Compressed Optical Imaging

Aurélien Bourquard

Thèse N° 5908 (décembre 2013)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Pierre Vandergheynst, *président*
Prof. Michael Unser, *directeur de thèse*
Prof. Jalal Fadili, *rapporteur*
Prof. Jean-Christophe Olivo-Marin, *rapporteur*
Prof. Dimitri Van De Ville, *rapporteur*

École polytechnique fédérale de Lausanne—2013

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with LaTeX

Available at **http://bigwww.epfl.ch/**

Abstract

We address the resolution of inverse problems where visual data must be recovered from incomplete information optically acquired in the spatial domain. The optical acquisition models that are involved share a common mathematical structure consisting of a linear operator followed by optional pointwise nonlinearities. The linear operator generally includes lowpass filtering effects and, in some cases, downsampling. Both tend to make the problems ill-posed. Our general resolution strategy is to rely on variational principles, which allows for a tight control on the objective or perceptual quality of the reconstructed data. The three related problems that we investigate and propose to solve are

1. *The reconstruction of images from sparse samples.* Following a non-ideal acquisition framework, the measurements take the form of spatial-domain samples whose locations are specified a priori. The reconstruction algorithm that we propose is linked to PDE flows with tensor-valued diffusivities. We demonstrate through several experiments that our approach preserves finer visual features than standard interpolation techniques do, especially at very low sampling rates.

2. *The reconstruction of images from binary measurements.* The acquisition model that we consider relies on optical principles and fits in a compressed-sensing framework. We develop a reconstruction algorithm that allows us to recover grayscale images from the available binary data. It substantially improves upon the state of the art in terms of quality and computational performance. Our overall approach is physically relevant; moreover, it can handle large amounts of data efficiently.

3. *The reconstruction of phase and amplitude profiles from single digital holographic acquisitions.* Unlike conventional approaches that are based on demodulation, our iterative reconstruction method is able to accurately recover the original object from a single downsampled intensity hologram, as shown in simulated and real measurement settings. It also consistently outperforms the state of the art in terms of signal-to-noise ratio and with respect to the size of the field of view.

The common goal of the proposed reconstruction methods is to yield an accurate estimate of the original data from all available measurements. In accordance with the forward model, they are typically capable of handling samples that are sparse in the spatial domain and/or distorted due to pointwise nonlinear effects, as demonstrated in our experiments.

Keywords—Compressed sensing, digital holography, generalized sampling theory, image interpolation, image reconstruction, inverse problems, iterative algorithms, multigrid optimization, partial differential equations, phase retrieval, quantization, regularization, variational methods.

Résumé

Nous nous intéressons à la résolution de problèmes inverses où de l'information visuelle doit être reconstruite à partir de mesures incomplètes effectuées dans le domaine spatial à l'aide d'un système optique. Les modèles mathématiques correspondant à l'acquisition des mesures ont une structure commune qui consiste en un opérateur linéaire parfois suivi de nonlinéarités agissant séparément en chaque point. L'opérateur linéaire inclut généralement un effet de filtrage passe-bas ainsi que de sous-échantillonage dans certains cas ; ces deux phénomènes tendent à rendre les problèmes mal posés. Notre stratégie globale de résolution est fondée sur des principes variationnels, ce qui permet un contrôle relativement fin de la qualité objective ou perceptuelle des données à reconstruire. Les trois problèmes correspondants que nous proposons d'étudier et de résoudre sont

1. *La reconstruction d'images à partir d'un faible nombre d'échantillons.* Dans le cadre d'un paradigme d'acquisition non-idéal, les mesures correspondent à des échantillons dont les positions sont pré-déterminées dans le domaine spatial. L'algorithme de reconstruction que nous proposons est lié à des équations aux dérivées partielles, celles-ci étant elles-mêmes associées à des flux dont les paramètres de diffusivité sont tensoriels. Nous montrons à travers plusieurs expériences que notre approche permet de présever des détails visuels plus fins comparé aux méthodes d'interpolation usuelles, en particulier dans des régimes où le taux d'échantillonage est très faible.

2. *La reconstruction d'images à partir de mesures binaires.* Tout en se basant sur des principes optiques, le modèle d'acquisition qui nous considérons s'inscrit dans la ligne de ce que l'on appelle le ≪compressed sensing≫. Nous

développons un algorithme permettant de reconstruire des images en niveaux de gris à partir des mesures binaires disponibles. Cet algorithme constitue une amélioration significative par rapport à l'état de l'art, aussi bien en terme de qualité que de performance de calcul. Notre approche globale d'acquisition et de reconstruction est ainsi pertinente sur le plan physique, tout en étant également capable de gérer d'importants volumes de données.

3. *La reconstruction de profils de phase et d'amplitude à partir d'une seule acquisition en holographie digitale.* Contrairement aux approches conventionnelles basées sur la démodulation, notre méthode de reconstruction itérative permet une estimation précise de l'objet initial à partir d'un seul hologramme sous-échantillonné, comme cela est démontré par des simulations et des expériences portant sur des données réelles. Notre méthode permet également d'augmenter sensiblement le rapport signal sur bruit ainsi que la taille du champ de vision effectif par rapport à l'état de l'art.

L'objectif commun des méthodes de reconstruction que nous proposons est de fournir une estimée fiable des données initiales à partir de toutes les mesures disponibles. Comme cela est démontré expérimentalement, ces méthodes sont capables de prendre en compte des configurations où les échantillons disponibles sont parcimonieux dans le domaine spatial et/ou altérés par des effets non-linaires agissant point par point, en accord avec le modèle d'acquisition.

Mots clés—Acquisition comprimée, algorithmes itératifs, équations différentielles partielles, holographie digitale, interpolation d'images, méthodes variationnelles, optimisation multiéchelle, problèmes inverses, quantification, reconstruction d'images, reconstruction de phase, régularisation, théorie de l'échantillonage généralisée.

To Sophie, to Joël, and to my parents

Acknowledgement

First of all, I would like to thank my thesis supervisor, Prof. Michael Unser, for his guidance throughout all these years. His invaluable insight, his inexhaustible energy, and his contagious passion for research are unique. These qualities make him not only a great scientist, but also, and most importantly in my opinion, a great collaborator. I can only hope to continue doing research in such an atmosphere of creativity, team spirit, and conviviality in the future.

I then wish to thank Dr. Philippe Thévenaz, our first research assistant who is also an unforgettable jogging partner, for his collaboration and for his constant support and advices on scientific article writing. I owe my taste for well written papers to him. My cordial thanks also go to Dr. Daniel Sage, whose kindness and helpfulness on any difficult software issue are unequaled, and to our secretary, Manuelle Mary, who, besides being very sociable, has always been ready to provide us all the necessary assistance in administrative matters.

I also thank my direct collaborators and former advisors who are currently working or have worked at the Biomedical Imaging Group, namely, Dr. Arash Amini, Dr. Cédric Vonesh, Emrah Bostan, Dr. François Aguet, Dr. Hagai Kirshner, Dr. John Paul Ward, Julien Fagot, Katarina Balać, Ramtin Madani, Dr. Sathish Ramani, Dr. Stamatis Lefkimmiatis, and, last but not least, Ulugbek Kamilov, whose professionalism is only equaled by his refined taste for scotch. My thoughts also go to all my other colleagues and friends from this lab who are not mentioned here.

Besides, I am very much indebted to Profs. Kunal N. Chaudhury and Pierre Vandergheynst for the fruitful discussions on PDE-based interpolation, to Arnaud Mader, Volker Zagolla, Dr. Eric Tremblay, and Profs. Dimitri Van de Ville, Chandra Sekhar Seelamantula, and Christophe Moser for their help and suggestions regarding the behavior of optical devices, and to Dr. Nicolas Pavillon and Prof.

Christian Depeursinge for their collaboration on our project related to digital holographic reconstruction.

This thesis would not have been possible without the support of our industrial partners Christian Bovet, Jean-Paul Cano, Jean-Marc Fournier, and Anthony Saugey. I am greatly indebted to them for their openness to new ideas, for their practical insight, and for their kind guidance throughout all these investigations. In that regard, my sincere thanks and appreciation also go to Essilor International. Finally, I would like to express my deepest gratitude and thanks to my family and my friends. They have always been there to encourage me and inspire me.

Contents

Chapter 1

Introduction

1.1 Preamble

Over the last decades, the exponential growth of computing power predicted by Moore's law [1] has allowed to innovate in the way image data are acquired. As early as in 1975, for instance, a computer implementation of the fast Fourier-transform (FFT) algorithm [2] enabled researchers to devise the first practical imaging modality based on nuclear magnetic resonance (NMR) signals [3, 4]. From a general perspective, the availability of digital processing gave rise to a novel imaging paradigm where data can be first captured in non-trivial form to be subsequently visualized through numerical reconstruction.

The complexity of the numerical-reconstruction process essentially depends on the type of acquisition setting, on the amount of data, and on the desired reconstruction accuracy. Based on a mathematical description of the acquisition procedure under consideration, the simplest algorithms typically aim at recovering the original data through one single application of the adjoint (or of an approximate inverse) of the forward model on the available measurements. For instance, NMR reconstruction is efficiently performed through an inverse FFT, while tomographic reconstruction involves filtered back-projection [5].

In contrast to direct approaches, most of the recent algorithms used in imaging are iterative and use prior knowledge on the solution to improve the quality of the

reconstruction. These more sophisticated techniques are especially useful when the data are noisy or incomplete, which is usually the case in practical applications [6]. Although iterative algorithms are computationally more costly than direct methods, the currently available computational power allows to treat large amounts of data efficiently or even in real time following this approach [7]. As explained below, the work of this thesis involves the development of iterative approaches for the resolution of several imaging problems.

1.2 Organization of the thesis

This thesis addresses three distinct problems where visual data have to be recovered from incomplete information acquired in the spatial domain. As discussed in Section 1.3, the acquisition models that are involved share the same overall mathematical structure. These problems thus involve similar concepts and are also associated with common characteristics and behaviors. Variational approaches will be proposed for their resolution.

1. In Chapter 2, we introduce a novel high-quality method to interpolate images from sparse samples [8]. Following a generalized sampling framework [9], the measurements consist in spatial-domain samples whose locations are specified a priori. The proposed reconstruction algorithm is based on variational principles and is linked to partial-differential-equation (PDE) flows with tensor-valued diffusivities. Several experiments demonstrate that this interpolation approach preserves finer visual features compared to the state of the art, especially at very low sampling rates.

2. In Chapter 3, we deal with the development of a method to acquire and reconstruct images through the sole use of binary measurements [10, 11]. The proposed acquisition model relies on optical principles and follows a compressed-sensing framework [12, 13]. The reconstruction algorithm that is developed accordingly allows to recover grayscale images from the obtained binary data, and substantially improves upon the state of the art [14] in terms of quality and computational performance.

3. In Chapter 4, we describe how the quality of digital holographic reconstruction can be improved compared to standard techniques using an implicit inverse-problem formulation [15]. The proposed reconstruction algorithm consistently

increases the signal-to-noise ratio (SNR) as well as the effective size of the field of view (FOV) of the reconstructions. It is also able to accurately recover both phase and amplitude profiles from downsampled acquisitions, as shown in several simulated and real measurement settings.

We conclude our thesis in Chapter 5. In this last part, we discuss the implications of our work and of our experimental observations from a more general perspective. We also mention possible theoretical and experimental investigations that can be conducted in the future.

1.3 Background

Let us consider a continuous two-dimensional (2D) spatial signal f defined over some bounded domain Ω. This signal is assumed to be of finite energy and thus belongs to the compact normed space $L_2(\Omega)$. In full generality, the scalar value of f at every continuous 2D coordinate $\mathbf{x} = (x_1, x_2)$ inside Ω is denoted by $f(\mathbf{x}) \in \mathbb{C}$. In Chapters 2 and 3, the profile of f is real-valued and corresponds to a 2D grayscale image. In Chapter 4, this profile is complex-valued and contains visual information associated with the phase and/or the amplitude components.

1.3.1 Signal acquisition

Based on these mathematical definitions, the linear effects of the forward models corresponding to the above imaging problems first yield an intermediate measurement vector of the form

$$\mathbf{g} = \mathcal{A}f, \tag{1.1}$$

where $\mathcal{A} : L_2(\Omega) \to \mathbb{C}^M$ is a linear operator mapping the original continuous-domain two-dimensional signal f to the measurement vector $\mathbf{g}$ composed of M distinct elements. The effects of the operator $\mathcal{A}$ in our problems correspond to one or several continuous-domain convolution operations followed by sampling.

The operator $\mathcal{Q}$ that is subsequently involved models nonlinear effects that are deterministic and intrinsic to the physics of the acquisition process. It acts on each component of $\mathbf{g}$ separately—as described in the following chapters—and corresponds either to the identity or to a particular quantization operation defined

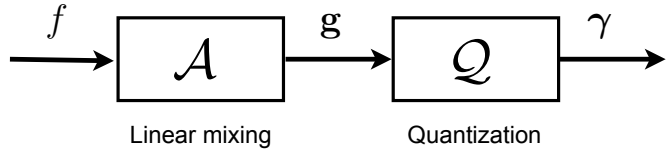


Figure 1.1: General forward model.

over the field of real or complex numbers. Accordingly, the output of our forward models correspond to a measurement vector $\boldsymbol{\gamma}$ that is defined componentwise as

$$\gamma_i = \mathcal{Q}((\mathcal{A}f)_i, \tau_i), \tag{1.2}$$

where $\mathcal{Q} : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ is a pointwise-separable nonlinear operator acting on each input vector component independently. The quantities τ_i act as fixed parameters. The measurements in (1.2) thus consist in M samples whose real values are concatenated into the vector $\boldsymbol{\gamma} \in \mathbb{R}^M$. The generic relation (1.2) between the original signal and the measurements is represented as a block diagram in Figure 1.1. Note that, in certain experimental settings, the real values of the measurements γ_i are subjected to noise with finite variance and are thus not deterministic. The presence of noise is taken into account in our problems whenever it is relevant.

In each problem mentioned in Section 1.2, the information acquired in $\boldsymbol{\gamma}$ is *incomplete*, in the sense that it is insufficient to guarantee the perfect reconstruction of the original data f. The loss of information is mainly due to downsampling and/or to the pointwise nonlinear effects of $\mathcal{Q}$. The linear mixing operator $\mathcal{A}$ always has a nontrivial nullspace because it is defined as a mapping from a continuous function to a finite number of elements [16]. In the sparse-interpolation problem of Chapter 2, the available measurements are heavily downsampled in the spatial domain but are not subjected to any nonlinear distortion (*i.e.*, $\mathcal{Q}$ reduces to the identity operator). In the binary compressed-sensing framework of Chapter 3 and in the holographic problem of Chapter 4, the loss of information occurs by default through the effect of $\mathcal{Q}$, although downsampling can also be involved. In these two latter problems, nonlinear effects are thus to be properly taken into account for the reconstruction. These effects correspond either to binarization or to absolute-value operations that act on each component of $\mathbf{g}$ up to an additive constant.

For all types of reconstruction problems under consideration, we shall demon-

strate that the spatial-domain mixing effect that $\mathcal{A}$ exerts over the original signal f plays a significant role in making the measurements robust to the influence of $\mathcal{Q}$ or of downsampling. In other words, in the context of (1.2), the presence of linear dependencies between the unknowns can prove more suitable for the estimation of f than in the degenerate configuration where $\mathcal{A}$ is purely injective (*i.e.*, in the ideal sampling case). As a matter of fact, the benefits of linear coupling are well known and well studied in the context of Shannon's sampling theory, where the use of an appropriate antialiasing filter potentially allows to perfectly reconstruct the original signal from its samples over some frequency bandwidth.

In physical acquisition devices such as cameras, linear coupling naturally occurs due to sensor-integration effects taking place before sampling [9]. In that setting, the lowpass nature of the underlying filtering process typically reduces the amount of aliasing artifacts in the reconstruction, as shown in our interpolation results of Chapter 2. In order to improve the quality of reconstruction in downsampled or quantized measurement regimes, linear mixing can also be increased and suitably specified in an acquisition device through the use of specific optical components, such as phase masks [12, 17, 18, 19, 11]. Accordingly, in the binary imaging framework that we propose in Chapter 3, the structure of $\mathcal{A}$ is specifically designed to make the acquired data spatially delocalized and robust to binarization through $\mathcal{Q}$ and to downsampling[1]. Although such an approach is unconventional and counter-intuitive a priori, it relies on solid theoretical background [20, 12, 13], and allows to substantially improve the recovery of visual information compared to the case without any prior linear mixing. This paves to way to optical compression approaches based on binary sensors [11]. Finally, in the context of digital holography, linear mixing phenomena are intrinsic to the acquisition process due to the physical nature of light propagation[2]. In particular, the properties of the Fresnel transform potentially allow to reconstruct objects from downsampled acquisitions [21]. In Chapter 4, holographic reconstruction in downsampled settings is also demonstrated with our algorithm [15] in the case where a single intensity hologram is available.

While the linear coupling effects that take place in our problems are potentially advantageous in terms of reconstruction quality, they substantially increase

[1] The linear part of the system must be carefully specified so as not to destroy any information in f prior to quantization. This issue is addressed in detail in Chapter 3.

[2] In the context of holography, the Fresnel operator is especially convenient because it is unitary and fully determines the linear dependencies based on its single propagation-distance parameter. From a practical standpoint, this suppresses the requirement of calibration.

the complexity of the reconstruction procedure from both algorithmic and computational points of view. Indeed, the general model (1.2) implies in such cases that each measurement γ_i depends linearly or nonlinearly on several values of f. Fortunately, and despite the high dimensionality of image data, the resolution of such problems is possible with the currently available computational power. However, specific algorithms must be developed for each setting in order to guarantee a satisfactory reconstruction performance. Thus, in the following chapters, we shall deal not only with the specification of the reconstruction problem for each type of imaging setting, but also with the development of adapted reconstruction methods.

In this thesis, the proposed algorithms map to one or several successive minimization operations. In every case, the overall reconstruction procedure exploits the known measurements, the mathematical description of the forward model, and some a priori information on the properties of the acquired data.

1.3.2 Signal reconstruction

In the context of our reconstruction problems, the acquired signal f is unknown, and the available data consists in the measurements $\boldsymbol{\gamma}$. The aim in each case is thus to provide an estimate $\tilde{f}$ that is as close as possible to the original signal f, the closeness between two signals being typically measured in terms of mean squared error (MSE). In order to deal with finite amounts of computation and storage, the continuous-domain estimate $\tilde{f}$ must have an equivalent representation in the discrete domain. In this thesis, we thus assume the representation space of $\tilde{f}$ to be spanned by translates of an analog generating kernel φ [8]. Accordingly, we have the generic expression

$$\tilde{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \tilde{c}[\mathbf{k}]\varphi\left(\frac{\mathbf{x}}{\Delta_\mathrm{c}} - \mathbf{k}\right), \tag{1.3}$$

where $\mathbf{k} = (k_1, k_2)$ is a discrete 2D coordinate, where $\tilde{c}$ is a sequence of N coefficients[3] defined on a rectangular unit cartesian grid $\Omega^0 \subset \mathbb{Z}^2$, and where Δ_c is the coefficient-grid spacing. In our framework, the solution $\tilde{f}$ is deemed suitable only if it is *consistent* with $\boldsymbol{\gamma}$ [9, 19, 8, 11, 15]. This means that, when substituting the unknown signal f with the estimate $\tilde{f}$ into (1.2), the resulting vector $\tilde{\boldsymbol{\gamma}}$ must be equal,

[3] In this thesis, the sums and integrals involving the unknown data are associated with computations over bounded domains, *e.g.*, Ω^0 and Ω, but are written in more generic form for convenience.

or at least close in some sense, to the known measurements $\boldsymbol{\gamma}$. This consistency constraint is typically relaxed in experimental settings due to the presence of noise. Actually, the expansion (1.3) of $\tilde{f}$ defines a fixed discrete linear mapping between its coefficients $\tilde{c}$ and the components of the vector $\tilde{\mathbf{g}}$ that is defined in accordance with (1.1) as

$$\tilde{\mathbf{g}} = \mathcal{A}\tilde{f}. \tag{1.4}$$

This mapping is fully described by a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ and can be written as

$$\tilde{\mathbf{g}} = \mathbf{A}\tilde{\mathbf{c}}. \tag{1.5}$$

In this thesis, the vectors and matrices are denoted by bold lowercase and capital symbols, respectively. The sequences represented in vectorized form use the same letters but in bold format, and vice versa. Accordingly, the N-vector $\tilde{\mathbf{c}}$ in the above relation (1.5) corresponds to the coefficient sequence $\tilde{c}$ in (1.3), each vector element being mapped to a sequence element[4] at a particular coordinate $\mathbf{k} \in \Omega^0$. The mapping used for vectorization (*e.g.*, lexicographic concatenation) is of the form $\mathcal{V} : \mathbb{N}^* \to \mathbb{Z}^2$. It needs not be specified explicitly because the derivations in the following chapters do not require the component values of the vectors or matrices to be determined elementwise. Nevertheless, the concept of vectorization provides an exact equivalence between the spatial-domain and the vector or matrix representations. For instance, matrices acting as a vector pre-multipliers are associated with image-processing operations such as convolution or pointwise multiplication. The matrix notation is especially well suited to the derivation of reconstruction algorithms because it is compatible with the formalism of linear algebra.

Given (1.2) and (1.5), the general form of the forward models used in our reconstruction problems for the estimation of f corresponds to the relation

$$\tilde{\gamma}_i = \mathcal{Q}((\mathbf{A}\tilde{\mathbf{c}})_i, \tau_i). \tag{1.6}$$

Note that, interestingly, the form of (1.6) is linked to the so-called *generalized linear models* that are studied in the field of statistics [22]. This equation indeed involves a linear operation that introduces dependencies between the unknown variables followed by pointwise nonlinearities.

[4]Predefined sets of samples are sometimes removed from sequences through masking. In such cases, the corresponding vector representations only include the non-masked values.

Since the action of all linear effects on our solution estimate $\tilde{f}$ is entirely defined through $\mathbf{A}$ given the assumption (1.3), we shall not refer to the operator $\mathcal{A}$ explicitly in the following chapters. Instead, we shall derive the structure of the matrix $\mathbf{A}$—and define the nonlinear operator $\mathcal{Q}$—according to the corresponding forward models. Finally, according to the expansion (1.3), the continuous-domain solution $\tilde{f}$ can be computed with arbitrary precision once $\tilde{c}$ is estimated.

Chapter 2

Image reconstruction from sparse non-uniform samples

2.1 Introduction

In this chapter, our goal is to reconstruct images from a given subset of samples based on interpolation. Our forward model consists in a continuous-domain prefilter φ_0 acting on the original image f by convolution followed by an ideal sampler. The prefilter corresponds to the nonideal impulse response of the optical device. Accordingly, the measurements that are obtained in this setting correspond to so-called *generalized samples*, the operator $\mathcal{Q}$ in (1.2) being reduced to the identity.

Assuming an expansion of the form (1.3), we express our solution $\tilde{f}$ in the continuous domain, considering consistent resampling as a data-fidelity constraint. To make the problem well-posed and ensure edge-preserving solutions, we develop an efficient anisotropic regularization approach that is based on an improved version of the anisotropic edge-enhancing diffusion (EED) equation. Following variational principles, our reconstruction algorithm minimizes successive quadratic cost functionals. To ensure fast convergence, we solve the corresponding sequence of linear problems by using multigrid iterations that are specifically tailored to their sparse structure.

We conduct illustrative experiments and discuss the potential of our approach

both in terms of algorithmic design and reconstruction quality. In particular, we present results that use as little as two percent of the image samples[1].

2.2 Overview

Shannon's sampling theorem [23] states that a bandlimited signal can be perfectly reconstructed from its samples, provided that Nyquist's criterion is satisfied. In that case, the solution can be found by sinc interpolation. However, in an imaging context, bandlimited functions do not correctly match the physical reality [24]. This classical assumption can thus lead to inaccurate results for such classes of problems. Specifically, optical acquisition systems do deviate from an ideal sampler in practice as they involve filtering associated with their impulse response prior to sampling. When taken into account, those effects impose *consistency constraints* on generalized samples [9], which makes the reconstruction more intricate. Several works have successfully dealt with this non-ideality [25, 26, 27, 19, 28, 29], yielding results that are visually sharper as compared to the standard sampling paradigm. Nevertheless, these approaches are typically focused on pure magnification cases, and have not been applied to sparse-interpolation problems.

In this chapter, we introduce a novel interpolation approach that simultaneously handles generalized and sparse image sampling. The objective of our method is to reconstruct a continuous image from a subset of its generalized samples. As a first step towards specifying our problem, we define a data-fidelity measure that is based on consistent resampling. The unknowns being under-constrained, regularization is needed to find a suitable solution.

Variational formulations are commonly employed for regularization in the literature. In particular, quadratic regularization has been previously considered to develop fast sparse-interpolation approaches [30, 31]. Despite their efficiency, these linear methods tend to blur image contours [25]. In order to produce edge-preserving reconstructions, nonquadratic functionals are required [25]. A classical solution is the total-variation (TV) semi-norm [32], which is also associated with fast algorithms such as primal-dual approaches [33, 34]. However, the behavior of standard edge-preserving regularization techniques is not adapted to sparse interpolation for it produces singular points [35]. Thus, recent works on image interpolation involve more advanced formulations, especially when the sampling rate

[1] This chapter is based on our paper [8].

is low, or when the available data consist in a reduced set of samples as in our problem. Promising results have been obtained in the variational framework using a nonlocal approach [36]. Unfortunately, the associated computation time tends to be prohibitive.

Some of the most efficient regularization strategies in the area of inpainting and sparse interpolation involve partial differential equations (PDEs) that are based on *anisotropic diffusion*[2] [35, 37, 38]. The behavior of these methods can be fine-tuned via the specification of diffusion tensors. A high-quality technique based on EED has been devised by Galić *et al.* to interpolate sparse image samples [35]. This approach enjoys remarkable edge-reconstruction performance even at very high sparsity levels.

Nonlinear PDEs can be solved using explicit or semi-implicit schemes that are based on finite-difference approximations of the original formulation [39]. An alternate approach involving *lagged-diffusivity fixed-point iterations* has been also investigated for the TV flow, and subsequently for other types of isotropic diffusion equations [40, 41]. In order to obtain rapid convergence, each iteration involves a linearized version of the flow where the diffusivity terms are fixed according to some current solution estimate. Despite their increased complexity [42], tensor diffusivities can also be handled efficiently in large-scale problems. In that regard, a novel class of algorithms based on fast explicit diffusion (FED) has been devised in [43]. This approach follows a coarse-to-fine strategy, and allows to implement advanced PDE-based methods such as [35].

The above PDE-based regularization approaches are most efficient for interpolation and inpainting problems. Variational approaches, on the other hand, do result in efficient implementations for a larger class of inverse problems, including image restoration [44]. In particular, variational formulations are most adapted to our extended interpolation model. They allow to efficiently handle our specific data-fidelity constraint that involves an analysis kernel before sampling.

The distinction between both types of methods, however, is not clear-cut. Indeed, regularization-based methods are related to PDE formulations through the Euler-Lagrange equation [32]. Consequently, the basic steepest-descent method applied to the given cost functional is equivalent to the corresponding gradient flow. Similarly, the iteratively reweighted least-squares (IRLS) technique [45] that is used

[2] In this chapter, *anisotropy* of a given diffusion process is understood in the sense of [35], implying tensor diffusivities. Though nonquadratic, the TV functional only acts as an isotropic regularizer following that definition.

for nonquadratic regularization is associated with linearized versions of the gradient of the original functional [41, 46]. This provides an interpretation that relates IRLS to lagged-diffusivity fixed-point iterations.

In this chapter, we develop a hybrid regularization framework that combines advantageous aspects of both PDE and variational formulations using similar principles based on lagged diffusivities. The specificity of our design is that it stems directly from the definition of an anisotropic-diffusion equation. As a consequence, our regularizer consists in a series of quadratic functionals that are based on successive tensor-valued diffusivity estimates. While being adapted to our particular problem, this cost-functional approach exhibits similarities with the FED method of [43] where first-order approximations of the underlying diffusion process are taken. Regarding the actual specification of the flow, our central contribution is to propose our own PDE as an extension of the EED solution considered in [35]. In particular, we redefine the associated tensor diffusivities so as to further improve edge-reconstruction capability on natural images.

Although not strictly originating from a variational formulation, our regularization approach yields an anisotropic version of the IRLS technique [45]. Starting from a quadratic data-fidelity constraint, our reconstruction algorithm called *anisotropic IRLS* (AIRLS) entails the partial resolution of successive weighted linear problems. Since the diffusivity estimation of our method is constrained to weight updates, they do not affect the overall algorithmic performance significantly. We also devise a fast multigrid solver that is adapted to the sparse structure of our linear problems and that is inspired from previous works [30, 31]. Note that the obtained AIRLS framework can then be used to implement distinct regularization PDEs as well, including the one of [35].

