.5 Variation

In this section, we introduce a variation operator. When applied to the previously selected members, it will produce the next generation of offspring. This new generation does not include any members from the previous one, instead, all individuals are descendants of the selected solutions. This can be accomplished using a single-parent scheme. In contrast to the mating or the cross over technique, each individual is mutated from only one parent. Therefore, it can only possess the characteristics of a single parent. To be able to create such an offspring, we use a small amount of zero mean Gaussian perturbation as the mutation operator, and apply this variation to the parent through addition. In other words, the offspring is generated by adding a small zero mean Gaussian integer vector to the parent vector. We allow each parent to have two children so that the new generation will have the same population size as before. The Gaussian variation is defined as

$$\tilde{\mathbf{d}}^p \sim N(0, \sigma(g)), \tag{7.38}$$

and the mutation operator $\mathbf{d}^p$ is the truncated integer version of $\tilde{\mathbf{d}}^p$. The new variable $g = 0, 1, \ldots, G$, denotes the generation number. In Equation (7.38), the standard deviation σ is written as a function of g, which implies that the variation is not constant across different generations. More specifically, this type of varying mechanism can be classified as a dynamic Gaussian mutation as described in [48]. The justification

of this operation is based on the fact that as the generation grows, the population becomes more adapted to its environment, therefore, a smaller amount of mutation is required. Suppose we let the variation be proportional to the number of generations, this leads to

$$\sigma(g) \propto \frac{1}{g},\ g = 0, 1, \ldots, G-1. \tag{7.39}$$

This equation tells us that as time goes by (more generations are created), the variation becomes less. A new generation is emerged from the old one with the help of the dynamic Gaussian mutation operator. This step can be written as

$$(\mathbf{l}^{(0,\alpha(j))}, \mathbf{d}^p) \ \Rightarrow \ \mathbf{l}^{(1,p)}, \tag{7.40}$$

where $\mathbf{l}^{(0,\alpha(j))}$ represents the selected individuals, $\mathbf{d}^p$ is the variation operator, and $\mathbf{l}^{(1,p)}$ is the new generation. The generation variable has been incorporated into this equation. For example, an arbitrary sample $l_n^{(g,p)}$ tells us that it is the n^{th} sample that belongs to the p^{th} member of the population from the g^{th} generation. The 0^{th} generation, i.e., $g = 0$, is used to represent the initialized population. Using the two offspring per single parent scheme, the actual members of this new generation (assume P is odd) are

$$\begin{aligned}
\mathbf{l}^{(1,0)} &= \mathbf{l}^{(0,0)} \\
\mathbf{l}^{(1,1)} &= \mathbf{l}^{(0,\alpha(0))} + \mathbf{d}^1 \\
\mathbf{l}^{(1,2)} &= \mathbf{l}^{(0,\alpha(0))} + \mathbf{d}^2 \\
\mathbf{l}^{(1,3)} &= \mathbf{l}^{(0,\alpha(1))} + \mathbf{d}^3 \\
\mathbf{l}^{(1,4)} &= \mathbf{l}^{(0,\alpha(1))} + \mathbf{d}^4 \\
\vdots \quad & \quad \vdots \quad \vdots \\
\mathbf{l}^{(1,P-2)} &= \mathbf{l}^{(0,\alpha(J))} + \mathbf{d}^{P-2} \\
\mathbf{l}^{(1,P-1)} &= \mathbf{l}^{(0,\alpha(J))} + \mathbf{d}^{P-1}.
\end{aligned} \tag{7.41}$$

Note that the first member is always the default measurement as described in previous sections. In Figure 7.9, new members are generated for an odd-number population,

in which every parent creates two offsprings. If an even number of the population is used, then only one child is produced from the last ranking solution.

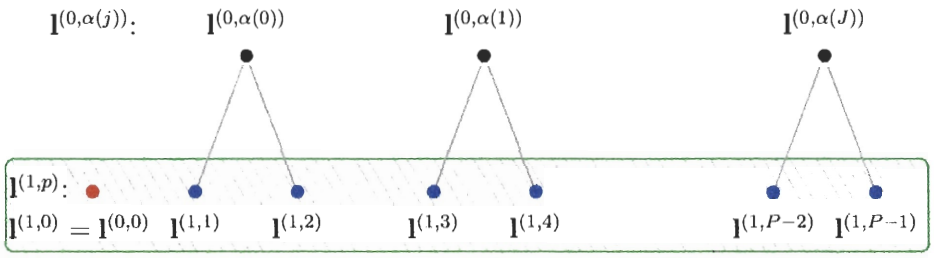


Figure 7.9: Creating the first generation when P is odd.

7.1.2.6 Stopping and Post-process

This is the final stage of the algorithm, in which the iteration stops after a certain number of generations. The simplest way to achieve the termination is to stop the program after a fixed number of generations. A more sophisticated method would involve setting up a stopping condition, with the program simply stopping when the condition is met. A third method is a combination of both techniques, which allows the program to terminate when either one of the two conditions is met. For example, the iteration stops after a maximum number of generations is reached or after a certain performance criterion is satisfied. This last method seems to be the best choice in terms of finding the most favorable solutions, and yet, escaping from the danger of letting the program fall into an infinite loop.

After the program has been successfully terminated, the fittest members, $\mathbf{l}^{(G-1,\alpha(j))}$, are selected according to the same selection rule described earlier. At this point, we could choose the best ranking individual to be our ultimate solution. Alternatively, we could use the average of the entire selection as the final answer. The second method is chosen because, as shown previously, the evaluation function can correctly estimate the top 50% of the population. In other words, we can identify the top

50% fittest members with a certain degree of confidence, but not each individual's ranking within that collection. In summary, at a given generation g, the selected interferograms corresponding to $\mathbf{l}^{(g,\alpha(j))}$ are

$$\mathbf{I}^{(g,\alpha(j))} = \mathbf{r}(\mathbf{l}^{(g,\alpha(j))}). \tag{7.42}$$

A single solution is found from the average of all interferogram estimates, and

$$\hat{\mathbf{I}}^g = \frac{1}{J}\sum_{j=0}^{J-1}\mathbf{I}^{(g,\alpha(j))}. \tag{7.43}$$

This solution together with the value of g are used to determine whether the stopping condition is satisfied:

$$\text{Does} \quad g < G \quad \text{or} \quad eval(\hat{\mathbf{I}}^g) < c_{stop} \ ?$$

If either one of these two conditions is reached, then the mean solution is

$$\mathbf{I}^{(G-1,\alpha(j))} = r_n(\mathbf{l}^{(G-1,\alpha(j))}) \tag{7.45}$$

$$\hat{\mathbf{I}}_m = \frac{1}{J}\sum_{j=0}^{J-1}\mathbf{I}^{(G-1,\alpha(j))}. \tag{7.46}$$

As described in Section 6.1.2, the ZPD value can be estimated from the spectrum DC offset. The spectrum can be obtained from

$$\hat{\mathbf{E}}_m = \mathcal{F}\{\hat{\mathbf{I}}_m\}. \tag{7.47}$$

After the DC offset correction, the new spectrum can be expressed as

$$\hat{E}_k = \hat{E}_{m,k} - \frac{\sum\limits_{k<w_1,\, k>w_2}\hat{E}_{m,\,k}}{N-(w_2-w_1+1)}, \ \forall\ k. \tag{7.48}$$

Thus, the final solution is

$$\hat{\mathbf{I}} = \mathcal{F}^{-1}\{\hat{\mathbf{E}}\}. \tag{7.49}$$

The complete structure of *Algorithm A* is shown in Figure 7.10. In the following section, a slightly different approach for reconstructing the nonuniform interferogram is presented.

