Derivative Compressive Sampling with Application to Inverse Problems and Imaging

by

Mahdi S. Hosseini

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Mahdi S. Hosseini

Abstract

In many practical problems in applied sciences, the features of most interest cannot be observed directly, but have to be inferred from other, observable quantities. In particular, many important data acquisition devices provide an access to the measurement of the partial derivatives of a feature of interest rather than sensing its values in a direct way. In this case, the feature has to be recovered through integration which is known to be an ill-posed problem in the presence of noises.

Moreover, the problem becomes even less trivial to solve when only a portion of a complete set of partial derivatives is available. In this case, the instability of numerical integration is further aggravated by the loss of information which is necessary to perform the reconstruction in a unique way. As formidable as it may seem, however, the above problem does have a solution in the case when the partial derivatives can be sparsely represented in the range of a linear transform. In this case, the derivatives can be recovered from their incomplete measurements using the theory of compressive sampling (aka compressed sensing), followed by reconstruction of the associated feature/object by means of a suitable integration method. It is known, however, that the overall performance of compressive sampling largely depends on the degree of sparsity of the signal representation, on the one hand, and on the degree of incompleteness of data, on the other hand. Moreover, the general rule is the sparser the signal representation is, the fewer measurements are needed to obtain a useful approximation of the true signal. Thus, one of the most important questions to be addressed in such a case would be of how much incomplete the data is allowed to be for the signal reconstruction to remain useful, and what additional constraints/information could be incorporated into the estimation process to improve the quality of reconstruction in the case of extremely under-sampled data. With these questions in mind, the present proposal introduces a way to augment the standard constraints of compressive sampling by additional information related to some natural properties of the signal to be recovered. In particular, in the case when the latter is defined to be the partial derivatives of a multidimensional signal (e.g. image), such additional information can be derived from some standard properties of the gradient operator. Consequently, the resulting scheme of *derivative compressive sampling (DCS)*

is capable of reliably recovering the signals of interest from much fewer data samples as compared to the case of the standard CS. The signal recovery by means of DCS can be used to improve the performance of many important applications which include stereo imaging, interferometry, coherent optical tomography, and many others. In this proposal, we focus mainly on the application of DCS to the problem of phase unwrapping, whose solution is central to all the aforementioned applications. Specifically, it is shown both conceptually and experimentally that the DCS-based phase unwrapping outperforms a number of alternative approaches in terms of estimation accuracy. Finally, the proposal lists a number of research questions which need to be answered in order to attach strong theoretical guarantees to the practical success of DCS.

Acknowledgements

I owe my deepest gratitude to my supervisor Professor Ravi Mazumdar for his endless supports. He has made available his support in a number of ways and this thesis would not have been possible without his encouragements and wise advises.

I would like to thank Dr. Oleg Michailovich for his great advises and help in the project. I have learned a lot from him.

I am heartily thankful to the graduate studies coordinator, Wendy Boles, for her kind and endless help throughout my graduation in the University of Waterloo.

I would like to specially thank my friends and colleagues who have helped and inspired me to enjoy of being in Waterloo. My deepest thanks go to Maaj Movahed, Darya Astaraaki, Fariborz Rahimi, Rashin Salehi, Masoud Barakati, Mehri Mehrjoo, Igor Solovy, Elad Shaked, Sudipto Dolui, Amir Goldan and Roham Farzami.

Finally, and the most important, I would like to express my deep appreciation to my parents for their endless support, patience, understanding and their true love. As always, I praise God for showing his beauty and power to me by having such wonderful mother and father.

to my mother Maryam and father Hossein

&

my dear love, Mahta

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

Reconstruction of signals from random incomplete samples is a task of considerable importance in signal processing, where it belongs to a general class of inverse problems. Factors such as signal corruption due to noises and technical limitations of the acquisition hardware give rise to corrupted or missing data samples, thereby necessitating the developments of methods for signal reconstruction from sub-critically sampled measurements. *Compressive sampling (CS)* is a framework for finding a solution for such problems by exploiting significant redundancy which may exist in digitally sampled signals.

Specifically, compressive sampling relies on two major concepts: sparsity and incoherency, which are properties of the signal of interest and the sensing modality, respectively. Sparsity exemplifies the degree to which the information contained in the signal can be concisely represented in a properly chosen basis Ψ. In other words, the number of non-zero coefficients in basis domain gives a measure of signal compressibility. Incoherency, on the other hand, provides a measure of the degree of similarity between the atoms of sensing (Φ) and representation (Ψ) dictionaries. This notion expands upon the idea of Heisenberg's uncertainty principle [1–11].

A compressed sensing scheme, which achieves a high degree of reconstruction accuracy, requires the signal of interest to be represented as sparsely as possible in the domain of Ψ.

The related representation domain required to be as incoherent as possible with respect to Φ. Unfortunately, finding proper bases Ψ and Φ is not a trivial task, since the definition of Φ is typically constrained by the nature of a data acquisition device at hand. This necessitates the use of specific sensing modality in compressed sensing and since the coherency of the latter should remain low with representation dictionary, finding a proper modal for sparse representation could be a challenge.

For instance, numerous applications are known in which one is provided with the measurements of the gradient of a multidimensional signal, rather than of the signal itself [12–18]. Central to such applications is, therefore, the problem of reconstruction of signals from their partial derivatives, subject to some a priori constraints which could be either probabilistic or deterministic in nature. The problem is further complicated when only a subset of the partial derivatives is provided via measurements. In this case the solution involves two concurrent inverse problems and the uncertainty of the scenario becomes more complicated. At the first glance, one might overcome the problem by demonstrating energy minimization methods in order to approximate the corrupted derivatives. Then applying least squares solution to bring approximated derivatives to the spatial domain to represent the estimated signal.

In many cases, the signals of interest are the functions of bounded variation whose distributional derivative is a set of locally finite measurement, i.e. $f \in BV(\Omega)$ where Ω be a open subset of $\mathbb{R}^n$ for multi-dimentional space. Therefore, the gradient values can be sparsely represented by choosing proper bases for encoding (e.g., DCT) and the question of whether the partial derivative of an image can be recovered through a compressive sampling (CS) scheme from partial observations must be asked. If so, the estimate of the original signal can be recovered from its recovered gradients by solving a convex optimization problem. The principal contribution of the proposed research is to demonstrate that the performance of such a reconstruction algorithm can be improved via introducing a priori knowledge which exist in the signal of recovery, e.g., cross-derivative equality in the case of partial derivative samples.

This proposal introduces a scheme called *derivative compressive sampling (DCS)*, which aims to solve the image recovery inverse problem in two steps. First, image gradients are

recovered from an incomplete data set using a modified version of the compressive sampling algorithm. Then, an estimate of the image is recovered from its gradients through solving a least squares problem. Figure 1.1 elaborates on the proposed DCS methodology, to be fully explained in subsequent sections. Among the advantages of such an approach are:

- *sparse representation* of image gradients in a proper basis, e.g., DCT. The sparser is the representation, the fewer data samples are needed to recover the estimate of the signal via compressive sampling.
- *cross-derivative constraints* is incorporated as a priori information to the signal to be recovered. The observed samples along the mentioned side information is an intriguing combination to improve the performance of proposed decoder.

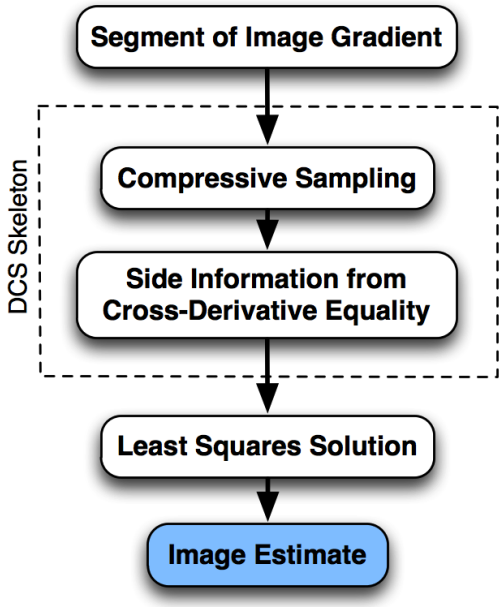


Figure 1.1: A general structure of DCS - based decoder

1.1.1 Concept of Sparse Representation

The image gradient needs to be compressible (sparsely represented) with respect to the properly chosen dictionary Ψ for decomposition. This intriguing challenge of transformation is critical since there is a trade-off between the number of measurements and sparsity level.

The most sparse the partial derivatives can be represented, fewer samples will be needed from the signal of interest. This proposal uses a discrete cosine transform (DCT) representation to present our preliminary results as an illustration. However, finding the optimal representation among all possible choices is a future research task to be accomplished.

1.1.2 Augmenting CS with Side Information

The domain to which the compressive sampling framework is applied is that of partial derivatives, which doubles the number of samples used to represent an image. Since the cross-derivative equality constraint must hold, this side information is added as a-priori knowledge of the recovery algorithm in DCS. Imposition of this constraint allows some of the observed samples to be discarded where they already can be incorporated from the cross-derivative constraints. This constraint prevents retaining redundant information where the half of the partial derivatives can be interpreted from the related equality.

1.2 Applications

Image interferometry is an important application of remote sensing which allows one to perform earth observation on terrain heights, depth sounding of coastal and ocean depths, weather monitoring, monitoring of glaciers in the Arctic and Antarctic regions, or mine detection [19, 20]. Such sensing makes it possible to remotely collect information from dangerous locations or otherwise inaccessible sources.

Aperture synthesis refers to the problem of recovering surfaces of an object using an interferometer. This method combines signals received from individual antennas to provide an image with a resolution declaring the maximum distance between the antennas. This is done by using correlation techniques where the image $i(x, y)$ then is restored by inverse Fourier transform of the measured function of the related coordinates [19].

For instance, interferometric synthetic aperture radar (InSAR) uses two or more antennas to collect the phase difference between antennas and terrain to infer the topography of such areas. In such acquisition systems, phase is measured via modulo-2π so called principle

phase value or wrapped phase, i.e., $\phi = \psi + 2k\phi$ where ϕ is the true phase value and ψ is the measured quantity. ψ is wrapped between $[-\pi, \pi]$ and $k \in \mathbb{Z}$ is an integer number of wavelength [21]. Phase unwrapping refers to the problem of recovering the true phase ϕ from wrapped phase ψ. However, the related task is an ill-posed problem if no further information is been provided. This information comes from the Itoh's condition [22] where it assumes that the absolute value of phase difference between neighbouring pixels is bounded by π. This assumption is no longer been guaranteed if there exist a discontinuity in true phase where it originates from insufficient sampling grids in steep terrain heights (Nyquist-Shannon sampling condition). It can also be originated from noisy measurements where in either cases, phase unwrapping becomes a very difficult problem to solve [21].

Methods which have been introduced to solve the problem of phase unwrapping use a variety of approaches which can be divided into two categories: Path following algorithms [23–25] and minimum L_p-norm solutions [26–33]. Path following algorithms use line integration schemes over wrapped phase image where it relies on the Itoh's condition to hold along the integration path. This condition along the all possible shortest paths, 2×2 neighbouring pixels, is been checked and if it get violated it refers to as inconsistent points so-called residual points. Although many efficient algorithms have been introduced for phase unwrapping in 2-D, most of them nevertheless struggle with the task of interpreting such residual points in their algorithms. These points present ambiguities to the algorithms causing them to fail in successfully unwrapping phase images when the total number of residual points increases in the wrapped images beyond a certain amount.

The wrapped version of the difference of wrapped phase is analogous to the derivative of true phase. Since the signal of interest in DCS is the derivative of the signal, this interesting analogy brings the idea whether the problem of phase unwrapping can be fitted to the derivative compressive sampling scheme or not. As mentioned before, residual points brings inconsistency to the phase unwrapping methods, so these points are in direct relationship with corrupted data in DCS scheme and the remaining points can be considered as observed samples. The next challenge is to represent such derivatives sparse in a domain that the dictionary used for sparse representation is highly incoherent with sampling dictionary which is dot-sampling in the case of phase unwrapping. Subsequently, this proposal provides a new

solution to the phase unwrapping problem, which compares favourably with other approaches based on the presented preliminary results for terrain height recovery.

1.3 Organization of the Proposal

The remainder of the proposal is organized as follows. Chapter 2 provides a literature review on two major subjects: 2-D phase unwrapping and compressive sampling. Phase unwrapping is discussed in Section 2.1. In this section, the problem of remote sensing and phase measurement is explained. Synthetic aperture radar (SAR) interferometry is stated as an example where the concept of the phase principle values, wrapped phase, is explained. Residual points and its practical implication is elucidated in Subsection 2.1.3. Since the residual points contaminate the measured phase values, Subsection 2.1.4 clarifies how to locate such points in wrapped phases. The quality of measured phase is discussed in Subsection 2.1.5 where it is used as a quality guide map in phase unwrapping process. Many unwrapping algorithms have been introduced in the literature. A short survey on the latter is been exemplified in Subsection 2.1.6.

The second portion of the Chapter 2 defines the theory of compressive sampling given in Section 2.2. The origination of the latter is explained and possible generalization of the theory is issued in Subsection 2.2.1. Subsection 2.2.2 provides a formal description of the compressive sampling problem. Stability analysis and recoverability of the CS problem is provided by two types of analysis, one with restricted isometry property (RIP) in Subsection 2.2.3 and the other with distortion of the kernel spaces in Subsection 2.2.4.

Chapter 3 presents the formulation of derivative compressive sampling (DCS) as well as the system architecture. Design of the method is explained in Section 3.1. Since the recovered signal via DCS is the image gradient, Section 3.2 applies the least squares solution to recover the image surface from its gradients. Section 3.3 clarifies the space of the solution and necessary number of samples to uniquely recover the signal of interest via DCS. The problem of redundant measurement in DCS is explained in Subsection 3.3.1 and proposes a solution to eliminate non-necessary samples. The DCS is applied on the problem of phase

unwrapping and preliminary results are demonstrated in Section 3.5. The following chapter is summarized in Section 3.6.

Finally, Chapter 4 outlines the conclusion and the research plan of the following thesis.

Chapter 2

Literature Review

Since the proposed methodology is derived as a symbiosis of two distinct areas of scientific research - phase unwrapping and compressive sensing - Section 2 provides an overview of existing literature on both fields. First, the problem of phase unwrapping is discussed. Phase unwrapping is a challenging task due to the existence of residual points where they impose ambiguities in unwrapping process. Path following, minimum L_p-norm, bayesian and parametric modelling methods are the main approaches for phase unwrapping algorithms. The second part focuses on the problem of compressed sensing where it expands the main ideas and results. This area has been vastly investigated both in theoretical and applicational aspects.

The following chapter is organized as follows. Section 2.1 explains the problem of phase unwrapping. Next, in subsection 2.1.3, the concept of residual points is introduced. These points are known as inconsistent points where subsection 2.1.4 explains the locating methodology in order to incorporate such information to prevent unstability of unwrapping process. The quality of the measured phase at a discrete level is affected by the residual points in wrapped phases. Subsection 2.1.5 gives an analogy to determine such quality by introducing quality maps and subsection 2.1.6 gives a short overview on the existing methodologies for phase unwrapping.

As mentioned in Section 1.2, the problem of phase unwrapping can be solved as a specific instance of compressive sampling scheme. The remainder of this chapter gives an overview

of the compressive sampling (CS) problem in Section 2.2. The basic intuitions of CS is given in subsection 2.2.2. Based on the analytical improvement of the field, two main categories have been separated in compressive sampling defined in Sections 2.2.3 and 2.2.4 by means of restricted isometry property (RIP) and kernel measurements, respectively. Finally Section 2.3 concludes the chapter.

2.1 Phase Unwrapping

Phase interferometry refers to the technique that infers the direction of the arrival of the signal collected by at least two separated antennas through measuring the difference in phase. This technique is used in many applications to estimate the amplitude differences from the ground. Among such applications are synthetic aperture radar (SAR) imaging [19], magnetic resonance imaging (MRI) [34], fringe pattern analysis [35], tomography and spectroscopy [36].

The phase is collected from the argument of complex functions of transferred signals used to record the data in such applications. The difference in phase magnitude is expressed as an integer number of wavelengths with addition of fraction of one wavelength, so the measured phase lyes between $(-\pi, \pi]$ [21].

2.1.1 Synthetic Aperture Radar (SAR) Interferometry

In this subsection, we exemplify the problem of phase unwrapping using the example of synthetic aperture radar (SAR) interferometry as an important application in remote sensing. This imaging technique is of great importance in geophysical monitoring [20]. It provides high resolution images at higher altitudes from the ground regardless of any climate conditions, day or night. It utilizes two or more reflected coherent signals from terrain to elicit relevant phase information through their interference. These signals are sent by an aircraft or satellite platform differ in the sensor flight track, acquisition time or used wavelengths [37]. Figure 2.1 demonstrates the geometry of synthetic aperture radar (SAR) interferometry by means of two antennas S_1 and S_2 separated by the baseline B. The phase difference between the two SAR images is referred as interferometric phase $\Delta\phi$.

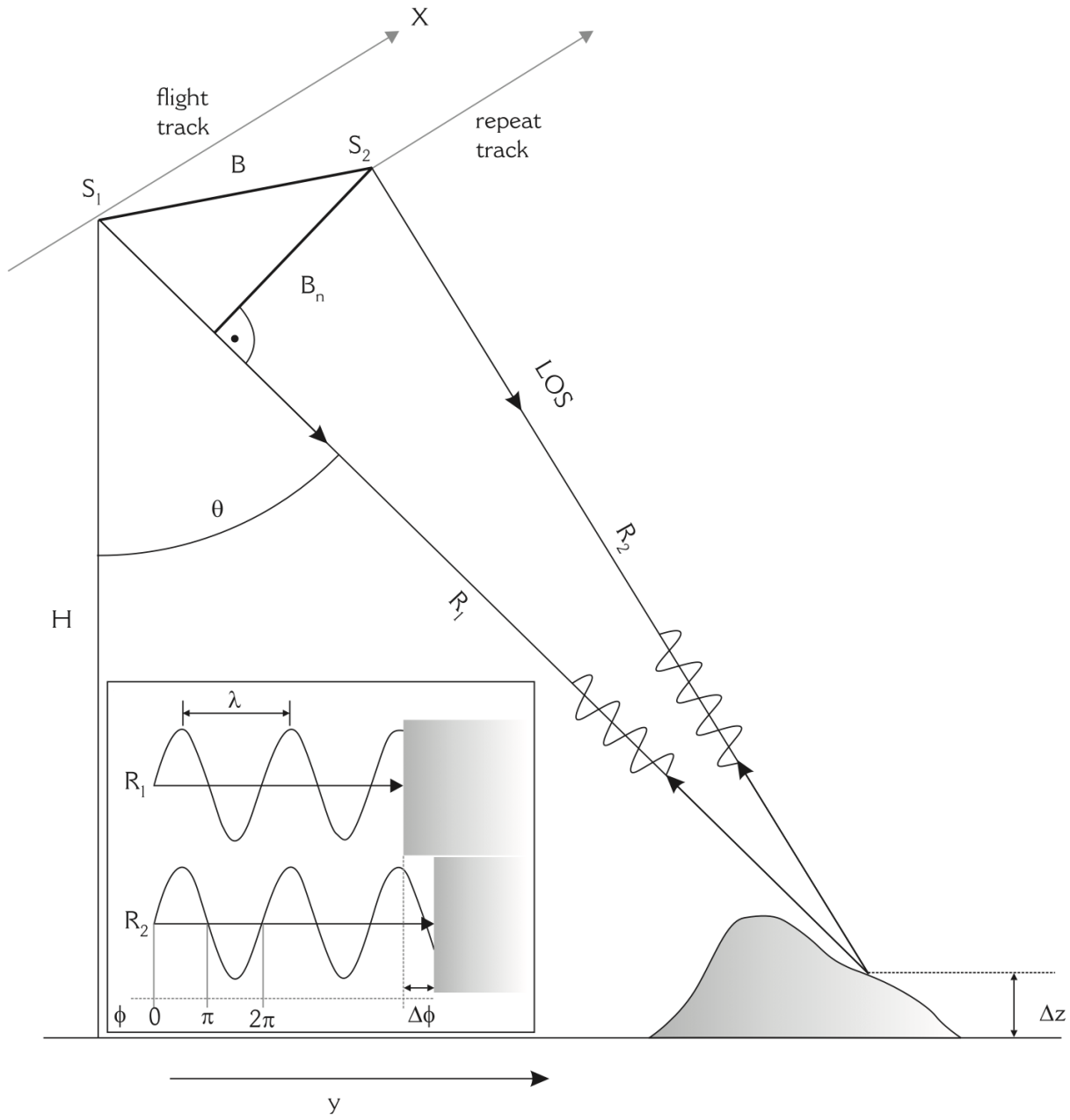


Figure 2.1: Geometry of SAR interferometry by means of two antennas. S_1 and S_2 are the antennas position. B refers to the baseline of the difference of two sources and B_n refers to the perpendicular baseline. LOS is the line of sight.