The chapter is organized as follows: In Section 2.3, we present our continuous interpolation framework where the unknowns are expressed as coefficients in a shift-invariant basis. In Section 2.4, we consider classical variational approaches to express our reconstruction problem. Our actual strategy resulting in an IRLS-type procedure is introduced in Section 2.5. The associated linear problems are then specified in Section 2.6, and their iterative resolution using our own multilevel approach is addressed in Section 2.7. In the experiments of Section 2.8, we consider distinct interpolation cases where our method is compared with the state of the art, both quantitatively and qualitatively. Implications of our results are finally discussed in Section 2.9.

2.3 Sampling and reconstruction

2.3.1 Forward model

As represented in Figure 2.1, the input signal of our model consists in the continuous-domain image f. Assuming generalized sampling, the latter is first convolved with a prefilter φ_0 that corresponds to the impulse response of the optical acquisition device [9]. The intermediate image is then sampled at integer intervals $\mathcal{M}$ along each dimension, which results in the sequence

$$f_1[\mathbf{k}] = (f * \varphi_0)(\mathbf{x})|_{\mathbf{x}=\mathcal{M}\mathbf{k}}, \tag{2.1}$$

where $*$ denotes continuous-domain convolution. Only a subset of the sequence f_1 is retained through the binary mask χ, which yields the M masked samples

$$g[\mathbf{k}] = \chi[\mathbf{k}] f_1[\mathbf{k}]. \tag{2.2}$$

The operator $\mathcal{Q}$ corresponds to the identity operator in this chapter. This implies that our measurement sequence γ is simply defined by the equality

$$\gamma[\mathbf{k}] = g[\mathbf{k}]. \tag{2.3}$$

Assuming that the sampling process and the binary mask are known, we define our interpolation problem as the task of accurately reconstructing the original image f from the available samples in γ. We elaborate on our reconstruction approach below.

2.3.2 Reconstruction space

Following the generalized sampling theory of [9], our reconstruction space is in the continuous domain and spanned by normalized translates of an analog generating kernel φ. Specifically, assuming the expansion (1.3) with a unit step Δ_c, the reconstructed image $\tilde{f}$ takes the form

$$\tilde{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \tilde{c}[\mathbf{k}] \varphi(\mathbf{x} - \mathbf{k}), \tag{2.4}$$

where $\tilde{c}$ is a discrete sequence of N coefficients that describes the solution exactly. The reconstruction is defined on a grid that is $\mathcal{M}$ times finer than the acquisition

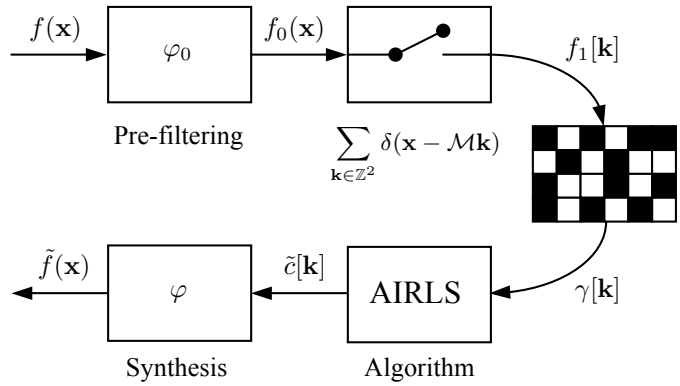


Figure 2.1: The continuously defined image f is convolved with φ_0 (the impulse response of the acquisition device) before being sampled at integer intervals $\mathcal{M}$ along each dimension. The resulting sequence f_1 is then masked, which yields γ. Starting from these masked samples, our algorithm outputs the coefficients $\tilde{c}$ of the reconstructed image. The continuously defined solution $\tilde{f}$ can be then be obtained from these coefficients according to (2.4).

in each dimension. In our implementation, the image data are defined over some rectangular domain Ω and are extended periodically outside.

The formulation (2.4) enables the solution to be computed and stored in terms of its coefficients, despite its continuous character. In this framework, we specify φ as a B-spline function [47] of order η, which makes straightforward sub-pixel post-processing (e.g., registration) of the reconstructed data possible, and allows to properly define our reconstruction approach. The interpolating B-spline is differentiable and has suitable approximation properties, such as reproduction of polynomials [47].

2.3.3 Constraints

In order to be accurate, the solution (2.4) has to be consistent with the available samples γ. While adopting the consistent-measurement principle of [9], we nevertheless want to accommodate for noise and model imperfections. Therefore, we

propose a soft form of this constraint, demanding that $\tilde{f}$ reintroduced in place of f into the generalized-sampling system of Figure 2.1 results in measurements $\tilde{\gamma}$ that are close to γ. The image $\tilde{f}$ and the measurements $\tilde{\gamma}$ are related in the same way as f and γ through (2.1), (2.2), and (2.3). Accordingly, we propose to define the data discrepancy measure as

$$\mathcal{D}(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \left|\tilde{\gamma}[\mathbf{k}] - \gamma[\mathbf{k}]\right|^2, \tag{2.5}$$

which is an implicit function of the expansion coefficients in (2.4). Note that, in the sequel, we shall use implicit functions of the solution coefficients when appropriate. As a soft constraint, we impose

$$\mathcal{D}(\tilde{c}) \leq \mathcal{K}_{\mathcal{D}}, \tag{2.6}$$

where $\mathcal{K}_{\mathcal{D}}$ is a positive constant. The sequence $\tilde{g}$ is derived from the above relations as

$$\tilde{g}[\mathbf{k}] = \chi[\mathbf{k}] \left(\varphi_0 * \sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{c}[\mathbf{m}]\varphi(\cdot - \mathbf{m})\right)(\mathbf{x})\Big|_{\mathbf{x}=\mathcal{M}\mathbf{k}}, \tag{2.7}$$

which[3] can be simplified as

$$\tilde{g}[\mathbf{k}] = \chi[\mathbf{k}]\{b \star \tilde{c}\}_{\downarrow\mathcal{M}}[\mathbf{k}], \tag{2.8}$$

where $\star$ denotes discrete convolution, and where the subscript $_{\downarrow\mathcal{M}}$ denotes downsampling by a factor of $\mathcal{M}$ in each dimension. The sequence b is defined as

$$b[\mathbf{k}] = \left(\varphi_0 * \varphi\right)(\mathbf{x})\big|_{\mathbf{x}=\mathbf{k}}. \tag{2.9}$$

In matrix notation, the relation (2.8) between the coefficients $\tilde{c}$ and the measurements $\tilde{g}$ is of the form (1.5), the measurement matrix $\mathbf{A}$ being defined as

$$\mathbf{A} = \boldsymbol{\chi}\mathbf{D}_{\mathcal{M}}\mathbf{B}. \tag{2.10}$$

[3]The symbol $\cdot$ denotes a dummy variable. It can be used to create new function definitions based on existing ones. In this case, for instance, $\varphi(\cdot - \mathbf{m})$ corresponds to the original function φ shifted spatially by $\mathbf{m}$.

The convolution matrix $\mathbf{B}$ is associated with the filter b in (2.9), and the matrix $\mathbf{D}_{\mathcal{M}}$ implements 2D $\mathcal{M}$-fold downsampling. Note that the transpose of $\mathbf{D}_{\mathcal{M}}$ corresponds to the upsampling matrix $\mathbf{U}_{\mathcal{M}}$. The matrix $\boldsymbol{\chi}$ is linked to the masking sequence χ. It corresponds to an identity matrix whose rows associated with the discarded measurements are suppressed [11].

Given (2.3), (2.8), and the binary nature of the mask χ, the data discrepancy measure (2.5) can be rewritten in the explicit form

$$\mathcal{D}(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi[\mathbf{k}] \left|\{b \star \tilde{c}\}_{\downarrow\mathcal{M}}[\mathbf{k}] - \gamma[\mathbf{k}]\right|^2 . \tag{2.11}$$

Therefore, we want any valid solution $\tilde{f}$ to have a low discrepancy measure $\mathcal{D}$. In this chapter, we assume that the prefilter φ_0 used in our generalized-sampling model (2.1) is nonnegative, which itself implies the non-negativity of b given the definition of φ as a B-spline.

Under (2.6), our reconstruction problem is still ill-posed. We thus have to define additional regularity constraints that make the problem well-posed. In the sequel, we discuss reconstruction approaches that satisfy this requirement while ensuring low data discrepancy.

2.4 Existing variational approaches

The variational framework lends itself well to the specification of our reconstruction problem. It allows the solution to satisfy a constraint of the form (2.6) under suitable regularity criteria. In this section, we review some classical regularization functionals that are used for image reconstruction. Their properties as well as their links with IRLS and PDE formulations also serve as background for our own reconstruction method introduced in Section 2.5. Their expressions are readily introduced within our reconstruction framework, which allows to predefine relevant quantities and relations for the sequel. In this variational setting, the generic reconstruction problem is to minimize the functional

$$\mathcal{J}(\tilde{c}) = \mathcal{D}(\tilde{c}) + \Lambda\mathcal{R}(\tilde{c}), \tag{2.12}$$

where $\mathcal{D}$ is the quadratic data-fidelity term defined in (2.11), and where $\mathcal{R}$ is a generic term that penalizes non-desired solutions. The constant $\Lambda > 0$ determines

an implicit $\mathcal{K}_{\mathcal{D}}$ value. Specifically, under suitable conditions, each $\mathcal{K}_{\mathcal{D}}$ is associated with a particular Lagrange multiplier Λ such that minimizing (2.12) is equivalent to minimizing $\mathcal{R}(\tilde{c})$ alone under the data-fidelity constraint (2.6). While $\mathcal{D}$ is fixed according to our forward model, the choice of $\mathcal{R}$ strongly determines the quality of the solution. Its proper specification is therefore extensively discussed below.

2.4.1 Quadratic regularization

When applied to our framework, an extended class of quadratic functionals can be written as the Sobolev-type norm

$$\mathcal{R}_{\mathrm{S}}(\tilde{c}) = \left\| \mathrm{L}\tilde{f} \right\|_{L_2}^2, \tag{2.13}$$

where L is a linear differential operator. These regularizers penalize high responses of L at each spatial location, which promotes regular solutions. Given (2.4) and the quadratic nature of both functionals (2.11) and (2.13), the associated minimization problems consist in discrete sets of linear equations.

When the data-fidelity term reduces to the denoising case (i.e., $\mathcal{M} = 1$, φ_0 is the Dirac distribution $\delta(\cdot)$, and $\chi = 1$), these regularization functionals are linked with the standard form of the so-called *smoothing splines*. Indeed, for appropriate L, the minimizer of (2.12) defined in the spline space (2.4) coincides with the optimum among all possible functions [48]. Note also that the noiseless magnification case (i.e., $\varphi_0 = \delta(\cdot)$, $\chi = 1$, and $\Lambda \to 0^+$) has been specifically addressed in [19]. In a similar framework, some authors have proposed fast linear solutions to interpolate very sparse samples [30, 31]. These approaches exploit the B-spline expansion of the solution in a multigrid fashion, yielding fast iterative algorithms.

2.4.2 Nonquadratic regularization

Edge-preserving reconstruction is achievable with nonquadratic regularizers. In this section, we review a class of such functionals described in [46, 49] and defined in our framework as

$$\mathcal{R}_{\mathrm{N}}(\tilde{c}) = \int_{\mathbb{R}^2} \Psi_{\mathcal{R}}(\|\nabla \tilde{f}(\mathbf{x})\|) \mathrm{d}\mathbf{x}, \tag{2.14}$$

where $\Psi_{\mathcal{R}} : \mathbb{R}_+ \to \mathbb{R}$ denotes a *potential function*. In addition, we discuss the associated IRLS technique that is widely used for reconstruction. As shown in the sequel, the structure of the latter is closely related to our approach developed in Section 2.5. The edge-preserving potential $\Psi_{\mathcal{R}}$ grows less fast than a quadratic function [46] unless $\mathcal{R}_{\mathrm{N}}$ degenerates to $\mathcal{R}_{\mathrm{S}}$ with $\mathrm{L} = \nabla$. For instance, TV regularization corresponds to the L_1-norm of the image gradient, *i.e.*, to the choice $\Psi_{\mathcal{R}}(t) = |t|$, in the sense of distributions. This case has already been considered in the context of generalized sampling for image magnification [25, 26, 33, 34].

In order to yield a tractable reconstruction problem, (2.14) is typically discretized before minimization. This approach is standard when dealing with sampled data. For instance, a discrete form of TV based on a graph model is formulated in [50]. Following a similar idea, the gradient values entering in the original continuous-domain definition are approximated using first-order difference filters in [51]. In our context, all discrete quantities are a natural outcome of the B-spline expansion (2.4) after replacing the integral (2.14) by a sum. Here, we select a configuration where the partial derivatives of the gradient are evaluated in-between the grid nodes. This avoids the creation of spurious oscillations or divergence of the solution[4]. Accordingly, we define

$$\mathcal{R}_{\mathrm{N}}^{0}(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Psi_{\mathcal{R}}(\|\mathring{\nabla}\tilde{f}(\mathbf{x})\|)_{\mathbf{x}=\mathbf{k}}. \tag{2.15}$$

The upper-ring notation modifies the gradient—and similar vector operators—as $\mathring{\nabla} = \mathcal{S}\nabla$, where $\mathcal{S}$ shifts a continuous-domain vector function $\mathbf{v}$ as

$$\mathcal{S}\mathbf{v}(\mathbf{x}) = (v_1(x_1 + 1/2, x_2), v_2(x_1, x_2 + 1/2)). \tag{2.16}$$

The explicit form of (2.15) in terms of the solution coefficients is then

$$\mathcal{R}_{\mathrm{N}}^{0}(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Psi_{\mathcal{R}}(\|(\tilde{c} \star \mathbf{r})[\mathbf{k}]\|), \tag{2.17}$$

where $\mathbf{r}$ is the discrete multivariate filter

$$\mathbf{r}[\mathbf{k}] = \mathring{\nabla}\varphi(\mathbf{x})|_{\mathbf{x}=\mathbf{k}}. \tag{2.18}$$

[4]Similar schemes are used in computational fluid dynamics to ensure numerical stability [52].

1) Initialize at $I = 0$ with the solution $\tilde{c}^{(0)}$
while $I < N_\mathrm{i}$ **do**
  a) Minimize $\mathcal{J}(\cdot|\tilde{c}^{(n)})$ with initialization $\tilde{c}^{(n)}$
    s.t. $\mathcal{J}(\tilde{c}^{(n+1)}|\tilde{c}^{(n)}) < \mathcal{J}(\tilde{c}^{(n)}|\tilde{c}^{(n)})$
  b) Store the solution $\tilde{c}^{(n+1)}$
  c) Establish the new bound $\mathcal{J}(\cdot|\tilde{c}^{(n+1)})$
  d) Count $I \leftarrow I + 1$
end

Algorithm 2.1: Generic IRLS procedure.

In order to minimize (2.12) with the nonquadratic term (2.15), we can resort to an IRLS approach, following a majorize-minimize (MM) strategy [53]. Starting from an initial solution estimate $\tilde{c}^{(0)}$, this strategy consists in the partial minimization of surrogate quadratic functionals $\mathcal{J}(\cdot|\tilde{c}^{(n)})$ that are based on the original $\mathcal{J}(\cdot)$ and successively updated according to the current solution estimate $\tilde{c}^{(n)}$. Following the *multiplicative form* of half-quadratic minimization [46, 54], we obtain the iterative procedure given in Algorithm 2.1. The surrogate functionals are defined as

$$\mathcal{J}_\mathrm{N}^0(\tilde{c}|\tilde{c}^{(n)}) = \mathcal{D}(\tilde{c}) + \Lambda\mathcal{R}_\mathrm{N}^0(\tilde{c}|\tilde{c}^{(n)}). \tag{2.19}$$

This form of minimization is called multiplicative because the structure of $\mathcal{R}_\mathrm{N}^0(\cdot|\tilde{c}^{(n)})$ involves multiplications with weights. Specifically, each surrogate regularizer is defined as the quadratic functional

$$\mathcal{R}_\mathrm{N}^0(\tilde{c}|\tilde{c}^{(n)}) = \frac{1}{2}\sum_{\mathbf{k}\in\mathbb{Z}^2} \theta_\mathrm{N}(\tilde{c}^{(n)}, \psi)[\mathbf{k}]\|(\tilde{c} \star \mathbf{r})[\mathbf{k}]\|^2, \tag{2.20}$$

where the weight sequence θ_N at each index $\mathbf{k}$ depends on the current solution $\tilde{c}^{(n)}$ as

$$\theta_\mathrm{N}(\tilde{c}^{(n)}, \psi)[\mathbf{k}] = \psi(\|(\tilde{c}^{(n)} \star \mathbf{r})[\mathbf{k}]\|). \tag{2.21}$$

The scalar function $\psi : \mathbb{R}_+ \to \mathbb{R}$ is derived from the potential function $\Psi_\mathcal{R}$ of the regularizer (2.17) through the constraint that the successive $\mathcal{J}(\cdot|\tilde{c}^{(n)})$ constitute valid upper bounds of $\mathcal{J}(\cdot)$. This constraint corresponds to

$$\mathcal{J}_\mathrm{N}^0(\tilde{c}|\tilde{c}^{(n)}) + \text{const.} \geq \mathcal{J}_\mathrm{N}^0(\tilde{c}), \forall\tilde{c}, \tag{2.22}$$

with equality at $\tilde{c} = \tilde{c}^{(n)}$; the scalar constant in (2.22) is independent from $\tilde{c}$. In the TV case where $\Psi_{\mathcal{R}}(t) = |t|$, we obtain $\psi(t) = (t^2 + \epsilon)^{-1/2}$, where the small positive parameter ϵ is added to overcome the non-differentiability of the original functional [41]. Note that the use of this constant is avoidable with nonsmooth solution initializations [51]. Minimizing an instance of (2.20) amounts to solving a linear problem. As specified in Algorithm 2.1, each quadratic cost need only be decreased slightly in IRLS. This implies that each of these linear problems must only be solved partially, which is doable using iterative methods.

As a prerequisite to our approach discussed in the next section, let us now draw the link between IRLS and the fixed-point interpretation discussed in [40, 41, 46]. Expanding the regularization part of (2.19), its minimum with respect to the coefficients $\tilde{c}$ satisfies the first-order condition

$$\Lambda^{-1} \frac{\partial}{\partial \tilde{c}} \mathcal{D}(\tilde{c}) + (\mathbf{r}[\cdot])^{\mathrm{T}} \star \Big(\theta_{\mathrm{N}}(\tilde{c}^{(n)}, \psi)(\tilde{c} \star \mathbf{r}) \Big) = 0. \tag{2.23}$$

The solutions of (2.23) depend on the current estimate $\tilde{c}^{(n)}$ through the weights θ_{N} defined from ψ in (2.21). Following the terminology of [40, 46], the latter quantities are identified as *lagged diffusivities*. By extension, the IRLS procedure of Algorithm 2.1 can be recast as a discretized *lagged-diffusivity fixed-point iteration*, which consists in alternating between the partial resolution of (2.23) and the update of $\tilde{c}^{(n)}$. Any sequence $\tilde{c}$ which satisfies

$$\Lambda^{-1} \frac{\partial}{\partial \tilde{c}} \mathcal{D}(\tilde{c}) + (\mathbf{r}[\cdot])^{\mathrm{T}} \star (\theta_{\mathrm{N}}(\tilde{c}, \psi)(\tilde{c} \star \mathbf{r})) = 0 \tag{2.24}$$

is thus a fixed point of the IRLS process. Given (2.18), the regularization part of (2.23) that involves the filter $\mathbf{r}$ corresponds to the spatially discretized form of

$$\operatorname{div} \Big(\psi(\|\nabla u^{(n)}\|) \nabla u \Big), \tag{2.25}$$

where u is a continuous-domain solution with its current estimate $u^{(n)}$. Remarkably, the above expression is directly related to isotropic-diffusion flows of the form

$$\partial_{\mathrm{t}} u = \operatorname{div} \left(\psi(\|\nabla u\|) \nabla u \right), \tag{2.26}$$

where isotropy is defined in the sense of [39]. Specifically, the divergence term (2.25) is similar to the right-hand-side term of (2.26), except that the diffusivities ψ are lagged in the former.

For general functions $\Psi_{\mathcal{R}}$, it holds from [46] that the left-hand-side terms of (2.23) correspond to the linearized gradient of the original nonquadratic functional, up to discretization. As a consequence, (2.26) corresponds to the gradient flow of the original regularizer as specified by the Euler-Lagrange equation. The IRLS method used for variational image reconstruction can thus be interpreted as fixed-point iterations where the successive surrogate regularization functionals are related to lagged versions of the corresponding PDE.

2.5 Proposed approach

The above discussion emphasizes the theoretical pathway that relates the regularization part of the IRLS structure to the corresponding PDE formulation. In our approach developed below, we first specify our own continuous-domain PDE, and then transpose it into an IRLS-type framework using similar concepts. Note that this section mainly deals with our regularization strategy; after derivation, our method shall involve a series of regularization functionals combined with the same data-fidelity term as in (2.19).

2.5.1 Edge-enhancing anisotropic diffusion

The EED equation has first been applied to interpolation problems by Galić *et al.* [55, 35]. This PDE is divergence-based as in (2.26), and involves tensor-valued diffusivities that are determined from a smoothed gradient map $\mathcal{G}u$ of the current solution u. As compared to the regularizers presented above, the anisotropic character of EED is associated with better reconstruction properties. Equations of a similar structure have been proposed for multichannel images by Roussos and Maragos [28, 44], considering denoising and magnification applications with a generalized-sampling model. The use of tensor-driven diffusion equations based on the divergence or on the trace operator[5] has also been investigated by Tschumperlé and Deriche for general imaging problems [56, 37]. In this chapter, we consider the original EED definition, which we write as

$$\partial_{\mathrm{t}} u = \operatorname{div}\left(\mathbf{T}(\mathcal{G}u, \psi)\nabla u\right), \tag{2.27}$$

[5]The forms of divergence-based and trace-based PDEs are closely related, as described in [56].

where $\mathbf{T} \in \mathbb{R}^{2\times 2}$ denotes a symmetric and positive-definite tensor-diffusivity function. The first argument of $\mathbf{T}$ is a smoothed gradient $\mathcal{G}u(\mathbf{x})$, while its last one is the scalar function ψ. The operator $\mathcal{G}$ denotes a modified gradient that includes additional smoothing. As discussed in Section 2.5.3, our contribution is to extend the definition of $\mathcal{G}$ so as to better preserve certain image features. The definition of $\mathbf{T}$ distinguishes EED from the other divergence-based anisotropic PDEs considered in [56, 28, 44]. According to [35],

$$\mathbf{T}(\mathbf{v}, \psi) = \psi(\|\mathbf{v}\|)\mathcal{P}_{\mathbf{v}} + \mathcal{P}^{\perp}{}_{\mathbf{v}}, \tag{2.28}$$

where $\mathcal{P}$ and $\mathcal{P}^{\perp}$ are projectors onto the subscripted vector and the perpendicular directions, respectively. While anisotropic flows linked to an energy function (*e.g.*, Beltrami flow) have been studied for imaging applications[6] [44, 56, 57], there is not any known energy interpretation of (2.27). For example, EED does not comply with the structure of [44] where the diffusion tensor involves two distinct convolutions with the same kernel. This absence of global interpretation is common in the literature [49].

In order to ensure the positive-semidefiniteness of the diffusivities and the stability of EED, we impose the function $\psi(t)$ to be nonnegative and nonincreasing in t with $\psi(0) = 1$. Given these constraints, the tensor $\mathbf{T}(\mathbf{0}, \psi)$ is well-defined and corresponds to the identity matrix. Similar to its role in (2.26), the purpose of ψ in (2.27) is to reduce smoothing across edges. In this tensor case, however, the associated flow modification is anisotropic. The action of $\mathbf{T}$ at each position is to decompose the corresponding gradient $\nabla u(\mathbf{x})$ into the sum of two orthogonal vectors that are respectively parallel and perpendicular to $\mathcal{G}u(\mathbf{x})$. The magnitude of the parallel part of this gradient is reduced by multiplication with $\psi(\|\mathcal{G}u(\mathbf{x})\|)$, while the perpendicular one is left untouched; this permits stronger diffusion along edges, hence the EED effect. From Definition (2.28), the elements T_{ij} of the tensor $\mathbf{T}$ are expressed as

$$\begin{aligned} T_{11}(\mathbf{v}, \psi) &= \|\mathbf{v}\|^{-2} \left(\psi(\|\mathbf{v}\|)v_1^2 + v_2^2\right), \\ T_{22}(\mathbf{v}, \psi) &= \|\mathbf{v}\|^{-2} \left(v_1^2 + \psi(\|\mathbf{v}\|)v_2^2\right), \\ T_{12}(\mathbf{v}, \psi) &= \|\mathbf{v}\|^{-2} \left(\psi(\|\mathbf{v}\|) - 1\right) v_1 v_2, \end{aligned} \tag{2.29}$$

[6] These flows can remain anisotropic (*e.g.*, tensor TV in [44]) or degenerate [56, 57] when applied to single-channel data.

where v_1 and v_2 are the coordinates of the vector $\mathbf{v}$.

2.5.2 Specification of the penalty function

In our approach, we consider three diffusivity functions ψ among those proposed in the literature. The first is the *Charbonnier diffusivity* used in [35] that is defined as

$$\psi_0(t) = (1 + t^2/{\beta_1}^2)^{-1/2}, \tag{2.30}$$

where $\beta_1 \in \mathbb{R}_+^*$ is a constant. The second one is linked to the Huber potential through the multiplicative form of half-quadratic minimization [54], and is defined as

$$\psi_1(t) \quad = \quad \begin{cases} 1, & |t| \leq \beta_1, \\ \beta_1 |t|^{-1}, & \text{otherwise.} \end{cases} \tag{2.31}$$

Since $\psi_1(t) \propto |t|^{-1}$ for large t, it tampers cross-edge smoothing the same way as TV. Meanwhile, this function can restore smoothly varying regions, because the associated potential is quadratic for $|t| < \beta_1$. The functions ψ_0 and ψ_1 are closely related because their values coincide when t tends to zero or to infinity. The third alternative that we propose is the Perona-Malik diffusivity

$$\psi_2(t) = \exp(-t^2/{\beta_2}^2), \tag{2.32}$$

where $\beta_2 \in \mathbb{R}_+^*$ is a constant. This function is well-known for its contrast-enhancing properties in the case of isotropic [58] as well as of anisotropic diffusion [39].

2.5.3 Specification of the modified gradient operator

As discussed above, the map $\mathcal{G}u$ is used to specify anisotropic diffusivities. The operator $\mathcal{G}$ serves in the EED equation (2.27) as an edge-information estimate that is more robust than the standard gradient. In [35], this operator corresponds to the Gaussian-smoothed gradient

$$\mathcal{G}_0 = \nabla * \varphi_{\mathrm{g}}, \tag{2.33}$$

where φ_{g} is an isotropic Gaussian filter of standard deviation l_σ. The associated results reported in [35] are very promising, even when dealing with very sparse data. Note that, although $\mathcal{G}u$ only enters in the definition of $\mathbf{T}$, it nonetheless determines the character of the edge-enhancing effect in terms of flow regulation. As shown in Section 2.8, the potential differences in terms of reconstruction behavior are important, which suggests introducing some better estimates.

Anisotropic flows are obtainable without the requirement of gradient smoothing in general [44]. This operation is nevertheless necessary in our PDE to guarantee the anisotropy of the diffusivities[7]. The smoothing process generally tends to wipe out fine-scale edge information. Our contribution is to specify an operator $\mathcal{G}_1$ that yields a *directionally smoothed* version of the image gradient, which better preserves fine-scale information compared to the Gaussian solution $\mathcal{G}_0$. The goal of $\mathcal{G}_1$ is thus to be able to retain more accurate edge-orientation information than with Gaussian smoothing while remaining robust to potential local disturbances. In order to determine the corresponding smoothing directions, we compute orientation estimates $\upsilon \in [0, \pi[$ that are matched with the local edge features of the image argument u at each position. These estimates are obtained as the solution of the optimization problem described below.

Let us consider the class of segment cross-sections of constant length l_{s}, centered at positions $\mathbf{x}$, and with orientations $\upsilon_0 \in [0, \pi[$. Given u, we associate this class to the local oriented-mean measure

$$\Sigma^*(u, \mathbf{x}, \upsilon_0) = {l_{\mathrm{s}}}^{-1} \int_{-l_{\mathrm{s}}/2}^{l_{\mathrm{s}}/2} u(\boldsymbol{\vartheta}(t)) \mathrm{d}t, \tag{2.34}$$

where

$$\boldsymbol{\vartheta}(t) = \mathbf{x} + t(\cos(\upsilon_0), \sin(\upsilon_0)). \tag{2.35}$$

The corresponding variance measure is given as

$$\mathrm{Var}^*(u, \mathbf{x}, \upsilon_0) = {l_{\mathrm{s}}}^{-1} \int_{-l_{\mathrm{s}}/2}^{l_{\mathrm{s}}/2} (u(\boldsymbol{\vartheta}(t)) - \Sigma^*(u, \boldsymbol{\vartheta}(t), \upsilon_0))^2 \mathrm{d}t. \tag{2.36}$$

Given (2.36), we choose our estimates υ to match the minimum-variance orientations of the image. In that context, l_{s} can be interpreted as a scale parameter that

[7] Replacing $\mathcal{G}$ by ∇ in (2.27) would cause EED to degenerate to the isotopic flow (2.26).

is approximately determined from the characteristic oriented-feature size. At each position, the solution thus corresponds to a local edge-orientation estimate that is expressed as

$$\upsilon(u,\mathbf{x}) = \arg\min_{\upsilon_0} \mathrm{Var}^*(u,\mathbf{x},\upsilon_0), \tag{2.37}$$

which satisfies translation and rotation invariance with respect to u.