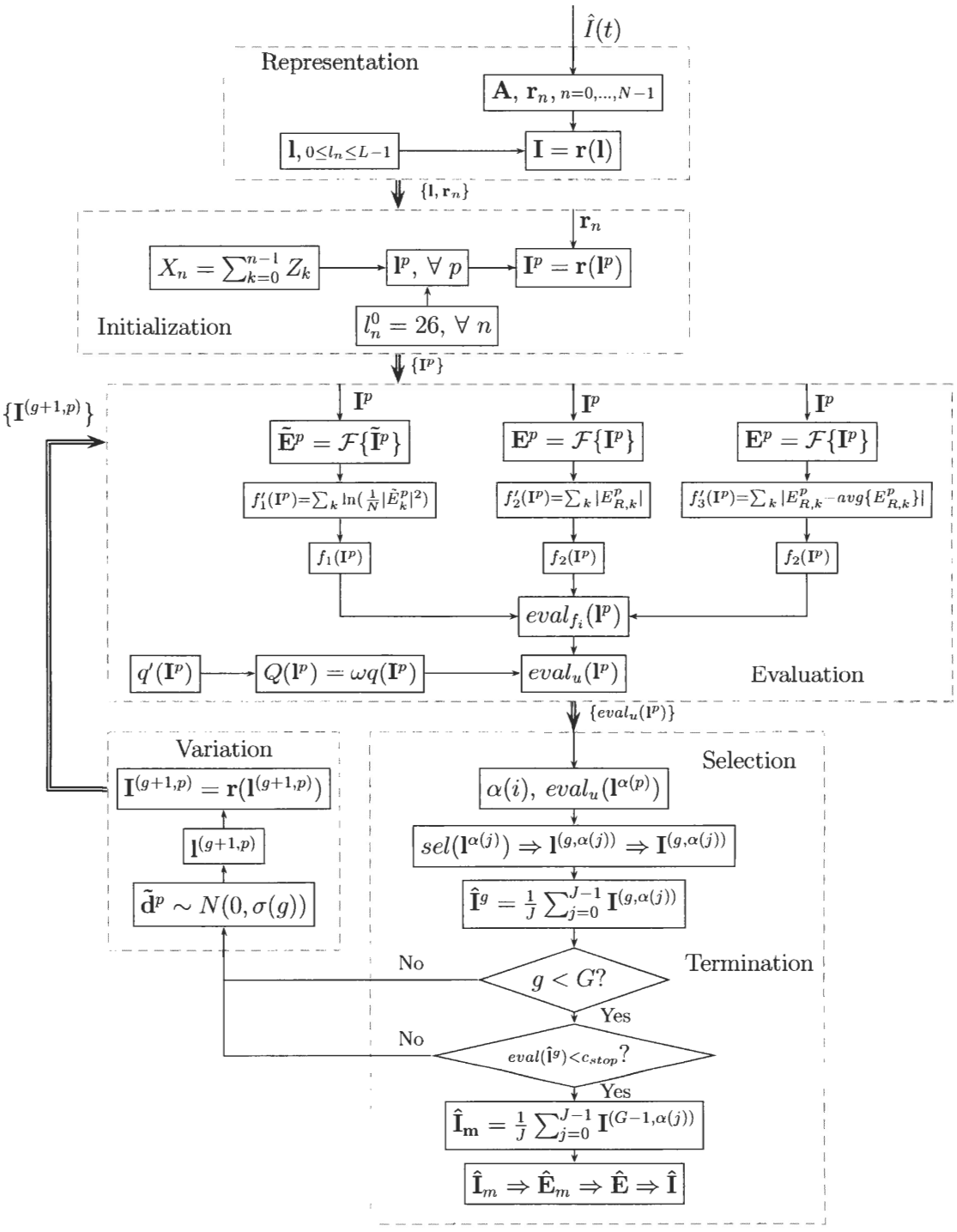


Figure 7.10: The structure of *Algorithm A*.

7.1.3 Interferogram Reconstruction: Algorithm B

The second reconstruction algorithm emphasizes the estimation of interferogram values instead of sampling locations. In Equation (5.6), the vector to be optimized is stated as the data value of each measurement. In this algorithm, only half (a single scan) of the interferogram is being estimated, the other half is simply set equal to the first scan in order to satisfy the symmetry constraint. We name this routine *Algorithm B*. In the previous section, we have discussed the advantage of having a soft symmetry constraint, however, in two special cases the hard constraint imposed in *Algorithm B* is a better choice. The first case occurs when only a single scan is available. The second case occurs when one scan is significantly better than the other scan; the amount of the difference between these two scans is determined experimentally, and will be discussed later in this Chapter. The structure of this second method is very similar to that of *Algorithm A*; the procedures of these two methods are almost identical except for the representation and initialization sections.

In *Algorithm B*, the matrix B of size $\frac{N}{2} \times L$ represents all possible interferogram sample values in a single scan, which can also be written as row vectors $\mathbf{s}_n$:

$$\underset{\frac{N}{2}\times L}{\mathbf{B}} = \begin{bmatrix} b_{00} & b_{01} & \dots & b_{0(L-1)} \\ b_{10} & b_{11} & \dots & b_{1(L-1)} \\ \vdots & \vdots & \dots & \vdots \\ b_{(\frac{N}{2}-1)0} & b_{(\frac{N}{2}-1)1} & \dots & b_{(\frac{N}{2}-1)(L-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_0^T \\ \mathbf{s}_1^T \\ \vdots \\ \mathbf{s}_{\frac{N}{2}-1}^T \end{bmatrix}, \tag{7.50}$$

and

$$\mathbf{s}_n = \left[b_{n0},\ b_{n1},\ \dots,\ b_{n(L-1)}\right]^T,\ \text{for } n = 0, \dots, \frac{N}{2}. \tag{7.51}$$

The definitions of $\mathbf{B}$ and $\mathbf{s}_n$ are somewhat different from those of Equation (7.1). Since only half of the data samples are being estimated, at a given point n, all nearby sample values from the two sample points in the original measurement that are symmetrical about the ZPD, i.e., n and $N-n$, need to be bounded by the elements

of $\mathbf{s}_n$. For instance, from *Algorithm A*, the vector $\mathbf{r}_n$ represents all possible values in the neighborhood of point n, whereas $\mathbf{r}_{N-n}$ stores possible sample values around point $N-n$, which is the symmetrical counterpart of point n. Then it would be desirable for vector $\mathbf{s}_n$, with the same length as $\mathbf{r}_n$, to span the feasible range of interferogram sample values at sample points n and $N-n$. We set the largest and smallest values for both $\mathbf{r}_n$ and its symmetrical counterpart $\mathbf{r}_{N-n}$ for all n as follows

$$\beta_{n,max} = \max\{\mathbf{r}_n, \mathbf{r}_{N-n}\}, \quad \text{and} \tag{7.52}$$

$$\beta_{n,min} = \min\{\mathbf{r}_n, \mathbf{r}_{N-n}\}, \quad n = 0, \ldots, \frac{N}{2}. \tag{7.53}$$

With $\beta_{n,max}$ and $\beta_{n,min}$, the elements of $\mathbf{s}_n$ are set equal to 53 equally spaced samples within this range, which are written as

$$\mathbf{s}_n = \frac{(\beta_{n,max} - \beta_{n,min})}{L}\,\mathbf{t} \quad \text{for } n = 0, \ldots, \frac{N}{2}, \tag{7.54}$$

where $L = 53$ and

$$\mathbf{t} = [1,\ 2,\ 3,\ \ldots,\ L-1, L]^T. \tag{7.55}$$

The representation of *Algorithm B* is

$$\mathbf{l} = [l_0,\ l_1,\ \ldots,\ l_{\frac{N}{2}}]^T, \tag{7.56}$$

and

$$\begin{aligned} \mathbf{s}(\mathbf{l}) &= \left[s_0(l_0),\ s_1(l_1),\ \ldots,\ s_{\frac{N}{2}}(l_{\frac{N}{2}})\right]^T \\ &= \left[I_{h,0},\ \ldots,\ I_{h,\frac{N}{2}}\right]^T \\ &= \mathbf{I}_h. \end{aligned} \tag{7.57}$$

The subscript h reminds us that only half of the interferogram is being estimated. The values of the representation vector $\mathbf{l}$ are the same as in *Algorithm A*, i.e., l_n is an integer between 0 and $L-1$, however, the meaning of $\mathbf{l}$ is no longer the same. Recall in *Algorithm A*, the elements of $\mathbf{l}$ indicate the relative sampling locations, therefore,

we could estimate these locations and evaluate them based on their corresponding sample values. But in *Algorithm B*, this "location-value" information is lost. Instead, we estimate the sample values directly by finding a data range and petition it into L equally spaced amplitude values. Here the elements of $\mathbf{l}$ are merely the indices of these samples, which do not represent the sampling locations.

The initialization steps are very similar to those in *Algorithm A*, except that the total number of samples is $\frac{N}{2}+1$ instead of N. The population $\mathbf{l}^p$ is initialized as

$$\mathbf{l}^0 = [l_0^0,\ l_1^0,\ \ldots,\ l_{\frac{N}{2}}^0]^T; \tag{7.58}$$

$$\mathbf{l}^p = [l_0^p,\ l_1^p,\ \ldots,\ l_{\frac{N}{2}}^p]^T,\ \text{for } p = 1,\ \ldots,\ P-1. \tag{7.59}$$

The corresponding interferogram vectors are

$$\mathbf{I}_h^0 = \mathbf{s}(\mathbf{l}^0), \tag{7.60}$$

and

$$\begin{aligned} \mathbf{I}_h^p &= \mathbf{s}(\mathbf{l}^p) && (7.61)\\ &= [I_{h,0}^p,\ I_{h,1}^p,\ \ldots,\ I_{h,\frac{N}{2}}^p]^T && (7.62)\\ &= [s_0(l_0^p),\ s_1(l_1^p),\ \ldots,\ s_{\frac{N}{2}}(l_{\frac{N}{2}}^p)]^T,\ p = 1,\ \ldots,\ P-1. && (7.63) \end{aligned}$$

In Figure 7.11, vectors $s_n(l_n^p)$ and $s_n(l_n^0)$ represent an arbitrary individual and the default measurement, respectively. Note that the elements of vector $\mathbf{l}^0$ are no longer centered at index number 26 as in *Algorithm A*, this is because the sampling location information of the original measurement is not in $\mathbf{l}^0$ any more. Consequently, an extra step is necessary to determine the values of $\mathbf{l}^0$, which can be derived from the corresponding locations of the measured interferogram sample values in $\mathbf{s}_n$. First, we must decide which half of the measured interferogram to use as our reference. Suppose we separate the interferogram into two parts, and extend each part to the full length