For instance, in the case of two coherent signals they can be approximated as

$$\begin{aligned} G_1(x,y) &\cong A_1(x,y)\exp\left(\tfrac{j4\pi R_1(x,y)}{\lambda}\right), \\ G_2(x,y) &\cong A_2(x,y)\exp\left(\tfrac{j4\pi R_2(x,y)}{\lambda}\right), \end{aligned} \tag{2.1}$$

where A_i is the complex terrain reflectivity, R_i is the range from satellite i to the point (x,y) and λ is the microwave's wavelength [21]. The reflectivity terms are usually highly correlated and can be considered to be equal, i.e., $A_1(x,y) = A_2(x,y) = A(x,y)$. The interferometric phase is related to the difference in the propagation path of the two transmitted signals, so the interfered of the two images can be expressed by conjugate multiplication,

$$G_1(x,y)G_2^*(x,y) \cong |A(x,y)|^2 \exp\left\{j\frac{4\pi}{\lambda}\left[R_1(x,y) - R_2(x,y)\right]\right\}, \tag{2.2}$$

The phase difference is measured by the argument function operating on the complex quantity in (2.2). This provides wrapped version of the phase, i.e.,

$$\mathcal{W}\left\{\underbrace{(4\pi/\lambda)\left[R_1(x,y) - R_2(x,y)\right]}_{\Delta\phi}\right\} \in (-\pi,\pi]\,. \tag{2.3}$$

The wrapping operator $\mathcal{W}$ here adds a piecewise constant function to the original interferometric phase $\Delta\phi$ resulting in

$$\mathcal{W}\left[\Delta\phi\right] = \Delta\phi + 2\pi k,\ k \in \mathbb{Z}. \tag{2.4}$$

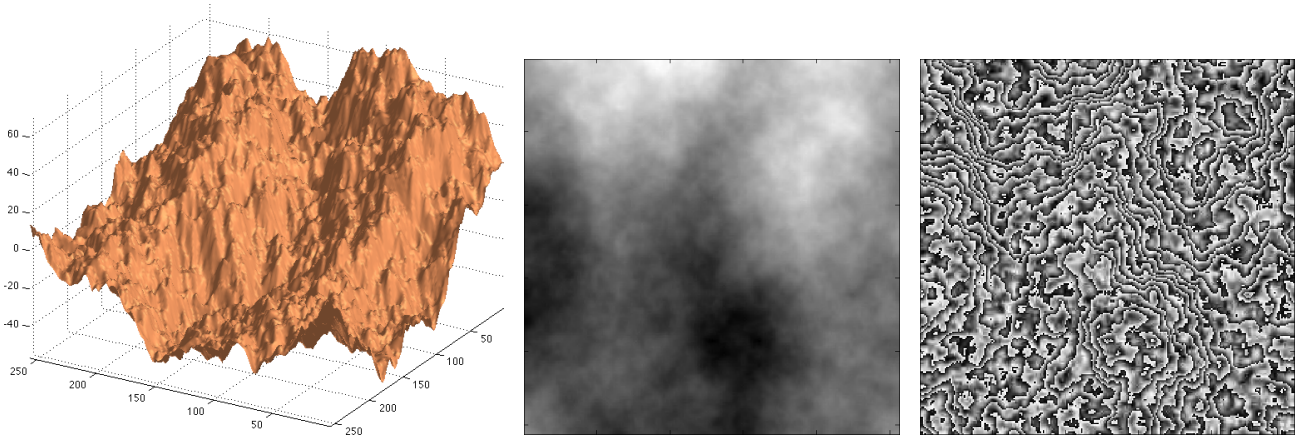
So, in conclusion the wrapped phase lyes between $-\pi \leq \mathcal{W}\left[\phi\right] \leq \pi$. The difference in phase originates from the elevation change in the ground where the relation between these two variation builds the topographic mapping, i.e.,

$$\Delta\phi = \Delta z \frac{4\pi}{\lambda}\frac{B_n}{R\sin\theta} \tag{2.5}$$

where θ here is the direction of arrival signal and Δz is the difference in elevation from the ground [38].

2.1.2 Principal (Wrapped Phases)

As mentioned before, there are different application in phase interferometry that the generated images are wrapped between $\pm\pi$. Let $F(x,y)$ be an arbitrary continuously differentiable



(a) (b) (c)

Figure 2.2: (a) Original phase; (b) the same phase represented as gray-scale image; (c) Its corresponding wrapped phase.

function defined over a closed subset of the real plane $\mathbb{R}^2$. If F happens to be the phase of a complex-valued function, it can only be measured in its *wrapped* form, i.e. *modulo* 2π. Formally, the process of phase wrapping can be represented by its associated operator $\mathcal{W} : \mathbb{R}^2 \to (-\pi, \pi]$. In this notation, the wrapped *principal* phase R is given as $R = \mathcal{W}[F]$. Specifically, the operator $\mathcal{W}$ adds to F a piecewise-constant function $K : \mathbb{R}^2 \to \{2\pi\, k\}_{k\in\mathbb{Z}}$ resulting in $R = \mathcal{W}[F] = F + K$ that obeys [39]:

$$-\pi < \mathcal{W}[F(x,y)] \leq \pi,\ \forall (x,y) \in \mathbb{R}^2. \tag{2.6}$$

In complex notation, the gradients of F and R can be defined by applying the following operator

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j}, \tag{2.7}$$

where $\mathbf{i}$ and $\mathbf{j}$ denote the unit vectors associated with the x- and y-axis, respectively. Consequently the gradient of R is given by

$$\nabla R = \nabla \mathcal{W}[F] = \nabla F + \nabla K. \tag{2.8}$$

Finally, applying the wrapping operator $\mathcal{W}$ one more time to both sides of (2.8) results in

$$\mathcal{W}[\nabla \mathcal{W}[F]] = \mathcal{W}[\nabla R] = \nabla F + \nabla K + K'. \tag{2.9}$$

Due the property of operator $\mathcal{W}$ to produce the values in interval $[-\pi, \pi]$, the term $K + K'$ vanishes as long as [21]

$$-\pi < \nabla F \leq \pi. \tag{2.10}$$

Therefore, as long as the condition (2.10) above holds, the gradient of the original phase F can be unambiguously recovered from the gradient of the corresponding principal phase R according to

$$\nabla F = \mathcal{W}\left[\nabla R\right]. \tag{2.11}$$

Provided (2.11) holds, in 1-D, the original phase can therefore be recovered through integrating the wrapped differences of wrapped phases done by Itoh's method [22]. The notion can be extended for higher dimensions e.g. 2-D signals (images) simply considering a path for integration along the phase gradients. In this case, the original phase is estimated by applying an optimization problem, given a measured wrapped phase R, an estimate $\hat{F}$ of the original phase F could be obtained as a solution to the following minimization problem

$$\hat{F} = \arg\min_{F} \int\int \|\nabla F - \mathcal{W}\{\nabla R\}\|^2 \, dx \, dy, \tag{2.12}$$

which amounts to solving a Poisson equation subject to appropriate boundary conditions. Unfortunately, situations are rare in which the condition (2.10) can be a priori guaranteed. In this case, the estimate of ∇F as $\mathcal{W}[\nabla R]$ is contaminated by, so called, residuals, which cause the solution of (2.11) to be of little practical value.

2.1.3 Residue Theorem and Its Practical Implications

As mentioned, the concept of Itoh's integration method [22] for phase unwrapping can be extended to the case of N dimensions, where by integrating from an initial point $\mathbf{r}_0$ any point $\mathbf{r}$ can be carried out through a line integral [21],

$$F(\mathbf{r}) = \int_{\partial S} \nabla F \cdot \mathbf{dr} + F(\mathbf{r}_0), \tag{2.13}$$

where ∂S is any path in N-dimensional space connecting the points $\mathbf{r}_0$ and $\mathbf{r}$ and ∇F is the phase gradient field. Assuming $\nabla F \in C^1$ is a differentiable vector field defined over

a piecewise C^2 bounded oriented surface $S \in \mathbb{R}^3$ whose boundary ∂S has the inherited orientation, then the line integral in (2.13) is equivalent to the surface integral from Stoke's theorem

$$\int_{\partial S} \nabla F \cdot d\mathbf{r} = \int \int_S (\nabla \times (\nabla F)) \cdot d\mathbf{a}. \tag{2.14}$$

This theorem is also regarded as curved version of Green's theorem in the plane [40]. An special case of this theory states that if the surface S is a closed surface ($\partial S = 0$) then the path integral of gradient field in (2.13) becomes path independent, i.e.,

$$\int \int_S (\nabla \times (\nabla F)) \cdot d\mathbf{a} = 0. \tag{2.15}$$

By substituting $\nabla F = \mathbf{i}\partial F/\partial x + \mathbf{j}\partial F/\partial y$ and $\mathbf{dr} = \mathbf{i}dx + \mathbf{j}dy$ where $\mathbf{i}$ and $\mathbf{j}$ are coordinate directions in two dimension, the integral in (2.13) along any path ∂S becomes

$$I = \int_{\partial S} \left(\frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy \right), \tag{2.16}$$

Recalling from the vector calculus, the curl of the gradient vanishes if the cross-derivatives are equal, i.e.,

$$\nabla \times (\nabla F) = \nabla \times \left\{ \mathbf{i}\frac{\partial F}{\partial x} + \mathbf{j}\frac{\partial F}{\partial y} \right\} = 0\mathbf{i} + 0\mathbf{j} + \left(\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) \right) \mathbf{k} = 0, \tag{2.17}$$

which happens if and only if

$$\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right). \tag{2.18}$$

The cross-derivative constraint in (2.18) holds for any differentiable function (gradient field here) $\nabla F \in C^1$ where it can be violated in isolated points. In this case, the path integral in (2.16) becomes path dependent. In conclusion, the integral in (2.16) will turn to zero through any closed contour (path) if, and only if, the condition in (2.18) holds,

$$\oint \left(\frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy \right) = 0, \tag{2.19}$$

If this condition holds for any closed path, the integration in (2.13) becomes path independent and the wrapped phase can be unwrapped by evaluating the related integral, starting from any initial point. In the case of $2-D$ phase unwrapping, the condition for path independency (2.18) can be violated if the condition in (2.10) does not hold. The failure of such condition in

practical implementation can be motivated by noisy measurements and sub-nyquist sampling problem where it can affect the process of phase unwrapping algorithms [21].

As was mentioned before, the closed loop integration in (2.19) can be non-zero for specific paths if it violates the condition in (2.18). The smallest closed path in two dimensional discrete case can be defined by a 2×2 pixel neighbourhood. This small loop can help to locate phase inconsistency in all over the sampled image, where it is called the "*residues*" by Goldstein et al [23]. The "residue theorem" for two-dimensional phase unwrapping is introduced by the following integration

$$\oint \nabla F \cdot \mathbf{dr} = 2\pi \times (\text{sum of enclosed path residue charges}), \tag{2.20}$$

where it defines that the closed path integral around any residue will equal some integer multiple of 2π radians. The line integral around a balanced residue is equal to zero for any simple path. Thus, two-dimensional phase unwrapping is possible if, and only if, balanced residues lye in all integral paths and the integration do not encircle any unbalanced residues.

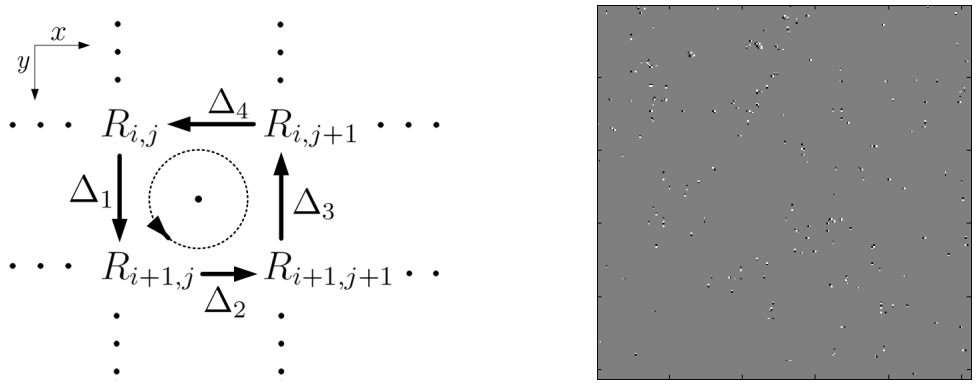
2.1.4 Locating Residues in Two-Dimensional Arrays

Locating the residues is a crucial step of phase unwrapping due to the inconsistence information contain in such points. Let $R_{i,j}$ be the wrapped counterpart of the original phase $F_{i,j}$. Recalling Equation (2.11), the wrapped difference of wrapped phase is analogous to the original phase difference, except for the phase residues. Thus, evaluating Equation (2.20) on all phase gradient fields will provide balanced and unbalanced phase residual information for all 2×2 pixel closed paths in the image. Lets assume that we have access to the sampled of wrapped phase arrays $R_{i,j}$. A small portion of such portion can be modelled as shown in Figure 2.3(a) and the wrapped difference of wrapped phase can be formulated by,

$$\begin{aligned} \Delta^y_{i,j} &= \mathcal{W}\left\{R_{i+1,j} - R_{i,j}\right\}, \quad i \in \{0, \ldots, M-2\}, \quad j \in \{0, \ldots, N-1\} \\ \Delta^y_{i,j} &= 0, \qquad\qquad\qquad\qquad \text{otherwise}, \end{aligned} \tag{2.21}$$

and

$$\begin{aligned} \Delta^x_{i,j} &= \mathcal{W}\left\{R_{i,j+1} - R_{i,j}\right\}, \quad i \in \{0, \ldots, M-1\}, \quad j \in \{0, \ldots, N-2\} \\ \Delta^x_{i,j} &= 0, \qquad\qquad\qquad\qquad \text{otherwise}, \end{aligned} \tag{2.22}$$



(a) (b)

Figure 2.3: (a) Closed contour evaluated of 2×2 pixel neighbourhood of wrapped phase ; (b) A residue maps with its positive and negative pixels indicating the positive and negative residues, respectively.

where reflective boundary conditions are used. x an y superscripts refer to wrapped difference in the j and i indexes, respectively. Integration of the gradient along 2×2 pixel closed path, shown in Figure 2.3(a), can be evaluated by summing the phase differences around the closed path. Referring to Figure 2.3(a),

$$\begin{aligned}
\Delta_1 &= \Delta^y_{i,j} = \mathcal{W}\{R_{i+1,j} - R_{i,j}\} \\
\Delta_2 &= \Delta^x_{i+1,j} = \mathcal{W}\{R_{i+1,j+1} - R_{i+1,j}\} \\
\Delta_3 &= -\Delta^y_{i,j+1} = \mathcal{W}\{R_{i,j+1} - R_{i+1,j+1}\} \\
\Delta_4 &= -\Delta^x_{i,j} = \mathcal{W}\{R_{i,j} - R_{i,j+1}\} \\
&\Rightarrow \text{Residue Charge} = \textstyle\sum_{k=1}^{4} \Delta_k
\end{aligned} \tag{2.23}$$

The closed path integral along any 2×2 pixel is equal to zero if the cross-derivative condition in (2.18) holds, otherwise there will exist an unbalanced residue of the closed integral with $\pm 2\pi$ value. Figure 2.3(b) demonstrates the location of residual point of wrapped phase in Figure 2.2(c). Positive and negative charges are shown by white and black pixels, respectively. Residues mark near origination/termination of end points of disconnected lines in wrapped phase image along which the true phase gradient exceeds $\pm\pi$. Such disconnected lines are also known as fringe-patterns.

2.1.5 Quality Maps and Masks

Residues make the process of phase unwrapping path dependent and the problem turns to be ill-posed. In fact, given a wrapped phase, one can retrieve infinite number of possible corresponding unwrapped images. In many algorithms the performance of unwrapping process depends on how such residues are going to be incorporated in the procedure. For example, in many path-following algorithms, the performance of the methods depends on the path chosen for unwrapping. Goldstain's brach cut algorithm [23], known as a classical algorithm, is one of fastest methods which connects nearby residues to make them balanced and minimize the length of branch cuts. However, this method is not an optimal solution. In several path following methods [41,42] and weighted L^p-norm solutions [26–29], the measured phase is qualified to guide the unwrapping procedure. This is done by quantifying the quality information generated by residual points.

The quality maps are arrays of values which define the quality of given phase data. For each two dimensional wrapped phase we can define a quality map that measures the quality or "goodness" of each pixel in the image [21]. Amongst many maps, the following maps are commonly used in the field of phase unwrapping

- *Pseudocorrelation* is analogous to the correlation map of the complex-valued SAR images,

$$|Q_{m,n}| = \frac{\sqrt{\left(\sum \cos\psi_{i,j}\right)^2 + \left(\sum \sin\psi_{i,j}\right)^2}}{k^2} \tag{2.24}$$

  where the sum are evaluated over a $k \times k$ neighbourhood of each pixel (m, n) and $\psi_{i,j}$ is the arrayed measured phase. Figure 2.4(c) shows the related Pseudocorrelation map of the wrapped image in Figure 2.4(a).

- *Phase Derivative Variance* is the statistical variance of difference of wrapped phase,

$$Q_{m,n} = \frac{\sqrt{\sum \left(\nabla_y\psi_{i,j} - \overline{\nabla_y\psi}_{m,n}\right)^2} + \sqrt{\sum \left(\nabla_x\psi_{i,j} - \overline{\nabla_x\psi}_{m,n}\right)^2}}{k^2} \tag{2.25}$$

  where the sum is taken over $k \times k$ windows centred at (m, n) pixel in the image. $\nabla_y\psi_{i,j}$ and $\nabla_x\psi_{i,j}$ are the y and x gradients which can be approximated by (2.21) and (2.22),

respectively. Finally, $\overline{\nabla_y \psi}_{m,n}$ and $\overline{\nabla_x \psi}_{m,n}$ are the averages of y and x gradients defined on $k \times k$ windows, respectively. This quality map is shown in Figure 2.4(d).

- *Maximum Phase Gradient* defines the largest phase gradient in $k \times k$ windows, see Figure 2.4(e) and 2.4(f).