At given scale l_s, our variance measure quantifies the image fluctuations along each orientation υ_0. In that respect, the minimum argument of (2.37) is conceptually similar to the coherence direction defined in [39] that is based on structure tensors. The advantage of variance-based criteria is their ability to estimate the orientation even in the vicinity of a contrast change [59]. As shown in Figure 2.2, the map υ provides accurate data on the local feature orientations of the image. Adaptive filtering of the gradient map along those directions can thus reduce the loss of information associated with cross-edge smoothing. Because no closed-form solution of (2.37) exists to compute the weights, we propose to optimize υ among a discrete set of N_o orientations. Note that similar discretization approaches have been considered using candidate stencils for the evaluation of image variations along oriented paths [29].

Prior to filtering, we make the gradient map of u consistent with the estimated orientation map υ pointwise. Accordingly, we only keep the component of each gradient vector $\nabla u(\mathbf{x})$ which is perpendicular to the orientation $\upsilon(u,\mathbf{x})$. This projection operation yields the corrected gradients

$$\nabla^\mathrm{c} u(\mathbf{x}) = \big((\nabla u)(\mathbf{x})^\mathrm{T}\mathbf{e}^\perp(u,\mathbf{x})\big)\,\mathbf{e}^\perp(u,\mathbf{x}), \tag{2.38}$$

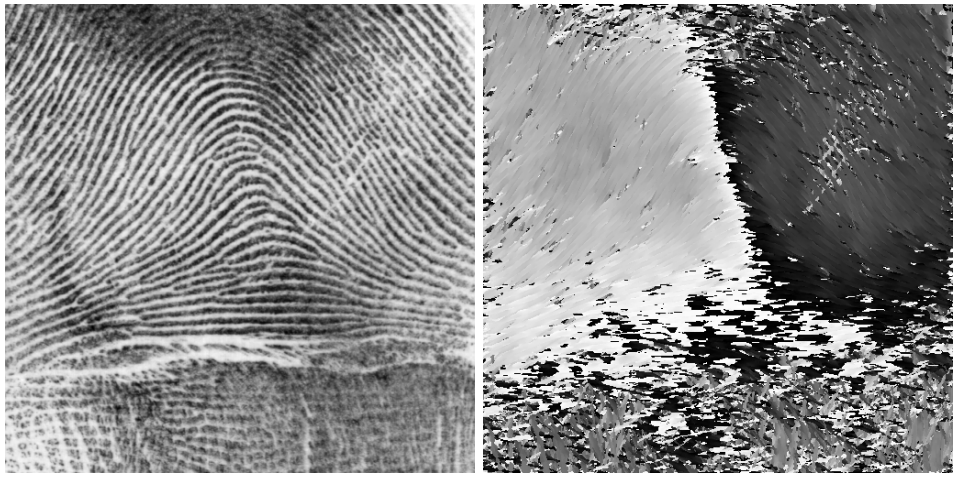
where

$$\mathbf{e}^\perp(u,\mathbf{x}) = (-\sin(\upsilon(u,\mathbf{x})),\cos(\upsilon(u,\mathbf{x}))). \tag{2.39}$$

We finally smooth these corrected gradients along the corresponding υ, using directional averaging filters with the same l_s as in the mean and variance measures of (2.34) and (2.36). The operator $\mathcal{G}_1$ thus acts as

$$\mathcal{G}_1 u(\mathbf{x}) = \Sigma^*(\nabla^\mathrm{c} u,\mathbf{x},\upsilon(u,\mathbf{x})). \tag{2.40}$$

The invariances of (2.34) and (2.37) and the pointwise character of (2.38) imply that $\mathcal{G}_1$ is intrinsically translation and rotation invariant as in $\mathcal{G}_0$. These important



(a) Original *Fingerprint* image (b) Directions of minimum variance

Figure 2.2: Illustration of our orientation-estimation method. The directions v of minimum variance are obtained from the corresponding image according to (2.37) with $l_s = 25$. The directions $[0, \pi[$ are mapped to the grayscale range $[\text{black}, \text{white}[$. This map contains two large zones that are associated with the two main feature directions of the original image. Note that abrupt black-white transitions appear because orientations are only defined modulo π.

characteristics guarantee that $\mathcal{G}$ is not biased towards particular orientations or positions in the image.

2.5.4 AIRLS algorithm

As previously discussed, the quadratic functionals of IRLS are linked to the lagged-diffusivity forms of the gradient flow. Similarly, we propose an IRLS procedure that is based on lagged versions of the EED flow, following the concepts introduced in Section 2.4. Despite the fact that there is no underlying maximization principle in our case, the successive functionals that we define constitute linear approximations

of (2.27). Given one current estimate $u^{(n)}$, the lagged EED equation is

$$\partial_t u = \operatorname{div}\left(\mathbf{T}(\mathcal{G}u^{(n)}, \psi)\nabla u\right). \tag{2.41}$$

According to the Euler-Lagrange equation, (2.41) is the gradient flow which originates from the functional

$$\mathcal{R}_{\mathrm{A}}(u|u^{(n)}) = \int_{\mathbb{R}^2} \left\langle \mathbf{T}(\mathcal{G}u^{(n)}(\mathbf{x}), \psi)\nabla u(\mathbf{x}), \nabla u(\mathbf{x}) \right\rangle \mathrm{d}\mathbf{x}, \tag{2.42}$$

up to a factor of 2 that we drop for convenience. Note that the regularizer proposed in [60] is of similar form; it includes structure tensors that are based on fixed estimates, but it is nonquadratic unlike (2.42).

Expression (2.42) can be rewritten in a more intelligible form. Indeed, defining $\nabla_{\cdot}$ and $\nabla^{\perp}{}_{\cdot}$ as directional derivatives[8] along the subscript vector argument and the direction perpendicular to it, respectively,

$$\begin{aligned} \langle \mathcal{P}_{\mathbf{v}} \nabla u(\mathbf{x}), \nabla u(\mathbf{x}) \rangle &= ((\nabla_{\mathbf{v}} u)(\mathbf{x}))^2, \\ \left\langle \mathcal{P}^{\perp}{}_{\mathbf{v}} \nabla u(\mathbf{x}), \nabla u(\mathbf{x}) \right\rangle &= ((\nabla^{\perp}{}_{\mathbf{v}} u)(\mathbf{x}))^2, \end{aligned} \tag{2.43}$$

which implies from (2.28) that $\mathcal{R}_{\mathrm{A}}(\cdot|u^{(n)})$ expands as

$$\begin{aligned} \mathcal{R}_{\mathrm{A}}(u|u^{(n)}) &= \int_{\mathbb{R}^2} \psi(\|\mathcal{G}u^{(n)}(\mathbf{x})\|)(\nabla_{\mathcal{G}u^{(n)}(\mathbf{x})} u)^2(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\quad + \int_{\mathbb{R}^2} (\nabla^{\perp}{}_{\mathcal{G}u^{(n)}(\mathbf{x})} u)^2(\mathbf{x})\mathrm{d}\mathbf{x}. \end{aligned} \tag{2.44}$$

Equation (2.44) gives further insight on how each quadratic regularizer—which is linked to the linearized form of the EED flow—penalizes u. In the first integral term, the directional derivatives of u that are parallel to $\mathcal{G}u^{(n)}$, i.e., perpendicular to the edge features, are weakly penalized given their multiplication with ψ. Edges are therefore well-preserved as in nonquadratic regularization. Simultaneously, regularity along those same edge estimates is strongly enforced in the second

[8] Accordingly, the expressions $\nabla_{\mathbf{v}} u$ and $\nabla^{\perp}{}_{\mathbf{v}} u$ appearing in (2.43) correspond to scalar fields.

integral term, which is akin to an L_2-norm. These two simultaneous constraints favor curvature regularity of the solution.

In order to obtain an IRLS procedure compatible with the framework of Algorithm 2.1, we discretize the cost (2.42) and combine it with our data term. Indeed, as for standard regularization techniques, computationally tractable approaches require a discretized version of the continuous quadratic costs of the form $\mathcal{R}_{\mathrm{A}}(\cdot|u^{(n)})$. Using the same discretization as in (2.15), the expression

$$\mathcal{R}^0_{\mathrm{A}}(\tilde{c}|\tilde{c}^{(n)}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \left\langle \mathbf{T}(\mathring{\mathcal{G}}\tilde{f}^{(n)}(\mathbf{x}), \psi)\mathring{\nabla}\tilde{f}(\mathbf{x}), \mathring{\nabla}\tilde{f}(\mathbf{x}) \right\rangle_{\mathbf{x}=\mathbf{k}} \tag{2.45}$$

is obtained for our solution coefficients. Along with (2.18), this definition allows to rewrite (2.45) as

$$\mathcal{R}^0_{\mathrm{A}}(\tilde{c}|\tilde{c}^{(n)}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} (\tilde{c}\star\mathbf{r})[\mathbf{k}]^{\mathrm{T}}\boldsymbol{\theta}(\tilde{c}^{(n)}, \psi)[\mathbf{k}](\tilde{c}\star\mathbf{r})[\mathbf{k}], \tag{2.46}$$

where the tensor weights $\boldsymbol{\theta}$ are determined as

$$\boldsymbol{\theta}(\tilde{c}^{(n)}, \psi)[\mathbf{k}] = \mathbf{T}(\mathring{\mathcal{G}}\tilde{f}^{(n)}(\mathbf{x}), \psi)|_{\mathbf{x}=\mathbf{k}}. \tag{2.47}$$

Equation (2.47) evaluates $\mathring{\mathcal{G}}\tilde{f}^{(n)}$ over a sequence of points $\mathbf{x} = \mathbf{k}$. The corresponding means and variances used for $\mathcal{G}_1$ in (2.37) and (2.40) can be determined with arbitrary precision, using the continuous-line integrals (2.34) and (2.36) for $\tilde{f}^{(n)}$ in the same way as for u in Section 2.5.3. This holds because $\tilde{f}^{(n)}$ is continuously defined from $\tilde{c}^{(n)}$ given our spline model (2.4). For computational reasons, we approximate all integrals by finite sums depending on uniformly spaced samples. This discretization also allows to compute (2.37) recursively: each of the N_{o} oriented-variance maps linked to $\mathring{\mathcal{G}}\tilde{f}^{(n)}$ can be estimated using parallel sliding windows of size l_{s} and orientation υ_0. The samples of $\tilde{f}^{(n)}$ and the local variance assigned to each window are then updated recursively[9]. Equations (2.11), (2.46), and (2.47) define our total quadratic functionals as

$$\mathcal{J}^0_{\mathrm{A}}(\tilde{c}|\tilde{c}^{(n)}) = \mathcal{D}(\tilde{c}) + \Lambda\mathcal{R}^0_{\mathrm{A}}(\tilde{c}|\tilde{c}^{(n)}). \tag{2.48}$$

[9]The computational performance is optimal when these updates are performed along rows or columns of values. We maintain this condition for $\upsilon_0 \notin \{0, \pi/2\}$ by applying approximate pre- and post-shearing transformations on the sample lattices.

According to the generic optimization framework of Algorithm 2.1, our algorithm is first initialized with the solution $\tilde{c}^{(0)}$. We define this sequence as the masked image samples γ upsampled by $\mathcal{M}$ using zero padding and smoothed by the filter $\frac{1}{4}[1\ 2\ 1]$ along each dimension. Subsequently, we partially minimize N_i successive costs (2.48) to obtain a solution. Each of these minimization steps corresponds to a fixed-point iteration where the next estimate $\tilde{c}^{(n+1)}$ is found from $\tilde{c}^{(n)}$ using fixed tensor diffusivities $\boldsymbol{\theta}$. In that regard, our approach is similar to the anisotropic technique proposed in [60] where a series of convex problems is solved to denoise images. In our case, however, the successive functionals to minimize are quadratic and thus easier to tackle. Moreover, since (2.46) is an ℓ_2-norm whose weights $\boldsymbol{\theta}$ depend on $\tilde{c}^{(n)}$, our reconstruction method is of IRLS type. We call it AIRLS according to the anisotropic nature of the tensor weights in (2.47). Each cost $\mathcal{J}_\mathrm{A}^0(\cdot|\tilde{c}^{(n)})$ has a unique minimum that satisfies the first-order condition

$$\Lambda^{-1}\frac{\partial}{\partial\tilde{c}}\mathcal{D}(\tilde{c}) + (\mathbf{r}[\cdot])^\mathrm{T} \star \Big(\boldsymbol{\theta}(\tilde{c}^{(n)},\psi)(\tilde{c}\star\mathbf{r})\Big) = 0. \tag{2.49}$$

Ultimately, any fixed point of our global iterative process satisfies

$$\Lambda^{-1}\frac{\partial}{\partial\tilde{c}}\mathcal{D}(\tilde{c}) + (\mathbf{r}[\cdot])^\mathrm{T} \star (\boldsymbol{\theta}(\tilde{c},\psi)(\tilde{c}\star\mathbf{r})) = 0. \tag{2.50}$$

In analogy with the case of Section 2.4, Condition (2.50) is the steady state of a discretized PDE whose regularization part corresponds to the EED flow (2.27). This shows that, in terms of asymptotic solutions, our approach is similar to conventional PDE-based methods. It is, however, more attractive computationally because the computation of the smoothed-gradient map as well as the nonlinear operations involved in the diffusivity estimations (2.47) are restrained to the reweightings. As confirmed in our experiments, satisfactory results are obtained with a small amount of reweightings N_i, precisely as in standard IRLS.

The AIRLS algorithm is specified by the parameters $\{\mathcal{G},\psi\}$. The first argument $\mathcal{G}$ corresponds either to the Gaussian-smoothed gradient $\mathcal{G}_0$ defined in (2.33), or to our modified operator $\mathcal{G}_1$ defined in (2.40). Similarly, the function ψ can be chosen as the Charbonnier diffusivity ψ_0 defined in (2.30), the Huber diffusivity ψ_1 defined in (2.31), or the Perona-Malik diffusivity ψ_2 defined in (2.32). Our specific EED settings denoted by EED_1 and EED_2 correspond to $\{\mathcal{G}_1,\psi_1\}$ and $\{\mathcal{G}_1,\psi_2\}$, respectively. When using $\{\mathcal{G}_0,\psi_0\}$, our algorithmic framework reproduces the PDE-based method of [35]; the related technique is then called EED_0.

2.6 Linear problems

Minimizing each weighted quadratic cost (2.48) amounts to solving a linear system. In this section, we derive the explicit form of those systems using matrix notation. Accordingly, the generic form of each matrix system is

$$\mathbf{S}^{(n)}\tilde{\mathbf{c}}^{(\min)} = \mathbf{y}, \tag{2.51}$$

where $\tilde{\mathbf{c}}^{(\min)}$ contains the lexicographically ordered coefficients of the minimizer, where $\mathbf{S}^{(n)}$ is the system matrix that depends on our estimate $\tilde{\mathbf{c}}^{(n)}$, and where $\mathbf{y}$ is a vector whose components are constant.

The first step towards specifying $\mathbf{S}^{(n)}$ and $\mathbf{y}$ is to reformulate (2.48) using matrix notation. Accordingly, and given the form of $\mathbf{A}$ in (2.10),

$$\begin{aligned}
\mathcal{D}(\tilde{\mathbf{c}}) &= \|\mathbf{A}\tilde{\mathbf{c}} - \boldsymbol{\gamma}\|_{\ell_2}^2 \\
&= (\boldsymbol{\gamma} - \boldsymbol{\chi}\mathbf{D}_{\mathcal{M}}\mathbf{B}\tilde{\mathbf{c}})^{\mathrm{T}}(\boldsymbol{\gamma} - \boldsymbol{\chi}\mathbf{D}_{\mathcal{M}}\mathbf{B}\tilde{\mathbf{c}}) \\
&= \tilde{\mathbf{c}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathcal{M}}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\chi}\mathbf{D}_{\mathcal{M}}\mathbf{B}\tilde{\mathbf{c}} - 2\,\tilde{\mathbf{c}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathcal{M}}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\gamma} + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\gamma} \\
&= \tilde{\mathbf{c}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathcal{M}}\mathbf{W}\mathbf{D}_{\mathcal{M}}\mathbf{B}\tilde{\mathbf{c}} - 2\,\tilde{\mathbf{c}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathcal{M}}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\gamma} + \text{const.},
\end{aligned} \tag{2.52}$$

where each vector is specified as discussed above, and where the diagonal matrix $\mathbf{W} = \boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\chi}$ is associated with point-wise multiplication with the weights χ used for masking in (2.2). Similarly, we can write the regularization term (2.46) in the compact form

$$\mathcal{R}_{\mathrm{A}}^{0}(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) = (\mathbf{R}\tilde{\mathbf{c}})^{\mathrm{T}}\boldsymbol{\Theta}(\tilde{\mathbf{c}}^{(n)}, \psi)\mathbf{R}\tilde{\mathbf{c}}, \tag{2.53}$$

where $\mathbf{R}$ and $\boldsymbol{\Theta}$ concatenate convolution and diagonal matrices, respectively. The rectangular matrix $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2)$ implements the gradient. Specifically, each convolution matrix $\mathbf{R}_i$ relates to the derivative component r_i of the multivariate filter $\mathbf{r}$ defined in (2.18). The square matrix $\boldsymbol{\Theta}$ is updated according to the current estimate $\tilde{\mathbf{c}}^{(n)}$; it decomposes as $(\boldsymbol{\Theta}_{11}\ \boldsymbol{\Theta}_{12}, \boldsymbol{\Theta}_{12}\ \boldsymbol{\Theta}_{22})$, where each diagonal matrix $\boldsymbol{\Theta}_{ij}$ is associated with point-wise multiplication with the corresponding scalar sequence $\theta_{\mathrm{A}ij}$ related to the tensor weights $\boldsymbol{\theta}$ of (2.47). Note that $\boldsymbol{\Theta}$ structurally extends its counterpart in IRLS, as it concatenates distinct and off-diagonal sub-matrices $\boldsymbol{\Theta}_{ij}$. Based on (2.52) and (2.53), we write the total cost as

$$\mathcal{J}_{\mathrm{A}}^{0}(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) = \mathcal{D}(\tilde{\mathbf{c}}) + \Lambda\mathcal{R}_{\mathrm{A}}^{0}(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}). \tag{2.54}$$

Since (2.54) is quadratic, its gradient with respect to $\tilde{\mathbf{c}}$ vanishes at $\tilde{\mathbf{c}}^{(\min)}$. Enforcing this condition, and using matrix differentiation, we obtain

$$\mathbf{S}^{(n)} = \mathbf{B}^{\mathrm{T}}\overline{\mathbf{W}}\mathbf{B} + \mathbf{R}^{\mathrm{T}}\overline{\mathbf{\Theta}}\mathbf{R}, \tag{2.55}$$

where $\overline{\mathbf{W}} = \mathbf{U}_{\mathcal{M}}\mathbf{W}\mathbf{D}_{\mathcal{M}}$, and where $\overline{\mathbf{\Theta}} = \Lambda\mathbf{\Theta}(\tilde{\mathbf{c}}^{(n)}, \psi)$. Finally, the vector $\mathbf{y}$ in (2.51) corresponds to

$$\mathbf{y} = \mathbf{B}^{\mathrm{T}}\mathbf{U}_{\mathcal{M}}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\gamma}. \tag{2.56}$$

The matrix system (2.55) is extremely large due to the considerable number of unknowns, which implies that Problem (2.51) cannot be solved exactly. However, as stated in Section 2.4.2, the corresponding quadratic functional (2.48) needs only be partially minimized with respect to the current solution estimate to yield $\tilde{\mathbf{c}}^{(n+1)}$. In Section 2.7, we propose to partially solve (2.51) iteratively, initializing the next solution $\tilde{\mathbf{c}}^{(n+1)}$ to the current estimate $\tilde{\mathbf{c}}^{(n)}$.

2.7 Iterative solution

As mentioned above, our approach allows to restrain the diffusivity estimations associated with the regularization problem in the reweightings. This allows to focus on the resolution of the subproblems derived in Section 2.6. In the sequel, we devise a fast iterative method to partially solve these linear systems, considering their sparse structure. In particular, the diagonal matrices entering the definition of (2.55) are well suited for multigrid solvers.

2.7.1 Multigrid approach

The multigrid strategy consists in solving problems by iterating not only at their nominal scale, but also at coarser ones, adapting their discretization accordingly. Multigrid iterative methods are beneficial for certain types of problems where the additional lower-resolution iterates are of negligible computational cost compared to their overall contribution in terms of convergence rates [61]. In particular, iterating at successively downscaled Cartesian grids is efficient for linear image reconstruction

from spatially sparse samples [30, 31]. In this context, according to the original formulation (2.51), we define our linear problems at N_{g} distinct grids, using the notation

$$\mathbf{S}^{\kappa}\tilde{\mathbf{c}}^{\kappa} = \mathbf{y}^{\kappa}, \tag{2.57}$$

where the superscripts $\kappa \in \{0, \ldots, N_{\mathrm{g}} - 1\}$ relate quantities to a specific grid Ω^{κ}. Each grid is constructed with a regular step 2^{κ} in each dimension, which means that the number of elements of $\tilde{\mathbf{c}}^{\kappa}$ scales as $4^{-\kappa}$.

From (2.57), we use Full-Multigrid V-cycles [61] as an iterative scheme to find a solution $\tilde{\mathbf{c}}^{(n+1)}$. This method is standard in the literature and involves transfer operations as well as iterations at each grid, as described in Appendix 2.10.1 with the relevant definitions. In order to maximize the performance of our approach, we use the obtained $\tilde{\mathbf{c}}^{\kappa}$ to initialize the next linear problem at all grids. The problem at grid Ω^0 is (2.51), which implies that $\mathbf{S}^0 = \mathbf{S}^{(n)}$, $\mathbf{y}^0 = \mathbf{y}$, and $\tilde{\mathbf{c}}^0 = \tilde{\mathbf{c}}^{(n+1)}$. At coarser grids, $\mathbf{S}^{\kappa}$ are scaled versions of $\mathbf{S}^0$, while $\mathbf{y}^h$ are the *residuals* produced from the iterative process itself.

Since the solution is expressed as coefficients in a B-spline basis, the corresponding *prolongation* and *restriction* operators $\mathbf{I}^{\triangle}$ and $\mathbf{I}^{\triangledown}$ exploit the two-scale relations [47]. Specifically, they correspond to the B-spline scaling filter h_2 of degree η following upsampling by 2, and to the B-spline scaling filter ${h_2}^{\mathrm{T}}$ followed by downsampling by 2, respectively. In matrix form, we write

$$\begin{aligned} \mathbf{I}^{\triangle} &= \mathbf{H}_2\mathbf{U}_2, \\ \mathbf{I}^{\triangledown} &= \mathbf{D}_2{\mathbf{H}_2}^{\mathrm{T}}. \end{aligned} \tag{2.58}$$

We now have to specify $\mathbf{S}$ at each grid. In order not to complexify the problem formulation, we impose similar matrix structures at all grids, decomposing $\mathbf{S}^{\kappa}$ as

$$\mathbf{S}^{\kappa} = {\mathbf{B}^{\kappa}}^{\mathrm{T}}\overline{\mathbf{W}}^{\kappa}\mathbf{B}^{\kappa} + {\mathbf{R}^{\kappa}}^{\mathrm{T}}\overline{\mathbf{\Theta}}^{\kappa}\mathbf{R}^{\kappa}, \tag{2.59}$$

where the separate terms at Ω^0 correspond to the ones of (2.55).

In order to specify the data part of $\mathbf{S}^{\kappa}$, we build a weight pyramid, starting from the available fine-scale matrix $\overline{\mathbf{W}}^0$. Simplifying the coarser-scale convolution matrices $\mathbf{B}^{\kappa}$ as identity, we express the diagonal elements of $\overline{\mathbf{W}}^1$ as

$$\overline{w}^1[\mathbf{k}] = \left\{\overline{w}^0 \star {h_2}^{\vee} \star b^{\vee}\right\}_{\downarrow 2}[\mathbf{k}], \tag{2.60}$$

where $^{\vee}$ flips a given sequence as $\cdot^{\vee}[\mathbf{k}] = \cdot[-\mathbf{k}]$. The convolutive effect of $\mathbf{B}$ is thus taken into account at this first weight level. For $\kappa > 1$, the expression of $\overline{w}^{\kappa}$ takes the simpler form

$$\overline{w}^{\kappa+1}[\mathbf{k}] = \left\{\overline{w}^{\kappa} \star {h_2}^{\vee}\right\}_{\downarrow 2}[\mathbf{k}]. \tag{2.61}$$

Regarding the regularization term, the components of the coarse-scale diagonal sub-matrices $\overline{\mathbf{\Theta}}_{ij}^{\kappa}$ involved in (2.59) are obtained as in (2.61) through the relations

$$\overline{\theta}_{ij}^{\kappa+1}[\mathbf{k}] = \left\{\overline{\theta}_{ij}^{\kappa} \star {h_2}^{\vee}\right\}_{\downarrow 2}[\mathbf{k}], \tag{2.62}$$

while $\mathbf{R}^{\kappa}$ is defined as

$$\mathbf{R}^{\kappa} = 2^{-\kappa}\mathbf{R}, \tag{2.63}$$

according to the scaling properties of the gradient operator.

2.7.2 Successive over-relaxation

The Full-Multigrid V-cycles involve iterations at distinct grids κ as part of two distinct cycle phases. According to the definitions of the prolongation and restriction operators, the current phase $\wp$ can be ascending or descending, with $\wp \in \{\triangle, \triangledown\}$. Based on the knowledge of κ and $\wp$, the iterative technique that we use to solve the linear problems (2.57) is parameterized with distinct numbers of iterations $N_{\mathrm{i}}^{*}(\kappa, \wp)$ and relaxation constants $\omega(\kappa, \wp)$. For convenience, we denote the relaxation constants by ω when referring to them in a generic sense.

Given the symmetry and positive-definitness of the $\mathbf{S}^{\kappa}$ in (2.57), a certain class of iterative methods can be used, including the well-known Conjugate Gradient (CG). The successive over-relaxation (SOR) technique [62] is especially efficient for our multigrid problem. It corresponds to a damped version of the Gauss-Seidel iterative method, and its convergence is guaranteed for $\omega \in [0, 2]$. Given $\mathbf{S}^{\kappa}$ and $\mathbf{y}^{\kappa}$, the SOR iterate at grid Ω^{κ} is defined as

$$\tilde{\mathbf{c}}^{\kappa} \leftarrow \tilde{\mathbf{c}}^{\kappa} + \omega(\mathbf{S}_{\mathrm{D}}^{\kappa} + \omega\mathbf{S}_{\mathrm{L}}^{\kappa})^{-1}\mathrm{R}(\Omega^{\kappa}), \tag{2.64}$$

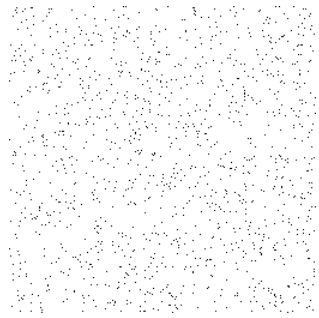


Figure 2.3: Masking process. The pseudo-random binary mask shown above is applied before interpolation and only keeps 2% of the original data; the average gap between the retained samples corresponds to 4.3 pixels.

where $\mathbf{S}_{\mathrm{D}}^{\kappa}$ and $\mathbf{S}_{\mathrm{L}}^{\kappa}$ stand for the diagonal and strictly lower triangular parts of the matrix $\mathbf{S}^{\kappa}$, respectively.

Unlike exact resolution, this iterative approach only involves partial matrix inversions. Structurally, each iteration is performed by updating the coefficient vector $\tilde{\mathbf{c}}^{\kappa}$ componentwise. The sparse structure of $\mathbf{S}^{\kappa}$ makes these updates correspond to space-domain operations of complexity $\mathcal{O}(N \log N)$ at most.

2.8 Experiments

In this section, we compare our AIRLS approach with respect to the state of the art, considering interpolation experiments on grayscale images[10]. Our implementation has been coded in Java, and run on Mac OS X with a Quad-Core Xeon 2×2.8 GHz and 4 GB of DDR2 memory. The computation of the optimal orientations in (2.37) for $\mathcal{G}_1$ and the SOR updates in (2.64) are parallelized using multithreading. The state-of-the-art methods that are considered for comparison are also based on

[10]The *CT* image is part of the Dicom stack CT HEAD-NK 5.0 B30s (Keith E. Blackwell, M.D.). The standard *Bird* image is found at `http://www2.isye.gatech.edu/~brani/images/bird.gif`, while the rest of the original data belongs to the GCF-BM3D set found at `http://www.cs.tut.fi/~foi/GCF-BM3D`. The image histograms are rescaled to $[0, 255]$ for all experiments.