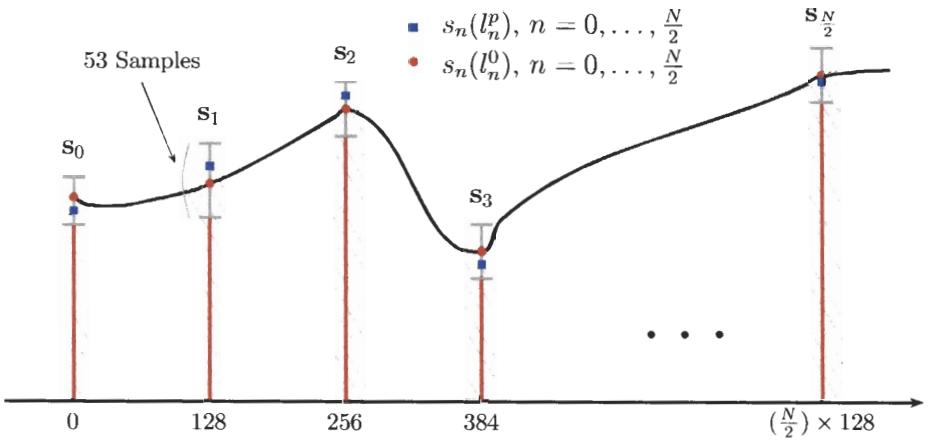


Figure 7.11: Generating the data matrix **B**.

by setting its second half to be equal to the first, then they can be expressed as

$$I_1(n\Delta t) := \begin{cases} I(n\Delta t),\ n = 0, 1, \ldots, \frac{N}{2} \\ I((N-n)\Delta t),\ n = \frac{N}{2} + 1, \ldots, N-1, \end{cases} \tag{7.64}$$

and

$$I_2(n\Delta t) := \begin{cases} I(n\Delta t),\ n = 0; \frac{N}{2}, \ldots, N-1 \\ I((N-n)\Delta t),\ n = 1, \ldots, \frac{N}{2} - 1. \end{cases} \tag{7.65}$$

Next, we apply the same evaluation function $eval_{f_i}(\cdot)$ to both interferograms, and select the one that performed better, then the chosen half, written as $\mathbf{I}_h^0$, is

$$\mathbf{I}_h^0 = sel\left(eval_{f_i}\left(I_1(n\Delta t),\ I_2(n\Delta t)\right)\right). \tag{7.66}$$

Now we need to determine what are the indices of the samples of $\mathbf{I}_h^0$ in $\mathbf{s}_n$, which are the desired values of $\mathbf{I}^0$. Since the elements of $\mathbf{s}_n$ do not represent the exact sample values of $\mathbf{I}_h^0$, a search method is needed to find the locations of the closest data values for all samples. This can be done by comparing every single sample in $\mathbf{I}_h^0$ with its

corresponding possible values in $\mathbf{s}_n$ as shown in Figure 7.12, then $l_n^0 \ \forall \ n$ is

$$l_n^0 = \min_l \left\{ |I_{h,n}^0 - b_{nl}| \right\} \tag{7.67}$$

$$= \min_l \left\{ |I_{h,n}^0 - b_{n0}|, \ |I_{h,n}^0 - b_{n1}|, \ \ldots, \ |I_{h,n}^0 - b_{n(L-1)}| \right\}. \tag{7.68}$$

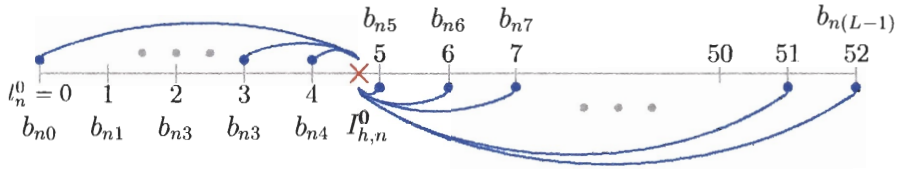


Figure 7.12: Estimating the initial individual $\mathbf{l}^0$.

Once $\mathbf{l}^p$ is initialized, the corresponding interferograms of full lengths are

$$I_n^p := \begin{cases} I_{h,n}^p, & n = 0, \ 1, \ \ldots, \ \frac{N}{2} \\ I_{h,(N-n)}^p, & n = \frac{N}{2} + 1, \ \ldots, \ N-1. \end{cases} \tag{7.69}$$

The following evaluation section is the same as in *Algorithm A*, except there is no constraint penalty factor in the evaluation function since the symmetry constraint is imposed by estimating only half of the interferogram. Same for the selection, variation and stopping steps with the exception of only half the data length is used. The complete structure of *Algorithm B* is shown in Figure 7.13.

7.1.4 Making a Choice: A or B, or A and B

Which one of these two algorithms is a better choice: A or B? The answer depends on the characteristics of the measurement. If one of the two measured interferogram scans is significantly better than the other according to the evaluation result, then B is superior to A. This is because in *Algorithm B*, we take the advantage of having

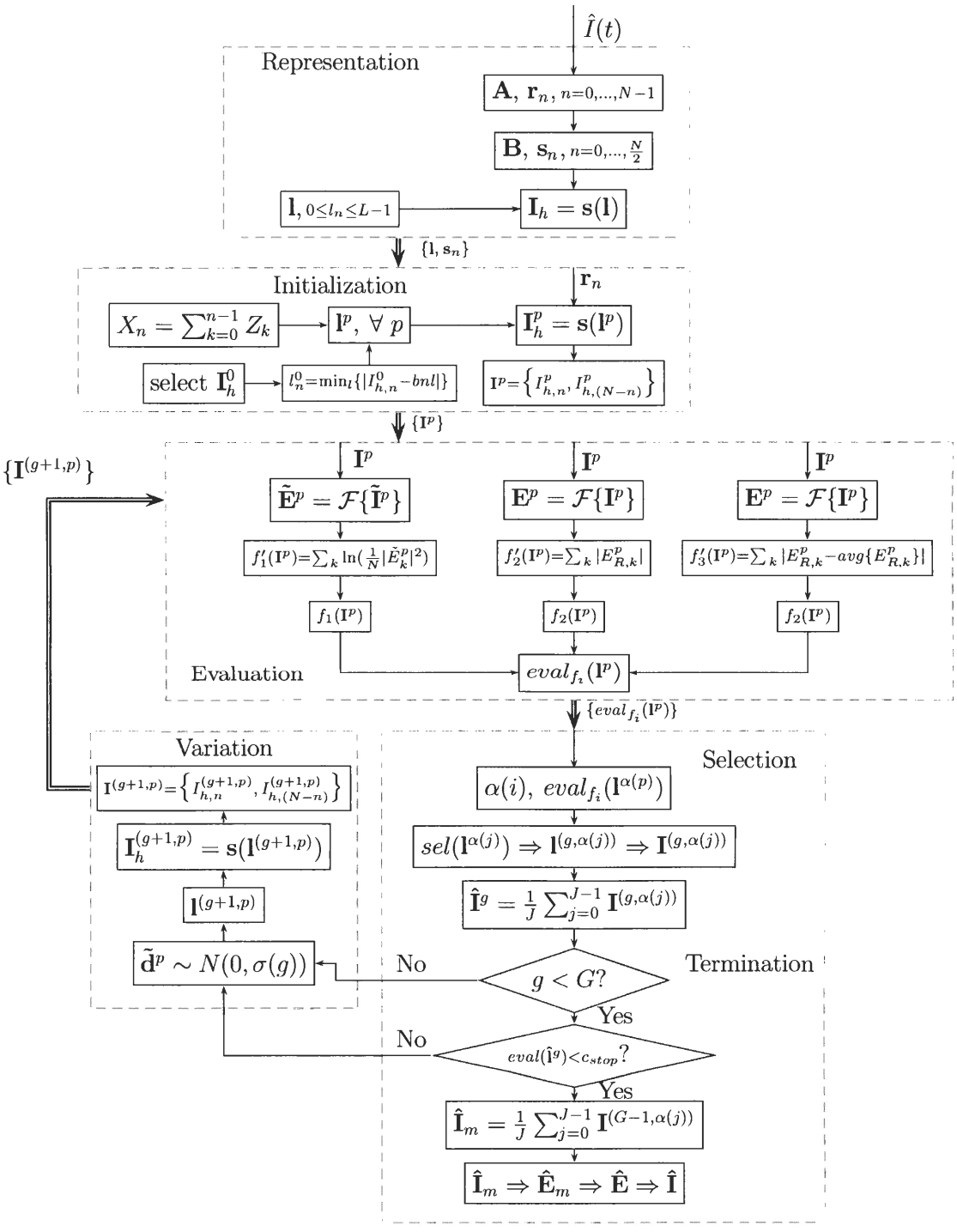


Figure 7.13: The structure of *Algorithm B*.