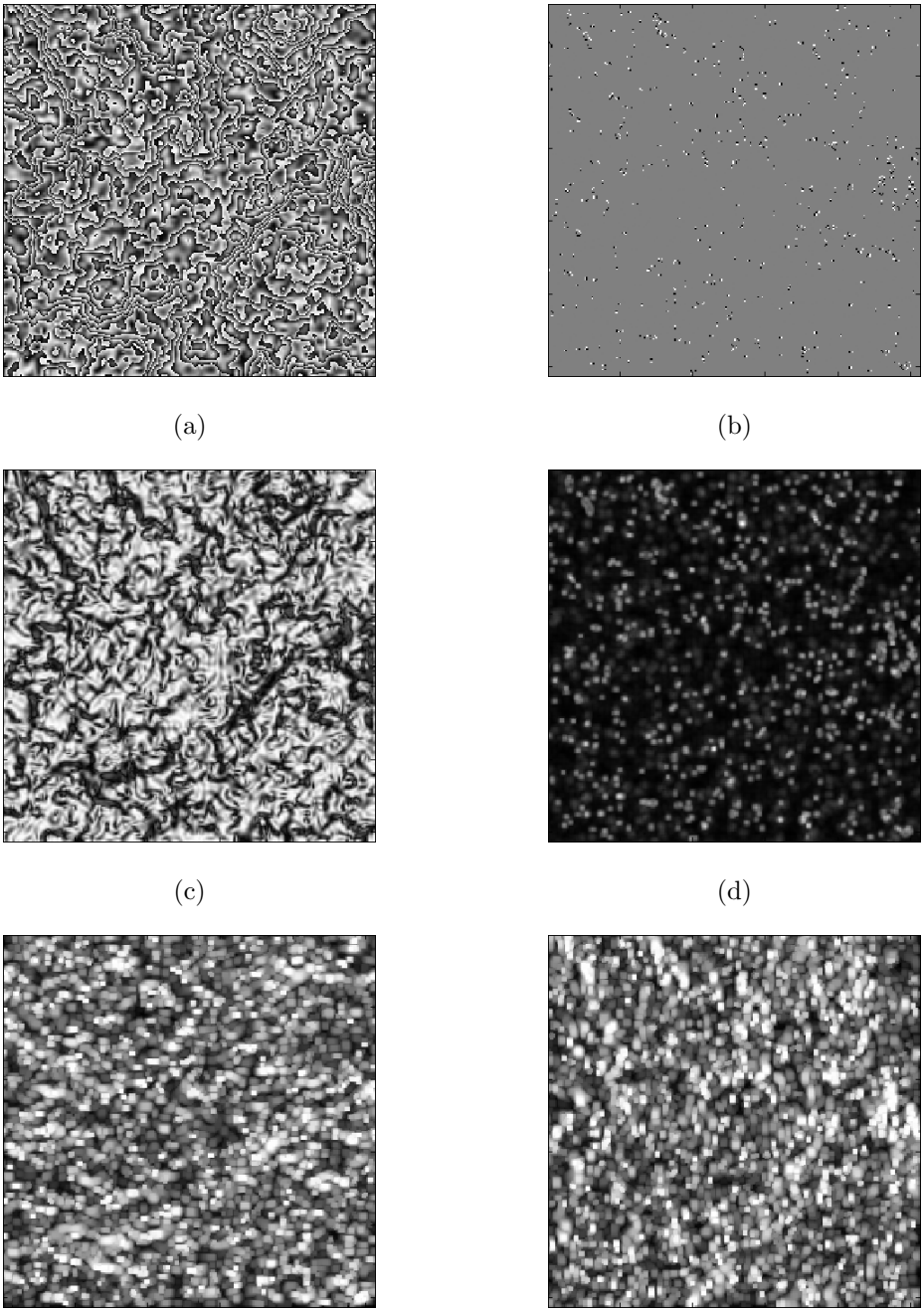
$$\begin{aligned} Q^y_{m,n} &= \max\left\{\left|\Delta^y_{i,j}\right|\right\} \\ Q^x_{m,n} &= \max\left\{\left|\Delta^x_{i,j}\right|\right\} \end{aligned} \tag{2.26}$$

On phase that remains in constant variation, the phase derivative variance is zero and it differs from pseudocorrelation measure, see Figure 2.4(c) and 2.4(d) for comparison. In addition to the mentioned three quality maps, there is an extra correlation map which is been calculated via tilted baseline, see Figure 2.1 in SAR imaging. This map is the best estimation of the quality of measured phase in SAR imaging since it measures the decorrelated phase caused by SAR layover. The baseline distance B between two antennas S_1 and S_2 and the tilt of this base line (see Figure 2.1) generates two complex-valued SAR images, $u_{i,j}$ and $v_{i,j}$, where it generates the correlation map, i.e.

$$Q_{m,n} = \frac{\sum u_{i,j} v^*_{i,j}}{\sqrt{\sum |u_{i,j}|^2 \sum |v_{i,j}|^2}}, \tag{2.27}$$

where $v^*_{i,j}$ is the complex conjugate of $v_{i,j}$. The sum in (2.27), called "multilook averaging," is evaluated over the $k \times k$ neighbourhood centred at (m, n) pixels. Among the three introduced quality maps, phase derivative variance highlight the same regions of decorrelated values in correlated map. This quality map is the most reliable one for phase quality measurement when the correlation map data is not available [21].

In order to depict the quality of measured phase, the arrays of the quality map is quantized and variates between $[0, 1]$ increasing from the poor to the best quality. Usually the unwrapping methods require mask for each pixel instead of the quality value in order to depict whether the related pixel is going to be considered in unwrapping procedure or not. So, in practice, the quality map are binarized by means of global thresholding to produce the related mask. Usually this threshold value is extracted by an adaptive method called quality re-mapping [43] or manually by a human factor [42], see Figure (2.5) for the binarized masks.

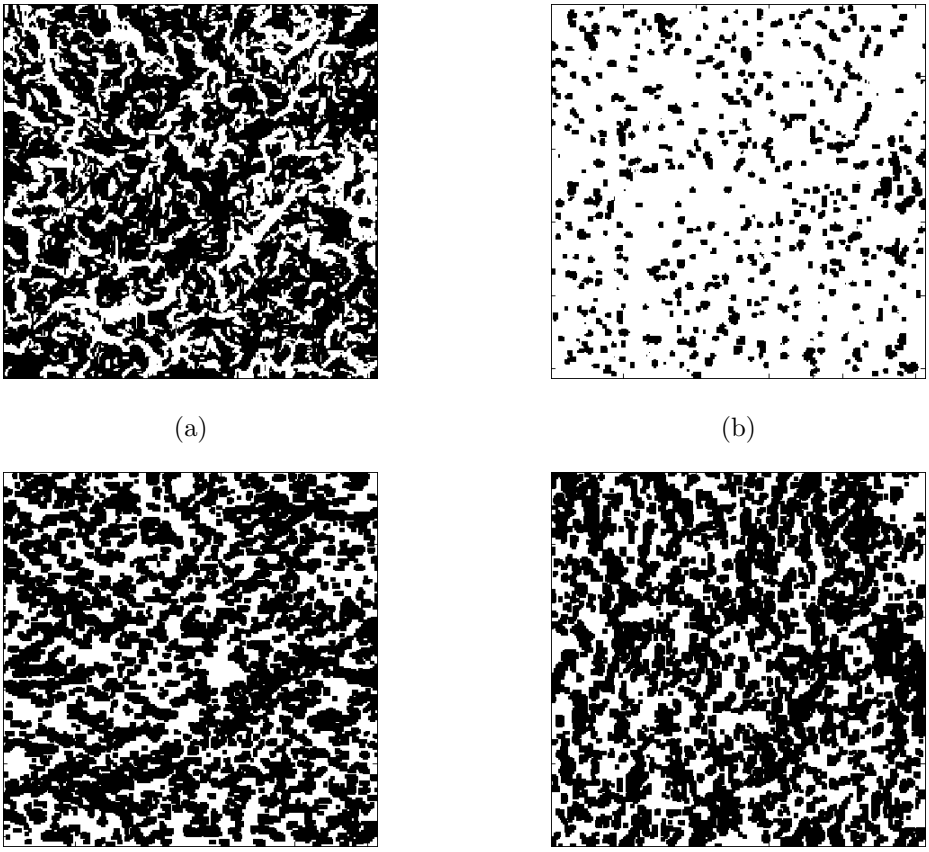


(a) (b)

(c) (d)

(e) (f)

Figure 2.4: (a) Wrapped phase; (b) Residues of the wrapped phase; (c) Pseudocorrelation map of the wrapped phase; (d) Phase gradient variance of the wrapped phase; (e) Maximum phase y-gradient map; (f) Maximum phase x-gradient map.



(a) (b)

(c) (d)

Figure 2.5: (a) Mask obtained by thresholding pseudocorrelation quality map; (b) Phase Gradient Variance Mask; (c) Maximum Phase Zy-Gradient Mask; (d) Maximum Phase Zx-Gradient Mask.

2.1.6 phase Unwrapping in 2D

An increasing number of research methods have been introduced over the past three decades for two dimensional phase unwrapping. These methods are mainly classified in two parts: path-following algorithms and L^p-norm solutions.

Branch cut algorithms by Goldstein et al [23] was introduced in 1988 which is regarded as classical path-following method. The aim of the method was to connect nearby residues with opposite polarities by branch cuts in order to minimize the sum of the cut lengths. The integration path in this method is not allowed to have cross-overs, so the closed loop integral of phase difference in the related path will vanish to zero. The algorithm generates minimum branch cuts and is extremely fast. However, Goldstein algorithm suffers from the noisy measurements since it generates branch-cuts that joins the residues in clumps rather than pairs. Some other branch cut algorithms are addressed in [24] and [25] overcome this problem by restricting the branch-cuts to dipole cuts.

The second approach for phase unwrapping is the minimum L^p-norm methods, where these methods try to match local derivatives with measured derivatives "as closely as possible" through minimizing the norm of the error in (2.12). Least squares solution (L^2-norm) first proposed by Fried and Hudgin [26, 27] in 1977 to minimize the sum of squares of gradients between the wrapped phase and reconstructed surface. Fornaro et al [28] improved the method by introducing a weighted mask on the least squares. Ghiglia and Romero [44] developed wighted least square method by combining fast cosine transform and iterative method to unwrap the phase. The L^2-nomr method has also been investigated by Pritt [43] where he used multigrid techniques to solve Gauss-Seidel relaxation schemes.

L^2-norm solutions are sensitive to the outliers, so the solution does not pass through the data points exactly. The minimum absolute error occurs where $p = 1$ where the outliers have lower impact on the solution. The solution to the problems is addressed by Flynn [29] and Costantini [30] where they used discontinuity modelling approaches and network flow for global minimization, respectively. In particular, for the case $0 \leq p < 1$ the discontinuity is preserved as feature of L^p-norm algorithms and enhanced by Chen [31]. Although some of these algorithms are more accurate and stable, but they lack from the computational

efficiency and practically heavy to implement.

L^0-norm optimized solutions are the most desirable methods in practice. These methods impose a constraint to the solution such that the phase difference of wrapped phase should matches with the unwrapped phase in the sense of minimum L^0-norm. During the process, the weighting coefficients from quality maps are required to mask the phase inconsistencies [21, 45]. Unfortunately, since the L^0-minimization method is non-convex, the search space is non-feasible and the unique minimized solution is not guaranteed. In addition, these methods need combinatorial searching and computationally heavy to implement.

Some other unwrapping methods have been introduce where they combine both path-following and minimum L^p-norm methods. Bioucas-Dias and Valado [32] proposed new energy minimization framework (PUMA) in 2007 on phase unwrapping based on graph cuts optimization technique, where the algorithm considers convex pairwise pixel interaction using classical minimum L^p-norm problems for $p \geq 1$. Later on, Bioucas-Dias [33] improved the methodology by means of adaptive local de-noising based on local polynomial approximation prior to their phase unwrapping algorithm.

2.2 Compressive Sampling

The idea of sparse signal recovery was first introduced by Donoho [46] in the form of the generalized uncertainty principle. In this initial setup, a band limited signal $f(t) \in \mathbb{L}_2(\mathbb{R})$ is used for transmission over a channel, in which it "loses" its values on a subset T. Formally, one can define

$$r\,(t) = (\mathbf{I} - \mathbf{P}_T)\, f(t) + n(t), \tag{2.28}$$

where $\mathbf{I}$ denotes the identity operator $(\mathbf{I}f)(t) = f$, $n(t)$ is observation noise, and $\mathbf{P}_T$ denotes the spatial limiting operator of the form

$$\mathbf{P}_T f(t) = \begin{cases} f(t), & t \in T \\ 0, & \text{otherwise} \end{cases} \tag{2.29}$$

The second operator used in [46] is a band limiting operator defined by

$$\mathbf{P}_\Omega f(t) \equiv \int_\Omega \hat{f}(\omega) e^{2\pi\imath\omega t} d\omega, \tag{2.30}$$

where $\hat{f}(\omega)$ denotes the Fourier transform of $f(t)$. The function f and its Fourier transformed $\hat{f}(\omega)$ are mainly concentrated on the measured subsets T and Ω, respectively, such that

$$\|f - \mathbf{P}_T f(t)\|_2 \leq \epsilon_T \tag{2.31}$$

and

$$\left\|\hat{f} - \mathbf{P}_\Omega \hat{f}(t)\right\|_2 \leq \epsilon_\Omega. \tag{2.32}$$

By the above assumption, the following theorem is introduced [46].

Theorem 1 *Let T and Ω be measurable sets and suppose there is a Fourier transform pair $(f, \hat{f})$, with f and $\hat{f}$ of unit norm, such that f is ϵ_T-concentrated on T and $\hat{f}(\omega)$ is ϵ_Ω-concentrated on Ω. Then the following inequality holds*

$$|1 - (\epsilon_T + \epsilon_\Omega)| \leq \|\mathbf{P}_\Omega \mathbf{P}_T\| \leq \sqrt{|\Omega| \cdot |T|}, \tag{2.33}$$

where for the bounded linear operator $\mathbf{P}_T \mathbf{P}_\Omega$ the norm of the operator is

$$\|\mathbf{P}_T \mathbf{P}_\Omega\| = \sup_{f \in \mathbb{L}_2(\mathbb{R})} \frac{\|\mathbf{P}_T \mathbf{P}_\Omega f\|}{\|\mathbf{P}_\Omega f\|}. \tag{2.34}$$

The operator norm $\|\mathbf{P}_T \mathbf{P}_\Omega\|$ measures how a band-limited function (i.e., $g \in \mathbf{B_2(\Omega)}$ implies $\mathbf{P_\Omega}$g=g) can be concentrated on T. The inequality $\|\mathbf{P}_\Omega \mathbf{P}_T\| \leq \sqrt{|\Omega| \cdot |T|}$ implies that there is a limited "energy" concentration on T for band limited signals. For example, if $|\Omega| \cdot |T| = 0.5$ then it implies no band limited function can be located on T with more than 50% of its "energy". The same theory applies for the finite dimensional signals where the sets T and Ω become index set.

The main goal in [46] is to reconstruct the transmitted signal f from the noisy received signal r. The possibility of such a recovery is assured by Theorems 1 above and Theorem 4 in [46] asserting that if $|\Omega|\,|T^c| < 1$, where T^c indicates the zero measure set of the function (complement of the set T), then there exists a linear operator $\mathbf{Q}$ and a constant p such that

$$\|f - \mathbf{Q}[r]\| \leq p\|n\|, \tag{2.35}$$

where $p \leq \left(1 - \sqrt{|T^c|\,|\Omega|}\right)^{-1}$. Specifically, the reconstruction operator $\mathbf{Q}$ is given by

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P}_T\mathbf{P}_\Omega)^{-1} = \sum_{k=0}^{\infty} (\mathbf{P}_T\mathbf{P}_\Omega)^k . \tag{2.36}$$

Moreover, the resulting solution is unique, and it can be approximated by computing a truncated Neumann series for some finite $k = N$. This interesting result indicates that if a band-limited signal corrupts in the receiver such that $|\Omega|\,|T^c| < 1$ then the signal can be recovered via iterative method introduced by Neumann series. Such condition is analogues to the Heisenberg inequality where the support of the signal in spatial and frequency domains have inverse relation with each other.

Over the past two decades, the idea of sparse signal recovery has been generalized and investigated with wavelets analysis. Sparse representation of the signals by means of multi-resolutional analysis brought many attentions to the field over the past decade where it originated the concept of compressed sensing to recover signals with fewer sample rates. This concept is in exact opposite definition of Nyquist-Shannon condition, where it indicates that in order to avoid aliasing in signal recovery, the signal should be sampled with twice the frequency exist in the signal. But, compressive sampling propose the idea that if the signal can be represented sparse in a domain then with much fewer sampling rate, the recovery is possible.

2.2.1 Possible Generalization of CS Problem

The theory of "*Compressive Sampling (CS)*", also referred as compressed sensing, is a novel pattern of sampling strategy which is against the conventional approach of data acquisition. The theory asserts that one can recover signals of interest with an incomplete measurements compared to traditional method used for signal recovery. This theory has been attracted a lot of attention both in mathematics and application. Generally speaking, this idea refers to recovering n-dimensional signals f approximately from linear measurements $\langle x, \phi_i \rangle$ where $\phi_i \in \mathbb{R}^N$ may form an orthonormal basis,

$$y_i = \langle f, \phi_i \rangle , \quad i \in \Omega \subset \{1, \ldots, n\} , \tag{2.37}$$

where $m = |\Omega|$ is the cardinality of Ω defines the number of partial indices, usually picked uniformly at random, such that $m < n$. Let Φ_Ω be the $n \times m$ matrix whose columns are restricted to ϕ_i for which $i \in \Omega$. Then, assuming that the signal f can be sparsely represented by a dictionary, for example usually wavelet domains, as $f = \Psi\, c$ for some coefficient vector $c \in \mathbb{R}^n$, the signal can be exactly recovered by solving combinatorial optimization problem

$$(P_0) \quad \min_c \|c\|_0 \;\; s.t. \;\; \Phi_\Omega^T \Psi c = y, \tag{2.38}$$

where $y \in \mathbb{R}^m$ stands the vector of m measurements in (2.11) and $\|c\|_0$ is the number of nonzero components in c. $K = \#\{c_i \neq 0, i \in \{1, \ldots, n\}\}$. However, this minimization is a non-convex problem and the searching space is unfeasible, so the unique minimization is not guaranteed. Instead, a more computationally efficient strategy for recovering c from its measurements can be carried out by solving convex L_1-minimization problem

$$(P_1) \quad \min_c \|c\|_1 = \textstyle\sum_{k=1}^n |c_k|, \;\; s.t. \;\; \Phi_\Omega^T \Psi c = y, \tag{2.39}$$

The above problem can be solved by means of linear programming which, among all solutions obeying the measurement constraint $\Phi_\Omega^T \Psi c = y$, picks the one that has the *sparsest* representation in the domain of Ψ as measured by the ℓ_1-norm of c.

One can show that, under certain conditions, the problem P0 and P1 are equivalent. For instance , Donoho et al showed in [1] that unique sparse representation $\hat{c}$ to the solution of P_1 is equivalent to P_0 if, and only if,

$$k = \|\hat{c}\|_0 < (1 + 1/M)/2 \tag{2.40}$$

where M is the mutual coherency of the overcomplete dictionary $[\Phi,\ \Psi]$,

$$M = \max_{\substack{1 \leq i,j \leq n \\ i \neq j}} |\langle \phi_i, \psi_j \rangle| \tag{2.41}$$

This result provides a bound on the number of elements of sparse vectors that can be recovered via linear programming LP (P_1) and allow to construct a matrices $[\Phi,\ \Psi]$ with $k \asymp \sqrt{n}$. Recent surveys in compressed sensing [2, 3] provide results in the existence of

matrices with $k \asymp n/\log(m/n)$ which is substantially larger than $\sqrt{n}$. Extensive research has been done in the field to extract bounds for stable recoveries. The sampling rate versus sparsity of the vectors is the main issue to guarantee unique and efficient recoveries.

2.2.2 Formal Definitions and Underlying Principles

The idea of compressive sampling was first formulated by Candes in [4] showing that certain classes of vectors can be recovered from their partial measurements via solving the minimization problem P1. Formally speaking, the necessary and sufficient condition for c to be a unique minimizer of (P_1) is dependent on the existence of a trigonometric polynomial function $P(t)$. Suppose our sampling is limited to partial information on $\hat{c}$ such that any solution to (P_1) should obey,

$$A_\Omega^T c = A_\Omega^T \hat{c}, \quad \Omega \subset \mathbb{Z}_n \tag{2.42}$$

where $A_\Omega \in \mathbb{R}^{n\times m}$. Lets define a subset of vectors $u \in X$ which has the following features,

1. The support of the vector obeys $supp(u) = T$
2. There exist a polynomial function $P(t)$: $\begin{cases} = sign(u(t)), & t \in T \\ |P(t)| < 1, & t \in T^c \end{cases}$
3. The polynomial functioned defined be u exist in the kernel of the sensing matrix restricted to the complement of the sampled indices Ω^c, i.e., $P(t) \in \mathcal{N}\left(A_{\Omega^c}^T\right)$

where $supp(.)$ denotes the support of a function by obtaining the indices where the function is non-zero. The main theorem in [4] can now be stated as follows.