Setting	Ideal Interpolation					Generalized Interpolation		
Method	GREY	EED	EED_0	EED_1	EED_2	EED_0	EED_1	EED_2
Runtime [s]	6.09*	134.90	**3.62**	5.73	5.00	**4.22**	6.52	6.60
Bird	19.86	20.45	20.77	**21.02**	20.42	**23.28**	22.98	22.49
Cameraman	17.37	17.68	17.58	**17.90**	17.40	19.58	**19.73**	19.50
CT	20.66	21.21	21.14	**21.88**	21.28	22.76	**23.68**	23.59
House	18.47	18.76	18.71	**19.73**	19.60	21.07	**21.62**	21.31
Lena (crop)	18.38	19.39	19.34	**20.22**	19.61	21.12	**21.87**	21.70
Montage	17.66	18.57	18.52	**18.67**	18.15	20.11	**20.38**	20.17
Peppers	19.00	18.97	19.12	**19.15**	18.49	20.77	**21.21**	20.80

* These average runtimes correspond to distinct implementations.

Table 2.1: Numerical results (PSNR) for the interpolation experiments.

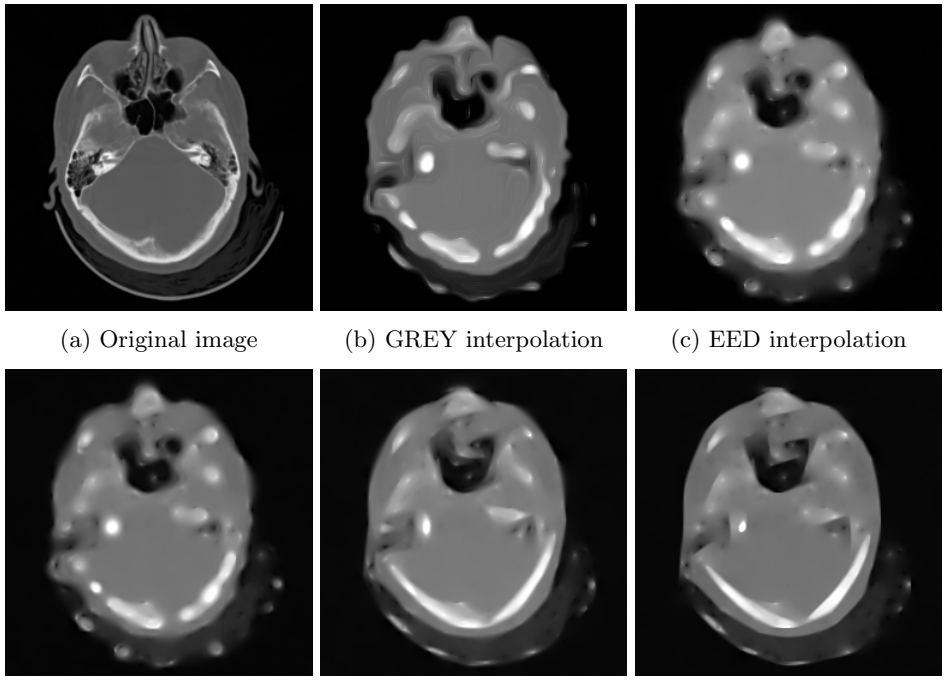
parallel implementations[11].

We consider B-splines of order $\eta = 2$, and compute $N_\text{o} = 16$ orientations when using the operator $\mathcal{G}_1$ in (2.27). Following the generic IRLS procedure of Algorithm 2.1, our algorithm reconstructs images by solving $N_\text{i} = 10$ successive linear problems that are defined on $N_\text{g} = 4$ grids in (2.57). For each linear problem, the iteration and relaxation constants common to all experiments are $\omega(\kappa, \triangle) = 1.5, N_\text{i}^*(\kappa, \bigtriangledown) = 2, N_\text{i}^*(\kappa, \triangle) = 1, \forall\kappa$, and $\omega(\kappa, \bigtriangledown) = 1.5, \forall\kappa > 0$. The constants β_1 and β_2 involved in the diffusivites (2.31) and (2.32) are scaled according to the dynamic range DR of the image under consideration. For convenience, we display unit-resampled versions of the continuous reconstructions [27]. The PSNR measures are evaluated on the central portion of the images (80%) so as to discard the influence of the boundary conditions.

2.8.1 Sparse interpolation of ideal samples

In these experiments, we interpolate grayscale images from 2% of their samples, according to one realization of a binary random mask. The sampling is ideal, meaning that the prefilter φ_0 used in our generalized model (2.1) reduces to the Dirac distribution $\delta(\cdot)$. The other parameters are $\Lambda = 0.01, \mathcal{M} = 1, l_\text{s} = 25, l_\sigma = 4, N_\text{v} = 1, \omega(0, \bigtriangledown) = 1.95, \beta_1 = 2 \cdot 10^{-3}\text{DR}$, and $\beta_2 = 8 \cdot 10^{-2}\text{DR}$. We consider sets

[11] The implementations that are distinct from our method sometimes run on different platforms, and can differ in their level of optimization and parallelization. Caution should therefore be exerted to not overinterpret the runtime results reported in Tables 2.1 and 2.2.



(a) Original image (b) GREY interpolation (c) EED interpolation

(d) EED_0 interpolation (e) EED_1 interpolation (f) EED_2 interpolation

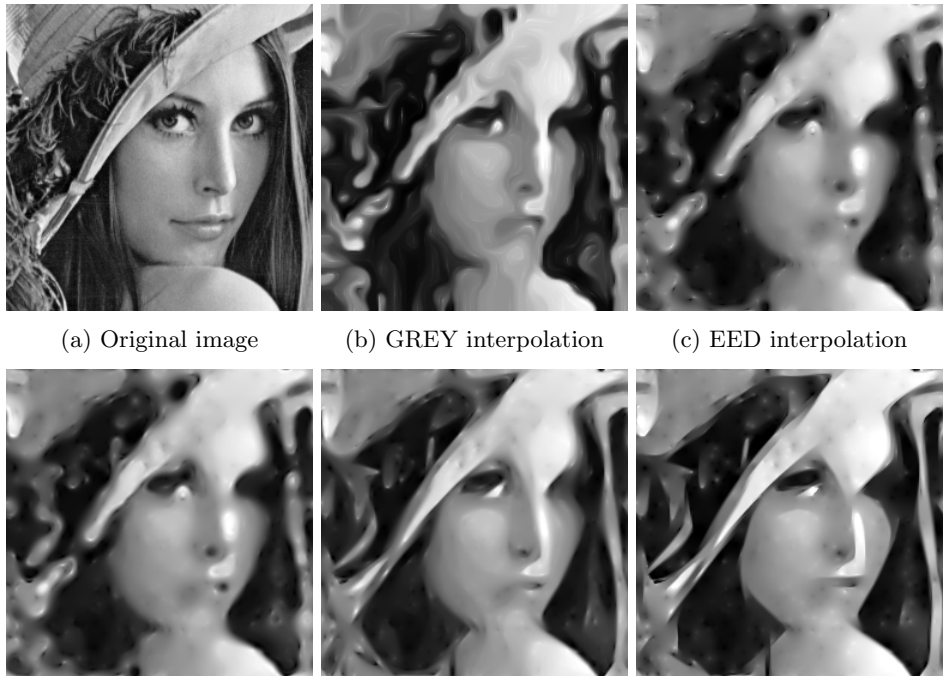
Figure 2.4: Ideal interpolation of *CT* (256×256 crop) from 2% of samples.

of 256×256 images, the corresponding binary mask being shown in Figure 2.3.

We compare our three EED_i methods with the fast PDE-based approach of Tschumperlé [37] implemented in version 2.9 of *GREYCstoration*[12] (GREY). We have also implemented a PDE-based version of the EED flow using explicit time steps; this approach is referred to as EED in the sequel.

Quantitative and visual results for these algorithms are provided for the interpolation of several images in Table 2.1 and in Figures 2.4 and 2.5, respectively.

[12]This code is run with 15 iterations under default settings. The resulting PSNR tends to degrade when iterating further.

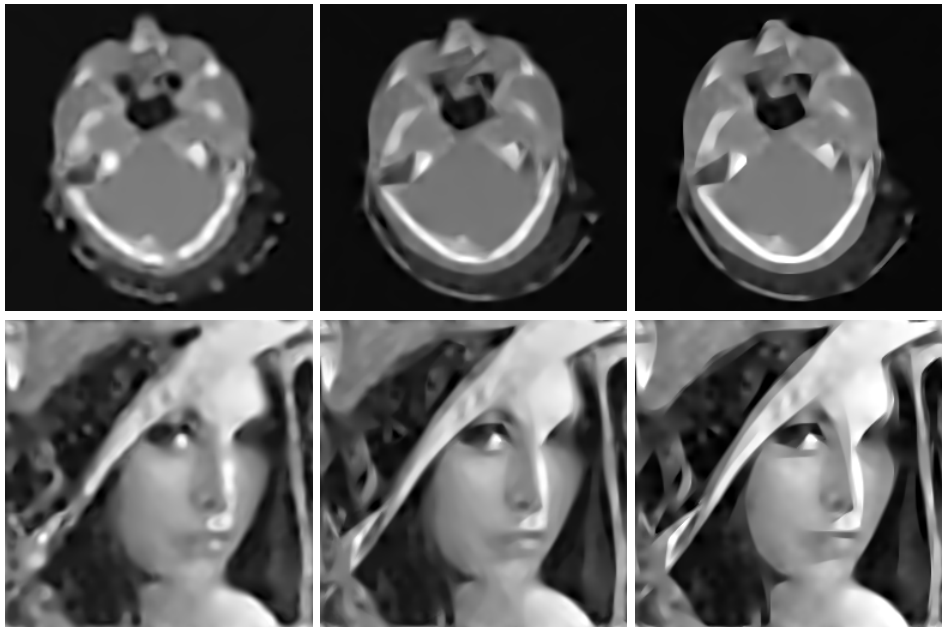


(a) Original image (b) GREY interpolation (c) EED interpolation

(d) EED_0 interpolation (e) EED_1 interpolation (f) EED_2 interpolation

Figure 2.5: Ideal interpolation of *Lena* (256×256 crop) from 2% of samples.

Observe that our specific tensor-estimation approach in EED_1 restores directional features better than GREY, EED, and EED_0, and yields higher SNR values. Our EED_2 method is quantitatively inferior to EED_1 but restores very sharp edges as can be seen in Figures 2.4 and 2.5. Its regularization behavior is consistent with the properties of the Perona-Malik diffusivity mentioned in Section 2.5.2. Our PDE-based implementation of EED requires 5000 time steps for convergence. It is much slower than EED_0, as shown in Table 2.1, but yields similar results in terms of PSNR and visual appearance. This corroborates the fixed-point interpretations discussed in Sections 2.4 and 2.5.



(a) EED_0 interpolation (b) EED_1 interpolation (c) EED_2 interpolation

Figure 2.6: Generalized interpolation of *CT* and *Lena* (256×256 crop) from 2% of samples acquired with a prefilter φ_0.

These results demonstrate the suitability of our EED_i methods to restore geometrical information from few image samples. In general, our approach is less successful at restoring textures because, in several cases, the latter are composed of repetitive patches rather than well-defined oriented features. To some extent, this discrepancy between geometrical-information and texture restoration is due to the fact that image structures and textures correspond to dual functional spaces [63]. Note that, although our EED_i methods are able to restore a substantial amount of geometrical features in an image, they may not recover certain types of 2D junctions because the associated diffusion tensors can only represent gradient information in

Method	Quad.	TV	R&M	EED_0	EED_1	EED_2
Runtime [s]	**1.31**	5.32*	1.88*	5.46	7.45	7.59
Bird	27.22	28.16	28.31	27.99	28.43	**28.68**
Cameraman	21.73	22.37	22.16	21.99	22.18	**22.42**
CT	27.11	27.67	28.75	27.53	**28.80**	28.67
House	25.74	26.46	**26.60**	25.98	26.55	26.54
Lena (crop)	24.81	25.19	25.70	25.54	**25.85**	25.70
Montage	21.85	22.20	22.04	22.03	22.28	**22.30**
Peppers	25.36	25.80	26.21	25.85	26.38	**26.54**

* These average runtimes correspond to distinct implementations.

Table 2.2: Numerical results (PSNR) for the magnification experiments.

a symmetric form [64]. This issue remains open to future research.

2.8.2 Sparse interpolation of generalized samples

In this second part, we wish to reconstruct the same images from 2% of their samples, considering here generalized sampling $\varphi_0 = \frac{1}{49}\,\text{rect}(\frac{\cdot}{7})$. The EED approach of [35] is not applicable here; it can be substituted with our EED_0 method as implemented in our more general AIRLS framework.

Using the same mask and parameters as above, the numerical results are provided in Table 2.1 and the visual results are shown in Figure 2.6 for our three EED_i methods. The improvement of EED_1 relative to the other methods is comparable with the ideal case in terms of PSNR and visual quality. Remarkably, all results are better in terms of PSNR than their counterparts in the ideal-sampling setting. As a matter of fact, the analysis function φ_0 acts as an anti-aliasing filter before sampling, which causes the overall reconstructed features to be more consistent with the original image. This emphasizes the interest of using generalized sampling for sparse interpolation.

2.8.3 Image magnification

In our framework, image magnification corresponds to a particular instance of sparse interpolation where the sparsity is regular and typically low. Our approach is well suited to that problem because the presence of a prefilter before sampling is inherent in practical acquisition devices [9]. For that case, we compare our EED_i

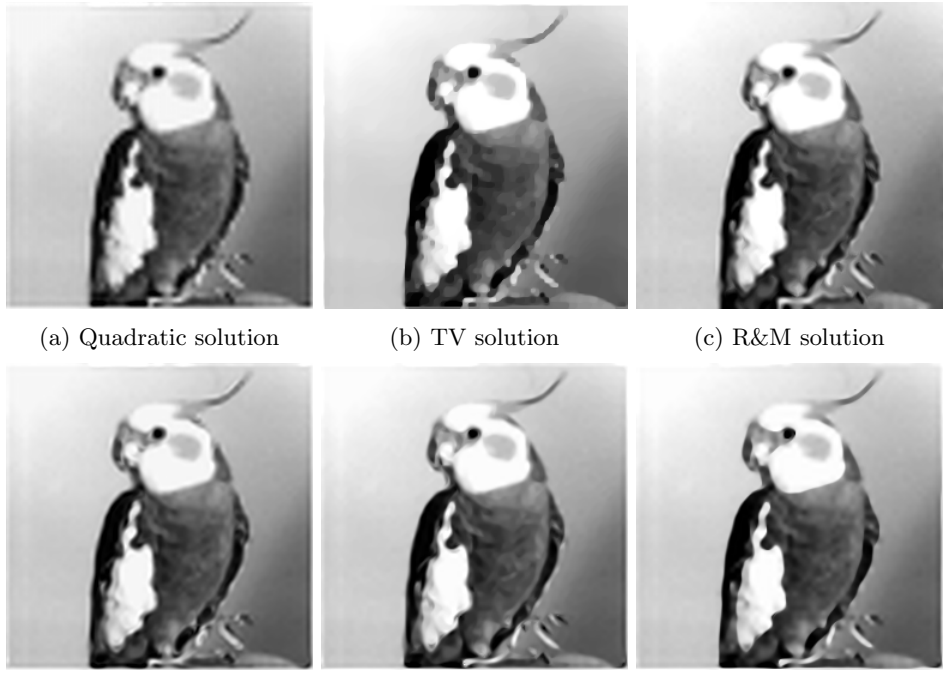
methods with quadratic regularization as well as with distinct magnification algorithms that also handle generalized sampling. The quadratic approach regularizes the L_2-norm of the image gradient. It is implemented in AIRLS as the limit case $\psi = 1$ where one single unweighted linear problem has to be solved. We further consider iterative TV reconstruction with an implementation[13] of the primal-dual method [34] as well as the PDE-based method of Roussos and Maragos (R&M) provided by Getreuer as an online demo [65]. Choosing a magnification factor of $\mathcal{M} = 4$, we model the sensor integration φ_0 as a 2D Gaussian of standard deviation $0.35\mathcal{M}$ in each dimension as can be specified in the R&M demo. The parameters specific to AIRLS are $\Lambda = 0.01$, $l_\mathrm{s} = 9, N_\mathrm{v} = 2$, $\omega(0, \bigtriangledown) = 1.5$, $\beta_1 = 2 \cdot 10^{-3}\mathrm{DR}$, and $\beta_2 = 2 \cdot 10^{-2}\mathrm{DR}$. The other algorithms are used with their default settings.

The available data consists in images that are primarily downsampled according to the above generalized-sampling settings. The corresponding reconstructions are obtained by magnification and compared with the known oracles; the results are shown in Table 2.2 and Figure 2.7. In Figure 2.8, we perform a similar magnification experiment on an image that is provided as such without any prior downscaling.

We observe that the combination of consistent-resampling constraints with edge-preserving regularization yields high-quality reconstructions in terms of feature preservation, as discussed in [26]. This emphasizes the interest of taking the sensor integration into account for image magnification. The results of Figures 2.7 and 2.8 demonstrate that all nonlinear methods restore sharper edges than quadratic regularization.

As compared to the isotropic TV solution, the anisotropic EED_0, EED_1, and R&M methods better preserve certain fine structures as well as the curvature of the objects, thanks to the associated flows. They also avoid staircasing artifacts but introduce some edge smearing. The EED_2 method yields the highest-quality results because it benefits from the desirable properties of anisotropic diffusion while preserving image sharpness nearly at the same level as TV. It also yields the highest PSNR values for several images as shown in Table 2.2. Note that, interestingly, recent works have proposed extended forms of TV that are based on

[13]This Matlab implementation is based on a publicly available source code of Y. Chen and T. Pock, Graz University of Technology, Austria. We have modified the original version so as to handle the φ_0 under consideration. While the code structure is not optimized for multithreading, most low-level Matlab functions that are involved are intrinsically parallelized (*e.g.*, elementary operators). The algorithm is run until 400 iterations are reached, or until the PSNR increase per iteration is lower than 10^{-3}.



(a) Quadratic solution (b) TV solution (c) R&M solution

(d) EED_0 solution (e) EED_1 solution (f) EED_2 solution

Figure 2.7: Magnification of *Bird* after it was downsampled by a factor 4 along each dimension.

higher-order differential operators, which avoids staircasing artefacts. For instance, the authors of [66] have introduced the concept of *total generalized variation* (TGV). The extent to which TGV-type methods may compete with EED-based ones in the image-magnification setting remains an open topic of research.

(a) Quadratic solution (b) TV solution (c) R&M solution

(d) EED_0 solution (e) EED_1 solution (f) EED_2 solution

Figure 2.8: Magnification of *Peppers* (64×64 crop) by a factor 4 along each dimension.

2.9 Conclusions

We have designed a method that reconstructs continuous images from a sparse set of generalized samples. Combined with consistent data-fidelity constraints, our anisotropic regularization approach was designed to preserve the edge information accurately, and to be functional at high sparsity levels. In the experiments, promis-

ing results have been obtained with nonuniform interpolation of images as well as with consistent image magnification.

From an algorithmic perspective, the low computational cost of our method demonstrates that approaches based on IRLS can successfully handle anisotropic regularization in sparse inverse problems. This low cost legitimates the use of linear-multigrid approaches. As in IRLS, we have successfully maximized the algorithmic performance by restricting the diffusivity estimations to the reweightings. This has led to a simple and efficient design that drastically reduces redundancy in terms of operations. Our algorithm has been optimized for sparse interpolation. It is comparable to state-of-the-art implementations in terms of computational efficiency.

Because it is based on IRLS, our approach is modular. Its data term can potentially be redefined so as to solve several types of inverse problems, such as multi-image super-resolution or deconvolution. The structure of the proposed regularization framework is also able to handle multivariate data or alternate definitions of the anisotropic flow. Extensions to higher dimensions are straightforward because the edge-enhancing anisotropic diffusion that we formulated in Section 2.5 is generic. This emphasizes the flexibility of our method in terms of applicability and leaves room for further improvements.

2.10 Appendix

2.10.1 Full-Multigrid V-Cycles

Each Full-Multigrid V-cycle FMG(κ) parameterized by $N_{\mathrm{v}} \in \mathbb{N}^*$ corresponds to the recursive function shown in Algorithm 2.2, the residual at a given grid Ω^κ being defined as

$$\mathrm{R}(\Omega^\kappa) = \mathbf{y}^\kappa - \mathbf{S}^\kappa \tilde{\mathbf{c}}^\kappa. \tag{2.65}$$

The expression $\mathbf{I}^\triangledown$ denotes a *restriction operator*, which transfers a sequence from a grid Ω^κ towards a coarser grid $\Omega^{\kappa+1}$, while $\mathbf{I}^\triangle$ denotes a *prolongation operator*, which transfers a sequence from a grid $\Omega^{\kappa+1}$ towards a finer grid Ω^κ. The operator V(κ) described in Algorithm 2.3 performs one V-Cycle at grid Ω^κ, which is itself a recursive function. The two arguments parameterizing our SOR iterative method correspond to the current phase $\wp$ of the V-Cycle followed by the current grid level

if $\kappa < N_{\mathrm{g}} - 1$ **then**
  a) Update $\mathbf{y}^{\kappa+1} \leftarrow \mathbf{I}^{\triangledown}\mathrm{R}(\Omega^{\kappa})$
  b) Run FMG$(\kappa + 1)$
  c) Correct $\tilde{\mathbf{c}}^{\kappa} \leftarrow \tilde{\mathbf{c}}^{\kappa} + \mathbf{I}^{\triangle}\tilde{\mathbf{c}}^{\kappa+1}$
  d) Count $I \leftarrow I + 1$
end
2) Run V(κ) N_{v} times

Algorithm 2.2: Full-Multigrid V-Cycle FMG(κ).

κ. They determine the number of iterations $N_{\mathrm{i}}^{*}(\kappa, \wp)$ and the relaxation constant $\omega(\kappa, \wp)$ that have to be used in each case.

2.10.2 Properties of the system matrix

The positive-definiteness of $\mathbf{T}$ and Λ imply that, $\forall \mathbf{k}$:

$$(\overline{\theta}_{11}^{0}[\mathbf{k}]\ \overline{\theta}_{12}^{0}[\mathbf{k}], \overline{\theta}_{21}^{0}[\mathbf{k}]\ \overline{\theta}_{22}^{0}[\mathbf{k}]) \geq 0. \tag{2.66}$$

From (2.62) and (2.66), given the positivity of the scaling filters h_2, and given that the set of positive-definite matrices is closed under summation, we infer the more general set of inequalities

$$\forall \kappa : (\overline{\theta}_{11}^{\kappa}[\mathbf{k}]\ \overline{\theta}_{12}^{\kappa}[\mathbf{k}], \overline{\theta}_{21}^{\kappa}[\mathbf{k}]\ \overline{\theta}_{22}^{\kappa}[\mathbf{k}]) \geq 0. \tag{2.67}$$

Similarly, from the non-negativity of the finest-scale data weights, given (2.60) and (2.61), and given the non-negativity of the B-spline scaling filter h_2 and of the sequence b defined in (2.9),

$$\forall \kappa : \overline{w}^{\kappa}[\mathbf{k}] \geq 0. \tag{2.68}$$

From (2.67) and (2.68), and given the eigenvalue-decoupling (sub)diagonal structure of the corresponding weight matrices, we obtain

$$\forall \kappa : \overline{\mathbf{W}}^{\kappa} \geq 0, \overline{\mathbf{\Theta}}^{\kappa} \geq 0. \tag{2.69}$$

1) Iterate at Ω^κ given the parameters κ and $\wp = \bigtriangledown$
if $\kappa < N_\mathrm{g} - 1$ **then**
  a) Update $\mathbf{y}^{\kappa+1} \leftarrow \mathbf{I}^{\bigtriangledown}\mathrm{R}(\Omega^\kappa)$
  b) Run $\mathrm{V}(\kappa + 1)$
  c) Correct $\tilde{\mathbf{c}}^\kappa \leftarrow \tilde{\mathbf{c}}^\kappa + \mathbf{I}^{\bigtriangleup}\tilde{\mathbf{c}}^{\kappa+1}$
end
2) Iterate at Ω^κ given the parameters κ and $\wp = \bigtriangleup$

Algorithm 2.3: V-Cycle $\mathrm{V}(\kappa)$.

Now, in the context of our reconstruction problem, we assume that the intersection between the nullspaces of $\mathbf{B}^{\kappa\mathrm{T}}\overline{\mathbf{W}}^\kappa\mathbf{B}^\kappa$ and $\mathbf{R}^{\kappa\mathrm{T}}\overline{\mathbf{\Theta}}^\kappa\mathbf{R}^\kappa$ in (2.59) is empty. Accordingly, from (2.69), given the symmetry of the weight matrices, and given that all quantities are real, (2.59) corresponds to a sum of Cholesky decompositions. The whole system matrix $\mathbf{S}^\kappa$ is therefore symmetric and positive definite at all scales.

2.10.3 AIRLS connections

Our reconstruction algorithm is linked with classical variational techniques. In certain settings, indeed, the successive quadratic regularizers of AIRLS can degenerate to simpler expressions that are associated with the minimization of one global energy functional. For convenience, we consider the continuous form (2.44) in our discussion; the same reasonings hold in the discrete case.

Connection with isotropic regularization

Parameterizing AIRLS with $\{\nabla, \psi\}$, and choosing a function $\psi \leq 1$ that is linked to a convex potential $\Psi_\mathcal{R}$ through the multiplicative form of half-quadratic minimization (*e.g.*, ψ_1), each regularizer $\mathcal{R}_\mathrm{A}(\cdot|u^{(n)})$ reduces to

$$\mathcal{R}_\mathrm{A}{}'(u|u^{(n)}) = \int_{\mathbb{R}^2} \psi(\|\nabla u^{(n)}(\mathbf{x})\|)\|\nabla u(\mathbf{x})\|^2 \mathrm{d}\mathbf{x} + \mathcal{R}_\mathrm{A}^+(u|u^{(n)}), \tag{2.70}$$

where

$$\mathcal{R}_\mathrm{A}^+(u|u^{(n)}) = \int_{\mathbb{R}^2} (1 - \psi(\|\nabla u^{(n)}(\mathbf{x})\|))(\nabla^\perp{}_{\nabla u^{(n)}(\mathbf{x})} u)^2(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.71}$$

using the properties of the gradient ∇. According to the definition of ψ, the first integral term of (2.70) upper bounds the isotropic regularizer

$$\mathcal{R}_{\mathrm{A}}{}'(u) = \int_{\mathbb{R}^2} \Psi_{\mathcal{R}}(\|\nabla u(\mathbf{x})\|)\mathrm{d}\mathbf{x}. \tag{2.72}$$

As for the second term of (2.70), the inequality $\psi \leq 1$ and the relations between the directional derivative $\nabla^{\perp}.$ and the gradient ∇ yield, $\forall u^{(n)}$:

$$\begin{aligned} \mathcal{R}_{\mathrm{A}}^{+}(\cdot|u^{(n)}) &\geq 0, \\ \mathcal{R}_{\mathrm{A}}^{+}(u^{(n)}|u^{(n)}) &= 0. \end{aligned} \tag{2.73}$$

Relations (2.73) imply that (2.70) defines successive upper bounds of (2.72). These bounds are as valid but looser than the conventional ones because of the additive integral term. In this setting, AIRLS thus minimizes one single functional whose regularizer is a discretized form of the isotropic cost $\mathcal{R}_{\mathrm{A}}{}'$.

Connection with quadratic regularization

In the limit case $\psi = 1$, the properties of the directional derivatives imply that the successive quadratic functionals $\mathcal{R}_{\mathrm{A}}(\cdot|u^{(n)})$ of AIRLS degenerate to

$$\mathcal{R}_{\mathrm{A}}{}''(u) = \|\nabla u(\mathbf{x})\|_{L_2}^2 . \tag{2.74}$$

In that case, AIRLS minimizes one single quadratic functional whose regularizer is a discretized form of $\mathcal{R}_{\mathrm{A}}{}''$. Since the latter is independent from $\tilde{c}^{(n)}$, only one iteration is required for convergence. Note that (2.74) is one particular instance of the quadratic regularizers of Section 2.4.1 where $\mathrm{L} = \nabla$.

Chapter 3

Image reconstruction from binary measurements

3.1 Introduction

Our goal here is to reconstruct images from binary measurements. In order to obtain results of satisfactory quality within reasonable computational time, we propose to design a framework that is adapted to visual data and that follows compressed-sensing principles. From a general perspective, compressed-sensing strategies are based on forward models that allow to substantially reduce the number of samples required for signal acquisition compared to more conventional approaches, at the expense of an additional reconstruction procedure. These strategies can also provide robust results with quantized measurements, including in our one-bit setting.