the better scan in the population at the starting point, which can positively influence all other members in the group. This may occur when only a single side of the interferogram is available or when one scan experiences much more sampling offset errors than the other, which may be the case in a mechanical failure. However, if both scans are comparable in quality, then there is no gain in working on only one of them. Furthermore, it is better to use the entire interferogram in that case, because while one scan may have better measurements for certain samples, the other scan may do a better job at measuring some other samples. Therefore a compromising solution would combine the best of both ways, if a judgment can be made on the interferogram based on a designated measure beforehand, then we can decide which algorithm to use. To design a decision making device, we evaluate both scans with the same function $eval_{f_i}(\cdot)$, the results are

$$\epsilon_1 = eval_{f_i}\big(I_1(n\Delta t)\big), \text{ and} \tag{7.70}$$

$$\epsilon_2 = eval_{f_i}\big(I_2(n\Delta t)\big). \tag{7.71}$$

If one scan is as much as twice better than the other, then *Algorithm B* is selected; otherwise, *Algorithm A* is the better choice. The factor of two is selected based on the simulation performed on several interferograms. The results have shown that when the evaluation score of one scan is two times better than the other scan, *Algorithm B* works better than A; otherwise, A is superior to B. This decision scheme can be written as

$$Algorithm: \begin{cases} B, & \text{if } \left|\frac{\epsilon_1}{\epsilon_2}\right| > 2 \text{ or } < 0.5 \\ A, & \text{otherwise.} \end{cases} \tag{7.72}$$

This decision making process in shown in Figure 7.14.

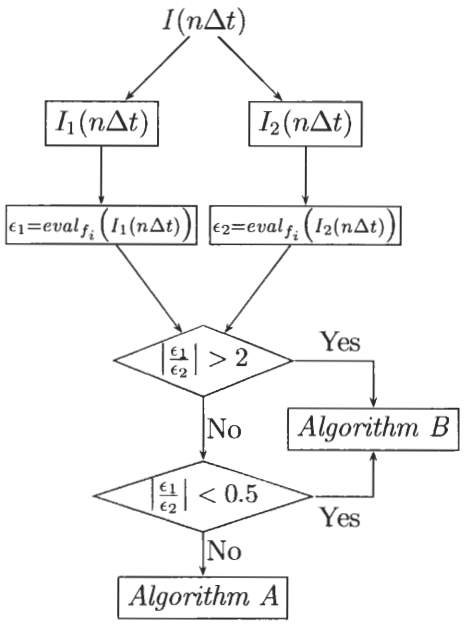


Figure 7.14: The algorithm selection rule.

7.2 Multi-Dimensional Reconstruction

For an array of detectors, the simplest approach is to apply the reconstruction scheme designed for the one-dimensional problem to each interferogram in the array independently. But this method does not take advantage of having multiple measurements, with each detector experiencing the same sampling offsets. Another useful observation is that many of the interferograms are highly correlated because of the similarities in atmospheric features. Therefore, a better solution would be a parallel correction algorithm that evaluates a population based on a member's fitness at all detectors. The fittest individuals are selected to be those who can perform the best over all array elements. Therefore, a more sophisticated ranking system is required in order to identify the best solutions in the population. Another issue is concerned with the choice of the algorithm to be used. If each array element is treated individually, then

the method described in Section 7.1.4 is sufficient, i.e., either one of the two techniques can be applied; the choice is made based on the decision measure described earlier. If all interferograms are corrected simultaneously, then *Algorithm A* is the better choice since it estimates the offset locations that are universal to all array elements. *Algorithm B* would not be a suitable method because it primarily estimates the values of the interferogram, thus, a single estimate update cannot be applied to the entire array. Now we discuss how to modify *Algorithm A* for reconstructing multiple interferograms in parallel.

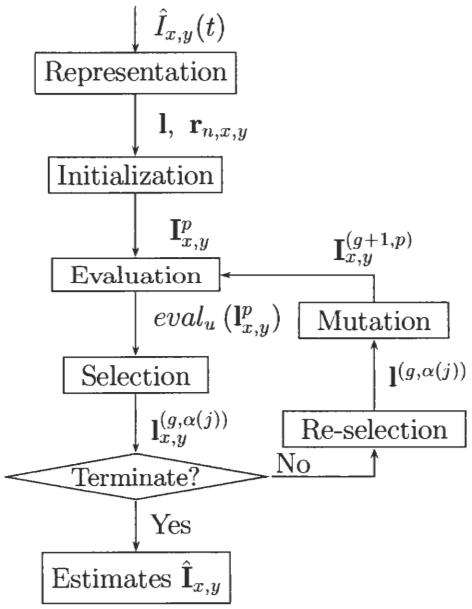


Figure 7.15: The algorithm outline for multi-dimensional reconstructions.

The same procedures as introduced in *Algorithm A* are applied to the multi-dimensional case with the exception of an extra selection-mutation transitional step. The outputs at each stage are outlined in Figure 7.15. Note that subscripts x, y denote the coordinates of the array elements, and x, $y = -64, \ldots, -1, 1, \ldots, 64$. A "re-selection" step is necessary for reconstructing an array of pixels. To see the reason behind this step, we first review the case for a single interferogram. We

initialize P individuals that are represented by sampling location vectors $\mathbf{l}^p$, $p = 0, \ldots, P-1$. Then we evaluate P interferograms based on the information provided by $\mathbf{l}^p$. After the evaluation step, we select the J fittest interferograms in the population, from which, we could obtain the corresponding J sampling location vectors $\mathbf{l}^{\alpha(j)}$ for $j = 0, \ldots, J-1$. In the multi-dimensional case, we initialize the same set of individuals $\mathbf{l}^p$ as for the one-dimensional case. However, in the evaluation stage, $P \times 128^2$ interferograms or P interferogram data cubes are generated, from which, the $J \times 128^2$ fittest interferograms are selected. Similarly, this will give us $J \times 128^2$ corresponding sampling location vectors $\mathbf{l}^{\alpha(j)}$. Now we must take an extra selection step to find J sampling location vectors, which will be used to create P individuals for the next generation. Thus, an additional system is required to re-select J members out of all $128 \times 128 \times J$ selections, i.e.:

$$\mathbf{l}_{x,y}^{(g,\alpha(j))} \Rightarrow \mathbf{l}^{(g,\alpha(j))} \; \forall \; x, \; y. \tag{7.73}$$

Note that the $128 \times 128 \times J$ individuals are not all unique solutions, for instance, some of the fittest members for one interferogram may also be the fittest solutions for other interferograms. Then a reasonable assumption is that the fittest individual among all members in the population is the one that has the most frequent occurrences as a top ranking solution based on its evaluation results at all interferograms. For example, in Table 7.4, the values of the ranking variable $\alpha(j)$ is listed for four pixels elements, ranging from pixel number (1,1) to (2,2). The first row tells us that for pixel (1,1), the fittest member is individual number 1; the second fittest member is individual number 3; the third is number 5, and so on. The same interpretation applies to the rest of the pixels as well. Suppose all elements in Table 7.4 are sorted by the number of occurrences as shown in Table 7.5. Then we could say that individual number 7 is the fittest member in the population for all interferograms, and there is a tie for second place between solutions 3 and 5. Since there are 4 individuals having the same rank of 3, we simply pick the first 2 individuals in order to maintain the same

Table 7.4: An example of the selected individuals for x, $y = 2$.

$\alpha_{1,1}(j)$	1	3	5	4	7
$\alpha_{1,2}(j)$	2	3	7	6	5
$\alpha_{2,1}(j)$	5	3	7	8	2
$\alpha_{2,2}(j)$	7	4	6	8	9

Table 7.5: The number of occurrences and ranking for each individual in the population.

Individual	0	1	2	3	4	5	6	7	8	9
Occurrences	0	1	2	3	2	3	2	4	2	1
Rank	5	4	3	2	3	2	3	1	3	4

population size.

Alternatively, we could use a "total-scoring" system to select the required J sampling location vectors. An example of this reselection method is shown in Table 7.6. The evaluation score of a particular individual $\mathbf{l}^p$ at a given pixel (x, y) is written as $\epsilon^p_{x,y}$. For all individuals in the population, their evaluation scores at all pixels are listed in the table. These scores are summed up to produce the last row ϵ^p for $p = 0, \ldots, P-1$. An element in the last row is an indication of how well does an individual perform over all pixels. In the next step, we rank ϵ^p, and select J individuals with the highest scores, i.e., $\epsilon^{\alpha(j)}$ for $j = 0, \ldots, J-1$. Finally, we find their corresponding sampling location vectors $\mathbf{l}^{\alpha(j)}$.

7.3 Implementation and Results

This section is concerned with several implementation issues related to each algorithm, which include the algorithm complexity, parameter selections, its performance under noisy conditions, and the effect of different sampling rates. The test results obtained from MATLAB simulations are also presented and analyzed.

Table 7.6: A reselection example for the multi-dimensional reconstruction problem.