Theorem 2 *The necessary and sufficient condition for the solution c to be the solution to (P_1) is that $c \in X$. If this holds, then adding any vector from the kernel of the operator $\forall \eta(t) \in ker(A_\Omega)$ will increase the norm, i.e. $\|c+\eta\|_1 \geq \|c\|_1$. Moreover, if $A_{T\Omega}$ is injective, then $\|c+\eta\|_1 = \|c\|_1$ which means the minimizer $\hat{c}$ to (P_1) is unique and is equal to c.*

$A_{T\Omega}$ here denotes A_Ω whose rows are restricted to T. If $T \leq \Omega$ and the matrix $A_{T\Omega}$ is of full row rank, then it is injective operator. This theorem states that certain class of vectors

which obeying the three conditions above can be the unique minimizer to (P_1) if the sensing matrix holds injectivity. Analogous results have been driven from uncertainty principles in [46], where it states that if the sparsity level and sampling rate hold $|T||\Omega| < n/2$, then (P_1) uniquely recovers the signal of interest. Donoho et al [47] expressed this classical results in more generalized format that c is the unique minimizer to (P_1) if, and only if,

$$\sum_{t\in\mathbb{Z}_n} |c(t)+\eta(t)| > \sum_{t\in\mathbb{Z}_n} |c(t)|\,, \quad \forall\eta\neq 0, \quad \eta\in\ker(A_\Omega) \tag{2.43}$$

by partition of the left sides of the inequality to T and T^c and applying triangle inequality,

$$\sum_{t\in\mathbb{Z}_n} |c(t)+\eta(t)| = \sum_{t\in T} |c(t)+\eta(t)| + \sum_{t\in T^c} |\eta(t)| \geq \sum_{t\in T} |c(t)| - \sum_{t\in T} |\eta(t)| + \sum_{t\in T^c} |\eta(t)|\,. \tag{2.44}$$

Hence, the sufficient condition for c to be a unique minimizer of (P_1) is

$$\sum_{t\in T} |\eta(t)| < \sum_{t\in T^c} |\eta(t)|\,, \quad \forall\eta\neq 0, \quad \eta\in\ker(A_\Omega). \tag{2.45}$$

By adding $\sum_{t\in T}|\eta(t)|$ to both sides in (2.45), consequently, the inequality can be redefined as,

$$\|\eta_T\|_1 < \frac{1}{2}\|\eta\|_1\,, \quad \forall\eta\neq 0, \quad \eta\in\ker(A_\Omega) \tag{2.46}$$

where η_T is a vector in kernel of the operator A_Ω whose indices are restricted to the support of the vector c. The analogy is now clear, which tells that c is the unique minimizer of (P_1) if we can not concentrate half of the l_1-norm of a vector in the kernel of sensing matrix A_Ω (missing information) on a small set T. This results, as mentioned above, connects the main theorem 2 in [4] with results of uncertainty principle in [46] by searching for the existence of a polynomial function $P(t)$ in second condition discussed before. This limits the choice of vectors to be sparse and unique minimizer to the (P_1). Candes defined this polynomial by,

$$P \triangleq A_\Omega A_{T\Omega}^* \left(A_{T\Omega}A_{T\Omega}^*\right)^{-1} \mathrm{sgn}(c_T) \tag{2.47}$$

where it satisfies $P(t) = \mathrm{sgn}(c(t))$ for $t\in T$. If the set of observation Ω is chosen unformly at random, such that

$$m = |\Omega| \geq C_M^{-1}\cdot|T|\cdot\log n \tag{2.48}$$

where C_M is an accuracy parameter, then

- Invertibility of $A_{T\Omega}A^*_{T\Omega}$ holds by means of the injectivity of $A_{T\Omega}$ with overwhelming probability of $1 - \mathcal{O}(n^{-M})$
- The polynomial function $P(t)$ will obey $|P(t)| < 1$ for T^c with the same probability

Several measurement models have been studied in [4] to analyze the stability of the recovery. The above conditions are defined as Exact Reconstruction Principle (ERP) in [5] in order to establish exact reconstruction of $\hat{c}$ in (P_1) by truncating c, keeping only T largest values.

The choice of sampling basis is of importance in CS since the linear measurement combine correlation of the atoms in dictionary with the signal of interest by means of inner product in (2.37). If this correlated value remains high it means the signal is observable by the atoms of the sampling dictionary, otherwise the amount of information measured by the related inner product will remain low and nothing can be done. Discrete Fourier dictionary for sampling basis has been studied by Candes and Romberg in [6] to justify the principles of exact recovery in CS. They have derived a bound on the number of measurements in the frequency domain (which chosen uniformly at random) versus sparsity in any orthogonal basis

$$|T| + |\Omega| \asymp (\log n)^{-1/2} \cdot n. \tag{2.49}$$

The result in (2.49) states that if the signal c is supported on T, then less than half of the energy of the signal in Fourier domain, $A^T_\Omega c$, will concentrate on Ω.

Fourier basis is not the only sampling/representation system which can be used in compressive sampling. Gaussian white noise and Bernoulli random matrices with ± 1 values distributed uniformly at random are also studied in [5, 8, 48, 49] where it has been shown that the sparse signals can be recovered with near-minimal number of measurements. As a matter of fact we are not in liberty to define the measurement ensembles, where there are two circumstances that are imposed to limit our choice in practice. First, the modalities used for data acquisition, e.g., magnetic resonance imaging (MRI) where we have control over the Fourier coefficients for sampling or tomographic imaging where the machinery measures the radon slices. The second drawback is the computational burden where the random measurement ensembles are widely numerically; for large scale problem (image reconstruction), the storage of atoms in dictionary are nearly impossible. So the goal is to find sampling matrices

that provide the same recovery bounds and can be quicklly applied in practice. Such restrictions have been taken to the account in compressed sensing by Candes and Romberg [7]. They have shown that the relationship between the sensing modality (Φ) and signal model (Ψ) affects the number of samples for sparse signals reconstruction. Simply, the recovery is possible if the number of observations exceeds

$$|\Omega| \geq C \cdot \mu^2(\Phi, \Psi) \cdot |T| \cdot \log n, \tag{2.50}$$

where C is a positive constant and $\mu(\Phi, \Psi)$ is the mutual incoherence between two dictionaries Φ and Ψ by measuring similarities of the atoms in two bases

$$\mu(\Phi, \Psi) = \max_{i,j \in \mathbb{Z}_n} |\langle \phi_i, \psi_j \rangle|. \tag{2.51}$$

Stability of the recovery is an important issue in the field of compressive sampling in order to analyze the performance of the minimization algorithms. The remainder of this chapter addresses two approaches for performance analysis: *restricted isometry property (RIP)* introduce by Candes et al [8, 9] and width of finite dimensional sets, called *distortion of a subspace*, introduce by Kashin, Garnaev and Gluskin [50, 51].

2.2.3 Restricted Isometry Property (RIP)

Suppose we can recover sub-sampled version of a signal f which is sparse (compressible) in the domain of Ψ, i.e. $c = \Psi^T f$. The stability of such recovery was first reported by Candes et al [8, 9], introducing restricted isometry property (RIP) of the sensing matrix $A = \Phi^T \Psi$. A is said to obey the RIP of order k, RIP^k, if there is a $0 < \delta_k < 1$ such that

$$(1 - \delta_k) \|c\|_2^2 \leq \left\| A_\Omega^T c \right\|_2^2 \leq (1 + \delta_k) \|c\|_2^2 \tag{2.52}$$

holds for all T of cardinality at most k, i.e., $|T| \leq k$. Property (2.52) gives,

$$(1 - \delta_k) \leq \lambda_{\min} \left(A_{T\Omega} A_{T\Omega}^T \right) \leq \lambda_{\max} \left(A_{T\Omega} A_{T\Omega}^T \right) \leq (1 + \delta_k). \tag{2.53}$$

So, the condition number of the sensing matrix is bounded by

$$\kappa \left(A_{T\Omega} \right) = \frac{\lambda_{\max}}{\lambda_{\min}} \leq \frac{1 + \delta_k}{1 - \delta_k} \tag{2.54}$$

As $\delta_k \to 0$ the bounds in (2.52) become more tight and the condition number of $A_{T\Omega}$ decreases to 1 and makes the matrix $A_{T\Omega}A_{T\Omega}^T$ invertible. Candes and Tao [10] proved if the sensing matrices A_Ω meets the condition $\delta_{2k} + \delta_{3k} < 1$ for the sparsity level of $k \asymp n/\log m/n$ then c will be the unique minimizer to (P_1). They also prove in [9] that if $\delta_{3k} + \delta_{4k} < 2$ then the error of recovery is bounded by

$$\|\hat{c} - c\| \leq C \cdot k^{-1/2}\sigma_k(c)_1 \tag{2.55}$$

where C is some positive constant and $\sigma_k(c)_1$ is the best k term approximation of vector c in l_1-norm,

$$\sigma_k(c)_1 \triangleq \min_{u\in\mathbb{R}^n:\ |supp(u)|\leq k} \|c - u\|_1 . \tag{2.56}$$

Further results on RIP analysis has been provided by Cohen et al [11] where they proved that if the sensing matrix satisfies RIP of order $2k$ with $\delta_{2k} < \delta < 1/3$ then the error of recovery is bounded by

$$\|\hat{c} - c\| \leq \frac{2 + 2\delta}{1 - 3\delta}\sigma_k(c)_1. \tag{2.57}$$

Several methods conducted the RIP condition as their framework of performance analysis for the sparse signal recovery [52–55]. The RIP condition also has been reshaped in generalized format, called GRIP [56], with equivalent term preserved by angels between sparse vectors. Although RIP analysis is a fundamental platform for performance measurements of (P_1) algorithm, but there are some restrictions that makes this framework arguable. For instance, Chander [57] raised negative results on explicit matrices whose entries are all 0 and 1 and he reported poor performance from RIP property.

Using RIP property for measuring the stability of (P_1) recovery is basically matrix dependent [58], where by multiplying the sensing matrix A_Ω with a fullrank matrix G can change the characterization of such stability analysis. In particular, any designed model for recovery, using the system of equation $A_\Omega^T c = y$, contains the whole information about the signal that is going to be recovered. The equation $A_\Omega^T c = y$ is equivalent to be multiplied from both sides to a non singular matrix $G \in \mathbb{R}^{m\times m}$, i.e., $GA_\Omega^T c = Gy$. From numerical considerations, $\left(GA_\Omega^T, Gb\right)$ carries the same amount of information that (A, b) does. However, the RIP properties of two equivalent equations can be vastly different such that the RIP of

GA to be bad no matter how good the RIP of A is. To show this, Equation (2.52) can be revised as

$$(1-\delta_k) \leq \frac{\left\|A_\Omega^T c\right\|_2^2}{\|c\|_2^2} \leq (1+\delta_k), \tag{2.58}$$

so, the bounds in (2.58) controls the the matrix norm $\left\|A_{T\Omega}^T\right\|_2^2$. Lets define

$$\lambda_{\min}^k = \min_{\|c\|_0=k} \frac{\left\|A_\Omega^T c\right\|_2^2}{\|c\|_2^2} = 1-\delta_k, \tag{2.59}$$

$$\lambda_{\max}^k = \max_{\|c\|_0=k} \frac{\left\|A_\Omega^T c\right\|_2^2}{\|c\|_2^2} = 1+\delta_k \tag{2.60}$$

and

$$\Gamma_k(A_\Omega^T) = \frac{\lambda_{\max}^k}{\lambda_{\min}^k} = \frac{1+\delta_k}{1-\delta_k} \tag{2.61}$$

Equation (2.61) can be solved for δ_k, i.e.,

$$\delta_k\left(A_\Omega^T\right) = \frac{\Gamma(A_\Omega^T)-1}{\Gamma(A_\Omega^T)+1}. \tag{2.62}$$

Lets assume A_Ω is stacked by two sub-matrices: $A_\Omega = [A_1;\ A_2]$ where $A_1 \in \mathbb{R}^{m\times m}$ is a non-singular matrix. Set $G = BA_1^{-T}$ for some non-singular $B \in \mathbb{R}^{m\times m}$. Then $GA_\Omega^T = \left[B,\ GA_2^T\right]$. Let $B_1 \in \mathbb{R}^{m\times k}$ corresponds to the first k columns of B and let the condition number of $\kappa\left(B_1^T B_1\right) = \lambda_1/\lambda_k$ where λ_1 and λ_k are the maximum and minimum eigenvalue of $B_1^T B_1$, respectively. Then $\kappa\left(B_1^T B_1\right) \leq \Gamma_k(GA_\Omega^T)$ and we have

$$\delta_k\left(GA_\Omega^T\right) = \frac{\Gamma(GA_\Omega^T)-1}{\Gamma(GA_\Omega^T)+1} = 1 - \frac{2}{\Gamma(GA_\Omega^T)+1} \geq 1 - \frac{2}{\kappa\left(B_1^T B_1\right)+1}. \tag{2.63}$$

Suppose one can choose B_1 such that $\kappa\left(B_1^T B_1\right)$ would be arbitrary large and $\delta_k\left(GA_\Omega^T\right) \to 1$. As long as the matrix B remains non-singular, the recoverability and stability of the system of equation $GA_\Omega^T c = Gy$ for l_1-minimization will remain exactly the same as (P_1). But since the RIP property of GA_Ω^T varies with G, it implies that the recoverability and stability of the decoder will vary as well. Beside the theoretical implication, the RIP based analysis remains conservative in practice for different sensing modalities, e.g. Gaussian or the ± 1 Bernoulli matrices [58]. These issues make the implication of RIP conditions weak for stability analysis and not reliable to guarantee the recoverability of the implemented decoders.

2.2.4 Kernel Measurements

Suppose we are interested to define conditions on which the problem (P_1) admits a unique solution. An alternative approach for stability and recoverability analysis of the problem, instead of RIP, is analyzed by the theory of spherical section property of the subspace generated by the sensing operator in (P_1) [58–60].

Suppose the sensing matrix is defined by $A_\Omega \in \mathbb{R}^{n\times m}$, where Ω denotes the indices referring to the columns of A. for any measurement vector $y = A_\Omega^T \hat{c}$ the class of all observable signals, which fit the same measurement model, will belong to the affine space $\mathcal{F}(y)$ defined as

$$\mathcal{F}(y) \triangleq \left\{c :\ A_\Omega^T c = y\right\}. \tag{2.64}$$

So, any vector belong in $\mathcal{F}(y)$ can be expressed as the sum of two vectors, namely, a vector in the affine space $\hat{c} \in \mathcal{F}(y)$ and a vector in the null space $\mathcal{N}(A_\Omega^T) = \left\{\eta :\ A_\Omega^T \eta = 0\right\}$.

$$c = \hat{c} + \eta \tag{2.65}$$

The vector in the null space η, also referred as the error vector, is generated in the process of recovery using (P_1) decoder. So, any solution to the problem exists in the radius ball centred by the unique minimizer with the radius vector introduced by the null space. The geometric structure of this null space, $X \subset \mathbb{R}^n$, can be characterized with the classical results introduced by Kashin [50], Figiel et al [61] and Garnaev and Gluskin [51], such that the ratio of the l_1-norm and l_2-norm is bounded and varies from 1 to $\sqrt{n}$, for every $\eta \in X$,

$$1 \le \frac{\|\eta\|_1}{\|\eta\|_2} \le \sqrt{n}, \quad \forall \eta \in X \setminus \{0\} \tag{2.66}$$

where the right inequality follows from Cauchy-Schwartz,

$$\|\eta\|_1 = \sum_{i\in\mathbb{Z}_n} |\eta_i| = \sum_{i\in\mathbb{Z}_n} \eta_i \cdot \mathrm{sgn}\eta_i \le \left[\sum_{i\in\mathbb{Z}_n} \eta_i^2\right]^{1/2} \left[\sum_{i\in\mathbb{Z}_n} (\mathrm{sgn}\eta_i)^2\right]^{1/2} = \sqrt{n}\,\|\eta\|_2 \tag{2.67}$$

Furthermore, the Cauchy-Schwartz inequality holds with equality for scalers. The ratio in (2.66) is relatively small for sparse vectors and as a matter of fact this ratio in most subspaces have much larger lower bound than 1. This implies that most subspaces do not contain excessively sparse vectors [58]. Thus, the following definition is used by many authors, e.g. [50, 51, 58–60, 62, 63]

Definition 1 *Let m, n be two positive integer where $m < n$. Let $X \subset \mathbb{R}^n$ be an $(n-m)$ dimensional subspace. The distortion of the subspace X is defined by*

$$\Delta(X) \triangleq \sup_{\eta \in X \setminus \{0\}} \frac{\sqrt{n}\,\|\eta\|_2}{\|\eta\|_1} \tag{2.68}$$

In particular $\Delta(X)$ lies between $[1, \sqrt{n}]$. Kashin [50] and [51] demonstrated that for random matrices, with i.i.d. Bernoulli entries $\Delta(X) \leq \sqrt{(n/m) \cdot \log n/m}$ holds with high probability. Similar results have been also reported in [62–64]. The important question to be asked here is, how the distortion measurement relates to the compressive sampling problem. The operator A_Ω in (P_1) maps sparse vector $A_\Omega^T : \mathbb{R}^n \mapsto \mathbb{R}^m$ from an n-dimensional space to an m-dimensional subspace. Since the sensing matrix is full rank, its kernel is a non-empty subspace $X \in \mathbb{R}^{n-m}$ which maps any vectors from this subspace to zero.

Let $T = \text{supp}(x) \leq k$. Any measurement vector c, which is a solution to (P_1), satisfies $A_\Omega^T c = y$ and lies in the affine space $\mathcal{F}(y)$ in (2.65). From the classical argument, c is the unique minimizer to (P_1) if (2.43) holds. From Definition 1, $\Delta(X)$ is bounded by

$$\Delta(X) \geq \frac{\sqrt{n}\,\|\eta\|_2}{\|\eta\|_1} \tag{2.69}$$

The condition of unique minimization in (2.43) is equivalent to the condition expressed for the kernel of the operator in (2.46). Lets assume η_T the null vector restricted to the indices of T. Then, the l_1-norm of η_T is bounded by

$$\|\eta_T\|_1 = \sum_{i \in T} \eta_i \cdot \text{sgn}\eta_i \leq \left[\sum_{i \in T} \eta_i^2\right]^{1/2} \left[\sum_{i \in T} (\text{sgn}\eta_i)^2\right]^{1/2} = \sqrt{|T|}\,\|\eta_T\|_2 \leq \sqrt{|T|}\,\|\eta\|_2 \tag{2.70}$$

By substituting (2.69) in (2.70),

$$\|\eta_T\|_1 \leq \sqrt{|T|}\,\|\eta\|_2 \leq \frac{\Delta(X)\sqrt{|T|}}{\sqrt{n}}\,\|\eta\|_1 \tag{2.71}$$

If the right term of the inequality in (2.71) is bounded by $\frac{1}{2}\|\eta\|_1$ from (2.46), then exact recovery is guaranteed via (P_1) decoder, i.e.,

$$\frac{\Delta(X)\sqrt{|T|}}{\sqrt{n}}\,\|\eta\|_1 < \frac{1}{2}\,\|\eta\|_1 \iff \frac{\Delta(X)\sqrt{|T|}}{\sqrt{n}} < \frac{1}{2}. \tag{2.72}$$

The following theorem [59] defines conditions for exact reconstruction via (P_1) minimization, which provides a bound to recover any sparse vector c measured by a linear mapping A_Ω^T whose kernel has low distortion.

Theorem 3 *Let $c \in \mathbb{R}^n$ with cardinality at most k , i.e., $|supp(c)| \leq k$. c is the unique solution to (P_1) if, and only if,*

$$k < \frac{n}{4\Delta(X)^2} \tag{2.73}$$

where $X = \mathcal{N}(A_\Omega^T)$ is the null space of X

The theorem specifies the necessary number of measurements for exact reconstruction. The error of the recovery is bounded by the kernel measure given by Temlyakov [59]. In similar fashion, Vavasis [60] proved the necessary condition of recoverability by restricting kernel members in (2.43) to best S-term approximation indices of a solution to (P_1) in order to derive the error bounds. Related analysis can also be found in [48] on the range space of the sensing operator and it has been proven that if the operator satisfies three condition (CS1-CS3), then there exists a constant ρ such that the recovery will be exact via l_1 minimization for sparse vectors, where $|T| < \rho \cdot |\Omega| / \log n$. Zhang [58] used the same concept of the null space property to show stability and recoverability of (P_1) in the presence of prior information which exist in some signals of interest.

2.3 Summary

The problem of phase unwrapping is still a challenge when the number of residual points increases in the wrapped images. This affects the existing unwrapping methods and increase the error of recovery. Interpretation of Such inconsistent points in phase unwrapping algorithms is still a challenge to the criteria. Path-following methods perform fast computations, however the accuracy of these algorithms is greatly influenced in the presence of noisy measurements. Quality guided maps ease the problem by masking the candidates of residual points in wrapped phases in order to prevent inconsistencies in unwrapping process.