Our forward model describes data acquisition and follows physical principles. It entails a series of random convolutions performed optically on the original signal f followed by sampling and binary thresholding. According to (1.2), the latter effect is modeled by the operator $\mathcal{Q}$; our overall acquisition model is thus nonlinear unlike in Chapter 2. Meanwhile, as in the previous chapter, the binary samples that are obtained can be either measured or ignored according to predefined functions.

Based on these measurements, we express our reconstruction problem as the minimization of a compound convex cost that enforces the consistency of the so-

lution $\tilde{f}$ with the available binary data under TV regularization. Then, relying on convex-optimization principles, we derive an efficient reconstruction algorithm that complies with the high dimensionality of image data. Finally, we conduct several experiments on standard images and demonstrate the practical interest of our approach[1].

3.2 Overview

In the context of compressed sensing, the amount of data to be acquired can be substantially reduced as compared to conventional sampling strategies [12, 18, 67, 68, 69, 70]. The key principle of this approach is to compress the information before it is captured, which is especially beneficial when the acquisition process is expensive in terms of time or hardware. For instance, in their previous work [13], Boufounos *et al.* investigated the performance of compressed sensing in the binary case where the extreme coarseness of the quantization must typically be compensated by taking more numerous measurements than in the classical case. The original signal can then be recovered from the available measurements through numerical reconstruction, whose computational complexity exhibits a strong dependance on the structure of the forward model. Consequently, specialized acquisition approaches are required for compressed sensing when dealing with large-scale data such as images. For instance, we were able to extend in [10] the central principles of [13] to image acquisition and reconstruction. Our associated forward model generates binary measurements that are based on random-convolution principles [12]. Though demonstrating satisfactory reconstruction capability for image data, this method tends to create spatial redundancy in the associated measurements, which is suboptimal from the perspective of information content.

In this chapter, our first contribution is to propose a general framework for the binary compressed sensing of images. Based on [10], we devise an extended forward model that can take several binary captures of a given grayscale image. Each of these acquisitions corresponds to one distinct convolution performed by an optical system. The flexibility of our approach allows us to improve the statistical properties of the associated binary data, which ultimately increases the quality of reconstructions.

[1] This chapter is based on our papers [10, 11].

Using a variational formulation to express our reconstruction problem, our second contribution is a fast reconstruction algorithm that uses bound-optimization principles. This proposed algorithm yields an IRLS procedure that is easily parameterized. Note that, in the context of reconstruction from binary measurements, several distinct resolution strategies have been proposed in the literature, some of which bear some similarities with our approach, and some of which are based on non-convex constrained formulations [14, 71].

We introduce some preliminary theoretical background in Section 3.3. We then describe our forward model for image acquisition in Section 3.4. In Section 3.5, we express our reconstruction problem as the minimization of a compound cost functional. Based on convex optimization, we derive the reconstruction algorithm in Section 3.6. In Section 3.7, we perform several experiments on standard grayscale images. We extensively discuss them and conclude our chapter in Section 3.8.

3.3 Compressed-sensing strategy

As discussed in the sequel, our unknown signal f maps to a coefficient vector $\mathbf{c}$. Accordingly, any linear measurement system can then be modeled by a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, where the measurements correspond to $\mathbf{g} = \mathbf{A}\mathbf{c}$, and where M, N are the numbers of measurements and the number of unknowns, respectively.

The theory of compressed sensing guarantees that $\mathbf{c} \in \mathbb{R}^N$ can be recovered from a small amount of measurements if it is sufficiently *sparse* in some appropriate linear basis $\mathbf{\Phi}$. By sparsity, it is meant here that $\mathbf{c}$ must consist of enough negligible entries once it has been represented in the basis $\mathbf{\Phi}$. As it turns out, natural images are often sparse in some transformed domains (e.g., wavelets). The importance of $\mathbf{\Phi}$, in the general theory, is its mere *existence*; its specific layout reflects the considered class of signals. The measurement matrix $\mathbf{A}$ bears no direct relation with $\mathbf{\Phi}$, except that, in order to be suitable, it must be *incoherent*—in the statistical sense—with that basis. This means that the bases of the measurement and sparse-representation domains of the signal must be uncorrelated with overwhelming probability [72].

If the measurements are not quantized, the standard reconstruction problem (P_0) in compressed sensing is to find the sparsest solution $\mathbf{c}$ leading to these same measurements (up to some imprecision $\mathcal{K}_{\mathcal{D}}$), given the system matrix $\mathbf{A}$. When $\mathbf{\Phi}$ is orthonormal, this can be expressed as

$$(\mathrm{P}_0) : \min_{\mathbf{c}} \left\| \mathbf{\Phi}^{\mathrm{T}} \mathbf{c} \right\|_{\ell_0} \quad \text{s.t.} \quad \|\mathbf{g} - \mathbf{A}\mathbf{c}\|_{\ell_2} \leq \mathcal{K}_{\mathcal{D}}. \tag{3.1}$$

The left term minimizes the solution sparsity, while the right one ensures fidelity to the available measurements. This problem is non-convex and *NP-hard*. Let us now consider the alternate convex-optimization problem

$$(\mathrm{P}_1) : \min_{\mathbf{c}} \|\mathbf{g} - \mathbf{A}\mathbf{c}\|_{\ell_2}^2 + \Lambda \left\| \mathbf{\Phi}^{\mathrm{T}} \mathbf{c} \right\|_{\ell_1}, \tag{3.2}$$

where $\Lambda \in \mathbb{R}_+^*$ is a constant. Interestingly, it has been shown that, when $\mathbf{c}$ is sufficiently sparse, (P_0) and (P_1) yield solutions that are equivalent in some sense. Moreover, unlike (P_0), the problem (P_1) is tractable and can be solved using convex-optimization techniques.

Besides the aforementioned properties, it has been shown that compressed-sensing measurements are robust to quantization as well [73]. The corresponding problem can thus be treated as a variation of the classical one in which the linear measurements $\mathbf{g}$ are further quantized through a pointwise nonlinear operator $\mathcal{Q}$, and are typically more numerous than the amount of unknows. In this chapter, we are going to deviate from the traditional compressed-sensing framework by first considering such quantized measurements as mentioned in Section 3.2, which requires the use of a modified data term in (3.2), and second by using a non-unitary regularization matrix $\mathbf{R}$ that corresponds to TV, and which promotes piecewise-smooth solutions [32].

3.4 Forward model

3.4.1 General structure

In this section, we establish a convolutive physical model that generates L binary measurement sequences γ_i from a given 2D continuously defined image f of unit square size. Following a design similar to the one of [10], each of these sequences is obtained through optical convolution of f with a distinct pseudo-random filter h_i followed by acquisition through binary sensors.

Specifically, each convolved image $f * h_i$ is sampled and binarized by a uniform 2D CCD-like array of $M_0 \times M_0$ sensors, the specific form of h_i being defined in Section 3.4.2. The actual sampling process is regular but nonideal, meaning that

each sensing area of side ${M_0}^{-1}$ has some pre-integration effect modeled by some spatial filter φ_0. Therefore, the global convolutive effect of our model before sampling corresponds to the spatial kernels $h_i * \varphi_0$, yielding the pre-filtered intermediate images

$$f_{0i}(\mathbf{x}) = (f * h_i * \varphi_0)(\mathbf{x}), \tag{3.3}$$

where the vector $\mathbf{x} \in \mathbb{R}^2$ denotes the 2D spatial coordinates. Then, the sensor array samples each image f_{0i} with a step ${M_0}^{-1}$, which produces the sequences f_{1i} defined for each index $\mathbf{k} \in \mathbb{Z}^2$ as

$$f_{1i}[\mathbf{k}] = f_{0i}(\mathbf{x})|_{\mathbf{x}=\mathbf{k}{M_0}^{-1}}. \tag{3.4}$$

Unlike [10], we allow for a finite-differentiation process to take place before the final quantization step. Denoting the corresponding discrete filters as ζ_i, the non-quantized measurements g_i are obtained as

$$g_i[\mathbf{k}] = (f_{1i} \star \zeta_i)[\mathbf{k}], \tag{3.5}$$

where $\star$ denotes a discrete convolution. These operations can be efficiently performed by the sensor array itself, for instance using voltage comparators. As discussed in the experimental section, finite differentiation brings improvements in terms of reconstruction quality and simplifies the calibration of the system. No finite differentiation occurs when taking ζ_i to be the discrete unit sample $\delta[\cdot]$.

Defining τ as a common threshold value, the measurements $g_i[\mathbf{k}]$ are finally binarized at the sensor level to the signs $\gamma_i[\mathbf{k}] = \mathcal{Q}(g_i[\mathbf{k}], \tau)$. The nonlinear operator $\mathcal{Q}$ is defined accordingly as

$$\mathcal{Q}(t, \tau) = \begin{cases} +1, & t \geq \tau \\ -1, & \text{otherwise.} \end{cases} \tag{3.6}$$

The measurements γ_i can be selectively stored according to discrete spatial indicator functions χ_i. Each $\gamma_i[\mathbf{k}]$ is actually kept and counted as a measurement if and only if the value $\chi_i[\mathbf{k}] \in \{0, 1\}$ is unity for the same $\mathbf{k}$. Note that, before binarization, every measurement g_i is a mere linear functional of f.

The successive operations that are involved in our forward model simplify to one single convolution in the continuous domain without subsequent discrete filtering,

as summarized in Figure 3.1. The equivalent spatial impulse response h_i' of the filter corresponds to

$$h_i'(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \zeta_i[\mathbf{k}]\ (h_i * \varphi_0)(\mathbf{x} - \mathbf{k}/M_0). \tag{3.7}$$

To sum up, our forward model yields $M = KL{M_0}^2$ binary measurements of the continuously defined image f in the form of L distinct binary sequences, where K is the storage ratio associated to the functions χ_i. These captured sequences are complementary, as they are associated with distinct random convolutions before sampling, binarization, and masking through the χ_i. Since the latter process allows to decrease M, the resolution M_0 can be kept constant. This avoids high-frequency losses due to coarse-sensor integration.

Besides reducing data storage, the process of binary quantization potentially consumes far less power than standard analog-to-digital converters, and is less susceptible to the nonlinear distortion of analog electronics [14]. Binary sensors are also associated with very high sampling rates in general [13]. In that regard, the selective subsampling that we specify by χ_i may also lead to further reductions of the acquisition time if fewer measurements are required; the acquisition of the selected samples can indeed be performed efficiently through randomly addressable image sensors[2].

3.4.2 Pseudo-random optical filters

As mentioned in Section 3.4.1, the L filters h_i are associated to optical convolution operations. Accordingly, we make each h_i correspond to a distinct spatially invariant point-spread function (PSF) that is generated by the same optical model. In our setup shown in Figure 3.2, the image f is associated with light intensities defined on a plane. For each of the L acquisitions, the intensities measured by the sensor array after optical propagation correspond to the convolution $f * h_i$ up to geometrical inversion.

The specific form of h_i depends on the profile of the central plane of the system called the *Fourier plane* [75]. In our model, this plane transmits light through a

[2]Image sensors that are based on the complementary-metal-oxide-semiconductor (CMOS) technology allow for parallel and random access, as opposed to other architectures that can only perform sequential readout [74]. While the potential benefits of binary sensors further motivate our imaging model, the proper development of such elements for optics remains to be addressed.

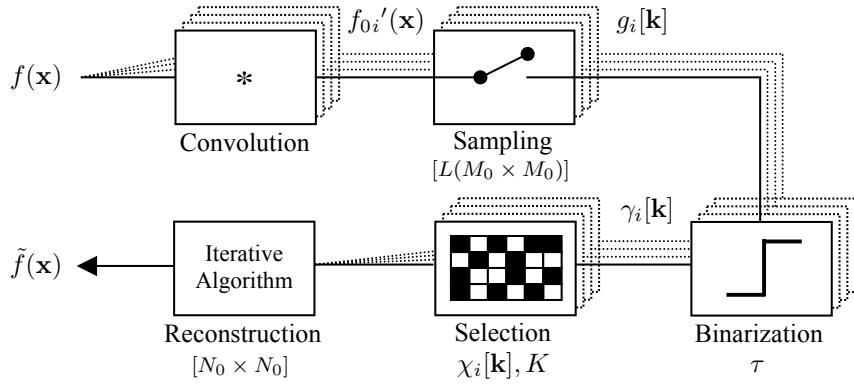


Figure 3.1: General framework. The unknown continuously defined image f is first convolved with L distinct kernels h_i', producing the intermediate images $f_{0i}{}' = f * h_i'$. Each $f_{0i}{}'$ is then sampled with step ${M_0}^{-1}$ to obtain the sequences g_i. The last acquisition step consists in pointwise binarization with threshold τ, resulting in the binary measurements γ_i. When retained, the latter constitute the available information on the original data. Based on these selected measurements, and assuming that the forward model is known, our reconstruction algorithm produces an estimate $\tilde{f}$ of the original image that is defined in terms of $N_0 \times N_0$ coefficients.

circular area and is further equipped for each acquisition with one distinct instance of a phase-shifting plate whose effect is to multiply the transmitted-light amplitudes with pseudorandom phase values. The resulting profile q_i is modeled as a complex-valued function expressed in normalized spatial coordinates [75].

Considering phase functions μ_i composed of square zones, each zone associated with either a 0 or π phase shift, we obtain

$$\mu_i(\boldsymbol{\xi}) = \sum_{\mathbf{k} \in \mathbb{Z}^2} \nu_i[\mathbf{k}] \operatorname{rect}(\boldsymbol{\xi} - \mathbf{k}), \tag{3.8}$$

where the phases ν_i are independent and random variables that take values from the pair $\{0, \pi\}$ with equal probability, where rect is the 2D rectangle function, and where $\boldsymbol{\xi}$ denotes normalized spatial coordinates. The phase-shifting plates associated with the μ_i are of finite extent since they only operate inside the transmissive

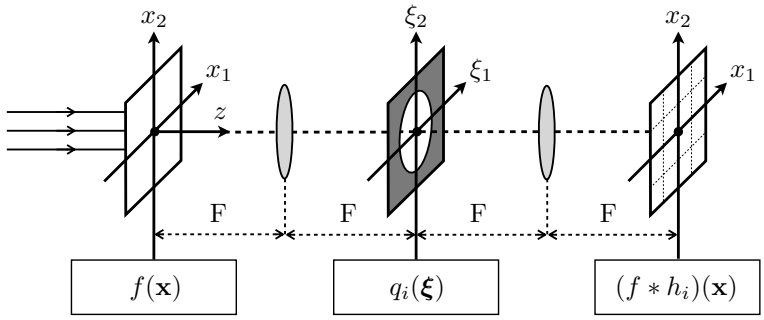


Figure 3.2: Optical setup. In our optical model, the image f maps to light-intensity values. Our optical device transforms this initial image wavefront using elements that are spaced by the same distance F. Following the direction z of light propagation, this system called 4F consists in the left plane where the image f lies, one first lens of focal length F, the central plane, one second lens identical to the first one, and the last plane containing the propagated wavefront to be captured by the sensor array.

circular area of Figure 3.2. The latter is designed such that the diameter of the circle covers N_d phase zones in the horizontal or vertical direction. It is thus specified by the function

$$\mathrm{circ}(\boldsymbol{\xi}) = \begin{cases} 1, & \|\boldsymbol{\xi}\| \leq N_\mathrm{d}/2 \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

The profile q_i combines the phase shifts of (3.8) with the transmissivities of (3.9). It is defined as

$$q_i(\boldsymbol{\xi}) = \mathrm{circ}(\boldsymbol{\xi}) \exp(-\mathrm{j}\, \mu_i(\boldsymbol{\xi})). \tag{3.10}$$

Due to the 4F placement of the lenses, the propagation of light implements a continuous Fourier transform [75]. The light amplitudes are also modulated by q_i in the Fourier plane. Accordingly, the impulse response of the system is defined up to scale as

$$h_i(\mathbf{x}) = |\mathcal{F}\{q_i\}(\mathbf{x})|^2, \tag{3.11}$$

where $\mathcal{F}$ denotes the Fourier transform

$$\mathcal{F}\{q_i\}(\mathbf{x}) = \int_{\mathbb{R}^2} q_i(\boldsymbol{\xi}) \exp(-\mathrm{j}\,\mathbf{x}^{\mathrm{T}}\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}. \tag{3.12}$$

The use of spatially incoherent illumination[3] and the fact that the measured quantities are light intensities results in a squared modulus in (3.11). Each filter h_i is thus nonnegative, and depends upon the corresponding μ_i defined in (3.8). The latter can be generated electronically by a spatial light modulator [12].

3.5 Reconstruction problem

For the general problem of binary compressed sensing, the authors of [14] have recently proposed a reconstruction technique that is based on binary iterative hard thresholding (BIHT), using the non-convex constraint that the solution signal lies on the unit sphere. This approach extends previous works [13, 76], and achieves better performance. The work of [71] uses a distinct strategy by formulating a convex reconstruction problem solvable by linear programming. An extension of this principle to the case of noisy measurements is also considered by the same authors in [77]. In the case of the proposed forward model, the use of phase masks produces random-like patterns in each of the binary-measurement sequences that are obtained. This closely relates our overall strategy to the aforementioned works, and, more specifically, to the compressed-sensing paradigm of [13].

3.5.1 Connection with compressed sensing

In order to comply with the compressed-sensing framework, we have to represent our signals in discrete form. Therefore, assuming (1.3) with normalized coefficient-grid spacing, we model the continuously defined estimate $\tilde{f}$ of f as the expansion

[3] Spatial incoherence means that the phases of the initial wavefront on the left plane of Figure 3.2 vary with time in uncorrelated fashions. This implies that the effective response of our optical system is linear in intensity rather than in amplitude [75].

$$\tilde{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \tilde{c}[\mathbf{k}]\varphi(\mathbf{x}-\mathbf{k}), \tag{3.13}$$

where the sequence $\tilde{c}$ corresponds to $(N_0 \times N_0)$ real coefficients placed on a regular grid, and where $\varphi(\mathbf{x}) = \varphi(x_1)\varphi(x_2)$ for $\mathbf{x} \in \mathbb{R}^2$ is the separable 2D B-spline of degree η. Given their small support and polynomial-reproduction properties, B-splines are especially adapted from both approximation and computational viewpoints. They thus constitute a suitable approach to represent continuous images [47].

Since our continuous image $\tilde{f}$ is modeled as a linear combination of B-spline basis functions, it is equivalently described through the corresponding coefficients. Then, given the expansion (3.13), the introduction of $\tilde{f}$ into our physical forward model naturally leads to a linear and discrete dependency between the image coefficients $\tilde{c}$ and the corresponding measurements $\tilde{g}$ that are obtained before quantization. Accordingly, the relation between $\tilde{c}$ and each corresponding sequence $\tilde{g}_i$ can be summarized into the *measurement matrix* $\mathbf{A} \in \mathbb{R}^{M\times N}$, whose structure is induced from our continuous-domain formulation. Using matrix notation, we obtain the matrix-vector relation of the form (1.5) between $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{g}}$, where $\tilde{\mathbf{c}}$ contains $N = {N_0}^2$ coefficients and where $\tilde{\mathbf{g}}$ contains M measurements. Our corresponding measurement matrix $\mathbf{A}$ generalizes [10] and vertically concatenates several terms $\mathbf{A}_i$ of similar structure as

$$\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_i, \ldots, \mathbf{A}_L). \tag{3.14}$$

These terms are associated with the sequences $\tilde{g}_i$. They depend from the corresponding kernels h_i' and from the rational sampling step ${M_0}^{-1}$. They are defined as

$$\mathbf{A}_i = \boldsymbol{\chi}_i \mathbf{D}_{\mathcal{N}} \mathbf{B}_i \mathbf{U}_{\mathcal{M}}, \tag{3.15}$$

where $\mathbf{D}_i$ and $\mathbf{U}_j$ denote downsampling-by-i and upsampling-by-j matrices. The integers $\mathcal{M}$ and $\mathcal{N}$ are such that the right-hand side of the equality $M_0/N_0 = \mathcal{M}/\mathcal{N}$ is in reduced form. Given periodic boundary conditions, the circulant matrix $\mathbf{B}_i$ is associated with the discrete impulse response

$$b_i[\mathbf{k}] = \frac{N}{\mathcal{N}{M_0}^2} \left(h_i' \left(\frac{N\cdot}{\mathcal{N}{M_0}^2} \right) * \varphi\left(\frac{\cdot}{\mathcal{M}} \right) \right) (\mathbf{x})|_{\mathbf{x}=\mathbf{k}}. \tag{3.16}$$

Finally, each matrix $\boldsymbol{\chi}_i$ is linked to χ_i. Specifically, it corresponds to an identity matrix whose rows associated with the discarded measurements are suppressed, if any. The overall structure of $\mathbf{A}$ will prove to be beneficial for the reconstruction in terms of computational complexity. The measurements are indeed related to the coefficients by mere discrete-Fourier-transform (DFT) and resampling operations.

In accordance with the principles discussed in Section 3.3, the compressed-sensing paradigm followed in this chapter requires the expansion of type (3.13) to be valid not only for the estimate $\tilde{f}$, but also for the unknown signal f itself. Accordingly, the relation between the coefficients c of f and the known measurements g is expressed in matrix notation as

$$\mathbf{g} = \mathbf{A}\mathbf{c}. \tag{3.17}$$

When the unknown vector $\mathbf{c}$ in (3.17) is sufficiently *sparse* in some adequate basis, the theory of compressed sensing offers guarantees on the quality of reconstruction in terms of robustness to measurement loss or quantization [73], provided that the measurement matrix is appropriate. As mentioned in Section 3.3, a common and suitable criterion for $\mathbf{A}$ is to be *statistically incoherent* with any fixed signal representation. This property has been shown theoretically to strictly hold for matrices consisting of independent and identically distributed (iid) Gaussian random entries [67, 70], and also to nearly hold for other random-matrix ensembles [78, 12, 68, 79, 69]. In the context of this chapter, we resort to an experimental validation of our measurement matrix for binary compressed sensing. In particular, we shall demonstrate in Section 3.7 that our model is suitable for the reconstruction of images from few data, and that the quality of the solution is linked to relatively simple criteria relying on the measurements themselves.

The appropriateness of $\mathbf{A}$ in our generalized model is tied to the set of discrete filters b_i defined in (3.16) and associated with the matrix terms $\mathbf{A}_i$. Indeed, they share similarities with the Romberg's random-convolution pulses proposed in [12] for compressed sensing. Firstly, their discrete Fourier coefficients also have phase values that are randomly distributed in $[0, 2\pi)$, given their relation with the profiles (3.8). Secondly, despite not being strictly all-pass as in [12], our filters are also spread-out in the spatial domain. Due to these properties, the form of b_i has been shown to yield satisfactory reconstructions in the binary case [10]. Besides being adequate individually, these filters also produce L distinct sequences g_i from the same image f because they are associated with L distinct pseudorandom phase-

mask profiles in (3.10). In some sense, our multi-acquisition framework is the reverse of multichannel compressed-sensing architectures where one single output sequence combines several source signals through distinct modulation or filtering operations [80, 81]. As will be discussed in Section 3.7, the subsequent thresholding operation (3.6) that is applied in our method yields binary measurements that follow an equiprobable distribution, as in [10]. The proper specification of the additional acquisition parameters of our system (including χ_i and L) will allow us to maximize the reconstruction performance while maintaining a high computational efficiency.

3.5.2 Variational approach

We propose to formulate our image-reconstruction problem in a variational framework. Specifically, our solution is expressed as the minimum of a convex functional that includes data-fidelity and regularity constraints. Using bound-optimization principles, the convexity of this functional is exploited in Section 3.6 to derive an efficient iterative-reconstruction algorithm. The latter can handle large-scale problems because, from a computational perspective, it involves the application of the forward model (whose form is essentially convolutive in our case) and of its adjoint inside each iteration as in other methods. Furthermore, besides quality considerations, the specific structure of our reconstruction problem will allow us to maximize iterative performance through preconditioning and Nesterov's acceleration [82].

The available data consist of the measurements γ_i obtained according to Section 3.4. In addition, we suppose that $\mathbf{A}$ is known. Its components can be deduced physically from the L impulse responses h_i produced by the optical system, or, more indirectly, from the phase-mask profiles μ_i. Based on that information, our goal is to reconstruct an accurate continuously defined estimate $\tilde{f}$ of the original image f according to a regularization functional $\mathcal{R}$ associated with some sparsity prior. Specifically, we demand our reconstructed coefficients $\tilde{c}$ to minimize

$$\mathcal{J}(\tilde{c}) = \mathcal{D}(\tilde{c}) + \Lambda\mathcal{R}(\tilde{c}). \tag{3.18}$$

The first scalar term $\mathcal{D}$ imposes the fidelity of the solution to the known binary measurements γ_i. Due to quantization, fidelity alone is in general under-constrained and accurate only up to contrast and offset. Then, the regularization term $\mathcal{R}$, weighted by Λ, encourages the sparsity of the reconstruction.

3.5.3 Data term

The role of our data-fidelity constraint is to ensure that the reintroduction of the reconstructed continuously defined image $\tilde{f}$ into the forward model results in a set of discrete values $\tilde{g}_i$ that are consistent with the known measurements γ_i, once binarized. The presence of noise is not considered in the framework of this chapter. In the context of 1-bit compressed sensing, the enforcement of sign consistency has been originally proposed in [13], where a one-sided quadratic penalty function was considered. Since the signs of the measurements do not provide amplitude information, any positive scalar multiple of the reconstructed signal, including the zero signal, is consistent with the measurements; trivial solutions were avoided by requiring that the signal lies on the unit $(N-1)$-sphere [13]. Here, as in [10], we introduce a variational consistency principle that preserves the convexity of the problem without requiring additional non-convex constraints. Note that, although convexity is not required to ensure nontrivial solutions, it is exploited for the development of our algorithm and to ensure its convergence, as described in Section 3.6. Regarding the data-fidelity term, our contribution is to propose a penalty function $\Psi_{\mathcal{D}}$ that is also suitable for bound optimization. We express our functional as

$$\mathcal{D}(\tilde{c}) = \sum_{i=1}^{L} \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi_i[\mathbf{k}] \Psi_{\mathcal{D}}(\tilde{g}_i[\mathbf{k}]\gamma_i[\mathbf{k}]), \tag{3.19}$$

The positive function $\Psi_{\mathcal{D}}$ is defined as

$$\Psi_{\mathcal{D}}(t) = \begin{cases} M^{-1} - t, & t < 0 \\ M^{-1}(M^2t^2 + Mt + 1)^{-1}, & \text{otherwise,} \end{cases} \tag{3.20}$$

where M is the total number of measurements. Besides penalizing sign inconsistencies, the rationale behind this definition is to yield nontrival solutions[4] while ensuring the convexity of the data term. The latter property holds because, according to (3.20), the Hessian of $\mathcal{D}$ is well-defined and positive semidefinite [83]. The function $\Psi_{\mathcal{D}}$ is itself $\mathcal{C}^2$-continuous and convex, its second derivative being always nonnegative. Moreover, this specific piecewise-rational polynomial function is suitable to the development of analytic upper bounds, as addressed in Section 3.6.

[4]Specifically, the part of the penalty function $\Psi_{\mathcal{D}}(t)$ defined for arguments $t \geq 0$ counteracts the effect of the regularization functional defined in (3.21) for which a trivial solution is a minimizer.

Given (3.6), negative arguments of $\Psi_{\mathcal{D}}$ correspond to sign inconsistencies. As shown in Figure 3.3, our penalty function is linear in that regime. In that regard, the authors of [14] have shown that, in the binary compressed sensing framework, such an ℓ_1-type penalty for consistency yields reconstructions that are of higher quality than with the ℓ_2 objective used in [13, 76]. To some extent, these results confirm similar observations mentioned in [10]. This type of penalty also relates to the so-called *hinge loss* which is considered a better measure than the square loss for binary classification [14, 84]. In our method, the values of the solution $\tilde{c}$ are defined up to a common scale factor, and also up to an additive constant because τ is not given. Non-constant solutions are favored by the contribution of the small nonlinear penalty that remains when the sign is correct. The transition between the linear and nonlinear regimes of $\Psi_{\mathcal{D}}$ is $\mathcal{C}^2$-continuous and takes place at the origin. The applied penalty vanishes for increasingly positive arguments.

3.5.4 Regularization term

Reconstruction algorithms frequently use TV [32] as a sparsifying transform when dealing with image data in inverse problems [20, 12, 85]. Yet, although suitable for regularization, the original form of TV is non-differentiable when the image gradient vanishes. While such a non-differentiability can be handled directly in primal-dual approaches [86], for instance, we opt in our problem setting for a smooth approximation of TV[5], as in the NESTA algorithm proposed in [88] for the recovery of sparse images. This implies that, asymptotically, our algorithm does not yield the exact same solutions as the ones obtained with the non-smoothed TV term [87].