Pixel (x, y)	Individual $\mathbf{I}^p$			
	$\mathbf{I}^0$	$\mathbf{I}^1$	...	$\mathbf{I}^{P-1}$
$(1,1)$	$\epsilon^0_{1,1}$	$\epsilon^1_{1,1}$	...	$\epsilon^{P-1}_{1,1}$
$(1,2)$	$\epsilon^0_{1,2}$	$\epsilon^1_{1,2}$	...	$\epsilon^{P-1}_{1,2}$
$(2,1)$	$\epsilon^0_{2,1}$	$\epsilon^1_{2,1}$	...	$\epsilon^{P-1}_{2,1}$
$(2,2)$	$\epsilon^0_{2,2}$	$\epsilon^1_{2,2}$	...	$\epsilon^{P-1}_{2,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(128,128)$	$\epsilon^0_{128,128}$	$\epsilon^1_{128,128}$	...	$\epsilon^{P-1}_{128,128}$
Sum	ϵ^0	ϵ^1	...	ϵ^{P-1}

7.3.1 Algorithm Complexity

The arithmetic complexity of *Algorithm A* is shown in Table 7.7. The explanation of each symbol can be found in Table 7.8. At each stage of the algorithm, the approximate number of multiplications and additions are listed for most functions. Certain operations such as sorting and ranking, do not have a countable number of "mults" and "adds", in which case, we adopt the o-notation defined in [18, 65] to express the asymptotically bound of the running time for these special functions. Note that for the evaluation, selection, and variation stages, the number of each particular operation is repeated for $P \times G$ times to account for all individuals in the population through all generations. Similarly, the complexity of *Algorithm B* is shown in Table 7.9, which has a similar number of required operation. For the multi-dimensional algorithm, the approximate complexity is the number of pixels to be reconstructed times the required operations listed for *Algorithm A*.

In Table 7.8, a typical set of values for the variables in the complexity calculation is shown. The numbers of multiplications and additions are computed for the three routines. The required mults and adds operations for *Algorithm A* is approximately equal to 10^7, whereas for *Algorithm B*, the requirement drops slightly to 7×10^6.

Table 7.7: A complexity analysis of *Algorithm A*.

Stage	Function	Number of Multiplications	Number of Additions	Running Time for Others
Representation	Interpolator	$NM \log_2(NM)$	$NM \log_2(NM)$	–
Initialization ($P\times$)	$\mathbf{I}^p$	$N \log_2(N)$	$N \log_2(N)$	–
Evaluation ($P \times G\times$)	FFT	$N \log_2(N)$	$N \log_2(N)$	–
	$f_1'(\mathbf{I}^p)$	N	$N-1$	$o(N \ln N)$
	$f_2'(\mathbf{I}^p)$	–	$W-1$	–
	$f_3'(\mathbf{I}^p)$	W	$(W-1)+W$	–
	$\sqrt{\sum_{p=0}^{P-1} \lvert f_i'(\mathbf{I}^p)\rvert^2}$	$3N$	$3(N-1)$	–
	$f_i(\mathbf{I}^p)$	$3N$	–	–
	$q'(\mathbf{I}^p)$	–	$N-3$	–
	$\sqrt{\sum_{p=0}^{P-1} \lvert q'(\mathbf{I}^p)\rvert^2}$	$N/2-1$	$N/2-2$	–
	$q'(\mathbf{I}^p)$	$N/2-1$	–	–
Selection ($P \times G\times$)	Sort, order, etc.	–	–	$o(P \log_2(P))$
Variation ($P \times G\times$)	Gaussian mutation	–	$N-1$	–
Stopping & Post-process	$\hat{\mathbf{I}}_m$	–	$N(J-1)$	–
	$\hat{\mathbf{E}}_m$	$N \log_2(N)$	$N \log_2(N)$	–
	$\hat{\mathbf{E}}$	W	$(W-1)+W$	–
	$\hat{\mathbf{I}}$	$N \log_2(N)$	$N \log_2(N)$	–

Table 7.8: A sample complexity calculation for all algorithms.

Parameter	Symbol	Algorithm A	Algorithm B	M-D Algorithm
Signal length	N	2048	2048	2048
Oversample rate	M	128	128	128
OOB length	W	185	185	185
Population size	P	10	10	10
Generation number	G	15	15	15
Selection size	J	5	5	5
Dimension size	D	–	–	128^2
Number of Multiplications		$\sim 10^7$	$\sim 7 \times 10^6$	$\sim 16384 \times 10^7$
Number of Additions		$\sim 10^7$	$\sim 7 \times 10^6$	$\sim 16384 \times 10^7$

7.3.2 Algorithm Parameter Tests

One of the strengths of an evolutionary algorithm is its flexibility in terms of selecting the appropriate parameter values; there are no rules or restrictions on the parameter choices. This unique feature offers us endless possibilities. However, at the same time, it makes the problem more difficult to solve because these parameters need to be determined empirically. This section describes the determination processes of several important variables introduced earlier.

In the initialization stage (Section 7.1.2), the choices of two parameters are particularly interesting, namely, the population size P and the standard deviation σ of the population. Each one of these two parameters is tested at four different values; the averaged results over five runs are shown in Table 7.10. The same nonuniform interferogram $I^1(n\Delta t)$ from Table 7.2 is used for these tests; without any correction being applied to it, $I^1(n\Delta t)$ has the SNR and error norm of 47.26dB and 1.04e5, respectively. Note that each parameter is tested at various values when others are held constant; these default settings are shown as shaded cells in Table 7.10.

As discussed in Section 7.1.2.2, the population size does not influence the results significantly. Tests are performed on groups that are composed of 10, 40, 70, and 100 individuals, the qualities of the reconstructed signals seem to be comparable; since the

Table 7.9: A complexity analysis of *Algorithm B.*

Stage	Function	Number of Multiplications	Number of Additions	Running Time for Others
Representation	Interpolator	$NM \log_2(NM)$	$NM \log_2(NM)$	–
	$\mathbf{s}_n$	$N/2+1$	$N/2+1$	–
Initialization ($P\times$)	$\mathbf{I}^p$	$N/2 \log_2(N/2)$	$N/2 \log_2(N/2)$	–
Evaluation ($P \times G\times$)	FFT	$N/2 \log_2(N/2)$	$N/2 \log_2(N/2)$	–
	$f_1'(\mathbf{I}^p)$	N	$N-1$	$o(N \ln N)$
	$f_2'(\mathbf{I}^p)$	–	$W-1$	–
	$f_3'(\mathbf{I}^p)$	W	$(W-1)+W$	–
	$\sqrt{\sum_{p=0}^{P-1} \lvert f_i'(\mathbf{I}^p)\rvert^2}$	$3N$	$3(N-1)$	–
	$f_i(\mathbf{I}^p)$	$3N$	–	–
Selection ($P \times G\times$)	Sort, order, etc.	–	–	$o(P \log_2(P))$
Variation ($P \times G\times$)	Gaussian mutation	–	$N-1$	–
Stopping & Post-process	$\hat{\mathbf{I}}_m$	–	$N(J-1)$	–
	$\hat{\mathbf{E}}_m$	$N \log_2(N)$	$N \log_2(N)$	–
	$\hat{\mathbf{E}}$	W	$(W-1)+W$	–
	$\hat{\mathbf{I}}$	$N \log_2(N)$	$N \log_2(N)$	–

smaller population size requires less computation, the most preferable choice would be somewhere between 10 and 40. The standard deviation of the initial population is tested at four different values as well; the best result is obtained when σ is set to 4, which indicates that the population should not be initialized overly diversified. However, at the same time, it should not be too concentrated around the center value either. The choices of parameters for the evaluation and selection stages are discussed in Sections 7.1.2.3 and 7.1.2.4. In the variation stage, we examine the results from using a constant Gaussian mutation at three fixed magnitudes, and compare them to the output obtained from applying a dynamic Gaussian mutation. The dynamic approach clearly out-performed any fixed mutation technique. Lastly, the algorithm is tested for different numbers of generations. As indicated in Table 7.10, the quality of the output improves as the generation number increases, however, at a decaying

Table 7.10: Parameter selection tests. (The results are average values over 5 trial runs .)

Initialization	Population size P	**10**	40	70	100
	SNR_E	**62.82**	66.65	66.26	64.36
	$\|\mathbf{e}\|$	**4.78e4**	3.96e4	4.04e4	4.44e4
	Std. Dev. σ	2	**4**	8	16
	SNR_E	54.00	**62.82**	57.54	57.35
	$\|\mathbf{e}\|$	7.45e4	**4.78e4**	6.24e4	6.30e4
Evaluation & Selection	Weighting factors w_i	See Table 7.1			
	Individual function $f_i'(\cdot)$	See Table 7.2			
	Selection rule $\alpha(i)$	See Table 7.3			
Variation	Mutation $\mathbf{d}^p$	$\sigma = 2$	$\sigma = 4$	$\sigma = 8$	**Dynamic $\sigma(g)$**
	SNR_E	49.77	53.05	48.47	**62.82**
	$\|\mathbf{e}\|$	9.21e4	7.82e4	9.82e4	**4.78e4**
Stopping	Generation G	5	**15**	25	50
	SNR_E	48.48	**62.82**	63.18	74.93
	$\|\mathbf{e}\|$	9.82e4	**4.78e4**	4.71e4	2.62e4

improvement rate. How many generations are sufficient and will the correct signal ever be recovered? The answer to the first question is entirely depended on the requirement for a particular application. Since a complete solution is computed at each generation, the iteration can be terminated after any number of generations, which depends on the desired quality of the final solution [48]. The second question is concerned with the convergence of the algorithm.