The conditions of exact recovery in compressive sampling are dependent on compressibility of the signal and the number of measurements. RIP condition and Kernel analysis properties are the present methods in literature to exam and analyze the performance of CS algorithms and state the condition of exact recoveries. Both methods suffer from some technical issues. RIP methods are matrix dependent and are not reliable to perform theoretical

analysis on the stability and recoverability of (P_1) decoder. On the other hand, kernel based analysis has difficulties in practical implementation where only some bounds are available on the distortion of sensing operators in (P_1), e.g., random matrices and Bernoulli matrices.

In some application, one can be provided with a priori knowledge where it carries extra information in addition to the system of equation in (P_1). This information limits the searching space in the decoder and hence can enhance recovery results. which is of great interest in the field of CS nowadays. Preliminary contribution is done by Zhang [58] where he showed that assuming the solution of the recovery to be close to a prior signal, the accuracy of the recovery increases .

Chapter 3

Proposed Method

As previously mentioned, there are many applications in which one is provided with the measurements of the gradient of an image rather than of the values of the image itself. Given the measurements of the gradient, the corresponding image can be reconstructed subject to an initial knowledge which comes from the boundary condition being considered for derivative approximation. One of such applications, which has been chosen to exemplify the major contribution of this proposal, is the problem of *phase unwrapping*.

In this proposal, we introduce a different solution to the problem of phase unwrapping which is based on the theory of compressive sampling. In particular, let $\Gamma \subset \mathbb{R}^2$ be a finite discrete subset over which the values of a phase F need to be recovered. Let further Γ_0 denote a subset of those points in Γ at which the condition (2.10) is known to hold, and hence at which the gradient ∇F estimated according to (2.11) can be assumed to be errorless. (Note that the subset Γ_0 can be identified by quality maps as detailed in 2.1.5). Subsequently, we first recover the values of ∇F over the whole Γ from its *incomplete* measurements over Γ_0, followed by estimating the original phase F using (2.11). Moreover, in addition to the standard constraints of compressive sampling, we propose to use the constraints stemming from the nature of the gradient as a potential field in (2.18). We refer to the problem of reconstruction of F from $\{\nabla F(x,y)\}_{(x,y)\in\Gamma_0}$ as the problem of *derivative compressive sampling* (DCS), and show that using (2.18) allows considerably reducing the cardinality of Γ_0, while preserving a predefined error rate.

Section 3.1 introduces the DCS method. This section defines how the derivative constraints can be used to improve the performance of CS. Since the signal recovered via DCS is an image gradient, Section 3.2 explains how the original image can be recovered from its derivatives through Least squares minimization. In Section 3.3, we introduce a procedure to find the linear dependency of the sensing matrix with cross-derivative matrix and specify a way to eliminate the unnecessary samples in Subsection 3.3.1. Preliminary results are demonstrated in Section 3.5.3 where the sampling indices for phase unwrapping are clarified by defining a quality map. This map uses two separate terms which is extracted from phase gradient variance method. Finally Section 3.6 concludes the chapter.

3.1 Derivative Compressive Sampling (DCS)

In the case when only partial derivatives of a signal of interest are available, the sampling operator of compressive sampling becomes the kernel of a derivative operator. In particular, in the 2-D case, we are given the measurements of $F_x = \partial F/\partial x$ and $F_y = \partial F/\partial y$. At this point, there are two possibilities to find F. The first would be to define Φ to be a discretized version of the 1st-order derivative operator. This choice, however, could result in relatively large values of the coherency $\mu(\Phi, \Psi) = 2\sqrt{2/n}$ for the case when Ψ is a DCT orthobasis (which is the choice in the present study). This would, in turn, increase the bound in (2.50), which could be unacceptable for practical considerations. On the other hand, one can define Φ to be the Dirac comb (i.e., $\Phi = \mathbf{I}$), so $\mu(\Phi, \Psi) = \sqrt{2/n}$. In this case, the partial derivatives can be recovered first, followed by integrating the latter using (2.12).

To proceed with the second of the above-mentioned possibilities, we turn to a discrete setup in which F, F_x and F_y are considered to be $n \times n$ matrices. In this case, the maximal possible number of measurements is equal to $2n^2$, and hence $\Gamma_0 \subset \{1, 2, \ldots, 2n^2\}$. Specifically, we are interested in the case when $m = \#\Gamma_0 < n^2$. In the two dimensional case, the partial derivatives F_x and F_y can be approximated according to

$$\begin{aligned} F_x &= \frac{\partial F}{\partial x} \cong F\mathrm{D}^T \\ F_y &= \frac{\partial F}{\partial y} \cong \mathrm{D}F, \end{aligned} \tag{3.1}$$

where D is two dimensional difference matrix given by (3.2) where, for the sake of concreteness, reflective boundary conditions have been assumed.

$$\mathrm{D} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{3.2}$$

For the sake of notational simplicity, let $\Phi^{\otimes} = \Phi \otimes \Phi$ and $\Psi^{\otimes} = \Psi \otimes \Psi$, where $\otimes$ stands for the Kronecker matrix product. Moreover, since the sampling sets for the x- and y-derivatives may be in general different, we denote the corresponding sampling matrices by $\Phi_x^{\otimes}$ and $\Phi_y^{\otimes}$, respectively. Hence, assuming that there exist sparse coefficients c_x and c_y which can be decomposed by Ψ basis such that $\mathrm{vec}(F_x) = \Psi^{\otimes}\,\mathrm{vec}(c_x)$ and $\mathrm{vec}(F_y) = \Psi^{\otimes}\,\mathrm{vec}(c_y)$ (with vec denoting the operation of matrix concatenation), the measurement constraints of the DCS problem are defined as

$$\begin{aligned} \Phi_x^{\otimes}\,\Psi^{\otimes}\,\mathrm{vec}\,(c_x) &= Y_x \\ \Phi_y^{\otimes}\,\Psi^{\otimes}\,\mathrm{vec}\,(c_y) &= Y_y, \end{aligned} \tag{3.3}$$

where Y_x and Y_y are the vectors of measured derivatives. In what follows, the constraints in (3.3) will be referred to as *primary*. As it will be discussed later, sparse representation of the image gradients is an important criteria where discrete cosine transform (DCT) is used here to approximately reach the sparsity level. Nonetheless, finding suitable dictionary is critical to the field since the sparsity is in direct relation with the sampling rate.

On the other hand, the cross-derivative (*secondary*) constraints in (2.18) can now be expressed as

$$\nabla_x \underbrace{\left\{\Psi\ c_y \Psi^T\right\}}_{Fy} = \nabla_y \underbrace{\left\{\Psi\ c_x \Psi^T\right\}}_{F_x}, \tag{3.4}$$

Using the approximation in (3.1), (3.4) can be expressed as

$$\Psi\ c_y \Psi^T \mathrm{D}^T = \mathrm{D}\Psi\ c_x \Psi^T, \tag{3.5}$$

Suppose the coefficients used in (P_1) minimization algorithm are concatenated as column-

wise. In the vectorized notation, the derivative constraint can be defined as

$$\left[\underbrace{\Psi \otimes (\mathrm{D}\Psi)}_{\mathbf{B_x}} \quad \underbrace{-(\mathrm{D}\Psi) \otimes \Psi}_{\mathbf{B_y}} \right] \underbrace{\begin{bmatrix} \mathrm{vec}\,(c_x) \\ \mathrm{vec}\,(c_y) \end{bmatrix}}_{\mathbf{c}} = 0 \tag{3.6}$$

The matrix $\mathbf{B} = [\mathbf{B_x},\ -\mathbf{B_y}]$ is of rank n^2-1 because the derivative operator D used in (3.2) is not full rank, i.e, $rank(\mathrm{D}) = n-1$. In the DCS formulation, this matrix of secondary (cross-derivative) constraints is combined with the primary constraints to result in the following optimization problem

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 = \sum_{k=1}^{n^2} |c_x(k)| + \sum_{k=1}^{n^2} |c_y(k)|, \tag{3.7}$$

subject to

$$\begin{bmatrix} \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \\ \mathbf{B} \end{bmatrix} \mathbf{c} = \begin{bmatrix} \mathbf{Y_4} \\ \mathbf{0} \end{bmatrix}, \quad \text{where} \quad \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} = \begin{bmatrix} \Phi_x^{\otimes}\,\Psi^{\otimes} & \mathbf{0} \\ \mathbf{0} & \Phi_y^{\otimes}\,\Psi^{\otimes} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} Y_x \\ Y_y \end{bmatrix} \tag{3.8}$$

The proposed L_1-norm minimization in (3.7) is a large scale problem due to the nature of Kronecker products. To alleviate the computational burden the large scale solutions have been found using the algorithm detailed in [65]. This algorithm seeks a minimum l_1-norm solution of an underdetermined least-squares problem by introducing a curve that traces the optimal trade-off between the least-squares fit. This curve is convex and differentiable in the feasible region, which is analogous to basis pursuit denoise (BPDN) fits.

3.2 Least-Squares Surface Reconstruction

Having estimated the partial derivatives F_x and F_y as $\Psi^{\otimes}c_x$ and $\Psi^{\otimes}c_y$, respectively, the phase F needs to be recovered next. We use least-squares solution to (2.12) to reconstruct the image of corresponding gradients introduced in (3.1). So the cost function can be written by

$$\mathrm{Cost} = \|\mathrm{D} \cdot f - d_y\|^2 + \left\|f \cdot \mathrm{D}^T - d_x\right\|^2, \tag{3.9}$$

where $f \in \mathbb{R}^{n \times n}$ denoted the original image that should be recovered, and d_x and d_y are the estimated x and y derivatives, respectively. The normal system corresponding to minimiza-

tion of Cost is given by

$$\mathrm{D}^T\mathrm{D}\cdot f + f\cdot \mathrm{D}^T\mathrm{D} = \mathrm{D}^T\cdot d_y + d_x\cdot \mathrm{D}. \tag{3.10}$$

This equation is known as Lyapunov or Sylvester equation which is linear respect to f and the solution to equation is basically dependent to boundary conditions for difference calculation in two dimensional case [66,67]. The boundary condition here is reflective boundary condition which is known also as Hudgin geometry in [68]. The term $\mathrm{D}^T\mathrm{D}$ in (3.10) can be diagonalized by DCT orthonormal matrix M in Fourier domain, i.e.,

$$\mathrm{D}^T\mathrm{D} = M\Lambda_{(\mathrm{D}^T\mathrm{D})}M^T. \tag{3.11}$$

By the definition of orthonormality, the transpose of M is equal to its inverse, i.e., $M^{-1} = M^T$ and alternatively $MM^T = M^TM = I$. So, by substituting (3.11) in (3.10), it yields

$$\Lambda_{(\mathrm{D}^T\mathrm{D})}M^TfM + M^TfM\Lambda_{(\mathrm{D}^T\mathrm{D})} = M^T(\mathrm{D}^Td_y + d_x\mathrm{D})M \tag{3.12}$$

As it turns, $M^T(\cdot)M$ is two dimensional DCT transform applied here on f and $\mathrm{D}^Td_y + d_x\mathrm{D}$ in (3.12). The left and right multiplication of M^TfM by $\Lambda_{(\mathrm{D}^T\mathrm{D})}$ is equivalent to multiplying every row and column of the two dimensional spectrum with the eigenvalue in the corresponding row and column of matrix $\Lambda_{(\mathrm{D}^T\mathrm{D})}$, respectively. So the final solution can be expressed by

$$f = M\left[\frac{M^T(\mathrm{D}^Td_y + d_x\mathrm{D})M}{\lambda u^T + u\lambda^T}\right]M^T, \tag{3.13}$$

where $\lambda = \mathrm{diag}(\Lambda_{(\mathrm{D}^T\mathrm{D})})$ and $u = [1, 1, \cdots, 1]^T$. Applying derivative operator D on an image will loose the bias information exist in the data (dc information) and will set the zero-frequencies to zero. So, In order to compare the original signal with the recovered version, both signals will be normalized between $[0, 1]$.

3.3 Space of Solutions and Its Analysis

Referring to the problem of (P_1), any solution is combined with two elements, one in the kernel space of the sensing operator and the other in the range space. In the case of derivative compressive sampling (DCS), the cross-derivative constraints in (3.6) have the form of $\mathbf{Bc} =$

$\mathbf{0}$ in addition to the system of equation $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{c} = \mathbf{Y}$. So, any solution to l_1-minimization in (3.7) exists in the intersection of the affine and kernel space, i.e.,

$$\mathbf{c} \in \mathcal{F}(\mathbf{Y}) \cap \mathcal{N}(\mathbf{B}). \tag{3.14}$$

In other words, any solution to l_1-minimization in (3.7) should belong to the kernel of $\mathbf{B}$ in addition to the affine subspace in (2.64). So, the general problem of derivative compressive sampling (DCS) in (3.7) can be interpreted as,

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathcal{F}(\mathbf{Y}) \cap \mathcal{N}(\mathbf{B})} \|\mathbf{c}\|_1, \tag{3.15}$$

In practice, data acquisition is contaminated by noisy measurements which makes the system of equations in (3.8) difficult to hold. As a consequence, in order to make this optimization problem more practical in implementation, a more robust and relaxed version of such minimization can be defined as,

$$\begin{aligned} &\min_{\mathbf{c}} \|\mathbf{c}\|_1 \\ \text{s.t.} \quad &(a): \ \left\|\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{c} - \mathbf{Y}\right\|_2 \leq \epsilon_1 \\ &(b): \ \|\mathbf{B}\mathbf{c}\|_2 < \epsilon_2 \end{aligned} \tag{3.16}$$

The total number of samples in *DCS* is twice the number of image pixels. This is because of the image gradients, F_x and F_y, being considered to be recovered by compressive sampling scheme. By randomly sampling the gradients, linear dependency of primary with secondary constraints (cross-derivative equality $\mathbf{Bc} = \mathbf{0}$) of sensing matrix in (3.7) is probable. Given the cross-derivative constraints, the following lemma defines the maximum number of samples that should be taken to recover the solution exactly.

Lemma 1 *Suppose the sampling matrix* $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \in \mathbb{R}^{|\mathbf{\Gamma_0}| \times \mathbf{2n^2}}$ *is defined by (3.8). Let the cross-derivative matrix* $\mathbf{B} \in \mathbb{R}^{(\mathbf{n^2-1}) \times \mathbf{2n^2}}$ *be of full rank. Then the condition that the solution* $\mathbf{c}$ *to (3.7) to be unique and exact is,*

- $|\Gamma_0| = dim\ \mathcal{N}(\mathbf{B})$
- $\left\|\mathcal{P}_{\mathcal{N}(\mathbf{B})}\left\{rng(\mathbf{A}_{\mathbf{\Gamma_0}})\right\}\right\| > 0$ *where* $\mathcal{P}_{\mathcal{N}}(\mathbf{B})$ *denotes the linear projection operator to the kernel space* $\mathcal{N}(\mathbf{B})$ *and* $rng(\mathbf{A}_{\mathbf{\Gamma_0}})$ *is the range space of the sensing matrix defined by* $rng(\mathbf{A}_{\mathbf{\Gamma_0}}) = \left\{\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^{|\mathbf{\Gamma_0}|}\right\}$

Proof Define the affine space V and the kernel space U by

$$\begin{aligned} V &= \left\{ \mathbf{c} \mid \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{c} = \mathbf{Y} \right\} \\ U &= \mathcal{N}(\mathbf{B}) = span \left\{ \theta_{\mathbf{k}} \in \mathbb{R}^{\mathbf{2n^2}} \right\}_{\mathbf{k=1}}^{\mathbf{n^2+1}} \end{aligned} \tag{3.17}$$

Where θ_k is a basis to expand the kernel space of $\mathbf{B}$. The solution exist in the intersection of the nullspace U and the affine space V. The necessity of these two conditions implies that any solution $\mathbf{c}$ in affine space can be expressed by a linear combination of the basis functions $\mathbf{\Theta} = \left\{ \theta_{\mathbf{k}} \in \mathbb{R}^{\mathbf{2n^2}} \right\}_{\mathbf{k=1}}^{\mathbf{n^2+1}}$ of U, i.e.,

$$\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{\Theta} \alpha = \mathbf{Y} \tag{3.18}$$

The dimension of both matrices $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}$ and $\mathbf{\Theta}$ are $[|\Gamma_0| \times 2n^2]$ and $[2n^2 \times n^2 + 1]$, respectively. The unique minimizer $\hat{\alpha}$ which satisfies (2.42) can be found without any minimization problem if the number of sampling exceeds $|\Gamma_0| \geq n^2 + 1$ since the system of equation in (2.42) becomes overcomplete and we just need to calculate the pseudo inverse of the matrix. In the case $|\Gamma_0| < dim\, \mathcal{N}(\mathbf{B}) = n^2 + 1$ the dimension of the kernel of the matrix in (3.18) will be non-zero, i.e., $dim\, \mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{\Theta}\right) > 0$ and any solution to the system of equation in (3.18) can also exist in the related kernel space where by applying the minimization problem in (3.7) there would be a probability that the solution will not be exact. When the number of samples is equal to the dimension of the kernel space of $\mathbf{B}$, i.e.,

$|\Gamma_0| = n^2 + 1$ then the matrix $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{\Theta}$ will be a $n^2 + 1$ squared matrix. However, there is a probability that this matrix will not be full rank if there will be non-empty intersection between the range space of $\mathbf{A}_{\mathbf{\Gamma_0}}$ and the range space of $\mathbf{B}^{\mathbf{T}}$. So the samples should be selected in a scheme that to avoid such correlation. Among the available samples $\Gamma_0 \subset \{1, 2, \ldots, 2n^2\}$ there is only one combination that prevents such intersection since the number of samples been taken $(n^2 + 1)$ and the dimension of the range space of $\mathbf{B}^{\mathbf{T}}$ completes the space, i.e.,

$$\underbrace{dim\, \mathcal{N}\left(\mathbf{B}\right)}_{|\Gamma_0|} + \underbrace{dim\, rng\left(\mathbf{B}^{\mathbf{T}}\right)}_{n^2-1} = 2n^2 \tag{3.19}$$

As we can conclude form Lemma 1, the total number of samples necessary for l_1-minimization problem in (3.7) is at most the half of the gradient samples in the image such that the primary constraint and the secondary constraint remain independent from

each other i.e.,

$$|\Gamma_0| < dim\, \mathcal{N}(\mathbf{B}). \tag{3.20}$$

3.3.1 The Problem of Redundant Measurements and Its Solutions

The second condition in Lemma 1 indicates that the primary and secondary constraints in (3.8) should not be linearly dependent, otherwise the problem of redundant will occur. This will impose an ambiguity to the problem of minimization in (3.7). To avoid such linear dependency, one can exclude the dependent rows of primary from secondary constraints or the other way around. This can be done by projecting each columns of $\mathbf{A_{\Gamma_0}}$ or $\mathbf{B^T}$ on the null-space of $\mathbf{B}$ or $\mathbf{A^T_{\Gamma_0}}$, respectively, to determine dependent columns. Algorithm 1 and 2 explain the procedure of excluding dependent rows for both cases.