In order to guarantee the well-posedness of the problem, we also include an additional energy term in our expression, since the nullspace of $\mathbf{A}$ can indeed be nonempty depending on ζ_i. This additional term is quadratic and ensures the unicity of the solution and of the corresponding linear subproblems. Replacing the Huber integral by a sum, our regularizer $\mathcal{R}$ is defined in approximate form as

$$\mathcal{R}^0(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Psi_{\mathcal{R}}(\|(\tilde{c}\star\mathbf{r})[\mathbf{k}]\|) + \Lambda_{\mathrm{E}}\tilde{c}[\mathbf{k}]^2, \tag{3.21}$$

where Λ_{E} is a small positive constant. The corresponding functional $\mathcal{J}^0$ is of the

[5]This approximation is based on a Huber potential function [54]. The latter can be described as the *Moreau envelope* of the ℓ_1-norm [87].

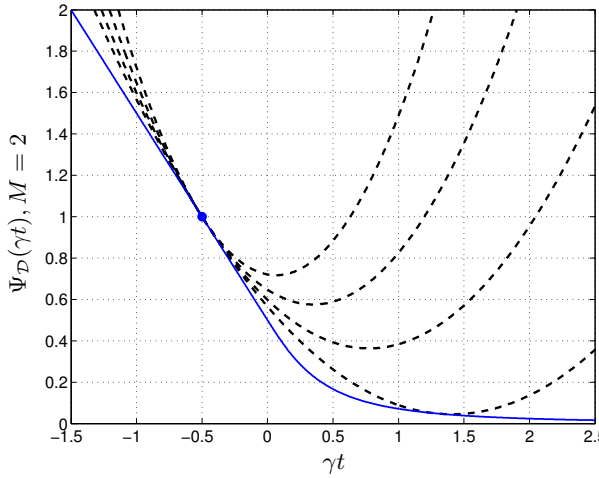


Figure 3.3: Shape of our penalty function. As discussed in Section 3.6 and further developed in Appendix 3.9.1, the values $\Psi_{\mathcal{D}}(\gamma t)$ (full line) can be bound from above by the quadratic function $\Psi_{\mathcal{D}}(t|\tilde{g}^{(n)}, \gamma)$ around $t = \tilde{g}^{(n)}$ (dot mark). Function values and derivatives must coincide at that point to satisfy (3.29). Among all possible parabolas (dashed lines), the solution $\Psi_{\mathcal{D}}(t|\tilde{g}^{(n)}, \gamma)$ under consideration is the upper bound with infimum second derivative.

form

$$\mathcal{J}^0(\tilde{c}) = \mathcal{D}(\tilde{c}) + \Lambda \mathcal{R}^0(\tilde{c}), \tag{3.22}$$

in accordance with the definition in (3.18). Based on a smoothing parameter ϵ, the scaled Huber potential $\Psi_{\mathcal{R}}$ is defined as

$$\Psi_{\mathcal{R}}(t) \quad = \quad \begin{cases} \epsilon^{-1} t^2, & |t| \leq \epsilon \\ 2|t| - \epsilon, & \text{otherwise.} \end{cases} \tag{3.23}$$

Each argument of $\Psi_{\mathcal{R}}$ in (3.21) corresponds to the norm of the gradient of $\tilde{f}$ evaluated at position $\mathbf{x} = \mathbf{k}$. More specifically, the components of $\tilde{c} \star \mathbf{r}$ are equivalent to

the spatial derivatives $\partial\tilde{f}/\partial x_1$ and $\partial\tilde{f}/\partial x_2$ of the solution sampled in-between the grid nodes. This type of discretization yields numerically stable solutions without oscillatory modes. It bears similarities with the so-called *marker-and-cell* methods used in fluid dynamics [89]. The signal expansion (3.13) determines the gradient filter $\mathbf{r}$ as

$$r_1[\mathbf{k}] = \frac{\partial}{\partial x_1}(\varphi(x_1+1/2)\varphi(x_2))\Big|_{\mathbf{x}=\mathbf{k}}, \tag{3.24}$$

$$r_2[\mathbf{k}] = \frac{\partial}{\partial x_2}(\varphi(x_2+1/2)\varphi(x_1))\Big|_{\mathbf{x}=\mathbf{k}}. \tag{3.25}$$

The first derivative of a B-spline φ has the symbolic expression given in [47].

3.6 Reconstruction algorithm

3.6.1 General approach

In this section, we derive an algorithm to efficiently minimize $\mathcal{J}^0(\cdot)$ and find the corresponding solution. Our main strategy is to recast the original formulation of the reconstruction problem as the partial minimization of successive quadratic costs $\mathcal{J}^0(\cdot|\tilde{c}^{(n)})$ that upper-bound $\mathcal{J}^0(\cdot)$ locally around the current solution estimate $\tilde{c}^{(n)}$. Each upper bound $\mathcal{J}^0(\cdot|\tilde{c}^{(n)})$ is then minimized using a specifically devised preconditioned conjugate-gradient method.

While sharing a common structure, every new quadratic cost is specified by the current solution. Its proper definition involves the pointwise nonlinear estimation of scalar quantities, which is a reweighting process akin to the one of IRLS. In our bound-optimization framework, each successive solution partially minimizes $\mathcal{J}^0(\cdot|\tilde{c}^{(n)})$ with respect to its current value at $\tilde{c}^{(n)}$. Finding this solution amounts to partially solving a linear problem with a given initialization. We propose to precondition each of these linear problems according to its particular structure and find an approximate solution using the linear conjugate-gradient (CG) method. This approach ensures the global convergence of our method without having to specify any step parameter.

According to Figure 3.4, the successive reweighting and linear-resolution steps

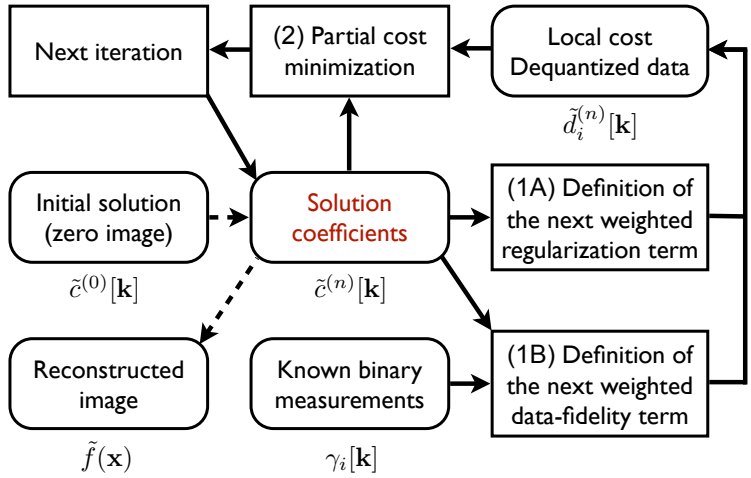


Figure 3.4: Overall principle of our reconstruction algorithm. The solution coefficients are first initialized to zero and then updated by minimizing successive quadratic-cost functionals. Using the current solution $\tilde{c}^{(n)}$, Steps (1A) and (1B) determine the next local cost. Each of these two steps is related to a deconvolution problem where the data $\tilde{d}_i^{(n)}$ to deconvolve correspond to dequantized versions of the available γ_i. An updated solution is found after minimization in Step (2). It determines the coefficients of the next solution. The overall convergence of the process is guaranteed because our global functional is convex.

can be interpreted as alternate dequantization and deconvolution operations, respectively.

3.6.2 Upper bound of the data term

In this part, we derive functionals of simpler form which upper-bound and approximate $\mathcal{D}(\cdot)$ around some initial or current estimate of the solution. Following a *majorization-minimization* (MM) approach [53], we build the local quadratic cost $\mathcal{D}^*(\cdot|\tilde{c}^{(n)})$ for the corresponding estimate $\tilde{c}^{(n)}$ such that

$$\begin{aligned} \mathcal{D}^*(\tilde{c}^{(n)}|\tilde{c}^{(n)}) &= \mathcal{D}(\tilde{c}^{(n)}), \\ \mathcal{D}^*(\tilde{c}|\tilde{c}^{(n)}) &\geq \mathcal{D}(\tilde{c}). \end{aligned} \tag{3.26}$$

For convenience, we bound the cost by the penalty $\Psi_{\mathcal{D}}(\cdot)$. This fixes the structure of $\mathcal{D}^*(\cdot|\tilde{c}^{(n)})$ as

$$\mathcal{D}^*(\tilde{c}|\tilde{c}^{(n)}) = \sum_{i=1}^{L} \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi_i[\mathbf{k}] \Psi_{\mathcal{D}}(\tilde{g}_i[\mathbf{k}]|\tilde{g}_i^{(n)}[\mathbf{k}], \gamma_i[\mathbf{k}]), \tag{3.27}$$

where $\tilde{g}_i^{(n)}$ is the current estimate of $\tilde{g}_i$ associated with the solution estimate $\tilde{c}^{(n)}$, and where $\Psi_{\mathcal{D}}(\cdot|\tilde{g}_i^{(n)}, \gamma_i)$ is a quadratic and scalar penalty function which takes the form

$$\Psi_{\mathcal{D}}(\tilde{g}_i|\tilde{g}_i^{(n)}, \gamma_i) = a_2(\tilde{g}_i^{(n)}, \gamma_i)\tilde{g}_i^2 + a_1(\tilde{g}_i^{(n)}, \gamma_i)\tilde{g}_i + a_0(\tilde{g}_i^{(n)}, \gamma_i), \tag{3.28}$$

where the $a_j(\tilde{g}_i^{(n)}, \gamma_i)$ are polynomial coefficients. The values of $\tilde{g}_i$ and γ_i depend on the solution estimate and the available binary measurements. Constraints (3.26) are then satisfied by fulfilling the simpler scalar conditions $\forall\gamma \in \{-1, 1\}$ and $\forall t \in \mathbb{R}$,

$$\begin{aligned} \Psi_{\mathcal{D}}(\tilde{g}^{(n)}|\tilde{g}^{(n)}, \gamma) &= \Psi_{\mathcal{D}}(\gamma\tilde{g}^{(n)}), \\ \Psi_{\mathcal{D}}(t|\tilde{g}^{(n)}, \gamma) &\geq \Psi_{\mathcal{D}}(\gamma t), \end{aligned} \tag{3.29}$$

where the subscripts have been dropped for convenience. These relations constrain the value of $\Psi_{\mathcal{D}}(\cdot|\tilde{g}_i^{(n)}, \gamma_i)$ and its derivative at $\tilde{g}^{(n)}$. As illustrated in Figure 3.3, further optimizing this upper penalty bound to best approximate $\Psi_{\mathcal{D}}(\gamma t)$ exhausts every remaining degree of freedom. This solution corresponds to the smallest positive a_2 in (3.28) that allows (3.29) to be satisfied. The particular definition that we have proposed for the penalty function $\Psi_{\mathcal{D}}(\cdot)$ allows for fast noniterative evaluation of the coefficients a_j. The actual expressions are derived in Appendix 3.9.1. The resulting coefficients then specify the quadratic cost $\mathcal{D}^*(\cdot|\tilde{c}^{(n)})$ as

$$\mathcal{D}^*(\tilde{c}|\tilde{c}^{(n)}) = \sum_{i=1}^{L} \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi_i[\mathbf{k}] a_2(\tilde{g}_i^{(n)}[\mathbf{k}], \gamma_i[\mathbf{k}]) \left(\tilde{g}_i[\mathbf{k}] - \tilde{d}_i^{(n)}[\mathbf{k}]\right)^2 + \mathcal{K}_+, \tag{3.30}$$

where the scalar $\mathcal{K}_+$ is constant with respect to $\tilde{c}$, and where $\tilde{d}_i^{(n)}$ is a sequence defined as

$$\tilde{d}_i^{(n)}[\mathbf{k}] \quad = \quad -\frac{1}{2}(a_2^{-1}a_1)(\tilde{g}_i^{(n)}[\mathbf{k}], \gamma_i[\mathbf{k}]). \tag{3.31}$$

Since the value of the constant $\mathcal{K}_+$ is irrelevant for minimization, we define the cost $\mathcal{D}(\cdot|\tilde{c}^{(n)})$ as $\mathcal{D}^*(\cdot|\tilde{c}^{(n)})$ minus that constant. Dropping the subscript I for convenience, its explicit form in matrix notation as a function of the coefficients reduces to

$$\mathcal{D}(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) = \sum_{i=1}^{L} \left\| \mathbf{W}_i^{\frac{1}{2}} \left(\mathbf{A}_i \tilde{\mathbf{c}} - \tilde{\mathbf{d}}_i^{(n)} \right) \right\|_{\ell_2}^2, \tag{3.32}$$

where $\mathbf{W}_i$ is a diagonal matrix with diagonal components $\chi_i[\mathbf{k}] a_2(\tilde{g}_i^{(n)}[\mathbf{k}], \gamma_i[\mathbf{k}])$ and where $\tilde{\mathbf{d}}_i^{(n)}$ is the vector associated with $\tilde{d}_i^{(n)}$.

3.6.3 Upper bound of the regularizer

The Huber-based convex functional $\mathcal{R}^0(\cdot)$ can be bound from above according to the same MM principles. The form of $\mathcal{R}^0(\cdot|\tilde{c}^{(n)})$ can be deduced from the results of [90]. Its matrix expression is

$$\mathcal{R}^0(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) = \Lambda_{\mathrm{E}} \|\tilde{\mathbf{c}}\|_{\ell_2}^2 + \left\| \mathbf{\Theta}^{\frac{1}{2}} \mathbf{R} \tilde{\mathbf{c}} \right\|_{\ell_2}^2, \tag{3.33}$$

where $\mathbf{\Theta}$ is a diagonal matrix with diagonal components

$$\theta_{\mathrm{N}}[\mathbf{k}] = \max(\|(\tilde{c}^{(n)} \star \mathbf{r})[\mathbf{k}]\|, \epsilon)^{-1}, \tag{3.34}$$

and where $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2)$ is the discretized-gradient matrix. Each term $\mathbf{R}_i$ is a circulant matrix associated with the filter r_i defined in (3.24).

3.6.4 Quadratic-cost minimization

Combining the data and regularization terms (3.32) and (3.33), we obtain the local quadratic cost

$$\mathcal{J}^0(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) = \mathcal{D}(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}) + \Lambda\mathcal{R}^0(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n)}). \tag{3.35}$$

In order to decrease $\mathcal{J}^0(\cdot)$, the new estimate $\tilde{\mathbf{c}}^{(n+1)}$ must decrease the upper bound $\mathcal{J}^0(\cdot|\tilde{\mathbf{c}}^{(n)})$ itself. In other words, we have to satisfy

$$\mathcal{J}^0(\tilde{\mathbf{c}}|\tilde{\mathbf{c}}^{(n+1)}) \leq \mathcal{J}^0(\tilde{\mathbf{c}}^{(n)}|\tilde{\mathbf{c}}^{(n)}). \tag{3.36}$$

Defining $\mathbf{I}$ as the identity matrix, the minimum of $\mathcal{J}^0(\cdot|\tilde{\mathbf{c}}^{(n)})$ is the solution of

$$\mathbf{S}\tilde{\mathbf{c}} = \mathbf{y}, \tag{3.37}$$

with the system matrix

$$\mathbf{S} = \sum_{i=1}^{L} {\mathbf{A}_i}^{\mathrm{T}}\mathbf{W}_i\mathbf{A}_i + \Lambda\sum_{i=1}^{2} {\mathbf{R}_i}^{\mathrm{T}}\boldsymbol{\Theta}\mathbf{R}_i + \Lambda\Lambda_{\mathrm{E}}\mathbf{I} \tag{3.38}$$

and the right-hand-side vector

$$\mathbf{y} = \sum_{i=1}^{L} {\mathbf{A}_i}^{\mathrm{T}}\mathbf{W}_i\tilde{\mathbf{d}}_i^{(n)}. \tag{3.39}$$

The huge matrix sizes entering into play require (3.37) to be solved iteratively. The positivity of the upper-penalty-bound coefficients a_2 and of the weights θ_{N} in (3.34) implies symmetry and positive-definiteness of $\mathbf{S}$, which allows for the CG method to be used. Initializing the latter at the current estimates, we guarantee the corresponding approximate solutions to comply with (3.36).

3.6.5 Preconditioning

We also take advantage of preconditioning to obtain an approximate solution $\tilde{\mathbf{c}}^{(n+1)}$ that is close to the exact minimum with fewer iterations. We impose our preconditioner $\mathbf{P}$ to be a positive-definite circulant matrix, and define the two-sided preconditioned system

$$\mathbf{S}' = \mathbf{P}^{-\frac{1}{2}}\mathbf{S}\mathbf{P}^{-\frac{1}{2}}. \tag{3.40}$$

It is associated with the modified linear problem

$$\mathbf{S}'\tilde{\mathbf{c}}' = \mathbf{y}', \tag{3.41}$$

where $\mathbf{y}'$ is predetermined as $\mathbf{y}' = \mathbf{P}^{-\frac{1}{2}}\mathbf{y}$ and where the actual solution $\tilde{\mathbf{c}}$ of the original problem is recovered as $\tilde{\mathbf{c}} = \mathbf{P}^{-\frac{1}{2}}\tilde{\mathbf{c}}'$. As a solution satisfying the above requirements, we consider

$$\mathbf{P} = \mathbf{F}^* \operatorname{diag}(\mathbf{F}\mathbf{S}\mathbf{F}^*)\,\mathbf{F}, \tag{3.42}$$

where $\mathbf{F}$ is the normalized DFT operator, where * denotes the adjoint, and where $\operatorname{diag}(\cdot)$ is a projector onto the diagonal-matrix space. Definition (3.42) corresponds to the optimal circulant approximation of $\mathbf{S}$ with respect to the Frobenius norm [91]. This solution is well-adapted to its convolutive nature as compared to diagonal preconditioning.

3.6.6 Minimization scheme

The successive quadratic bounds and the corresponding preconditioned linear problems being defined, we now describe the overall iterative minimization scheme that yields the solution $\tilde{\mathbf{c}}$, starting from an initialization $\tilde{\mathbf{c}}^{(0)}$. Our overall scheme is composed of two embedded iterative loops. The weight specification of the successive quadratic costs corresponds to external iterations with solutions $\tilde{\mathbf{c}}^{(n)}$.

Since our algorithm involves upper bounds that are partially minimized and that satisfy MM conditions of the form (3.26), it is part of the generalized MM (GMM) family [51]. In that regard, the continuity of our functional $\mathcal{J}^0(\cdot|\tilde{\mathbf{c}}^{(n)})$ implies that the MM sequence $\{\mathcal{J}^0(\tilde{\mathbf{c}}^{(0)}), \mathcal{J}^0(\tilde{\mathbf{c}}^{(1)}), \mathcal{J}^0(\tilde{\mathbf{c}}^{(2)}), \ldots\}$ converges monotonically to a stationary point of $\mathcal{J}^0(\cdot)$. The corresponding solution sequence $\{\tilde{\mathbf{c}}^{(0)}, \tilde{\mathbf{c}}^{(1)}, \tilde{\mathbf{c}}^{(2)}, \ldots\}$ also converges because each $\tilde{\mathbf{c}}^{(n)}$ is bounded [92]. The convexity of $\mathcal{J}^0(\cdot)$ and the Lipschitz continuity of the gradient also imply that the whole minimization process is compatible with Nesterov's acceleration technique [82], which we apply to update our estimates. This requires the use of auxiliary solutions that we mark with star subscripts, as well as the definition of scalar iterative-step values $\omega^{(I)}$. The steps of our global scheme yielding the solution $\tilde{c}$ are described in Algorithm 3.1.

We use N_i external iterations, each of which corresponds to a refined quadratic approximation $\mathcal{J}^0(\cdot|\tilde{\mathbf{c}}^{(n)})$ of the global convex cost $\mathcal{J}^0(\cdot)$. For the partial resolution of each internal problem, we apply CG on the modified system (3.41). Accordingly, the corresponding intermediate value $\tilde{\mathbf{c}}'$ is first initialized to the current solution

1) Initialize the solution $\tilde{\mathbf{c}}^{(0)}$ as the zero vector
2) Initialize $\tilde{\mathbf{c}}_*^{(0)} = \tilde{\mathbf{c}}^{(0)}$, $I = 0$, and $\omega^{(0)} = 1$
while $I < N_\mathrm{i}$ **do**
  a) Compute the matrix $\mathbf{S}$ and the vector $\mathbf{y}$ given $\tilde{\mathbf{c}}^{(n)}$
  b) Compute the preconditioner $\mathbf{P}$ linked to $\mathbf{S}'$
  c) Update the external-iteration counter as $I \leftarrow I + 1$
  d) Compute the vector $\mathbf{y}' = \mathbf{P}^{-\frac{1}{2}}\mathbf{y}$
  e) Initialize the vector $\tilde{\mathbf{c}}'$ as $\tilde{\mathbf{c}}' = \mathbf{P}^{\frac{1}{2}}\tilde{\mathbf{c}}^{(n-1)}$
  f) Update $\tilde{\mathbf{c}}'$ using N_i^* CG iterations on the linear problem $\mathbf{S}'\tilde{\mathbf{c}}' = \mathbf{y}'$
  g) Do the Nesterov's solution update $\tilde{\mathbf{c}}_*^{(n)} = \mathbf{P}^{-\frac{1}{2}}\tilde{\mathbf{c}}'$
  h) Do the Nesterov's step update $\omega^{(I)} = \sqrt{\left(\omega^{(I-1)}\right)^2 + 1/4} + 1/2$
  i) Compute $\tilde{\mathbf{c}}^{(n)} = \tilde{\mathbf{c}}_*^{(n)} + {\omega^{(I)}}^{-1}(\omega^{(I-1)} - 1)(\tilde{\mathbf{c}}_*^{(n)} - \tilde{\mathbf{c}}_*^{(n-1)})$
end
3) Compute the solution coefficients $\tilde{\mathbf{c}} = \tilde{\mathbf{c}}^{(n)}$

Algorithm 3.1: Minimization approach described in matrix notation.

estimate in the preconditioned domain, and then updated using N_i^* CG iterations each time. In accordance with (3.13), the final continuous-domain image is obtained from the coefficients $\tilde{c}$ as

$$\tilde{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \tilde{c}[\mathbf{k}]\varphi(\mathbf{x} - \mathbf{k}). \tag{3.43}$$

As demonstrated in Section 3.7.1, the use of Nesterov's technique and of preconditioning to solve the linear problems ensure the fast convergence of our method.

3.7 Experiments

We conduct experiments on grayscale images that are part of a standard test set. First, we evaluate the computational performance of our algorithm in Section 3.7.1 and show baseline results in Section 3.7.2. In Section 3.7.3, we propose an estimate of the acquisition quality based on the spatial redundancy of the available measurements. In Sections 3.7.4 and 3.7.5, we address cases where downsampling and

finite differentiation are used for data acquisition. In particular, we determine to what extent these strategies impact on the acquisition and reconstruction quality. We finally assess the optimal rate-distortion performance of our method for distinct amounts of measurements in Section 3.7.6.

The discretization of type (3.13) does not induce any loss because we match the square grid of $N_0 \times N_0$ spline coefficients to the resolution of each digital test image, choosing $\eta = 1$. Specifically, we determine c beforehand such that f interpolates the corresponding pixel values[6]. In order to maximize the acquisition bandwidth, the size $N_\mathrm{d} \times N_\mathrm{d}$ of the phase mask and the number $M_0 \times M_0$ of sensors are themselves set to $N_0 \times N_0$. The sampling prefilter φ_0 is defined as a 2D separable rectangular window. The threshold τ is set to the mean image intensity[7] when no finite differentiation is used, and to zero otherwise. The latter choice is a heuristic that directly yields equidistributed binary measurements γ_i from our data as in [10], without requiring any optimization or further refinement. For non-unit K, we consider identical spatial masks χ_i that correspond to horizontal and vertical subsampling, which allows for the proper display and evaluation of our measurements. Our reconstruction parameters are $\Lambda = 10^{-4}$, $\Lambda_\mathrm{E} = 10^{-5}$, $\epsilon = 5 \cdot 10^{-4}$, $N_\mathrm{i} = 20$, and $N_\mathrm{i}^* = 4$. The smoothing parameter ϵ chosen for our regularizer aims at approximating TV as in [88], while the small values of the constants Λ and Λ_E ensure that the reconstructions are consistent with the binary measurements with enough accuracy (*i.e.*, about 99% or above).

We have found that the most-consistent solutions are also the ones of highest quality, which corroborates the results of [14]. Knowing that each instance of (3.41) can be solved partially, the choice of N_i^* is meant to maximize computational performance, while the value of N_i is used as a stop criterion. Note that the values of ϵ and Λ cannot be reduced further without impacting negatively on the speed of convergence; choosing an arbitrarily small ϵ would actually make our algorithm tend to a subgradient-descent-type scheme associated with worse convergence rates.

In order to provide a quality assessment in terms of SNR, the mean and variance of the solution coefficients are matched to the reference signal. We also define a quantity called blockwise-corrected SNR (BSNR) where this same matching is performed blockwise using 8×8 blocks. As discussed in Section 3.7.4, the BSNR

[6] Given our forward model and the high values of N_0 involved in our experiments, the choice of η has no significant impact.

[7] This quantity corresponds to the mean component value of the vector $\mathbf{g}$. It is assumed to be known for reconstruction.

is consistent with visual perception.

3.7.1 Computational performance

To evaluate the computational performance of our algorithm, we perform a reconstruction experiment on a 256×256 test image using ${M_0}^2 = 256^2$, $L = 1$, $K = 1$, and no finite differentiation. The results are reported in Figure 3.5, including a comparison with the BIHT algorithm[8] introduced for reconstruction from binary measurements in [14]. These results demonstrate that Nesterov's acceleration method, as well as the preconditioning used in our algorithm, play a central role to obtain fast convergence. By contrast, we have observed that 3,000 iterations are required to ensure convergence with BIHT—which is used for the experiments of Section 3.7.4—as opposed to a total of $N_\mathrm{i} N_\mathrm{i}^* = 80$ internal iterations with our algorithm. This corresponds to an order-of-magnitude improvement in time efficiency.

3.7.2 Baseline results

Our framework can handle several measurement sequences unlike in [10]. Accordingly, the goal in this part is to reconstruct the 512×512 images *Lena* and *Barbara* from distinct numbers L of acquisitions with $K = 1$ and no finite differentiation. Each acquisition includes ${M_0}^2 = 512^2$ samples, the total number of measurements being multiplied by the corresponding L.

The binary acquisitions and the corresponding reconstructions with our algorithm are shown in the spatial domain in Figures 3.6 and 3.7. In both examples, the reconstruction quality substantially improves with L, one single acquisition being already sufficient to preserve substantial grayscale and edge information. The binary measurements of Figures 3.6 and 3.7 are not interpretable visually because the image information has been spread out through the filters h_i. These measurements follow a random distribution that originates from the pseudo-random phases

[8] We have adapted BIHT to our forward model, assuming sparsity in the Haar-wavelet domain. Besides its simplicity, the latter choice was observed to yield higher-quality results in our case than when using higher-order Daubechies wavelets, despite the generated block artifacts. Each iteration involves a gradient step scaled as $M^{-1/2}\|\mathbf{A}\|_2^{-1}$ and renormalization [14]. A zero-mean $\mathbf{A}$ is used in the algorithm to handle the case where τ is nonzero. The sparsity-level parameter specifying the assumed amount of nonzero wavelet coefficients has been determined experimentally for best reconstruction performance and set as 2,000. Both BIHT and the proposed algorithms have been implemented in MATLAB.

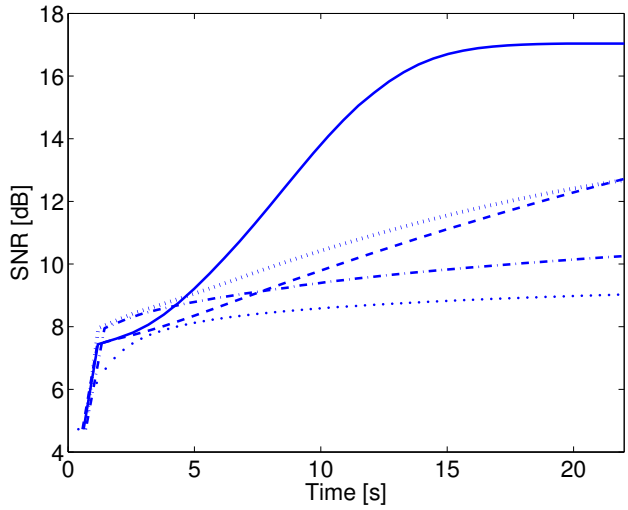
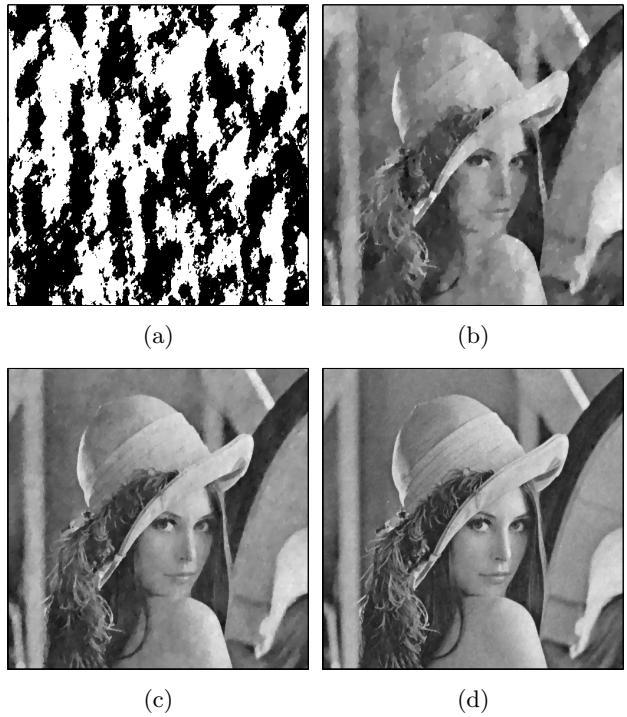


Figure 3.5: Reconstruction SNR as a function of time for *Montage* (256×256). For our reconstruction method, the sole use of preconditioning (dashed line) or Nesterov's acceleration (dotted line) already improves the convergence rates as compared to standard CG (mixed line). When both techniques are enabled (solid line), the performance of our algorithm improves substantially. For comparison, the reconstruction performance of BIHT is also shown for the same problem (bottom dots). In the latter case, each corresponding iteration lasts about half a second. The times that are given correspond to an execution of the algorithms on Mac OS X version 10.7.1 (MATLAB R2011b) with a Quad-Core Xeon 2 $\times$ 2.8 GHz and 4 GB of DDR2 memory.