The convergence analysis of an evolutionary algorithm can be categorized into two classes: the first type focuses on the convergence property of a single fittest solution in the population, thus, the convergence of this type is named as *point-wise*. The second type attempts to examine the convergence of an entire population, and is described as *population-wise* [32, 31]. Both types of analyses emphasize the operation of the variation/mutation in order to prove a guaranteed global convergence. In [56], authors concluded that with an infinite population size, the resulting solutions are distributed in the vicinity of the optimum. Furthermore, the mean of the resulting

population converges to the optimal point, which can be seen as a population-wise convergence. When the population size is finite, it has been shown that the fittest solution converges to the optimum over an infinite number of generations using the Borel-Cantelli Lemma [4, 5]. In this case, the convergence is point-wise since only a single fittest solution is considered. Recall in Equation 5.7, the optimization problem has the form of

$$f^* := f(\mathbf{I}^*) := \min\{f(\mathbf{I}) | \mathbf{I} \in \mathcal{F} \subseteq \mathbb{R}^n\}, \tag{7.74}$$

where f^* is the optimal solution, and $f(\mathbf{I})$ is defined as

$$f(\mathbf{I}) = \sum_{i=1}^{k} w_i f_i(\mathbf{I}).$$

The Borel-Cantelli Lemma [4, 5, 32, 31] is stated as follows: Let $p_g := P\{I_g \in L_{f^*+\epsilon}\}$ be the probability to reach the level set $L_{f^*+\epsilon}$ for an arbitrary $\epsilon > 0$ at generation g. If

$$\sum_{g=1}^{\infty} p_g = \infty$$

then $f(\mathbf{I}_g) - f^* \to 0$ as $g \to \infty$ for any initial point $\mathbf{I}_0 \in \mathcal{F}$. This can be expressed in the equivalent form of

$$P\{\lim_{g\to\infty}(f(\mathbf{I}_g)) - f^* = 0\} = 1. \tag{7.77}$$

Equation 7.77 states that when the number of generation approaches infinity, the probability of a guaranteed convergence equals one.

7.3.3 Reconstruction Test Results

In this section, the results from several reconstruction tests are presented. The first experiment is conducted to test the algorithm's performance under noisy conditions. The second test demonstrates the effect of the interferogram sampling rate on the outputs. Finally, the reconstruction results obtained from a 2×2-pixel array are shown.

Table 7.11: Reconstruction results (without references) from interferograms with additive detector noise.

Noise Condition	Signal Properties Without Corrections			Reconstructed Signals	
		SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
Without Noise	I^1	47.26	1.04e5	58.94	5.87e4
	I^2	47.15	1.73e5	62.20	5.01e4
	I^3	47.40	1.04e5	68.98	3.70e4
With Detector Noise	I_n^1	47.23	1.05e5	69.88	5.55e4
	I_n^2	47.14	1.74e5	60.98	5.28e4
	I_n^3	47.35	1.05e5	61.45	3.86e4

In the first test, three nonuniformly sampled interferograms are created by applying three separate sets of clock signals to a single interferogram. These signals are identical to the test interferograms that have been analyzed in Table 7.2. Note that only the detector noise has been added to the signals; the detector noise has the characteristic of a white (normally distributed) noise with a amplitude of $\sqrt{2N}$ [74], where N is the number of samples. In Table 7.11, the default error properties are shown for signals under noisy and noise-free conditions. The errors of the reconstructed signals are also presented in the table. The results from recovered signals reveal no major decline in the quality of the outputs when the detector noise is present. The

Table 7.12: Reconstruction results (without references) from interferograms sampled at different rates.

Signal		Near Nyquist Rate		4× Oversampled		8× Oversampled	
		SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
I^1	Original	47.26	1.04e5	72.15	3.01e4	103.94	6.13e3
	Corrected	58.94	5.87e4	88.41	1.48e4	112.06	4.09e3
I^2	Original	47.15	1.73e5	59.94	5.54e4	74.21	2.71e4
	Corrected	62.20	5.01e4	77.80	2.54e4	92.21	1.10e4
I^3	Original	47.40	1.04e5	58.56	5.92e4	71.50	3.11e4
	Corrected	68.98	3.70e4	76.52	2.48e4	84.90	1.59e4

next experiment investigates the reconstruction results obtained at different sampling rates. The nonuniform interferograms are sampled at near Nyquist rate, 4×, and 8× oversampling rates. The outputs from these tests are summarized in Table 7.12. The errors are shown for both the original (reconstructed without any corrections) and the corrected signals for the purpose of comparison. A conclusion can be drawn from these data outputs: as the signal sampling rate increases, the qualities of both the original and the corrected signals improve as well.

Table 7.13: Reconstruction results (without references) from a four-pixel array.

Pixel	Signal Properties		Reconstructed Signals	
(x, y)	SNR_E	$\|\mathbf{e}\|$	SNR_E	$\|\mathbf{e}\|$
(-1,-1)	38.0932	1.9162e5	69.6741	3.9504e4
(-1, 1)	38.0990	1.9179e5	69.8955	3.9071e4
(1, 1)	38.0968	1.9174e5	69.1203	4.0616e4
(1, -1)	38.0921	1.9159e5	69.1548	4.0545e4

Finally, we examine the results from a nonuniformly sampled four-pixel array. Note that the previous tests are performed on a single interferogram that has sampled nonuniformly at three distinct sets of clock signals. Here, the multi-dimensional test is conducted on a four-pixel array consisting of four nonuniform interferograms, which are created using only one set of clock signal. The test results are shown in Table 7.13 for pixels $(1, 1)$ to $(-1, -1)$. Because the multi-dimensional reconstruction algorithm corrects all interferograms in parallel, as a result, the errors of recovered signals have similar magnitudes.

Alternatively, for an array of pixels, we could use the single-pixel reconstruction algorithm on each interferogram, and repeat for the total number of pixels in the array. By doing so, the interferograms are corrected independently. However, the multi-dimensional reconstruction algorithm has several advantages over the single-pixel reconstruction method. First of all, if any pixel detector is experiencing higher

noise, the single-pixel algorithm may not be able to recover the desired signal. On the other hand, the multi-dimensional method works on all pixels in parallel since all pixels in the array experience the same sampling offset error. This will give us a chance to recover the desired interferogram based on the information provided by other pixels. Secondly, the multi-dimensional method should produce more stable results because the performance of each estimated solution is evaluated among all pixels. This is equivalent to performing the evaluation test for each estimate 128^2 times. The fittest individual should have the highest total evaluation score, and will be recognized by most pixels. However, from the simulation results presented earlier, there are no significant improvements when the multi-dimensional algorithm is applied. This is because only a small number of pixels are tested due to limited computation resource, in addition, the amount of the detector noise used in the simulation was the same for all pixels. If real data cube measurements were used in the test, the outcome may be different. However, this may require further analyses and comparison on the results obtained from real interferogram data cube measurements.

CHAPTER 8

CONCLUSIONS

A common problem that exists in FTS is concerned with how to compensate for sampling errors when an interferogram is sampled at nonuniform instants in the path-difference domain. These errors, resulting from various mechanical irregularities, are generally associated with a continuous scanning system, which samples the interferogram at either equal space or equal time intervals. In both systems, the accuracy of the reconstructed signal can be significantly compromised if no error correction is performed. For an IFTS, the off-axis effects also affect the accuracy of the sampling locations. Furthermore, if the nonuniform sampling locations are unknown, which is the case when a laser reference is not present, the reconstruction algorithm must be able to correct the sampling errors "blindly."

The current technique for solving this problem in the FTS industry involves a low-pass interpolation/resampling process, which has been applied only to a single detector problem, and it does not offer a solution when sampling locations are unknown. Other reconstruction methods for solving generic nonuniform sampling problems include a variety of iterative algorithms, which attempt to recover the desired signal via a series of updates on the sinc interpolation kernels and data outputs. Although these algorithms are capable of producing excellent results, they are extremely time-consuming and impractical for processing a large quantity of data as for the case of an IFTS. Both the iterative and non-iterative strategies rely on the knowledge of irregular sample positions, which are unavailable if the laser reference is not employed. Thus, a reconstruction algorithm that is capable of correcting the irregular sampling errors without the knowledge of sample positions must be developed. This is possible

only because additional information on the data itself allows us to impose a set of constraints to the solution.

8.1 Summary of Thesis Contributions

In this thesis, the nonuniform reconstructions have been categorized into two classes, namely, the with-reference and without-reference cases. Each case is discussed individually for both the single-pixel and array problems. The unique contributions of this thesis can be summarized as follows:

- When sampling positions are available, two alternatives are presented. The first method recovers the data using a truncated "sinc" interpolation, which is proven to be a better alternative than the existing techniques because it allows the option of choosing the desired data quality by setting a particular truncation window length.