Algorithm 1 Excluding dependent columns of $\mathbf{A_{\Gamma_0}}$ with $\mathbf{B^T}$

```
for k = 0 to |Γ0| do
    if ‖P_{N(B)}{col_k A_{Γ0}}‖ > ε then
        keep the k-th column of A_{Γ0}
    else
        exclude the k-th column of A_{Γ0}
    end if
end for
```

Algorithm 2 Excluding dependent columns of $\mathbf{B^T}$ with $\mathbf{A_{\Gamma_0}}$

```
for k = 0 to n^2 − 1 do
    if ‖P_{N(A^T_{Γ0})}{col_k B^T}‖ > ε then
        keep the k-th column of B^T
    else
        exclude the k-th column of B^T
    end if
end for
```

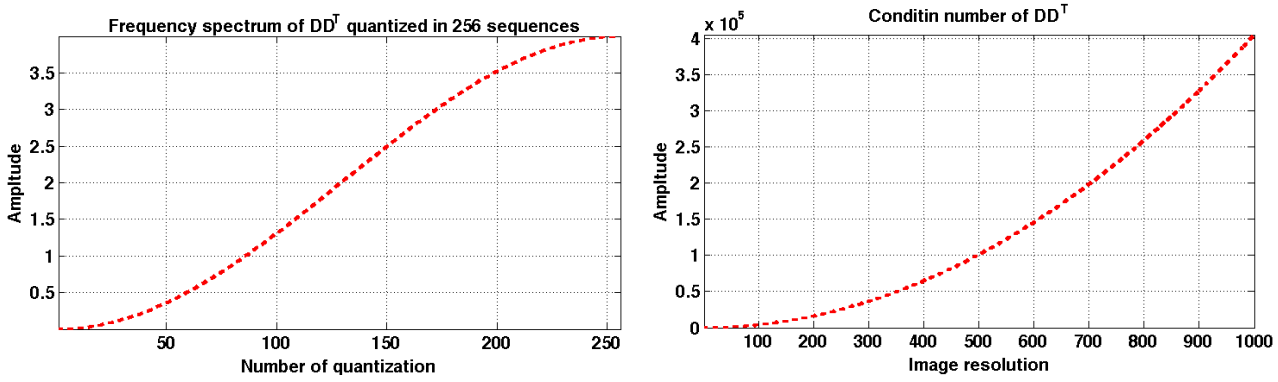
Since any solution to (3.8) exist in the kernel space of $\mathbf{B}$, the minimization problem in (3.7) finds the appropriate solution by projecting a vector to the the related null-space. This projection can be done by

$$\mathcal{P}_{\mathcal{N}(\mathbf{B})} = I - \mathbf{B^T}\left(\mathbf{BB^T}\right)^{-1}\mathbf{B} \tag{3.21}$$

Substituting $\mathbf{B}$ defined in (3.6) to $\mathbf{BB^T}$ in (3.21), yields

$$\begin{aligned}\mathbf{BB^T} &= \begin{bmatrix} I \otimes \mathrm{D} & -\mathrm{D} \otimes I \end{bmatrix} \underbrace{\begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix} \begin{bmatrix} \Psi^{\otimes^T} & 0 \\ 0 & \Psi^{\otimes^T} \end{bmatrix}}_{\mathbf{I}} \begin{bmatrix} I \otimes \mathrm{D}^T \\ -\mathrm{D}^T \otimes I \end{bmatrix} \\ &= \left[I \otimes \mathrm{DD}^T + \mathrm{DD}^T \otimes I \right] = \left[\mathrm{DD}^T \oplus \mathrm{DD}^T \right]_{(n^2-1)\times(n^2-1)},\end{aligned} \tag{3.22}$$

which ends up calculating the Kronecker sum of DD^T with itself. $\Psi^{\otimes}$ is an orthonormal basis and it turns to an identity when it multiplies to its transpose.The inverse of the Kronecker sum is equivalent to solve a sylvester equation, similar to what defined in (3.12) and (3.13). The term DD^T is diagonalizable in discrete Fourier domain using (3.11). The related diagonalized eigenvalues obeys the spectrum of $2-2\cos \pi t_n/n$ in n discrete sequences where it is sketched for 256×256 resolution in Figure 3.1(a). The condition number is calculated by $\lambda_{\max}/\lambda_{\min}$, where it increases in order of $\mathcal{O}(n^2)$ by increasing the discrete resolution n, see Figure 3.1(b).



(a) (b)

Figure 3.1: a) Diagonal spectrum $diag(\Lambda_{\mathrm{D}^T\mathrm{D}})$ of the matrix $\mathrm{D}^T\mathrm{D}$ in Fourier domain; b) Condition number of $\mathrm{D}^T\mathrm{D}$ for different image resolutions

Among the introduced algorithms for excluding dependent rows, one should be selected such that to make the system of equations in (3.8) well-conditioned. Since Algorithm 1 keeps all rows of $\mathbf{B}$, this algorithm will not be suitable because it imposes a high condition matrix to the system of equation in (3.8). But, Algorithm 2 excludes the corresponding dependent rows in $\mathbf{B}$ such that it increases the probability of being well-conditioned system of equations

in (3.8). To proceed with Algorithm 2, the rows of $\mathbf{B}$ can be projected to the null-space of $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}$ by applying

$$\mathcal{P}_{\mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)}\{\mathbf{B}^{\mathbf{T}}\} = \left[\mathbf{I} - \mathbf{A}_{\mathbf{\Gamma_0}}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{A}_{\mathbf{\Gamma_0}}\right)^{-\mathbf{1}}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right]\mathbf{B}^{\mathbf{T}}. \tag{3.23}$$

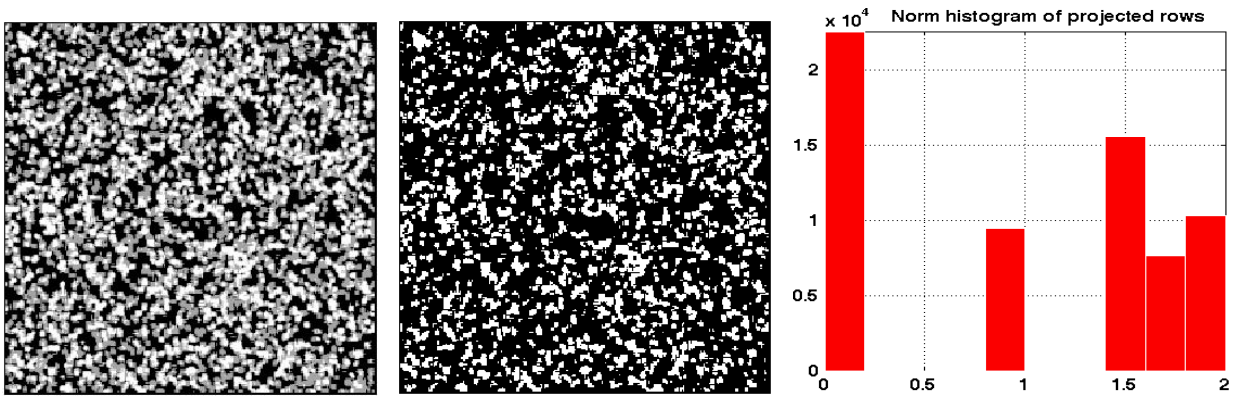
Substituting (3.8) to the term $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{A}_{\mathbf{\Gamma_0}}$ in (3.23), yields

$$\begin{aligned}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{A}_{\mathbf{\Gamma_0}} &= \begin{bmatrix}\Phi_x^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes}\end{bmatrix}\underbrace{\begin{bmatrix}\Psi^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes}\end{bmatrix}\begin{bmatrix}\Psi^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes T}\end{bmatrix}}_{\mathbf{I}}\begin{bmatrix}\Phi_x^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes T}\end{bmatrix}\\ &= \begin{bmatrix}\Phi_x^{\otimes}\Phi_x^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes}\Phi_y^{\otimes T}\end{bmatrix} = [\mathbf{I}]_{|\Gamma_0|\times|\Gamma_0|}.\end{aligned} \tag{3.24}$$

So, the inverse of $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{A}_{\mathbf{\Gamma_0}}$ is identity and (3.23) continues as follows,

$$\begin{aligned}\mathcal{P}_{\mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)}\{\mathbf{B}^{\mathbf{T}}\} &= \left[\mathbf{I} - \mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right]\mathbf{B}^{\mathbf{T}}\\ &= \left(I - \begin{bmatrix}\Psi^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes T}\end{bmatrix}\begin{bmatrix}\Phi_x^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes T}\end{bmatrix}\begin{bmatrix}\Phi_x^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes}\end{bmatrix}\begin{bmatrix}\Psi^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes}\end{bmatrix}\right)\mathbf{B}^{\mathbf{T}}\\ &= \begin{bmatrix}\Psi^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes T}\end{bmatrix}\left(I - \begin{bmatrix}\Phi_x^{\otimes T}\Phi_x^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Phi_y^{\otimes T}\Phi_y^{\otimes}\end{bmatrix}\right)\begin{bmatrix}\Psi^{\otimes} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes}\end{bmatrix}\mathbf{B}^{\mathbf{T}}\\ &= \begin{bmatrix}\Psi^{\otimes T} & \mathbf{0}\\ \mathbf{0} & \Psi^{\otimes T}\end{bmatrix} diag\left(\mathbf{1}_{\mathbf{\Gamma_0^c}}\right)\begin{bmatrix}I\otimes \mathrm{D}^T\\ -\mathrm{D}^T\otimes I\end{bmatrix}\end{aligned} \tag{3.25}$$

Here, $\mathbf{1}_{\mathbf{\Gamma_0^c}}$ indicates a vector with 1 values restricted to the complement indices of Γ_0. After projecting each rows of $\mathbf{B}$ to the related null-space, the norm of the produced vector provides a weight for each indices indicating the amount of the liner dependency. If the value is greater than a threshold then it will remain in the system of equation for minimization. Figure 3.2(a) shows l_2-norm value of the projected rows of $\mathbf{B}$ to the null-space of $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}$ for 256×256 pixel image. Each pixel value in the image corresponds to the row indices of $\mathbf{B}$. The primary samples observed here is 60% of the image gradients. The histogram of the related gray-value image is plotted in Figure 3.2(c). In order to define excluded samples, a threshold value should be determined to mask them out. Figure 3.2(c) demonstrates the thresholded image in Figure 3.2(a) by 1.7 value.



(a) (b) (c)

Figure 3.2: a) Norm of the projected rows of $\mathbf{B}$ to the null-space of $\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}$ shown in gray-value; b) Histogram of the norm of the projected rows

3.4 Generalization of DCS Problem

The problem of derivative compressive sampling, after excluding redundant measurements from cross-derivative constraints, can be generalized as follows,

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1, \tag{3.26}$$

subject to

$$\begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} \\ \mathbf{B}_{\mathbf{\Gamma_B}} \end{bmatrix} \mathbf{c} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix}$$

where, Γ_B is the indices corresponding to the remaining rows of $\mathbf{B}$ after excluding redundant measurements. So, both stacked matrices $\left[\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}; \mathbf{B}_{\mathbf{\Gamma_B}}\right]$ is a $(|\Gamma_0| + |\Gamma_B|)\times 2n^2$ full rank matrix. Subsequently, the sensing matrix $\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}$ and the sub-sampled cross-derivative matrix $\mathbf{B}_{\mathbf{\Gamma_B}}$ are defines as follows,

$$\begin{aligned} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} &= \begin{bmatrix} \Phi^{\otimes}_x & 0 \\ 0 & \Phi^{\otimes}_y \end{bmatrix} \begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix}; \\ \mathbf{B}_{\mathbf{\Gamma_B}} &= \Phi^{\otimes}_{\Gamma_B} \begin{bmatrix} I \otimes \mathrm{D} & -\mathrm{D} \otimes I \end{bmatrix} \begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix} \end{aligned} \tag{3.27}$$

where, $\Phi^{\otimes}$ and $\Psi^{\otimes}$ are the sampling and representation matrices, respectively, which are defined in 3.1. Reminding from Equation 2.65, any solution to the problem 3.26 is divided

to two parts, one, unique minimizer $\hat{\mathbf{c}}$ and the second is the error vector η, i.e., $\mathbf{c} = \hat{\mathbf{c}} + \eta$. Consequently, the error vector η is originated by the intersection of two kernel subspaces defined by sensing matrix and cross-derivative constraint. The intersection of both null-spaces in 3.28 is the smaller null-space of larger matrix stacked by $\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}$ and $\mathbf{B}_{\mathbf{\Gamma_B}}$ [69], i.e.,

$$\eta \in \left\{ \mathcal{N}\left(\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}\right) \cap \mathcal{N}\left(\mathbf{B}_{\mathbf{\Gamma_B}}\right) = \mathcal{N}\left(\begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} \\ \mathbf{B}_{\mathbf{\Gamma_B}} \end{bmatrix} \right) \right\}. \tag{3.28}$$

Projecting any vector $z \in \mathbb{R}^{2n^2}$ to the intersection of null-spaces can be carried out by excluding the range space of $\left[\mathbf{A}_{\mathbf{\Gamma_0}}, \mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}}\right]$ from the identity operator using

$$\eta = \mathbf{P}_n z = \left[\mathbf{I} - \begin{bmatrix} \mathbf{A}_{\mathbf{\Gamma_0}} & \mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} \end{bmatrix} \left(\begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} \\ \mathbf{B}_{\mathbf{\Gamma_B}} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{\mathbf{\Gamma_0}} & \mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} \\ \mathbf{B}_{\mathbf{\Gamma_B}} \end{bmatrix} \right] z. \tag{3.29}$$

The 2 cross 2 partitioned matrix in (3.29) can be redefined as,

$$\begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}} \\ \mathbf{B}_{\mathbf{\Gamma_B}} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{\mathbf{\Gamma_0}} & \mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}\mathbf{A}_{\mathbf{\Gamma_0}} & \mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}\mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} \\ \mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{A}_{\mathbf{\Gamma_0}} & \mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} \end{bmatrix}. \tag{3.30}$$

Each element of the partitioned matrix can be defined by the sampling and representation bases as follows,

$$\begin{aligned}
\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}\mathbf{A}_{\mathbf{\Gamma_0}} &= \begin{bmatrix} \Phi^{\otimes}_x & 0 \\ 0 & \Phi^{\otimes}_y \end{bmatrix} \begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix} \begin{bmatrix} \Psi^{\otimes^T} & 0 \\ 0 & \Psi^{\otimes^T} \end{bmatrix} \begin{bmatrix} \Phi^{\otimes^T}_x & 0 \\ 0 & \Phi^{\otimes^T}_y \end{bmatrix} = [\mathbf{I}]_{\mathbf{m}\times\mathbf{m}} \\
\mathbf{A}^{\mathbf{T}}_{\mathbf{\Gamma_0}}\mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} &= \begin{bmatrix} \Phi^{\otimes}_x & 0 \\ 0 & \Phi^{\otimes}_y \end{bmatrix} \begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix} \begin{bmatrix} \Psi^{\otimes^T} & 0 \\ 0 & \Psi^{\otimes^T} \end{bmatrix} \begin{bmatrix} I \otimes \mathrm{D}^T \\ -\mathrm{D}^T \otimes I \end{bmatrix} \mathbf{\Phi}^{\otimes^{\mathbf{T}}}_{\mathbf{\Gamma_B}} \\
&= \begin{bmatrix} \Phi^{\otimes^T}_x \left(I \otimes \mathrm{D}^T\right) \Phi^{\otimes^T}_{\Gamma_B} \\ -\Phi^{\otimes^T}_y \left(\mathrm{D}^T \otimes I\right) \Phi^{\otimes^T}_{\Gamma_B} \end{bmatrix}_{m\times|\Gamma_B|} \\
\mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{A}_{\mathbf{\Gamma_0}} &= \begin{bmatrix} \Phi^{\otimes}_{\Gamma_B} \left(I \otimes \mathrm{D}\right) \Phi^{\otimes}_x & -\Phi^{\otimes}_{\Gamma_B} \left(\mathrm{D} \otimes I\right) \Phi^{\otimes}_y \end{bmatrix}_{|\mathbf{\Gamma_B}|\times\mathbf{m}} \\
\mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{B}^{\mathbf{T}}_{\mathbf{\Gamma_B}} &= \mathbf{\Phi}^{\otimes}_{\mathbf{\Gamma_B}} \begin{bmatrix} I \otimes \mathrm{D} & -\mathrm{D} \otimes I \end{bmatrix} \begin{bmatrix} \Psi^{\otimes} & 0 \\ 0 & \Psi^{\otimes} \end{bmatrix} \begin{bmatrix} \Psi^{\otimes^T} & 0 \\ 0 & \Psi^{\otimes^T} \end{bmatrix} \begin{bmatrix} I \otimes \mathrm{D}^T \\ -\mathrm{D}^T \otimes I \end{bmatrix} \mathbf{\Phi}^{\otimes^{\mathbf{T}}}_{\mathbf{\Gamma_B}} \\
&= \Phi^{\otimes}_{\Gamma_B} \left[I \otimes \mathrm{D}\mathrm{D}^T + \mathrm{D}\mathrm{D}^T \otimes I\right] \Phi^{\otimes^T}_{\Gamma_B} = \Phi^{\otimes}_{\Gamma_B} \left[\mathrm{D}\mathrm{D}^T \oplus \mathrm{D}\mathrm{D}^T\right] \Phi^{\otimes^T}_{\Gamma_B}
\end{aligned} \tag{3.31}$$

The fact that the first row in Equation (3.31) turns to an identity is because $\Psi^{\otimes}$ is an orthonormal basis.

Since the pivot elements of the partitioned matrix in (3.30) are square matrices, the inverse of 2 cross 2 partitioned matrix can be calculated by [70],

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} (a - bd^{-1}c)^{-1} & -(a - bd^{-1}c)^{-1} bd^{-1} \\ -d^{-1}c\,(a - bd^{-1}c)^{-1} & d^{-1} + d^{-1}c\,(a - bd^{-1}c)^{-1} bd^{-1} \end{bmatrix} \tag{3.32}$$

where d and $(a - bd^{-1}c)$ are non-singular matrices. Substituting Equation (3.30) in (3.32) and proceeding with calculation of the null-space projection in (3.29) we have,

$$\begin{aligned}
\mathbf{P}_r &= \begin{bmatrix} \mathbf{A_{\Gamma_0}} & \mathbf{B^T_{\Gamma_B}} \end{bmatrix} \begin{bmatrix} \mathbf{A_{\Gamma_0}A^T_{\Gamma_0}} & \mathbf{A_{\Gamma_0}B^T_{\Gamma_B}} \\ \mathbf{B_{\Gamma_B}A_{\Gamma_0}} & \mathbf{B_{\Gamma_B}B^T_{\Gamma_B}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A^T_{\Gamma_0}} \\ \mathbf{B_{\Gamma_B}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A_{\Gamma_0}} & \mathbf{B^T_{\Gamma_B}} \end{bmatrix} \begin{bmatrix} \mathbf{M^{-1}} & -\mathbf{M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \\ -\left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}} & \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} + \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A^T_{\Gamma_0}} \\ \mathbf{B_{\Gamma_B}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A_{\Gamma_0}} & \mathbf{B^T_{\Gamma_B}} \end{bmatrix} \begin{bmatrix} \mathbf{M^{-1}A^T_{\Gamma_0}} - \mathbf{M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} \\ -\left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} + \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} + \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} \end{bmatrix} \\
&= \begin{array}{l} \mathbf{A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} - \mathbf{A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} - \mathbf{B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} \\ +\mathbf{B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} + \mathbf{B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}} \end{array}
\end{aligned} \tag{3.33}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{A^T_{\Gamma_0}P_BA_{\Gamma_0}}$. Assuming $\mathbf{P}_B = \mathbf{B^T_{\Gamma_B}} \left(\mathbf{B_{\Gamma_B}B^T_{\Gamma_B}}\right)^{-1} \mathbf{B_{\Gamma_B}}$, finally, the projection to the intersection of both null-spaces is defined by,

$$\begin{aligned}
\mathbf{P}_n &= \mathbf{I} - \mathbf{P}_r \\
&= \mathbf{I} - \mathbf{A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} + \mathbf{A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}P_B} + \mathbf{P_BA_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} - \mathbf{P_B} - \mathbf{P_BA_{\Gamma_0}M^{-1}A^T_{\Gamma_0}P_B} \\
&= \mathbf{I} - \mathbf{P}_B - \mathbf{A_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} (\mathbf{I} - \mathbf{P_B}) + \mathbf{P_BA_{\Gamma_0}M^{-1}A^T_{\Gamma_0}} (\mathbf{I} - \mathbf{P_B})
\end{aligned} \tag{3.34}$$

As the projection operator has the property $(\mathbf{I}-\mathbf{P}_B)^2=(\mathbf{I}-\mathbf{P}_B)$, Equation (3.34) can be proceed as follows,

$$\begin{aligned}\mathbf{P}_n &= (\mathbf{I}-\mathbf{P}_B)^2 - \mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}(\mathbf{I}-\mathbf{P}_{\mathbf{B}}) + \mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}(\mathbf{I}-\mathbf{P}_{\mathbf{B}}) \\ &= \left(\mathbf{I}-\mathbf{P}_B-\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}+\mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)(\mathbf{I}-\mathbf{P}_B) \\ &= \left(\mathbf{I}-\mathbf{P}_B-(\mathbf{I}-\mathbf{P}_B)\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)(\mathbf{I}-\mathbf{P}_B) \\ &= (\mathbf{I}-\mathbf{P}_B)\left(I-\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)(\mathbf{I}-\mathbf{P}_B)\end{aligned} \tag{3.35}$$

The above Equation (3.35) gives an analogy to the intersection of both null-spaces in (3.28). In comparison, both methods of compressive sampling and derivative compressive sampling can be generalized in Table 3.1.