ν_i of the masks, and that is heavily correlated spatially as in [10]. As a matter of fact, random-convolution measurements do not display strict statistical incoherence [12]. We investigate below how spatial correlation can be quantified and reduced to improve reconstruction.



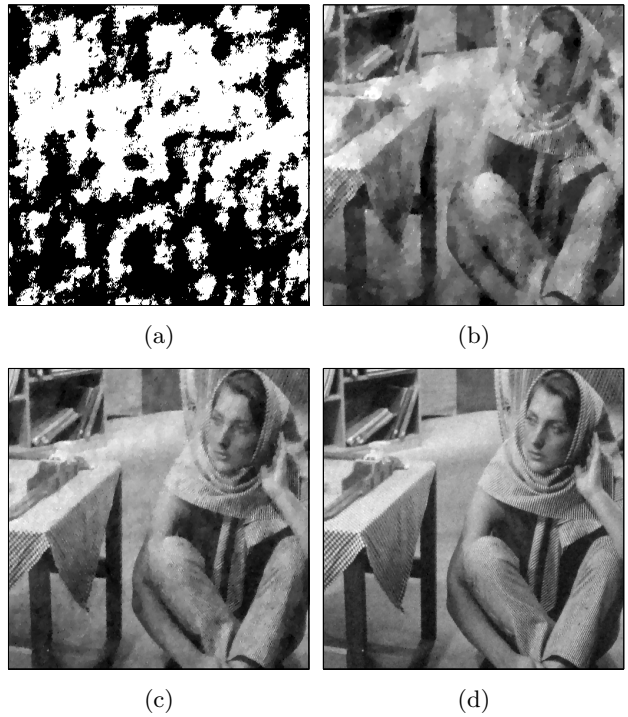
(a) (b)

(c) (d)

Figure 3.6: Results on *Lena* (512×512) for distinct numbers L of acquisitions using ${M_0}^2 = 512^2$ and $K = 1$ without finite differentiation ($M = L \cdot 512^2$ measurements in total). (a) First acquisition γ_1 obtained from our model (b)–(d) Reconstruction from one (M = 262,144, SNR: 17.49 dB, BSNR: 22.35 dB), two (M = 524,288, SNR: 22.42 dB, BSNR: 24.61 dB), and four (M = 1,048,576, SNR: 26.46 dB, BSNR: 27.13 dB) acquisitions.

3.7.3 Incoherence estimation

The potential quality of reconstruction depends on the appropriateness of $\mathbf{A}$ for binary compressed sensing. We assume our matrix to be suitable for the specific data in hand when the corresponding binarized measurements behave as independent



(a) (b)

(c) (d)

Figure 3.7: Results on *Barbara* (512×512) for distinct numbers L of acquisitions using ${M_0}^2 = 512^2$ and $K = 1$ without finite differentiation ($M = L \cdot 512^2$ measurements in total). (a) First acquisition γ_1 obtained from our model (b)–(d) Reconstruction from one ($M = 262{,}144$, SNR: 13.96 dB, BSNR: 16.09 dB), two ($M = 524{,}288$, SNR: 17.69 dB, BSNR: 17.74 dB), and four ($M = 1{,}048{,}576$, SNR: 20.3 dB, BSNR: 20.28 dB) acquisitions.

and identically distributed random variables. As a practical solution, we propose to estimate the "randomness" of the acquired γ_i through their autocorrelation [93]. We specifically infer a correlation distance l_c based on the L unnormalized autocorrelations R_i of our (possibly subsampled) binary sequences. This distance is used as a quality indicator, inasmuch as it measures the degree of spatial redundancy

arising in our measurements. To determine this value, we first compute the characteristic length $(l_c)_i$ of each autocorrelation peak, using the standard deviation of $|R_i|^4$ for the sake of robustness. The autocorrelation being symmetric and centered at the origin, we write that

$$(l_c)_i = \left(\frac{\sum_{\mathbf{k}} |R_i[\mathbf{k}]|^4 \|\mathbf{k}\|^2}{\sum_{\mathbf{k}} |R_i[\mathbf{k}]|^4} \right)^{1/2}. \tag{3.44}$$

Averaging $(l_c)_i$ over i then yields the final l_c. As shown in the sequel, this value strongly depends on the parameters of the forward model. In particular, it can be decreased compared to the case of Section 3.7.2 by enabling downsampling (*i.e.*, non-unit K) or finite differentiation in our framework. Note that, as in [10], our choice for the threshold τ ensures the uniformity of the binary distribution of the measurements.

3.7.4 Influence of acquisition modality

In this section, we investigate the performance of finite differentiation when used in our framework. To this end, we choose a fixed set of two perpendicular first-derivative filters whose convolution masks are $[1 \;\; 0 \;\; -1]$ and $[1 \;\; 0 \;\; -1]^{\mathrm{T}}$ for the horizontal and vertical orientations, respectively. Assuming an even L, the former filter is applied on acquisition sequences of even index, and the latter one is applied on the remaining indices. The operation of each filter ζ_i followed by zero thresholding is physically realizable by means of binary comparators that are connected to the two corresponding pixels. From a practical standpoint, such an approach eliminates the need of threshold calibration.

In order to compare the acquisition modalities with and without finite differentiation, we perform experiments on several 256×256 images. These experiments involve $M = 131{,}072$ measurements taken in $L = 2$ acquisitions, using ${M_0}^2 = 256^2$ and $K = 1$. Besides our own algorithm, BIHT is also considered for reconstruction in each case. The results are reported in Table 3.1, and shown in Figure 3.8 for *House*. The best numerical values are emphasized in the tables using bold notation.

Our qualitative and quantitative results demonstrate that finite differentiation globally yields the best reconstructions. These solutions consistently correspond to lower l_c values as well, which reflects itself visually in less-redundant binary measurements. Finite differentiation decreases redundancy because it spatially decorre-

Image	Modality	Proposed (TV) SNR	Proposed (TV) BSNR	BIHT (Haar) SNR	BIHT (Haar) BSNR	l_c
Bird	Standard	25.64	27.80	19.80	22.56	54.00
	Finite differences	**25.81**	**31.66**	15.17	23.88	**33.08**
Cameraman	Standard	20.65	20.96	15.95	16.32	64.99
	Finite differences	**22.63**	**24.04**	5.87	17.16	**16.04**
House	Standard	**25.67**	26.44	20.40	21.58	47.82
	Finite differences	24.38	**28.85**	13.83	22.30	**20.16**
Peppers	Standard	**20.16**	21.79	14.71	15.43	40.30
	Finite differences	18.21	**24.95**	7.15	15.61	**19.87**
Shepp-Logan	Standard	19.25	20.00	9.53	9.95	34.15
	Finite differences	**22.96**	**25.24**	5.72	12.26	**11.58**

Table 3.1: Acquisition modalities compared on 256×256 images using ${M_0}^2 = 256^2$, $L = 2$, and $K = 1$ ($M = 131{,}072$).

lates the image measurements g_i before quantization. Because finite differentiation senses the high-frequency content of the measurements, most visual features such as edges are indeed better restored as compared to the other acquisition modalities. In return, reconstructions tend to display slightly higher low-frequency error. Because of its cumulative nature, the latter may then cause substantial SNR deterioration in unfavorable scenarios. In such cases, however, the amount of visual details is still higher, as illustrated in Figure 3.8. For instance, fine details such as the house gutter are better preserved. We observe that the BSNR measure is consistent with visual impression, as it adapts to slow intensity drifts in the solution. For both acquisition modalities, our algorithm based on TV yields the best reconstructions. This confirms the suitability of TV for our problem, in accordance with the discussion of Section 3.5.4. Note, however, that proper adjustment of the sparsity level in BIHT is delicate. For instance, images that are sparser than the assumed level might lead to suboptimal reconstructions in Table 3.1.

3.7.5 Respective influence of K and L

The following experiments address how reconstruction quality can be maximized given a fixed measurement budget, using the same 256×256 images as above. Considering the finite-differentiation modality specified in Section 3.7.4, our strategy is to further decrease spatial redundancy by sharing the measurements between more

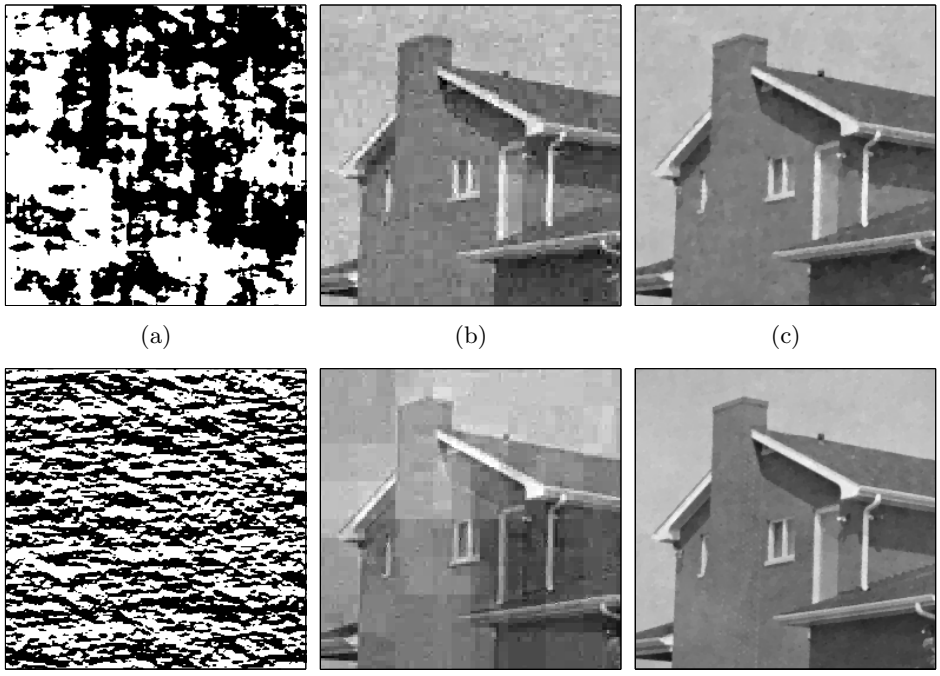


Figure 3.8: Acquisition modalities compared on *House* (256×256) using ${M_0}^2 = 256^2$, $L = 2$, and $K = 1$ ($M = 131{,}072$). (a) First acquisition γ_1 without finite differentiation (b)–(c) Reconstruction without finite differentiation using BIHT (SNR: 20.40 dB, BSNR: 21.58 dB) and our algorithm (SNR: 25.67 dB, BSNR: 26.44 dB) (d) First acquisition γ_1 with finite differentiation (e)–(f) Reconstruction with finite differentiation using BIHT (SNR: 13.83 dB, BSNR: 22.3 dB) and our algorithm (SNR: 24.38 dB, BSNR: 28.85 dB).

acquisitions. Choosing ${M_0}^2 = 256^2$ and $M = 32{,}768$ as constraints, we thus adapt the ratio K to the number of acquisitions as $K^{-1} = 2L$. On the one hand, minimizing L reduces to previous system configurations. On the other hand, maximizing it is highly inefficient, as it amounts to taking one single measurement per convolutive

Image	**Measure**	$L = 2$ $K = 1/4$	$L = 4$ $K = 1/8$	$L = 8$ $K = 1/16$	$L = 16$ $K = 1/32$	$L = 32$ $K = 1/64$
Bird	l_c	13.76	10.57	5.66	4.94	**2.31**
	SNR	22.62	22.77	24.30	25.35	**25.37**
	BSNR	28.89	29.25	29.41	**29.57**	29.49
Cameraman	l_c	5.91	3.77	1.92	1.59	**1.04**
	SNR	18.73	18.63	**19.91**	19.81	19.53
	BSNR	20.79	21.08	21.30	21.26	**21.38**
House	l_c	8.05	6.35	3.74	2.83	**1.78**
	SNR	20.71	21.10	24.01	24.05	**24.56**
	BSNR	26.34	26.51	26.81	26.88	**26.96**
Peppers	l_c	8.01	6.14	2.99	2.63	**1.49**
	SNR	15.09	15.68	18.95	19.01	**19.19**
	BSNR	21.29	21.98	22.28	22.42	**22.47**
Shepp-Logan	l_c	4.26	2.59	1.51	1.27	**0.93**
	SNR	16.88	16.84	17.20	17.48	**17.49**
	BSNR	19.42	19.50	19.60	**19.64**	19.58

Table 3.2: Influence of K and L evaluated on 256×256 images using ${M_0}^2 = 256^2$ and finite differentiation. The same number of measurements $M = 32{,}768$ is shared between distinct numbers of acquisitions ($M/N = 1/2$).

acquisition. A tradeoff has to be found between these two limits to improve the quality of the reconstructions while preserving the parallelism of our model.

Our numerical results are reported in Table 3.2, the measurements and reconstruction of *Peppers* being shown for two distinct settings in Figure 3.9. The values of Table 3.2 confirm that the correlation length l_c consistently decreases with K. Moreover, the SNR and BSNR improve by several decibels when increasing L. This is further corroborated by the visual results of Figure 3.9. In particular, grayscale information is more-finely preserved in the solution displayed on the right. Interestingly, the increase in quality starts saturating when l_c reaches near-optimal values, as shown in Table 3.2. The compression performance of our method is thus optimal or nearly optimal with $L \geq 8$ for a given amount of measurements. These results confirm the strong inverse correlation between measurement redundancy and reconstruction quality.

3.7.6 Rate-distortion performance

In this section, we confront our global acquisition and reconstruction framework (GF) described in this chapter with the simpler single-convolution framework (SF) that we had proposed in [10]. The following experiments allow us to evaluate their respective image-reconstruction performance in terms of the rate of distortion, defining the number of bits per pixel (bpp) as the ratio between M and the raw bitsize of the corresponding uncompressed 8-bit-grayscale image.

In order to decrease l_c within a reasonable amount of acquisitions, our forward model is parameterized with $L = 8$ and $K = L^{-1}B$, depending on the chosen bitrate B in bpp. The number of measurements taken on an $N_0 \times N_0$ test image is thus $M = B{N_0}^2$ since $M_0 = N_0$. Our method is evaluated with (D) and without (S) finite differentiation. In the SF case, the sensor resolution has to match M strictly, because one single convolution is performed without subsequent drop of samples. The forward model is configured accordingly, adapting the remaining parameters to the image size as in our method. That particular framework requires equal rational factors for resampling, which implies that certain bitrates cannot be evaluated. The reconstruction parameters are set as in corresponding experiments of [10].

Results on several test images are reported in Tables 3.3 and 3.4. They indicate that at least one version of our method always exceeds SF in terms of reconstruction quality. This confirms the relevance of sharing the acquired data between more acquisitions as a means of decreasing spatial redundancy. This strategy thus tends to compensate the non-ideal statistical properties of binary measurements that are based on random convolutions. In the case of SF, spatial redundancy cannot be decreased similarly since only one convolutive acquisition is used. As previously observed, the (D) modality of our method can yield worse SNR values in certain configurations, while displaying superior BSNR performance globally. Nevertheless, these complementary results reveal an advantageous SNR performance of (D) at higher bitrates.

The efficiency of our method at 1/8 bpp, which corresponds to a compression factor of 64, is illustrated for both modalities in Figure 3.10. Also shown is the plain JPEG version of the image compressed at similar bitrate. In this example, the GF framework with finite differentiation yields the best BSNR. We observe that the corresponding reconstruction contains fine details despite the low amount of measurements. It is also visually more pleasant than the JPEG solution. This experiment illustrates the highest compression ratio at which our method reconstructs

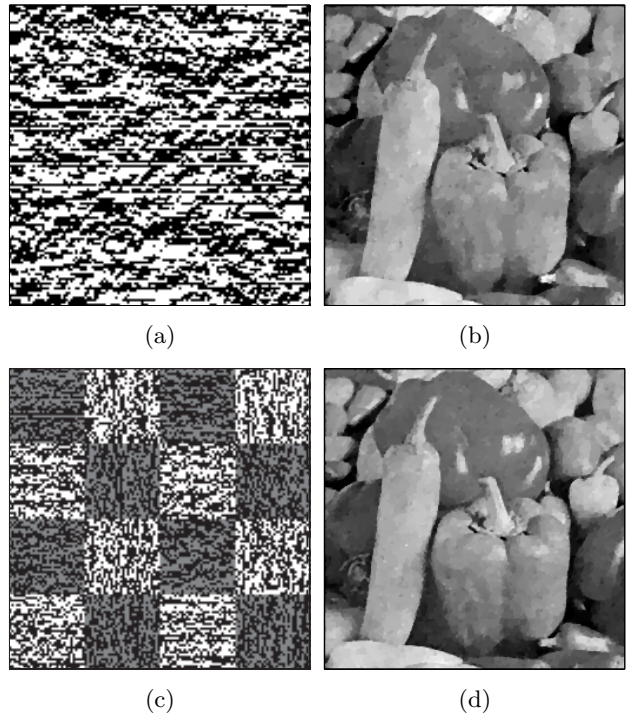


Figure 3.9: Results on *Peppers* (256×256) when sharing $M = 32{,}768$ measurements between distinct numbers of acquisitions with finite differentiation and ${M_0}^2 = 256^2$. (a)–(b) Acquisition and reconstruction for $L = 2$ and $K = 1/4$ with γ_1 (SNR: 15.09 dB, BSNR: 21.29 dB) (c)–(d) Acquisition and reconstruction for $L = 32$ and $K = 1/64$ with γ_1 to γ_{16} shown in concatenated form using a gray/white checkerboard-type display (SNR: 19.19 dB, BSNR: 22.47 dB).

images with reasonable quality. From a general standpoint, the results of this section demonstrate that, although generally inferior, the rate-distortion performance of binary compressed sensing can compete with JPEG at low bitrates. This can be deduced by comparing the plain-JPEG performance to the corresponding SNR values reported in Table 3.3 and corroborates the analysis of [94] where compressed

Bitrate B [bpp]		**1/16**	**1/8**	**1/4**	**1/2**	**1**	**2**	**4**
Sampling Ratio K^*		**1/128**	**1/64**	**1/32**	**1/16**	**1/8**	**1/4**	**1/2**
Bird	SF	18.35	-	21.27	-	22.95	-	23.78
(256 × 256)	GF (S)	**19.44**	**21.65**	**23.74**	**25.58**	27.13	28.36	28.58
	GF (D)	16.84	19.79	22.30	24.30	**27.34**	**30.67**	**33.17**
Cameraman	SF	13.86	-	16.34	-	18.14	-	19.33
(256 × 256)	GF (S)	**14.63**	**16.06**	**17.26**	18.54	19.78	21.27	22.67
	GF (D)	11.33	15.00	17.20	**19.91**	**21.96**	**23.73**	**25.72**
House	SF	17.36	-	20.74	-	23.21	-	24.60
(256 × 256)	GF (S)	**18.39**	**20.62**	**22.67**	**24.47**	25.74	27.11	27.78
	GF (D)	15.48	18.56	20.94	24.01	**26.62**	**28.78**	**30.39**
Peppers	SF	12.43	-	15.09	-	17.35	-	18.58
(256 × 256)	GF (S)	**14.31**	**15.99**	**17.44**	**19.02**	20.91	23.06	24.67
	GF (D)	10.84	13.40	16.11	18.95	**21.20**	**23.87**	**26.73**
Shepp-Logan	SF	7.28	-	12.57	-	17.33	-	22.33
(256 × 256)	GF (S)	**8.52**	**10.95**	13.34	15.53	17.78	19.52	21.37
	GF (D)	7.98	10.91	**14.14**	**17.20**	**20.16**	**22.94**	**25.57**
Barbara	SF	11.27	-	12.71	-	13.97	-	13.39
(512 × 512)	GF (S)	**14.56**	**15.71**	**16.54**	**17.23**	18.06	19.06	20.94
	GF (D)	8.51	10.38	12.50	15.79	**18.64**	**21.95**	**24.85**
Boat	SF	14.13	-	16.17	-	17.84	-	17.53
(512 × 512)	GF (S)	**16.28**	**17.70**	**19.21**	**20.83**	**22.54**	24.28	25.99
	GF (D)	12.72	14.85	16.42	19.16	22.39	**25.07**	**27.37**
Hill	SF	12.89	-	15.10	-	16.39	-	15.63
(512 × 512)	GF (S)	**16.28**	**17.74**	**18.96**	**20.33**	**21.62**	23.27	24.51
	GF (D)	8.51	10.15	12.45	16.33	19.85	**23.88**	**26.71**
Lena	SF	13.82	-	16.56	-	18.02	-	18.18
(512 × 512)	GF (S)	**18.03**	**19.63**	**21.32**	**23.00**	**24.75**	**26.45**	27.92
	GF (D)	11.36	13.04	15.49	18.38	21.35	25.40	**28.10**
Man	SF	13.03	-	15.47	-	16.96	-	16.50
(512 × 512)	GF (S)	**15.95**	**17.41**	**18.78**	**20.27**	21.76	23.46	24.97
	GF (D)	11.33	13.97	16.56	18.80	**21.94**	**24.83**	**27.65**

* This parameter is used for GF with the constant number of acquisitions $L = 8$.

Table 3.3: Rate-distortion performance of GF with (D) and without (S) finite differentiation compared to SF [10] in terms of SNR.

sensing is compared to traditional image-compression methods.

Bitrate B [bpp]		**1/16**	**1/8**	**1/4**	**1/2**	**1**	**2**	**4**
Sampling Ratio K		**1/128**	**1/64**	**1/32**	**1/16**	**1/8**	**1/4**	**1/2**
Bird	SF	21.85	-	24.44	-	26.05	-	27.14
(256 × 256)	GF (S)	21.94	23.40	25.04	26.41	27.76	28.96	29.57
	GF (D)	**23.56**	**25.65**	**27.65**	**29.41**	**31.19**	**33.07**	**34.54**
Cameraman	SF	14.79	-	17.02	-	18.94	-	20.34
(256 × 256)	GF (S)	15.03	16.03	17.22	18.42	19.68	21.22	22.73
	GF (D)	**16.11**	**17.84**	**19.57**	**21.30**	**23.01**	**24.63**	**26.09**
House	SF	20.39	-	23.08	-	25.10	-	26.31
(256 × 256)	GF (S)	20.40	21.87	23.32	24.73	25.85	27.11	27.87
	GF (D)	**21.25**	**23.47**	**25.15**	**26.81**	**28.23**	**29.80**	**31.10**
Peppers	SF	14.63	-	16.94	-	19.85	-	21.49
(256 × 256)	GF (S)	15.38	16.66	17.80	19.38	21.36	23.51	25.02
	GF (D)	**15.70**	**17.80**	**19.97**	**22.28**	**24.57**	**26.58**	**28.26**
Shepp-Logan	SF	8.73	-	13.88	-	18.24	-	22.89
(256 × 256)	GF (S)	10.21	12.12	14.18	16.17	18.30	20.13	22.25
	GF (D)	**11.74**	**14.49**	**17.09**	**19.60**	**22.22**	**24.84**	**27.6**
Barbara	SF	14.11	-	14.82	-	15.98	-	16.00
(512 × 512)	GF (S)	**14.61**	15.10	15.58	16.17	17.10	18.44	20.95
	GF (D)	14.45	**15.49**	**16.80**	**18.69**	**21.22**	**23.87**	**26.43**
Boat	SF	16.16	-	18.04	-	19.91	-	20.13
(512 × 512)	GF (S)	17.09	18.19	19.53	21.02	22.69	24.35	26.00
	GF (D)	**17.82**	**19.44**	**21.33**	**23.26**	**25.16**	**27.10**	**28.86**
Hill	SF	16.50	-	17.68	-	18.93	-	18.96
(512 × 512)	GF (S)	17.34	18.34	19.26	20.50	21.72	23.29	24.51
	GF (D)	**17.90**	**19.34**	**20.99**	**22.89**	**24.75**	**26.79**	**28.64**
Lena	SF	18.78	-	20.52	-	22.27	-	22.56
(512 × 512)	GF (S)	19.55	20.69	22.11	23.56	25.14	26.69	28.10
	GF (D)	**20.25**	**21.99**	**23.95**	**25.89**	**27.94**	**29.88**	**31.73**
Man	SF	16.33	-	17.98	-	19.61	-	19.83
(512 × 512)	GF (S)	16.97	18.01	19.20	20.48	21.92	23.57	25.08
	GF (D)	**17.42**	**19.03**	**20.89**	**22.89**	**25.03**	**27.11**	**29.09**

Table 3.4: Rate-distortion performance of GF with (D) and without (S) finite differentiation compared to SF [10] in terms of BSNR.

3.7.7 Influence of the optical system

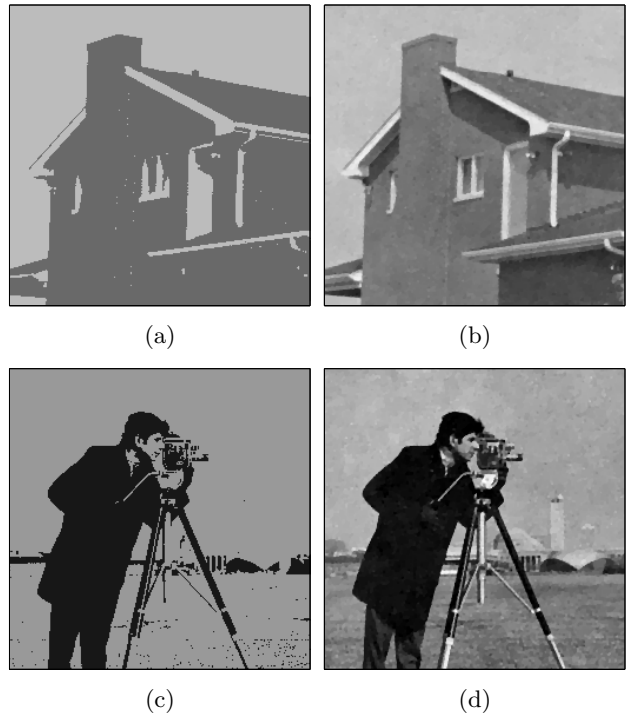
In this last experimental section of the chapter, our goal is to compare the acquisition and reconstruction approach that we propose [10, 11] with the conventional binarization setting where the operator $\mathcal{Q}$ is applied without any prior optical filter-

(a) GF (S) solution (b) GF (D) solution (c) JPEG solution

Figure 3.10: Reconstruction of *Bird* (256×256) at 1/8 bpp ($M = 8{,}192$) using three distinct methods. (a)–(b) GF using ${M_0}^2 = 256^2$, $L = 8$, and $K = 1/64$ without (SNR: 21.65 dB, BSNR: 23.4 dB) and with finite differentiation (SNR: 19.79 dB, BSNR: 25.65 dB) (c) JPEG (SNR: 19.68 dB, BSNR: 22.66 dB). The plain-JPEG compression is performed at its lowest quality settings, which approximately yields the same bitrate (the corresponding file size is 10,280 bits, including header data).

ing with the kernels h_i. For fair comparison, both settings have to extract the same number of bits—in form of binary measurements—from the data. Accordingly, we consider $M = 65{,}536$. In the conventional case, no reconstruction is required, but the binary threshold is optimized with respect to the mean-squared quantization error. The solution is provided by the Lloyd-Max (LM) algorithm [95, 96]. In the case where optical filtering occurs before quantization, we use our GF framework with finite differentiation and with parameters $L = 4$, ${M_0}^2 = 256^2$, and $K = 1/4$.