- A second method is designed for solving the nonuniform reconstruction problem by employing a linear interpolation based on the interferogram's symmetry property. This method is very simple to implement and computational efficient, and offers exceptionally accurate results when additive noises are kept low.

- In the case where sampling locations are unknown, which has no previous techniques for solving such a problem, an optimization problem with multiple objective and constraint functions is designed based on the spatial and spectral characteristics of the data measurement.

- The optimization problem is solved using an evolutionary approach, in which potential solutions are competing to be the fittest individual in a simulated natural environment. Two evolutionary algorithms are developed and compared to obtain the most desirable solution. The choice of which algorithm to use is made based on the characteristic of each particular interferogram. One is based

on the estimation of sampling offsets while the other attempts to recover the actual values of all correct samples.

- The extensions of these algorithms are made to solve the multi-dimensional array problem. These algorithms show promising results, when compared with the original erroneous interferograms, the errors of recovered signals have been reduced by at least 50% on average.

8.2 Further Research Topics

In this section, we present a list of potential research topics that can be further pursued. The list includes further algorithm developments and additional analyses on the reconstruction results. Further algorithm developments include:

- A real-time reconstruction algorithm that is suitable for on-board implementation. This algorithm may not produce the optimal solution, however, it will be able to operate under limited resources available in a real-time environment.

- A nonuniform reconstruction algorithm for recovering the super-channel data. The super-channel data is produced by integrating the interferogram spectrum over different bands. Therefore, the reconstruction algorithm must be able to recover the uniformly sampled spectrum for the selected frequency bands without having the entire wide-band spectrum available.

- A multi-dimensional algorithm that incorporates the off-axis effect correction routine into the nonuniform sampling reconstruction algorithm. For example, the off-axis correction can be embedded into any one of the interpolative reconstruction methods.

Further analyses on the reconstruction results include:

- The noise factor being considered in this thesis was simulated uncorrelated detector noise. However, for real interferogram measurements, other form of

noise will present. Therefore, further analyses on the reconstruction results are required under different noise conditions.

- Similarly, once the real measurements become available, the pros and cons of using the multi-dimensional algorithm will need to be further analyzed. Comparisons will be made between the routine that corrects interferograms in parallel and the algorithm that treats each pixel independently.

- The benefits of the nonuniform reconstruction algorithm on the final data product will need to be addressed. For example, after the reconstruction step, how much improvements do we see in the quality of the absorption data, moisture, wind, and other atmospheric profiles.

BIBLIOGRAPHY

[1] ABRAMS, M. C. *et al.*, "Remote sensing of the Earth's atmosphere from space with high-resolution Fourier-transform spectroscopy: development and methodology of data processing for the Atmospheric Trace Molecule Spectroscopy experiment," *Appl. Opt.*, vol. 35, pp. 2774–2786, June 1996.

[2] BÄCK, T., FOGEL, D. B., and MICHALEWICZ, Z., eds., *Evolutionary Computation 1: Basic Algorithms and Operators*. Bristol: Institute of Physics Publishing, 2000.

[3] BÄCK, T., FOGEL, D. B., and MICHALEWICZ, Z., eds., *Evolutionary Computation 2: Advanced Algorithms and Operators*. Bristol: Institute of Physics Publishing, 2000.

[4] BÄCK, T., RUDOLPH, T., and SCHWEFEL, H.-P., "Evolutionary programming and evolution strategies: Similarities and differences," in *Proc. of Second Annual Conf. on Evolutionary Programming*, pp. 11–22, 1993.

[5] BÄCK, T. and SCHWEFEL, H.-P., "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation*, vol. 1, pp. 1–23, 1993.

[6] BAILEY, R. B. *et al.*, "256×256 hybrid HgCdTe infrared focal plane arrays," *IEEE Trans. Electron Devices*, vol. 38, pp. 1104–1109, May 1991.

[7] BAJAJ, J., "HgCdTe infrared detectors and focal plane arrays," in *Proc. IEEE, Conference on Optoelectronic and Microelectronic Materials Devices*, pp. 23–31, 1999.

[8] BARR, S., "Special applications of spectral sensing," tech. rep., U. S. Army Topographic Engineering Center, Alexandria, Virginia, 1995.

[9] BEER, R., *Remote Sensing by Fourier Transform Spectrometry*, vol. 120 of *Chemical Analysis*. New York: John Wiley & Sons, 1992.

[10] BEST, F. A. *et al.*, "Calibration of the Geostationary Fourier Transform Spectrometer (GIFTS)," *Proceedings of SPIE*, vol. 4151, pp. 21–31, 2001.

[11] BEUTLER, F. J., "Error-free recovery of signals from irregularly spaced samples," *SIAM Rev.*, vol. 8, pp. 328–335, July 1966.

[12] BRAULT, J. W., "New approach to high-precision Fourier transform spectrometer design," *Appl. Opt.*, vol. 35, pp. 2891–2896, June 1986.

[13] BRIGGS, W. L., HENSON, V. E., and MCCORMICK, S. F., *A Multigrid Tutorial*. Philadelphia, PA: SIAM, 2nd ed., 2000.

[14] BROWN, J. M., *Molecular Spectroscopy*, vol. 55 of *Oxford Chemistry Primers*. Oxford: Oxford University Press, 1998.

[15] CAMPBELL, K., "Remote sensing." Submitted to the *Encyclopedia of Environmetrics*, 2002.

[16] CLARK, R. J. H. and HESTER, R. E., eds., *Spectroscopy in Environmental Science*, vol. 24 of *Advances in Spectroscopy*. Chichester, England: John Wiley & Sons, 1995.

[17] COHEN, D. L., "Performance degradation of a Michelson interferometer due to random sampling errors," *Appl. Opt.*, vol. 38, pp. 139–151, Jan. 1999.

[18] CORMEN, T. H., LEISERSON, C. E., and RIVEST, R. L., *Introduction to Algorithms*. New York: McGraw-Hill Book Co., 1990.

[19] COX, P. A., *Introduction to Quantum Theory and Atomic Structure*, vol. 37 of *Oxford Chemistry Primers*. Oxford: Oxford University Press, 1996.

[20] DAVIS, S. P., ABRAMS, M. C., and BRAULT, J. W., *Fourier Transform Spectrometry*. San Diego, California: Academic Press, 2001.

[21] DUCKETT, S. and GILBERT, B., *Foundations of Spectroscopy*, vol. 78 of *Oxford Chemistry Primers*. Oxford: Oxford University Press, 2000.

[22] DUIJNDAM, A. J. W., SCHONEWILLE, M. A., and HINDRIKS, C. O. H. H., "Reconstruction of band-limited signals, irregularly sampled along one spatial direction," *Geophysics*, vol. 64, pp. 524–538, Mar. 1999.

[23] FEICHTINGER, H. G. and CENKER, C., "Reconstruction algorithms for discrete non-uniform sampled bandlimited signals," in *15. ÖAGM Conf., ÖCG*, vol. 56, pp. 51–61, May 1991.

[24] FEICHTINGER, H. G. and GRÖCHENIG, K., "Theory and practice of irregular sampling," in *Wavelets: Mathematics and Applications* (BENEDETTO, J. and FRAZIER, M., eds.), CRC Press, 1993.

[25] FEICHTINGER, H. G., GRÖCHENIG, K., and STROHMER, T., "Efficient numerical methods in non-uniform sampling theory," *Numerische Mathematik*, vol. 69, pp. 423–440, 1995.

[26] FOGEL, D. B., *Evolutionary Computation: Principles and Practice for Signal Processing*, vol. TT43 of *Tutorial Texts in Optical Engineering*. Wellingham, WA: SPIE Press, 2000.

[27] FOGEL, D. B., "The advantages of evolutionary computation," in *Bio-Computing and Emergent Computation 1997* (LUNDH, D., OLSSON, B., and NARAYANAN, A., eds.), pp. 1–11, Singapore: World Scientific Press, 1997.

[28] FOWLES, G. R., *Introduction to Modern Optics.* Mineola, New York: Dover Publications, 2nd ed., 1989.

[29] GOLUB, G. H. and VAN LOAN, C. F., *Matrix Computations.* Baltimore: The Johns Hophins University Press, 3rd ed., 1996.

[30] GOODY, R. M. and YUNG, Y. L., *Atmospheric Radiation: Theoretical Basis.* Oxford: Oxford University Press, 1989.

[31] HART, W. E., "A convergence analysis of unconstrained and bound constrainted evolutionary pattern search," *Evolutionary Computation*, vol. 9, no. 1, pp. 1–22, 2001.

[32] HART, W. E., "A stationary point convergence theory for evolutionary algorithms," in *Foundatations of Genetic Algorithms 4* (BELEW, R. K. and VOSE, M. D., eds.), pp. 325–342, San Fransico, CA: Morgan Kaufmann Publishers, Inc., 1997.