Table 3.1: Generalizing the problem of derivative compressive sampling (DCS) compared to conventional compressed sensing (CS)

CS	DCS
$\{\hat{\mathbf{c}}\}=\arg\min_{\mathbf{c}\in\mathcal{F}}\|\mathbf{c}\|_1$	$\{\hat{\mathbf{c}}\}=\arg\min_{\mathbf{c}\in\mathcal{F}}\|\mathbf{c}\|_1$
$\mathcal{F}=\{\mathbf{c}\ :\ \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{c}=\mathbf{Y}\}$	$\mathcal{F}=\{\mathbf{c}\ :\ \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{c}=\mathbf{Y}\}\cap\mathcal{N}(\mathbf{B}_{\mathbf{\Gamma_B}})$
$\mathbf{c}=\hat{\mathbf{c}}+\eta,\ \ \text{where } \eta\in\mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)$	$\mathbf{c}=\hat{\mathbf{c}}+\eta,\ \ \text{where } \eta\in\mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)\cap\mathcal{N}(\mathbf{B}_{\mathbf{\Gamma_B}})$
$\eta=\left(\mathbf{I}-\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)z,\ z\in\mathbb{R}^{2n^2}$	$\eta=(\mathbf{I}-\mathbf{P}_B)\left(\mathbf{I}-\mathbf{A}_{\mathbf{\Gamma_0}}\mathbf{M}^{-1}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\right)(\mathbf{I}-\mathbf{P}_B)z,\ z\in\mathbb{R}^{2n^2}$
	$\mathbf{M}=\mathbf{I}-\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}}$
	$\mathbf{P}_B=\mathbf{B}_{\mathbf{\Gamma_B}}^{\mathbf{T}}\left(\mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{B}_{\mathbf{\Gamma_B}}^{\mathbf{T}}\right)^{-1}\mathbf{B}_{\mathbf{\Gamma_B}}$

The error of the recovery is controlled by the null-space of the sensing matrix in l_1-minimization algorithm. In the case of DCS, this null-space is limited to the intersection of the two null-spaces related to the sensing and cross derivative matrices. In order to have access to such error vector, η, in related space we need to calculate the inverse of $\mathbf{M}$ in (3.35). This is problematic in large-scale problem and it can not be used directly. This inverse can be approximated by the Neumann series formula if the norm of $\|\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}}\|<\mathbf{1}$. To show this, the results from Equation (3.31) can be substituted in $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}}$ where it yields,

$$\begin{aligned}\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{P}_{\mathbf{B}}\mathbf{A}_{\mathbf{\Gamma_0}} &= \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}\mathbf{B}_{\mathbf{\Gamma_B}}^{\mathbf{T}}\left(\mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{B}_{\mathbf{\Gamma_B}}^{\mathbf{T}}\right)^{-1}\mathbf{B}_{\mathbf{\Gamma_B}}\mathbf{A}_{\mathbf{\Gamma_0}} \\ &= \begin{bmatrix}\Phi_x^{\otimes} & 0\\ 0 & \Phi_y^{\otimes}\end{bmatrix}\mathbf{P}\begin{bmatrix}\Phi_x^{\otimes^T} & 0\\ 0 & \Phi_y^{\otimes^T}\end{bmatrix}\end{aligned} \tag{3.36}$$

where, $\mathbf{P}$ is a projection matrix to the rang space of $[\ (I \otimes \mathrm{D}) \quad -(\mathrm{D} \otimes I)\ ]^T$, i.e.,

$$\mathbf{P} = \Phi_{\Gamma_B}^{\otimes^T} \begin{bmatrix} I \otimes \mathrm{D}^T \\ -\mathrm{D}^T \otimes I \end{bmatrix} \left[\Phi_{\Gamma_B}^{\otimes} \left[\mathrm{D}\mathrm{D}^T \oplus \mathrm{D}\mathrm{D}^T \right] \Phi_{\Gamma_B}^{\otimes^T} \right]^{-1} \Phi_{\Gamma_B}^{\otimes} \begin{bmatrix} I \otimes \mathrm{D} & -\mathrm{D} \otimes I \end{bmatrix} \tag{3.37}$$

The norm of the projection matrix is $\|\mathbf{P}\| = \mathbf{1}$. The matrix $\Phi_x^{\otimes}$ and $\Phi_y^{\otimes}$ in (3.36) are the sub-sampled of the rows of identity matrix and multiplying both from right and left hand side will produce a sub-matrix of $\mathbf{P}$. The norm of a sub-matrix does not exceed the norm of the matrix so the norm of $\|\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{P}_{\mathbf{B}} \mathbf{A}_{\mathbf{\Gamma_0}}\| < \mathbf{1}$. The projection $\mathbf{P}_n$ matrix in (3.35) can now be approximated by the Neumann series approximation, i.e.,

$$\mathbf{P}_n = (\mathbf{I} - \mathbf{P}_B) \left(\mathbf{I} - \mathbf{A}_{\mathbf{\Gamma_0}} \sum_{\mathbf{i=0}}^{\infty} \left(\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \mathbf{P}_{\mathbf{B}} \mathbf{A}_{\mathbf{\Gamma_0}} \right)^{\mathbf{i}} \mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}} \right) (\mathbf{I} - \mathbf{P}_B) \tag{3.38}$$

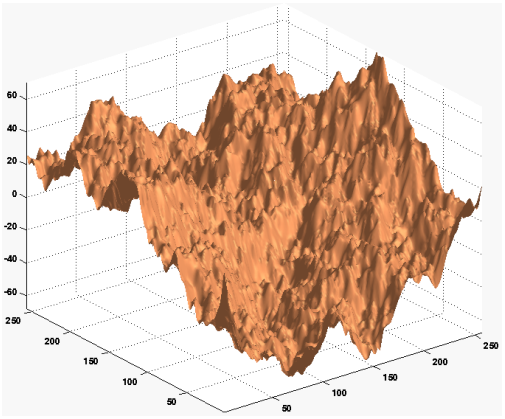
The above equation provides an analogy to access a vector member in the intersection of null-spaces for the case of derivative compressive sampling.

3.5 Phase Unwrapping by Means of DCS

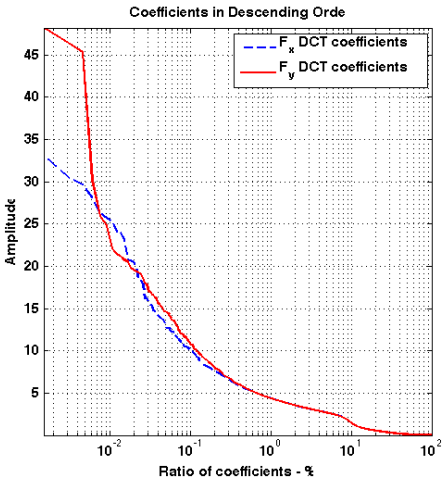
This section provides the preliminary results of the proposed solution to the problem of two-dimensional phase unwrapping. We demonstrate the performance of the proposed methodology using fractal landscapes (terrain) data. The synthetic fractal terrains are generated by midpoint displacement (diamond square algorithm). The terrain is created iteratively by performing diamond step for every level of the full terrain and applying square step afterwards. An example of the terrain data with 256×256 pixels and its wrapped version are shown in Figure 3.3(d) and (3.3), respectively.

3.5.1 Sparse Representation of Image Gradients

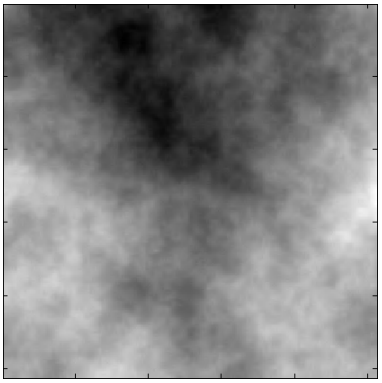
The compressibility of a signal implies that the latter is sparse in the domain of a linear transform. In the present case, the partial derivatives of a phase surface are assumed to be compressible in the domain of orthonormal bases Ψ. Finding an optimal sparse representation of the image gradients is of a great interest since it can reduce the necessary number of



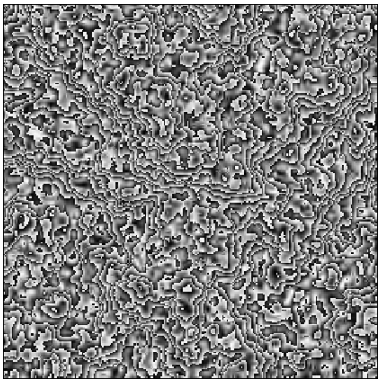
(a)



(b)



(c)



(d)

Figure 3.3: (a) Original phase in a "terrain view"; (b) DCT coefficients of the partial derivative of the phase sorted in descend order; (c) Original phase represented as grayscale image; (d) Its corresponding wrapped phase

samples for image recovery. This investigation is of future plan of this research, but nevertheless, as an initial step, we use discrete cosine transform (DCT) to approximately achieve the desired sparsity level. Figure 3.3(b) demonstrates the x and y derivative coefficients, where the images is approximately 10% sparse in DCT domain for both derivatives.

3.5.2 Data Classification using Quality Maps

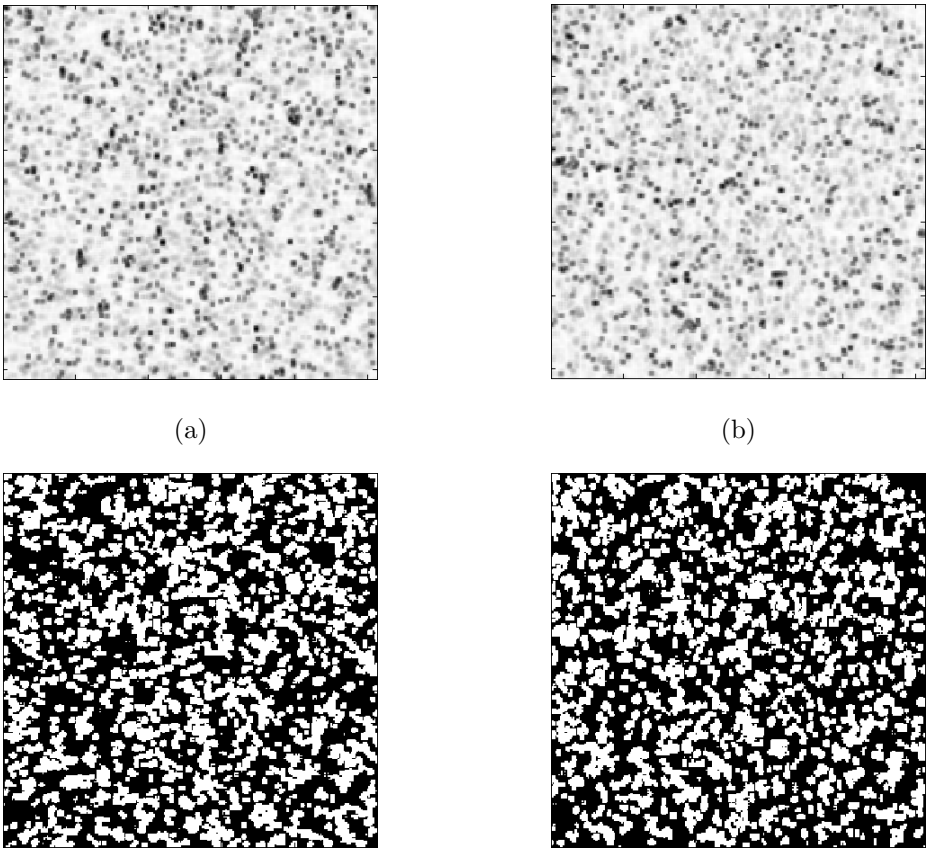
As it mentioned in Section 2.1.5, the quality maps can be used to indicate the quality of measured phase in each pixel. Phase gradient variance marks highly varying phases in terrain data (steep heights). Such pixels cause the integral in (2.19) to be path dependent. In DCS recovery, the sampling indices, $\Phi_x^{\otimes}$ and $\Phi_y^{\otimes}$ in (3.7), are analogues to the non-residual points in phase unwrapping. So, we can select derivative samples for each pixel in which good quality map is provided. We define two quality maps for x and y gradient domain by separating the gradient variance in (2.25) as follows,

$$\begin{aligned} Q^x_{m,n} &= \sqrt{\sum \left(\nabla_x \psi_{i,j} - \overline{\nabla_x \psi}_{m,n}\right)^2} \\ Q^y_{m,n} &= \sqrt{\sum \left(\nabla_y \psi_{i,j} - \overline{\nabla_y \psi}_{m,n}\right)^2} \end{aligned} \tag{3.39}$$

These two equations define the quality of measured phase, separately, in both derivative domains. Figure 3.4(a) and 3.4(b) demonstrates $Q^x_{m,n}$ and $Q^x_{m,n}$, respectively. Here, the gray-values increase from white to black and the sum is taken over 3 windows ($k = 3$). In order to specify the locations of reliable data samples we threshold the quality maps and generate a mask template, see Figure 3.4(c) and 3.4(d). These two masks, corresponding x and y phase gradient variances, combined together, produce the observed indices Γ_0 in (3.8). The threshold value here is applied manually of 0.13. This threshold value should be taken such that to prevent observing phase values contaminated by residual points.

3.5.3 Simulation Results

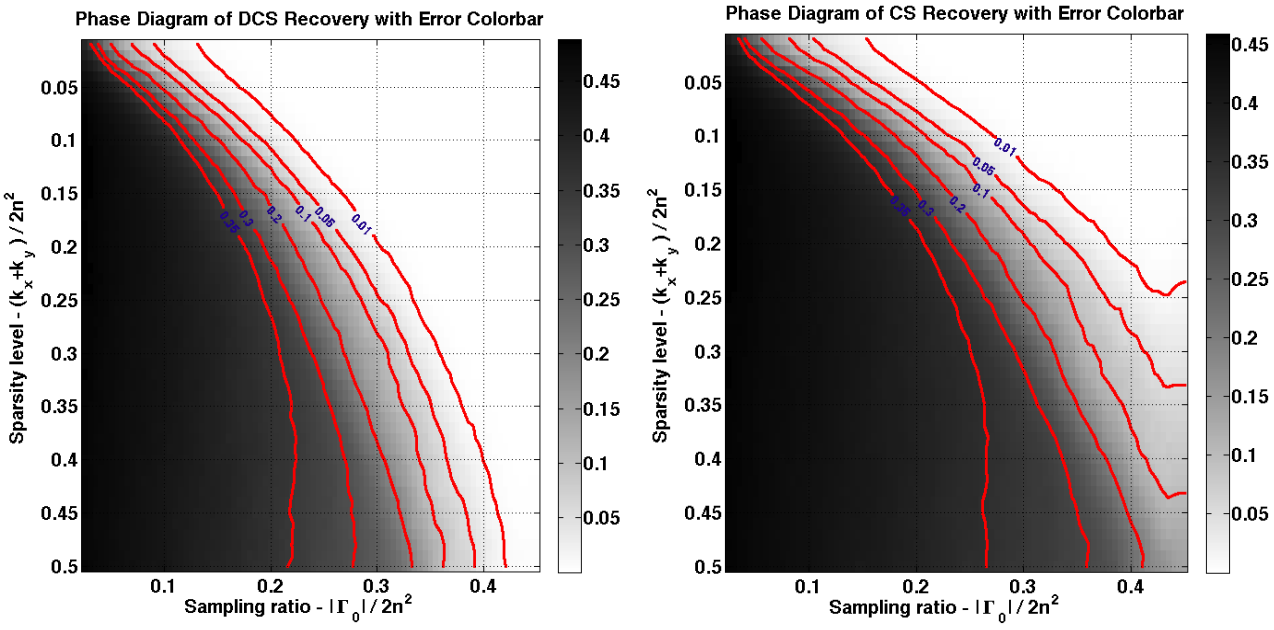
As stated previously, we use fractal terrain image in size of 256×256 pixel and it is approximately 10% sparse in gradient domain using Daubechies transform (Ψ) with 10 vanishing moments. We manually define the threshold value to binarize the quality map in (3.39) by



(a) (b)

(c) (d)

Figure 3.4: (a) Phase x-gradient VQM of wrapped phase R; (b) Phase y-gradient VQM of wrapped phase R; (c) x-gradient VQM after applying a threshold of 0.13 ; (d) y-gradient VQM after applying a threshold of 0.13.



(a) (b)

Figure 3.5: The error surface of phase reconstruction by : (a) DCS; (b) Standard CS; The surface are show as a function of visualized

extracting pixel coordinates which satisfies $\left|Q^x_{m,n}\right| > threshold$ and $\left|Q^y_{m,n}\right| > threshold$ for x and y indices, respectively. In order to have a small number of samples, a lower threshold should be applied to exclude more pixels that are likely to contain residual points.

We compared the performance of the DCS algorithm with that of the standard CS method. Figure 3.5(a) shows the phase diagram of error recovery through l_1-minimization problem using DCS method. This figure demonstrates a shaded surface contributing to the error of recovery with different sparsity levels on vertical axes and different sampling ratios in horizontal axes. Shaded attribute is the number of coordinate of reconstruction and displays its transition from perfect disagreement to perfect reconstruction. As it can be seen, the method exactly recovers the signal at 15% of sparsity level by reaching to its 30% of total number of samples. Since, the cross-derivative is provides as side information to the problem as an orthogonal complement of the sensing matrix, this method shows better performance on the recovery compared to the CS method, please see Figure 3.5(b).

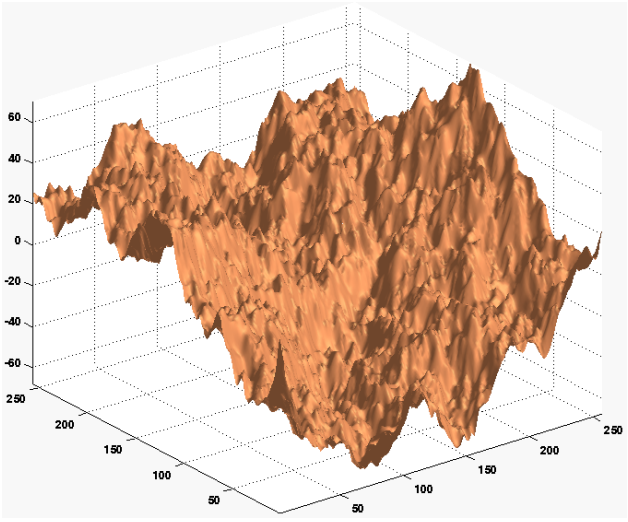
Figure 3.6 shows the recovery results of terrain height for both CS and DCS cases. Figure 3.6(c) demonstrates the unwrapped phase estimated using the DCS algorithm. The mean-

squared error (MSE) of the estimation was found to be 0.03%. As a comparison, the same solution was computed using the standard CS, whose MSE was found to be equal to 0.41% (see Figure 3.6(b)).

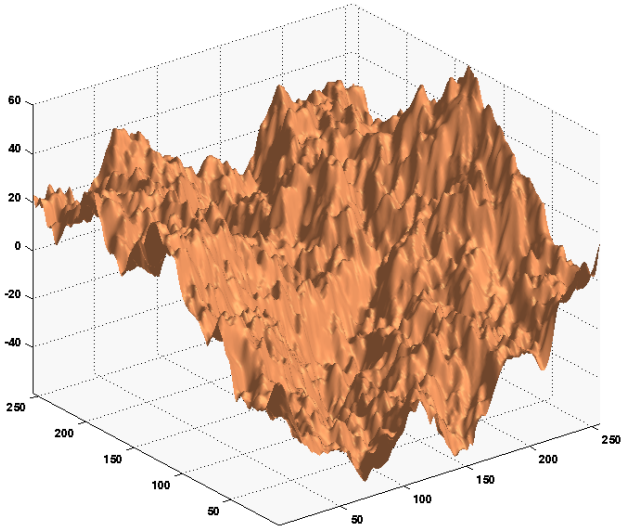
The performance of the proposed method is evaluated under the condition of noisy measurements which has been described in Equation 3.16 and shown in Figure 3.7.a. As it can be seen, the proposed DCS method performs better job by including secondary constraints and remain more stable compared to CS problem. Robustness of the proposed DCS has been also evaluated and compared with two distinctive phase unwrapping methods: Network-Flow [30] and PUMA [32, 71]. In addition, different number of residual points have been introduced in wrapped fractal terrains to outperform the robustness of phase unwrapping methods demonstrated in Figure 3.7.b. Each method is evaluated at its best performances on 20 different wrapped phases, where they have been recovered and averaged per-point. The number of residues is relaxed in lower rate where all of the methods recover the terrain with high accuracies. By increasing this number more samples are contaminated by residual points. DCS performs a better job in this case and remains more stable among the all implemented methods, where it retain its rest of the information from cross-derivative constraints. In the case of DCS, the number of sampling points cannot exceed the number that include residual points. This will effect the accuracy of recovered terrain and to prevent such ambiguity, the mask driven by thresholding quality map was defined by excluding the pixels which were found to be likely to violate the condition in (2.18).