[33] HAYES, M. H., *Statistical Digital Signal Processing and Modeling.* New York: John Wiley & Sons, Inc., 1996.

[34] HIGGINS, J. R., *Sampling Theory in Fourier and Signal Analysis Foundations.* New York: Oxford University Press, 1996.

[35] HOBBS, P. V., *Introduction to Atmospheric Chemistry.* Cambridge, United Kingdom: Cambridge University Press, 2000.

[36] JACOB, D. J., *Introduction to Atmospheric Chemistry.* Princeton, New Jersey: Princeton University Press, 1999.

[37] JACOBSEN, A., *Analysing Airborne Optical Remote Sensing Data from a Hyperspectral Scanner and Implications for Environmental Mapping and Monitoring.* PhD thesis, University of Copenhagen, 2000.

[38] KAUPPINEN, J. and PARTANEN, J., *Fourier Transforms in Spectroscopy.* Berlin, Germany: WILEY-VCH, 2001.

[39] KIDDER, S. Q. and VONDER HAAR, T. H., *Satellite Meteorology: An Introduction.* San Diego, California: Academic Press, 1995.

[40] LEVINSON, N., "On non-harmonic Fourier series," *Ann. Math.*, vol. 37, pp. 919–936, Oct. 1936.

[41] LIOU, K.-N., *An Introduction to Atmospheric Radiation.* New York, New York: Academic Press, 1980.

[42] MALLAT, S., *A Wavelet Tour of Signal Processing.* San Diego: Academic Press, 2nd ed., 1999.

[43] MARKS, R. J. II, ed., *Advanced Topics in Shannon Sampling and Interpolation Theory*. New York: Springer-Verlag, 1993.

[44] MARTIN, R. J., *Irregularly Sampled Signals: Theories and Techniques for Analysis*. PhD thesis, University College London, 1998.

[45] MARVASTI, F., "Nonuniform sampling theorems for bandpass signals at or below the Nyquist density," *IEEE Trans. Signal Processing*, vol. 44, pp. 572–576, Mar. 1991.

[46] MARVASTI, F., ANALOUI, M., and GAMSHADZAHI, M., "Recovery of signals from nonuniform samples using iterative methods," *IEEE Trans. Signal Processing*, vol. 39, pp. 872–878, Apr. 1991.

[47] MICHALEWICZ, Z., *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin: Springer-Verlag, 2nd ed., 1994.

[48] MICHALEWICZ, Z. and FOGEL, D. B., *How to Solve It: Modern Heuristics*. Berlin: Springer-Verlag, 2002.

[49] NASA, "Geostationary Fourier Transform Spectrometer (GIFTS) concept study report," tech. rep., NASA Langley Research Center, Hampton, VA, 1999.

[50] NASA/GODDARD SPACE FLIGHT CENTER, "The remote sensing tutorial," tech. rep., http://rst.gsfc.nasa.gov/, 2002.

[51] PALCHETTI, L., BARBIS, A., HARRIES, J. E., and LASTRUCCI, D., "Design and mathematical modelling of the space-borne far-infrared Fourier transform spectrometer for REFIR experiment," *Infrared Physics & Technology*, vol. 40, pp. 367–377, 1999.

[52] PALCHETTI, L. and LASTRUCCI, D., "Spectral noise due to sampling errors in Fourier-transform spectroscopy," *Appl. Opt.*, vol. 40, pp. 3235–3243, July 2001.

[53] PAPOULIS, A., "Error analysis in sampling theory," *Proc. IEEE*, vol. 55, pp. 1677–1681, 1967.

[54] PIERRET, R. F. and NEUDECK, G. W., *Advanced Semiconductor Fundamentals*, vol. 6 of *Modular Series on Solid State Devices*. Boston: Addison-Wesley, 1987.

[55] PRESS, W. H. *et al*, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press, 2nd ed., 1992.

[56] QI, X. and PALMIERI, F., "Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space part i: Basic properties of selection and mutation," *IEEE Trans. on Neural Networks*, vol. 5, pp. 102–119, Jan. 1994.

[57] RATHKE, C. *et al.*, "Realization of improved electronics for a commerical FTIR spectrometer dedicated to airborne radiation measurements," in *Proc. IRS^2 2000 Conference*, (Erfurt, Germany), AMA Service GmbH, 2000.

[58] RAWN, M. D., "A stable nonuniform sampling expansion involving derivatives," *IEEE Trans. Inform. Theory*, vol. 35, no. 6, 1989.

[59] REVERCOMB, H. E. *et al.*, "Radiometric calibration of IR Fourier transform spectrometers: solution to a problem with the High-Resolution Interferometer Sounder," *Appl. Opt.*, vol. 27, pp. 3210–3218, 1988.

[60] RICHARDS, J. A. and JIA, X., *Remote Sensing Digital Image Analysis: An Introduction.* Berlin: Springer-Verlag, 3rd ed., 1999.

[61] RODGERS, C. D., *Inverse Methods for Atmospheric Sounding: Theory and Practice*, vol. 2 of *Series on Atmospheric, Oceanic and Planetary Physics.* Singapore: World Scientific, 2000.

[62] SCHOWENGERDT, R. A., *Remote Sensing, Models and Methods for Image Processing.* San Diego, California: Academic Press, 2nd ed., 1997.

[63] SHEPHERD, G. G., *Spectral Imaging of the Atmosphere*, vol. 82 of *International Geophysics Series.* San Diego, California: Academic Press, 2002.

[64] SHINEBARGER, S., "Time resolved infrared spectroscopy using both step-scan and rapid-scan methods." Analytical spectroscopy, 2000.

[65] SKIENA, S. S., *The Algorithm Design Manual.* New York: Springer-Verlag, 1998.

[66] SMITH, B. C., *Fundamentals of Fourier Transform Infrared Spectroscopy.* Boca Raton, Florida: CRC Press, 1995.

[67] SMITH, G. D. and PALMER, R. A., "Fast time-resolved mid-infrared spectroscopy using an interferometer," in *Handbook of Vibrational Spectroscopy* (CHALMERS, J. M. and GRIFFITHS, P., eds.), vol. 1, pp. 625–640, John Wiley and Sons, Ltd., 2002.

[68] SMITH, W. L. *et al.*, "GIFTS-a system for wind profiling from geostationary satellites," in *Proceedings, the Fifth International Winds Workshop*, (Lorne, Australia), Feb. 2000.

[69] SMITH, W. L. *et al.*, "Geostationary Fourier Transform Spectrometer (GIFTS)-The New Millennium Earth Observing-3 Mission," in *Proc. of IRS 2000: Current Problems in Atmospheric Radiation*, (Hampton, Virginia), A Deepak Publishing, 2001.

[70] STEPHENS, G. L., *Remote Sensing of the Lower Atmosphere: An Introduction.* Oxford: Oxford University Press, 1994.

[71] STROHMER, T., *Efficient Methods for Digital Signal and Image Reconstruction from Nonuniform Samples.* PhD thesis, University of Vienna, 1993.

[72] TOBIN, D. *et al.*, "Simulation of GIFTS data cubes," tech. rep., UW-GIFTS, Madison, WI, 2001.

[73] UNSER, M., "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.

[74] UW-SSEC, "GIFTS simulated compression alogrithm dataset description document," tech. rep., UW-GIFTS 04-005 ver. 2.0, Madison, WI, 2001.

[75] UW-SSEC, "Signal and noise relationships for GIFTS simulated datacubes," tech. rep., UW-GIFTS Note 07-001, ver. 3.0, Madison, WI, 2001.

[76] VANDER, O. W. and CLEVERS, J. G. P. W., "Introduction to optical satellite remote sensing: applications of present and future optical remote sensing satellite sensors," tech. rep., Wageningen Univ. & Research LUW-GIRS-199801, Wageningen, The Netherlands, 1998.

[77] WEIDNER, H. and PEALE, R. E., "Time-resolved Fourier spectroscopy for activated optical materials," *Appl. Opt.*, vol. 35, pp. 2849–2856, June 1996.

[78] WELLS, N., *The Atmosphere and Ocean: A physical introduction.* New York: John Willey & Sons, 2nd ed., 1997.

VITA

Jialin Tian was born in the northern seaport city of Tianjin, China, on December 25, 1974. She received her B. S. and M. S. degrees in Electrical Engineering from North Carolina State University and Purdue University in 1997 and 1999, respectively. Since Fall 1999, she has been a doctoral student at the Georgia Institute of Technology. Her research topic is concerned with the problem of reconstructing nonuniformly sampled interferograms in Imaging Fourier Transform Spectrometry. Over the years, she has been working as a product engineer, a software engineer, and a design engineer at Motorola SPS, Baker Hughes INTEQ, INNOVA Electronics Inc, and Precision Drilling Ltd. While at Georgia Tech, she was a graduate research assistant at the Center for Signal and Image Processing from 2000 to 2003. During her spare time, she enjoys playing the piano works of Brahms, Chopin, and Rachmaninoff. She also loves painting, travel photographing, and webpage designing. Recently, she has developed a passion for baking sinfully delicious treats.