3.6 Summary

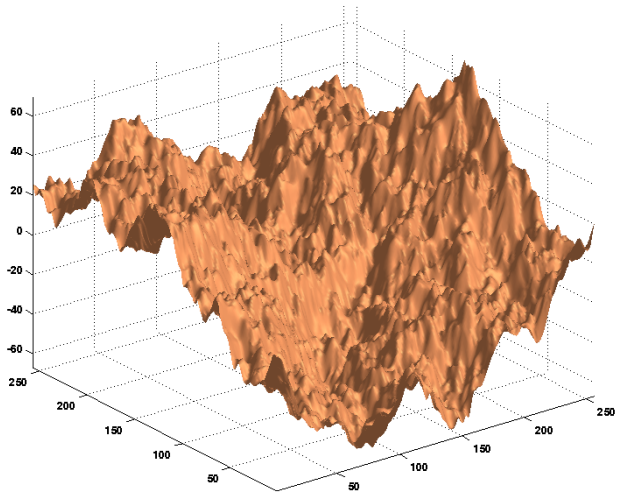
The main idea of the derivative compressive sampling is to reconstruct image gradients via its incomplete samples using compressive sampling scheme. The image gradients assumed to be sparse in an orthonormal basis Ψ. The sparsity of the image gradients is the key role to achieve higher accuracies in the presence of lower sampling rates. The second factor is introduce by cross-derivative equality which plays an important role here since it provides additional information (secondary constraint) beside the observed gradient samples (primary constraints). Since there is a correlation between these two constraints, primary and sec-



(a)

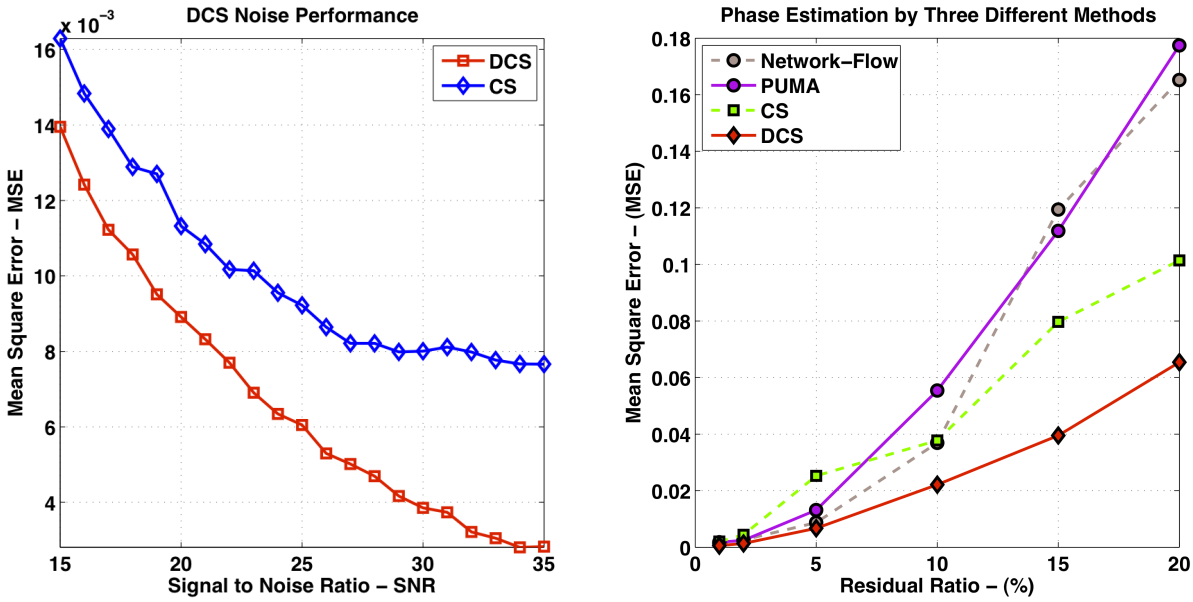


(b)



(c)

Figure 3.6: (a) Original Terrain; (b) estimate obtained by the standard CS (c) estimate obtained by the proposed DCS method.



(a) (b)

Figure 3.7: (a) MSE of phase estimation by DCS method in the presence of introduced noisy measurements varying from 15 db - 35 db. Here, the ratio of random sampling is 38.83% and the method is evaluated and averaged on 20 different noisy scheme for each SNR value; (b) MSE of phase estimation by three different methods: Network-Flow, PUMA, CS and DCS. The total percentage of residual points in measured wrapped phases increases from 1% - 20%

ondary, the linear dependency is removed to avoid redundant measurements. This is done by projecting the range space of two operators, $\mathbf{A}_{\mathbf{\Gamma_0}}^{\mathbf{T}}$ and $\mathbf{B}^{\mathbf{T}}$, to the null-space of each other, respectively, and exclude dependent vectors. The secondary constraint $\mathbf{B}$ contain derivative operator which makes it ill-conditioned matrix and can affect the l_1-minimization algorithm. So, the dependent rows in $\mathbf{B}$ is excluded to make the system of equation in (3.8) well-posed.

This side information from secondary constraint is interpreted to the recovery procedure to reduce the error of the estimation. This is because the searching space of the proposed l_1-minimization problem in (3.8) is limited to the intersection of two subspaces introduced by the kernel space of $\mathbf{B}$ and the range space of $\mathbf{A}_{\mathbf{\Gamma_0}}$. However, it should be defined whether this limitation contains excessively sparse vectors, otherwise it can harm the recovery and increase the error bound. The experimental results grantees such enhancement, but nevertheless it is a future plan of this research to investigate on the stability and recoverability analysis of the proposed DCS method. Consequently, the error vector produced in l_1-minimization belongs to the intersection of both primary and secondary's kernel spaces. This subspace has lower perturbation compared to CS problem.

The proposed DCS method provides different solution to the problem of phase unwrapping compared to what exist in the literature which is mostly dominated by path-follow and l_p-norm methods. Contaminated pixels by residual points in wrapped phase affects the unwrapping procedure where DCS method show better performance and reliability than the other methods, e.g. PUMA and Network-Flow.

Chapter 4

Conclusion and Future Plan

4.1 Introduction

To this point in this proposal we have done the following,

- Description of the DCS scheme and implementation issues
- Redundant measurements and excluding procedure
- Preliminary results of DCS on phase unwrapping and comparison of the results with PUMA and Network-Flow algorithms. Simulated terrain data were used as the framework to demonstrate and compare the output results.

4.2 Proposed Research Plan

The proposed research plan can be divided into three main categories,

1. The application of DCS on phase unwrapping should be optimized. The threshold level should be automatically defined instead in order to find pixel candidates for residual points in the wrapped image. The next criteria is to optimize the procedure used to find independent primary constraint from cross-derivative matrix $\mathbf{B}$. Finally, discovering different applications for DCS.

2. Since we used DCT transform for sparse representation of the derivatives in the domain, this framework could achieve the sparsity at an approximate level. Approximation of the sparsity will bring ambiguity to the error rate and will effect the recovery procedure. Optimal sparse representation of the image gradients can be found through Curvelet transforms [72], KSVD algorithms [73] or orthogonal matching pursuits [74].

3. The main principle of derivative compressive sampling (DCS) is the cross-derivative information provided as additional constraint to the problem of usual compressive sampling. The latter can be analyzed through stability and recoverability analysis to exemplify the improvement of the error recovery, capability of handling less sparse signals or need for lower rate of sampling. Two concepts in the literature are available to show such development by means of restricted isometry property (RIP) condition and distortion of kernel subspaces. The source of the recovery error in DCS is limited to the intersection of two kernel spaces introduced by $\mathcal{N}(\mathbf{B}) \cap \mathcal{N}\left(\mathbf{A}_{\mathbf{\Gamma_o}}^{\mathbf{T}}\right)$. Distortion of this subspace can provide an analogy that how much it contains sparse vectors. In fact by adding cross-derivative constraints this value is expected to decrease. All of the related analysis should be analytically expressed and derived.

4. More general question is how to interpret any other side information to the problem of the compressive sampling. Many criteria can be considered,

   - The signal is bandlimitted due to the communication channel is used in transmission, i.e., $\mathbf{c} \in \mathbf{B_2}(\mathbf{\Omega})$ where Ω here denotes the frequency limit of the channel.
   - In the case when we are provided with information that the steep of the terrain in DCS recovery is monotonically increasing or decreasing, then this means that the derivative coefficients are positive ($\mathbf{c} > \mathbf{0}$) or negative ($\mathbf{c} < \mathbf{0}$), respectively.

4.3 summary of Contributions and Publications

The preliminary research has resulted in a paper in the 17th European signal processing conference (EUSIPCO 2009) [75] which has been rated amongst the top 5% in 550 presented

papers. The results are invited to submit a manuscript in EURASIP journal SIGNAL PROCESSING at Elsevier. Most recent research results in Chapter 3, Sections 3.5.3 and 3.3, are going to be submitted as a journal paper to Transaction on Image Processing (TIP).

Bibliography

[1] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE TRANS. INFORM. THEORY*, vol. 52, no. 1, pp. 6–18, 2006.

[2] E. J. Cands, "Compressive sampling," in *International Congress of Mathematicians, Madrid, Spain, 2006*, vol. 3, pp. 1433–1452, August 22-30 2006.

[3] R. A. DeVore, "Optimal computation," in *International Congress of Mathematicians, Madrid, Spain, 2006*, vol. 1, pp. 187–215, August 22-30 2006.

[4] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, pp. 489–509, Feb. 2006.

[5] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *Information Theory, IEEE Transactions on*, vol. 52, pp. 5406–5425, Dec. 2006.

[6] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comput. Math.*, vol. 6, no. 2, pp. 227–254, 2006.

[7] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, June 2007.

[8] E. J. Cands and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, pp. 4203–4215, Dec. 2005.

[9] E. J. Cands, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, June 2006.

[10] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, pp. 4203–4215, Dec. 2005.

[11] A. Cohen, W. Dahmen, and R. Devore, "Compressed sensing and best k-term approximation," tech. rep., 2006.

[12] O. Monga and S. Benayoun, "Using partial derivatives of 3d images to extract typical surface features," *Comput. Vis. Image Underst.*, vol. 61, no. 2, pp. 171–189, 1995.

[13] N. Petrovic, I. Cohen, B. Frey, R. Koetter, and T. Huang, "Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–743–I–748 vol.1, 2001.

[14] J. Jackson, A. Yezzi, and S. Soatto, "Joint priors for variational shape and appearance modeling," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–7, June 2007.

[15] J. Courchay, J.-P. Pons, R. Keriven, and P. Monasse, "Dense and Accurate Spatio-Temporal Multi-View Stereovision IMAGINE Research Report 09-43." 07 2009.

[16] O. Michailovich and A. Tannenbaum, "Fast approximation of smooth functions from samples of partial derivatives with applications to phase unwrapping," *Signal Processing*, vol. 88, pp. 358–374, Aug. 2008.

[17] O. Michailovich and D. Adam, "Phase unwrapping for 2-d blind deconvolution of ultrasound images," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 7–25, Jan. 2004.

[18] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," pp. 513–531, 1989.

[19] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*. New York, USA: Wiley-Interscience, 1991.

[20] D. Massonnet and K. L. Feigl, "Radar interferometry and its application to changes in the earth's surface," *Rev. Geophys.*, vol. 36, no. 4, pp. 441–500, 1998.

[21] D. C. Ghiglia and M. D. Pritt, *Two-dimensional phase unwrapping: Theory, algorithms, and software*. New York, USA: Wiley-Interscience, 1998.

[22] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.*, vol. 21, no. 14, pp. 2470–2470, 1982.

[23] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, pp. 713–720, 1988.

[24] J. M. Huntley, "Noise-immune phase unwrapping algorithm," *Appl. Opt.*, vol. 28, no. 16, pp. 3268–3270, 1989.

[25] R. Cusack, J. M. Huntley, and H. T. Goldrein, "Improved noise-immune phase-unwrapping algorithm," *Appl. Opt.*, vol. 34, no. 5, pp. 781–789, 1995.

[26] D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Am.*, vol. 67, no. 3, pp. 370–375, 1977.

[27] R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *J. Opt. Soc. Am.*, vol. 67, no. 3, pp. 375–378, 1977.

[28] G. Fornaro, G. Franceschetti, R. Lanari, and E. Sansosti, "Robust phase-unwrapping techniques: a comparison," *J. Opt. Soc. Am. A*, vol. 13, no. 12, pp. 2355–2366, 1996.

[29] T. J. Flynn, "Two-dimensional phase unwrapping with minimum weighted discontinuity," *J. Opt. Soc. Am. A*, vol. 14, no. 10, pp. 2692–2701, 1997.

[30] M. Costantini, "A novel phase unwrapping method based on network programming," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 36, pp. 813–821, May 1998.

[31] C. Chen, *Statistical-Cost Network-Flow Approaches to Two-Dimensional Phase Unwrapping for Radar Interferometry*. PhD thesis, Stanford University, 2001.

[32] J. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," *Image Processing, IEEE Transactions on*, vol. 16, pp. 698–709, March 2007.

[33] J. Bioucas-Dias, V. Katkovnik, J. Astola, and K. Egiazarian, "Absolute phase estimation: adaptive local denoising and global unwrapping," *Appl. Opt.*, vol. 47, no. 29, pp. 5358–5369, 2008.

[34] G. H. Glover and E. Schneider, "Three-point dixon technique for true water/fat decomposition with B_0 inhomogeneity correction," *Magnetic Resonance in Medicine*, vol. 18, no. 2, pp. 371–383, 1991.

[35] J. Zhong and J. Weng, "Spatial carrier-fringe pattern analysis by means of wavelet transform: Wavelet transform profilometry," *Appl. Opt.*, vol. 43, no. 26, pp. 4993–4998, 2004.

[36] H. C. Hendargo, M. Zhao, N. Shepherd, and J. A. Izatt, "Synthetic wavelength based phase unwrapping in spectral domain optical coherence tomography," *Opt. Express*, vol. 17, no. 7, pp. 5039–5051, 2009.

[37] B. L. Burns and J. T. Cordaro, "Sar image-formation algorithm that compensates for the spatially variant effects of antenna motion," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (D. A. Giglio, ed.), vol. 2230, pp. 14–24, June 1994.

[38] H. Rott, "Advances in interferometric synthetic aperture radar (InSAR) in earth system science," *Progress in Physical Geography*, vol. 33, no. 6, pp. 769–791, 2009.

[39] A. V. Oppenheim and R. W. Schafer, *Discrete time signal processing*. London, UK: Prentice Hall, 1989.

[40] V. J. Katz, "The history of stokes' theorem," *Mathematics Magazine*, vol. 52, no. 3, pp. 146–156, 1979.

[41] D. Derauw, "Phase unwrapping using coherence measurements," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 2584, pp. 319–324, nov 1995.

[42] T. Flynn, "Consistent 2-d phase unwrapping guided by a quality map," in *Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future.', International*, vol. 4, pp. 2057–2059, May 1996.

[43] M. D. Pritt, "Phase unwrapping by means of multigrid techniques for interferometric sar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 728–738, may 1996.

[44] D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *J. Opt. Soc. Am. A*, vol. 11, no. 1, pp. 107–117, 1994.

[45] C. W. Chen and H. A. Zebker, "Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms," *J. Opt. Soc. Am. A*, vol. 17, no. 3, pp. 401–414, 2000.

[46] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[47] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, pp. 2845–2862, Nov 2001.

[48] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, pp. 1289–1306, April 2006.

[49] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 797–829, April 2006.

[50] B. S. Kašin, "The widths of certain finite-dimensional sets and classes of smooth functions," *Izv. Akad. Nauk SSSR Ser. Mat.*, vol. 41, no. 2, pp. 334–351, 478, 1977.

[51] A. Y. Garnaev and E. D. Gluskin, "The widths of a Euclidean ball," *Dokl. Akad. Nauk SSSR*, vol. 277, no. 5, pp. 1048–1052, 1984.

[52] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, pp. 589–592, May 2008.

[53] R. A. Devore, "Deterministic constructions of compressed sensing matrices," tech. rep., List of References November 2001 B.1 Complete IDL Listing OMG IDL B // File: CosActivity, 2007.

[54] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx*, vol. 2008, 2007.

[55] J. Blanchard, C. Cartis, and J. Tanner, "Decay properties of restricted isometry constants," *Signal Processing Letters, IEEE*, vol. 16, pp. 572–575, July 2009.

[56] J. Haupt and R. Nowak, "A generalized restricted isometry property," tech. rep., Department of Electrical and Computer Engineering, University of Wisconsin Madison, 2007.

[57] V. Chandar, "A negative result concerning explicit matrices with the restricted isometry property," 2008.

[58] Y. Zhang, "Theory of compressive sensing via l_1-minimization: A non-rip analysis and extensions," tech. rep., Department of computational and Applied Mathematics, Rice Univesity, 2008.

[59] B. Kashin and V. Temlyakov, "A remark on compressed sensing," *Mathematical Notes*, vol. 82, pp. 748–755, May 2007.

[60] S. A. Vavasis, "Derivation of compressive sensing theorems from the spherical section property," tech. rep., Department of Combinatorics and Optimization, University of Waterloo, 2009.

[61] T. Figiel, J. Lindenstrauss, and V. Milman, "The dimension of almost spherical sections of convex bodies," *Acta Mathematica*, vol. 139, no. 1, pp. 53–94, 1977.

[62] V. Guruswami, J. R. Lee, and A. Razborov, "Almost euclidean subspaces of ℓn1 via expander codes," in *SODA '08: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, (Philadelphia, PA, USA), pp. 353–362, Society for Industrial and Applied Mathematics, 2008.

[63] P. Indyk, "Uncertainty principles, extractors, and explicit embeddings of l2 into l1," in *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 615–620, ACM, 2007.

[64] S. Artstein-Avidan and V. Milman, "Logarithmic reduction of the level of randomness in some probabilistic geometric constructions," *Journal of Functional Analysis*, vol. 235, no. 1, pp. 297 – 329, 2006.

[65] E. v. Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," Tech. Rep. TR-2008-01, Department of Computer Science, University of British Columbia, Vancouver, January 2008. To appear in *SIAM J. Sci. Comp.*

[66] M. Harker and P. O'Leary, "Least squares surface reconstruction from measured gradient fields," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–7, June 2008.

[67] H. Ren and R. Dekany, "Fast wave-front reconstruction by solving the sylvester equation with the alternating direction implicit method," *Opt. Express*, vol. 12, no. 14, pp. 3279–3296, 2004.

[68] B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *J. Opt. Soc. Am.*, vol. 69, no. 3, pp. 393–399, 1979.

[69] G. Strang, *Introduction to Linear Algebra, Fourth Edition*. Wellesley-Cambridge Press, 2009.

[70] H. Lutkepohl, *Handbook of Matrices*. New York: John wiley & Sons, 1996.

[71] J. M. Dias and J. M. Leitao, "The zπm algorithm for interferometric image reconstruction in sar/sas," *IEEE Transactions on Image Processing*, vol. 11, 2002.

[72] J.-L. Starck, F. Murtagh, E. Candes, and D. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *Image Processing, IEEE Transactions on*, vol. 12, pp. 706–717, June 2003.

[73] M. Aharon, M. Elad, and A. Bruckstein, "K -svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 4311–4322, Nov. 2006.

[74] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," tech. rep., 2006.

[75] M. S. Hosseini and O. V. Michailovich, "Derivative compressive sampling with application to phase unwrapping," in *Proceed. of the 17th European Signal Processing Conference (EUSIPCO)*, August 2009.