The results on *House* and *Cameraman* are shown in Figure 3.11. On the one hand, we observe that the direct application of Lloyd-Max quantization perfectly preserves the main object contours. On the other hand, our approach produces compressed-sensing-type measurements that are robust to 1-bit quantization, which allows to recover grayscale information at the cost of the additional procedure of numerical reconstruction. Overall, the obtained SNR and BSNR values are higher when applying our GF acquisition and reconstruction framework than when applying direct quantization. Note that the compressed-sensing measurements that are obtained with our framework typically have substantially less spatial redundancy



(a) (b)

(c) (d)

Figure 3.11: Results on 256×256 images using direct quantization with minimum-error threshold versus our GF framework with ${M_0}^2 = 256^2$, $L = 4$, $K = 1/4$, and finite differentiation. (a) Lloyd-Max result on *House* (SNR: 16.59 dB, BSNR: 20.94 dB) (b) Corresponding GF (D) result (SNR: 23.99 dB, BSNR: 28.08 dB) (c) Lloyd-Max result on *Cameraman* (SNR: 14.78 dB, BSNR: 18.64 dB) (d) Corresponding GF (D) result (SNR: 20.68 dB, BSNR: 22.90 dB).

than the direct Lloyd-Max-quantization results shown in Figure 3.11. This observation corroborates our previous findings on the relationship between redundancy and potential reconstruction quality.

These results validate the 1-bit optical imaging concept that we propose. How-

ever, aspects that are related to practical realization (e.g., precise estimation of the system PSF) will have to be addressed in further research. Let us mention that, according to the example of the gigapixel camera [97], the binary-sensor array of our system could be produced using standard memory-chip technology.

3.8 Conclusions

We have proposed a binary compressed-sensing framework which is suitable for images. In our experiments, we have illustrated how measurement redundancy can be minimized by properly configuring our acquisition model. We have considered the single-acquisition case as well as a multi-acquisition strategy. In the two cases, our reconstruction algorithm has demonstrated state-of-the-art reconstruction performance on standard images. In particular, detailed features have been successfully recovered from small amounts of binary data. From a global perspective, our results confirm the 1-bit-compressed-sensing paradigm to be promising for imaging applications. In that regard, the specific interest of our method is to involve binary measurements that are suitable to convex optimization. We have proposed an iterative algorithm that combines preconditioning and Nesterov's approach to provide very efficient reconstructions of our measurements. Synthetic experiments demonstrate the potential of our method.

Related to this work, an interesting topic that is worth investigating from a theoretical standpoint in future research is the relationship between the spatial redundancy of the measurements obtained according to our forward model and the use of finite differentiation. Another aspect would be to determine to what extent alternate quality measures such as SSIM [98] are consistent with the BSNR measure introduced in this chapter.

3.9 Appendix

3.9.1 Coefficients of the penalty bounds

Formulation of the optimization task

The continuity of $\Psi_{\mathcal{D}}(\gamma t)$ and the upper-bound conditions on $\Psi_{\mathcal{D}}(t|\tilde{g}^{(n)},\gamma)$ impose that the value and first derivative of these two functions coincide at $t=\tilde{g}^{(n)}$. This

requires that

$$\begin{aligned} a_0 &= \Psi_{\mathcal{D}}(\gamma\tilde{g}^{(n)}) - \tilde{g}^{(n)}(\tilde{g}^{(n)}a_2 + a_1), \\ a_1 &= -2\tilde{g}^{(n)}a_2 + (\mathrm{d}\Psi_{\mathcal{D}}(\cdot)/\mathrm{d}\cdot)|_{\cdot=\gamma\tilde{g}^{(n)}}. \end{aligned} \tag{3.45}$$

The remaining degree of freedom $a_2 \in \mathbb{R}_+^*$ is optimized so as to best approximate $\Psi_{\mathcal{D}}(\cdot)$. The resulting optimal a_2 corresponds to the lowest positive value satisfying (3.29). In that configuration, the parabola $\Psi_{\mathcal{D}}(t|\tilde{g}^{(n)}, \gamma)$ touches one and only one distinct point of $\Psi_{\mathcal{D}}(\gamma t)$ at $t = t_{\|}$. The convexity of $\Psi_{\mathcal{D}}(\cdot)$ ensures the existence and uniqueness of the solution.

Solution

According to (3.28), the abscissas of the intersections between the functions $\Psi_{\mathcal{D}}(\cdot)$ and $\Psi_{\mathcal{D}}(\cdot|\tilde{g}^{(n)}, \gamma)$ are solutions of

$$a_2t^2 + a_1t + a_0 = \Psi_{\mathcal{D}}(\gamma t). \tag{3.46}$$

These solutions correspond to the set union

$$\mathcal{I} = \{t \leq 0 : P_1(t) = 0\} \cup \{t > 0 : P_2(t) = 0\}, \tag{3.47}$$

where $P_{1,2}(t) = 0$ gives the intersections between $\Psi_{\mathcal{D}}(\cdot|\tilde{g}^{(n)}, \gamma)$ and the linear and nonlinear parts of $\Psi_{\mathcal{D}}(\cdot)$. This corresponds to the separate formulas of (3.20) without the argument condition. Accordingly, the polynomials $P_{1,2}$ are expressed as

$$\begin{aligned} P_1(t) &= a_2t^2 + (a_1 + \gamma)t + (a_0 - M^{-1}), \\ P_2(t) &= (M^2t^2 + M\gamma t + 1)(a_2t^2 + a_1t + a_0) - M^{-1}. \end{aligned} \tag{3.48}$$

The optimal $\Psi_{\mathcal{D}}(t|\tilde{g}^{(n)}, \gamma)$ is tangent to $\Psi_{\mathcal{D}}(\gamma t)$ at $t = \tilde{g}^{(n)}, t_{\|} \in \mathcal{I}$, and intersects no other point. This causes the two double roots $\tilde{g}^{(n)}$ and $t_{\|}$ to appear in one of the two polynomials, be it jointly or not. Either of these two roots cancels the discriminant D of the associated polynomial. For the sake of conciseness, we define $t_0 = M\gamma\tilde{g}^{(n)}$ and consider two distinct cases.

Case where the point $\tilde{g}^{(n)}$ is in the nonlinear part of $\Psi_{\mathcal{D}}(\gamma t)$

In this case, where $t_0 \geq 0$, the coefficients a_0 and a_1 are expressed as

$$\begin{aligned} a_0 &= M^{-1}\left((t_0{}^2 + t_0 + 1)^{-1} - M^{-1}t_0{}^2 a_2 - \gamma t_0 a_1\right), \\ a_1 &= -\gamma\left(2M^{-1}t_0 a_2 + (2t_0+1)(t_0{}^2 + t_0 + 1)^{-2}\right). \end{aligned} \tag{3.49}$$

Then, the optimal parabola can be tangent at a distinct point of $\Psi_{\mathcal{D}}(\cdot)$ either in the same nonlinear part, or in the linear part. If $t_{\|}$ lies in the linear part, the corresponding polynomial P_1 contains one double root $t_{\|}$ for an optimal a_2. This first subcase corresponds to the solution

$$\begin{aligned} a_{2,1} &= \{a_2 \in \mathbb{R}_+^* : \mathrm{D}(P_1(\cdot)) = 0\} \\ &= \frac{1}{4}M\frac{t_0(t_0{}^2 + 2t_0 + 3)^2}{(t_0{}^2 + t_0 + 1)^3}. \end{aligned} \tag{3.50}$$

If $t_{\|}$ lies in the nonlinear part of $\Psi_{\mathcal{D}}(\cdot)$, the corresponding P_2 contains two double roots. Its discriminant is thus always zero regardless of a_2. Nevertheless, this same quantity divided by $(t-\tilde{g}^{(n)})^2$ is a viable indicator, as it only vanishes in the optimal case. This yields the solution for this second subcase as

$$\begin{aligned} a_{2,2} &= \left\{a_2 \in \mathbb{R}_+^* : \mathrm{D}((\cdot - \tilde{g}^{(n)})^{-2}P_2(\cdot)) = 0\right\} \\ &= \frac{1}{3}M\frac{(2t_0+1)^2}{(t_0{}^2 + t_0 + 1)^2}. \end{aligned} \tag{3.51}$$

Given its definition, the function $\Psi_{\mathcal{D}}(\cdot)$ corresponds to the maximum between its linear and nonlinear constituents. This determines our overall first-case solution as

$$\begin{aligned} a_2 &= \max(a_{2,1}, a_{2,2}) \\ &= \begin{cases} \frac{1}{3}M(t_0{}^2 + t_0 + 1)^{-2}(2t_0+1)^2, & 0 \leq t_0 \leq 1 \\ \frac{1}{4}M(t_0{}^2 + t_0 + 1)^{-3}t_0(t_0{}^2 + 2t_0 + 3)^2, & t_0 > 1. \end{cases} \end{aligned} \tag{3.52}$$

In this first case, the three coefficients are thus determined by combining (3.49) and (3.52) given t_0.

Case where the point $\tilde{g}^{(n)}$ is in the linear part of $\Psi_{\mathcal{D}}(\gamma t)$

In this case, where $t_0 < 0$, the coefficients a_0 and a_1 are expressed as

$$\begin{aligned} a_0 &= M^{-1}(M^{-1}{t_0}^2 a_2 + 1), \\ a_1 &= -\gamma(2M^{-1}t_0 a_2 + 1), \end{aligned} \tag{3.53}$$

the optimal parabola being always tangent at some distinct point in the nonlinear part of $\Psi_{\mathcal{D}}(\cdot)$. Since the corresponding polynomial P_2 contains one single double root in that configuration, the corresponding solution is

$$a_2 = \left\{ a_2 \in \mathbb{R}_+^* : \mathrm{D}\left(P_2(\cdot)\right) = 0 \right\}. \tag{3.54}$$

The scalar value a_2 corresponds to the positive and real root of the cubic polynomial

$$\begin{aligned} P_3(t) = \; & 12({t_0}^2 + t_0 + 1)^3 t^3 \\ & + (3{t_0}^5 + 68{t_0}^4 + 214{t_0}^3 - 24{t_0}^2 - 89t_0 + 8)Mt^2 \\ & + (14{t_0}^3 + 168{t_0}^2 - 66t_0 - 4)M^2 t \\ & + 27M^3 t_0, \end{aligned} \tag{3.55}$$

for which the analytical expression can be found [99]. The behavior of P_3 as a function of $t_0 < 0$ guarantees the uniqueness of the solution. The coefficients are obtained in this case by solving (3.55) and then using (3.53).

3.9.2 Convexity of the data term

The Hessian of $\mathcal{D}$ takes the form

$$\mathrm{H}(\tilde{\mathbf{c}}) = (\partial/\partial\cdot)\mathcal{D}(\cdot)(\partial/\partial\cdot)^{\mathrm{T}}|_{\cdot=\tilde{\mathbf{c}}}, \tag{3.56}$$

which corresponds to a real matrix. Applying the generalized derivative chain rule on (3.19), we obtain

$$\mathrm{H}(\tilde{\mathbf{c}}) = \sum_i {\mathbf{A}_i}^{\mathrm{T}} {\mathbf{\Gamma}_i}^{\mathrm{T}} \mathbf{\Psi}''_{\mathcal{D}i}(\tilde{\mathbf{c}}) \mathbf{\Gamma}_i \mathbf{A}_i, \tag{3.57}$$

where the diagonal matrices $\mathbf{\Gamma}_i$ and $\mathbf{\Psi}''_{\mathcal{D}i}(\tilde{\mathbf{c}})$ contain the diagonal terms γ_i and $\chi_i(\mathrm{d}\Psi_{\mathcal{D}}(t)/\mathrm{d}t^2)|_{t=\tilde{g}_i\gamma_i}$, respectively. Note that $\tilde{g}_i$ is an implicit function of $\tilde{\mathbf{c}}$. The twice-differentiated penalty is obtained from (3.20) as

$$\mathrm{d}\Psi_{\mathcal{D}}(t)/\mathrm{d}t^2 = 6M^2\left(M^2t^2+Mt+1\right)^{-3}(Mt+1)\max(t,0). \tag{3.58}$$

The above function is nonnegative, regardless of its argument. Therefore, every diagonal matrix $\mathbf{\Psi}''_{\mathcal{D}_i}(\tilde{\mathbf{c}})$ admits a unique Cholesky decomposition, which implies that the Hessian of the data term can itself be decomposed as

$$\mathrm{H}(\tilde{\mathbf{c}}) = \sum_i (\mathbf{\Psi}''^{1/2}_{\mathcal{D}_i}(\tilde{\mathbf{c}})\mathbf{\Gamma}_i\mathbf{A}_i)^{\mathrm{T}}(\mathbf{\Psi}''^{1/2}_{\mathcal{D}_i}(\tilde{\mathbf{c}})\mathbf{\Gamma}_i\mathbf{A}_i). \tag{3.59}$$

Since (3.59) is positive semidefinite, the data term $\mathcal{D}(\cdot)$ is convex, which is what we wanted to show.

3.9.3 Expression of the preconditioner

From (3.15) and (3.38), we first reformulate our system matrix $\mathbf{S}$ as

$$\mathbf{S} = \Lambda\Lambda_{\mathrm{E}}\mathbf{I} + \sum_{i=1}^{L}\mathbf{S}_{\mathcal{D}i} + \sum_{i=1}^{2}\mathbf{S}_{\mathcal{R}i}, \tag{3.60}$$

where

$$\begin{aligned}\mathbf{S}_{\mathcal{D}i} &= \mathbf{D}_{\mathcal{M}}{\mathbf{B}_i}^{\mathrm{T}}\overline{\mathbf{W}}_i\mathbf{B}_i\mathbf{U}_{\mathcal{M}},\\ \mathbf{S}_{\mathcal{R}i} &= {\mathbf{R}_i}^{\mathrm{T}}\overline{\mathbf{\Theta}}\mathbf{R}_i,\end{aligned} \tag{3.61}$$

and where the diagonal matrices $\overline{\mathbf{W}}_i$ and $\overline{\mathbf{\Theta}}$ are defined as

$$\begin{aligned}\overline{\mathbf{W}}_i &= \mathbf{U}_{\mathcal{N}}\mathbf{W}_i\mathbf{D}_{\mathcal{N}},\\ \overline{\mathbf{\Theta}} &= \Lambda\mathbf{\Theta}.\end{aligned} \tag{3.62}$$

Using the property of linearity, and associating each $\mathbf{P}_{\mathcal{D}i}$ and $\mathbf{P}_{\mathcal{R}i}$ to $\mathbf{S}_{\mathcal{D}i}$ and $\mathbf{S}_{\mathcal{R}i}$, respectively, we obtain

$$\mathbf{P} = \Lambda\Lambda_{\mathrm{E}}\mathbf{I} + \sum_{i=1}^{L} \mathbf{P}_{\mathcal{D}i} + \sum_{i=1}^{2} \mathbf{P}_{\mathcal{R}i}. \tag{3.63}$$

In the sequel, we also define the matrices

$$\begin{aligned} \mathbf{B}_i' &= \mathbf{F}\mathbf{B}_i\mathbf{F}^*, \\ \mathbf{R}_i' &= \mathbf{F}\mathbf{R}_i\mathbf{F}^*, \\ \overline{\mathbf{W}}_i' &= \mathbf{F}\overline{\mathbf{W}}_i\mathbf{F}^*, \\ \overline{\mathbf{\Theta}}' &= \mathbf{F}\overline{\mathbf{\Theta}}\mathbf{F}^*. \end{aligned} \tag{3.64}$$

The circulant matrices $\overline{\mathbf{W}}_i', \overline{\mathbf{\Theta}}'$ and the diagonal matrices $\mathbf{B}_i'$ are associated with the 2D filters $\overline{w}_i', \overline{\theta}'$ and with the pointwise multiplication map b_i', respectively. According to (3.42), the regularization parts are obtained as

$$\mathrm{diag}\,(\mathbf{F}\mathbf{S}_{\mathcal{R}i}\mathbf{F}^*) = \overline{\theta}'[\mathbf{0}]\mathbf{R}_i'^{\mathrm{H}}\mathbf{R}_i', \tag{3.65}$$

where $^{\mathrm{H}}$ denotes the Hermitian transpose, and where the zero-frequency value $\overline{\theta}'[\mathbf{0}]$ corresponds to the diagonal-term average of $\overline{\mathbf{\Theta}}$. This yields

$$\mathbf{P}_{\mathcal{R}i} = \overline{\theta}'[\mathbf{0}]\mathbf{R}_i{}^{\mathrm{T}}\mathbf{R}_i. \tag{3.66}$$

For the data parts, we first derive

$$\begin{aligned} \mathbf{F}\mathbf{S}_{\mathcal{D}i}\mathbf{F}^* &= \mathbf{F}\mathbf{D}_{\mathcal{M}}\mathbf{B}_i{}^{\mathrm{T}}\overline{\mathbf{W}}_i\mathbf{B}_i\mathbf{U}_{\mathcal{M}}\mathbf{F}^* \\ &= \mathbf{\Pi}^{\mathrm{T}}\mathbf{F}\mathbf{B}_i{}^{\mathrm{T}}\overline{\mathbf{W}}_i\mathbf{B}_i\mathbf{F}^*\mathbf{\Pi} \\ &= \mathbf{\Pi}^{\mathrm{T}}\mathbf{B}_i'^{\mathrm{H}}\overline{\mathbf{W}}_i'\mathbf{B}_i'\mathbf{\Pi}, \end{aligned} \tag{3.67}$$

where $\mathbf{\Pi} = \mathbf{F}\mathbf{U}_{\mathcal{M}}\mathbf{F}^*$ acts as a 2D $\mathcal{M}$-fold-replication operator, up to some scaling factor. Expressing each matrix $\mathrm{diag}\,(\mathbf{F}\mathbf{S}_{\mathcal{D}i}\mathbf{F}^*)$ through the corresponding sequence ${p_{\mathcal{D}i}}'$, we have that, from (3.67),

$${p_{\mathcal{D}i}}'[\mathbf{k}] = \mathcal{M}^{-2} \sum_{\mathbf{m}} b_i'^*[\mathbf{k} + \mathbf{m}N_0]\overline{b}_i'[\mathbf{k} + \mathbf{m}N_0]. \tag{3.68}$$

The sequences $\bar{b}'_i$ are defined as

$$\bar{b}'_i[\mathbf{k}] = \sum_{\mathbf{m}} \overline{w}'_i[\mathbf{m}N_0]\{b'_i\}_{\circ\,\mathbf{m}N_0}[\mathbf{k}], \tag{3.69}$$

where $\circ\,\mathbf{v}$ applies an integer 2D delay shift of $\mathbf{v} \in \mathbb{R}^2$ to the associated sequence. Considering the diagonal matrices $\bar{\mathbf{B}}'_i$ corresponding to the sequences $\bar{b}'_i$, and given the circulant matrices $\bar{\mathbf{B}}_i = \mathbf{F}^*\bar{\mathbf{B}}'_i\mathbf{F}$, we obtain

$$\mathbf{P}_{\mathcal{D}i} = \mathbf{D}_{\mathcal{M}}{\mathbf{B}_i}^{\mathrm{T}}\bar{\mathbf{B}}_i\mathbf{U}_{\mathcal{M}}, \tag{3.70}$$

whose structure is circulant despite the upsampling and downsampling matrix terms. Assuming that

$$\overline{w}'_i[\mathbf{k}N_0] \approx 0, \forall \mathbf{k} \neq \mathbf{0}, \tag{3.71}$$

we choose to approximate (3.70) by

$$\mathbf{P}_{\mathcal{D}i} \approx \overline{w}'_i[\mathbf{0}]\mathbf{D}_{\mathcal{M}}{\mathbf{B}_i}^{\mathrm{T}}\mathbf{B}_i\mathbf{U}_{\mathcal{M}}, \tag{3.72}$$

where the zero-frequency value $\overline{w}'_i[\mathbf{0}]$ corresponds to the diagonal-term average of $\overline{\mathbf{W}}_i$. Our preconditioner is thus fully determined from (3.63), (3.66), and (3.72).

Chapter 4

Complex field reconstruction from intensity measurements

4.1 Introduction

In this chapter, we describe a new technique for high-quality reconstruction of complex fields from single digital holographic acquisitions. Our goal is thus to estimate complex-valued object profiles from intensity measurements, including in downsampled regimes. The first part of our forward model consists in the convolution of the unknown signal f with a spatial filter h. Here, the signal f is a 2D complex field corresponding to an object profile in focus, while the filter h models the joint effects of frequency cut-off and light propagation that take place in the holographic device. The last part of our forward model consists in the addition of a reference field ρ followed by sampling and loss of the phase information. The latter effect occurs due to the fact that only light intensities are measured in our particular holographic setting. Our overall forward model is thus nonlinear as in Chapter 3, each nonlinearity being modeled by the operator $\mathcal{Q}$ according to (1.2).

In our reconstruction problem, the estimate $\tilde{f}$ of the complex object field is formulated as the mimimizer of a non-convex energy functional. The latter includes a data-fidelity term of particular form, and TV regularization terms that constrain the spatial amplitude and phase distributions of the reconstructed data. As in

the previous chapters, the algorithm that we derive tolerates downsampling, which allows to acquire substantially fewer measurements for reconstruction compared to the state of the art. We demonstrate the effectiveness of our method through several experiments on simulated and real off-axis holograms[1].

4.2 Overview

Given an intensity hologram that is acquired in a proper configuration, it is generally possible to reconstruct the complex field, commonly through demodulation, either in the temporal domain with the so-called phase-shifting approach [100], or in the spatial domain in an off-axis configuration, where complex information can be retrieved through Fourier filtering [101]. Several solutions have been proposed for reconstruction following these approaches [102, 103]. The corresponding algorithms are presently used in practice and incorporated in digital holographic microscopy for instance. Their common characteristic is that the spectral information that is actually used for reconstruction reduces to the first diffraction orders.

Besides direct Fourier-based reconstruction, iterative procedures that are based on more implicit (*i.e.*, inverse-problem) formulations have been proposed for the recovery of complex fields from intensity data. This includes the well known Gerchberg-Saxton algorithm used for phase retrieval in its original form [104] as well as more evolved strategies [105]. The principles of some of these phase-retrieval algorithms have been studied from the perspective of optimization theory, which has allowed to gain substantial insight into their behavior, shortcomings, and performance [106]. Recently, an approach called *PhaseLift* [107] has been proposed to recover complex signals from magnitude measurements based on simple convex programming. However, despite its attractive theoretical properties, this approach becomes computationally intractable when dealing with high-dimensional data.

The recent trend in the literature is also to formulate the specific task of digital holographic reconstruction as an inverse problem. For instance, Cetin *et al.* [108] have applied total-variation-type regularization to reconstruct amplitude profiles based on a linear physical model. Zhang and Lam have followed a similar approach for optical scanning holography [109]. In particular, they provided reconstruction results using model-matched phantom experiments. In their work, Bardy *et*

[1] This chapter is based on our paper [15].

al. [110] and Marim *et al.* [111] have demonstrated the possibility of reconstructing amplitude profiles from undersampled holograms in the context of Gabor and off-axis holography, respectively. Along the same line of research, Rivenson *et al.* [112] have investigated distinct undersampling strategies. Note that, although the aforementioned methods yield promising results compared to standard reconstruction techniques, they either require the acquisition of two or more holographic planes or assume a linearized model for reconstruction. Meanwhile, Sotthivirat and Fessler [113] have developed an optimization-transfer method for reconstruction that involves a nonlinear model and that also handles the case of a single acquisition. Focusing on image-plane holography, their work has shown substantial improvements on simulated data compared to the conventional approaches used for that particular setting.

In this chapter, we propose a new variational-reconstruction approach that can be applied to experimentally-acquired holographic data. Our technique accurately models the measurement system based on a nonlinear forward model and allows to recover the complex object field from one single off-axis intensity hologram. Compared to the state of the art, the main contribution of our reconstruction algorithm is to employ side information apart from the measured data to improve the quality of reconstruction. Specifically, we make prior assumptions on some shared object properties, and also use the knowledge of a reference correction hologram (RCH) to simplify the task of reconstruction and thus find suitable solutions. Another asset of our method is to exploit the complete image information at hand. The assumptions on the solution are used as a mean to regularize the amplitude and phase components of the complex field, which allows to perform reconstructions from spatially undersampled measurements. Note that an undersampling scheme was initially proposed in [21] for a particular off-axis protocol in low-level imaging conditions where the forward model can be linearized. However, unlike in our case, the protocol of [21] requires the complex amplitudes of the field to be first determined at the hologram plane through the combination of several intensity acquisitions. This is a step that can be avoided with our new nonlinear-model-based method presented in the sequel.

The chapter is organized as follows. In Section 4.3, we describe the nonlinear model that corresponds to our own holographic setup, and introduce the relevant notations. In Section 4.4, we introduce our formulation of the reconstruction problem, and discuss the relevance of this new approach for obtaining high-quality solutions in practice. We then derive our reconstruction algorithm in Section 4.5.

In Section 4.6, we conduct experiments on synthetic and real holograms, and compare our method with the state of the art from both qualitative and quantitative standpoints. We conclude our chapter in Section 4.7.

4.3 Forward model

4.3.1 General structure

The general holographic model that we consider is standard in the literature. It involves a coherent object field $o(\mathbf{x}, z)$ and a reference field $\rho(\mathbf{x}, z)$ that interfere to form the light intensities

$$|o(\mathbf{x}, 0) + \rho(\mathbf{x}, 0)|^2 \tag{4.1}$$

at each position of the sensor plane. The coordinate vector $\mathbf{x} = (x_1, x_2)$ corresponds to the spatial position parallel to the hologram plane, while the coordinate z denotes the signed distance from the hologram plane in the direction of light propagation. Note that the squared norm appearing in (4.1) makes the acquired light intensities depend nontrivially on the field o. This will cause our own reconstruction problem to be non-convex, as further discussed in Section 4.4.

Following the MKS system of units, the physical parameters of the optical acquisition system consist in the wavelength λ_0, the camera-sensor spacing Δ_s, the detector resolution M_0, the numerical aperture NA of the microscope objective, the object-space immersion medium n, and the magnification factor O. Assuming a diffraction-limited configuration, the spatial bandwidth of the object corresponds to $\Delta_\omega = 2\mathrm{NA}\Delta_\mathrm{s}M_0/(On\lambda_0)$ [114]. In the case of an off-axis hologram, a demodulated and non-aberrated object field can potentially be retrieved by using a sampled estimate $\tilde{\rho}[\mathbf{k}]$ of the reference wave $\rho(\mathbf{x}, 0)$ employed for recording, up to a multiplicative factor w_ρ. One common way to get $\tilde{\rho}$ is to use a calibration measurement with an empty field of view which accounts for the phase terms induced by the optical system, denoted RCH [115]. The amplitudes of the estimate $\tilde{\rho}$ are spatially averaged in our reconstruction approach.

We assume in this chapter that the holographic device measures an infinitely thin object profile. The field o is thus expected to be focused at some negative-valued distance $z = z_\mathrm{f}$, the corresponding planar profile being denoted as f. While

the in-focus distance z_{f} is classically independent from the actual object-field reconstruction, it is relevant to our method because we shall introduce some prior knowledge on f in Section 4.4. Accordingly, we consider f as our unknown, and reformulate (4.1) as

$$\mathcal{Q}((\mathcal{W}\mathcal{H}f)(\mathbf{x}), \rho(\mathbf{x}, 0)). \tag{4.2}$$

The linear operators $\mathcal{W}$ and $\mathcal{H}$ denote forward light propagation with distance $-z_{\mathrm{f}}$ and optical filtering that is due to the spatial-frequency cut-off of the microscope objective (MO), respectively. These operators behave as a function of the parameters of the holographic setup, and their mathematical expressions are provided in Appendix 4.8.1. Given the form of these expressions, the action of the compound operator $\mathcal{W}\mathcal{H}$ on the signal f corresponds to a single convolution operation with a continuous-domain spatial filter h. This filter is defined as $h = \mathcal{W}\mathcal{H}\delta(\cdot)$, where $\delta(\cdot)$ denotes the Dirac distribution. Then, the operator $\mathcal{Q}$ is defined $\forall t \in \mathbb{C}$ and $\forall \tau \in \mathbb{C}$ as

$$\mathcal{Q}(t, \tau) = |t + \tau|^2. \tag{4.3}$$

As represented geometrically in Figure 4.1, the effect of $\mathcal{Q}$ can be interpreted as a quantization operation acting over the complex plane. Based on a reference $\tau \in \mathbb{C}$, the operator $\mathcal{Q}(t, \tau)$ acts on the input value $t \in \mathbb{C}$ as a quantizer. Specifically, the complex plane defined by t is tiled into circles centered around $-\tau$; any point t_i belonging to the same circle is quantized to the same corresponding squared radius length $l_i^2 = \mathcal{Q}(t_i, \tau)$. As illustrated by the dotted arrows in the same figure, each circle can span complex values with distinct arguments. Note that, according to our quantization interpretation, the second parameter $\tau \in \mathbb{C}$ in (4.3) can be considered as a generalized and spatially-varying version of the real-valued threshold that is used in Chapter 3.

Finally, according to the pixel size Δ_{s} of the sensor array, the intensities in (4.2) are sampled to

$$\gamma[\mathbf{k}] = \mathcal{Q}(g[\mathbf{k}], \rho(\mathbf{x}, 0)|_{\mathbf{x}=\mathbf{k}\Delta_{\mathrm{s}}}), \tag{4.4}$$

where the sequence g is defined as

$$g[\mathbf{k}] = (f * h)(\mathbf{x})|_{\mathbf{x}=\mathbf{k}\Delta_{\mathrm{s}}}. \tag{4.5}$$

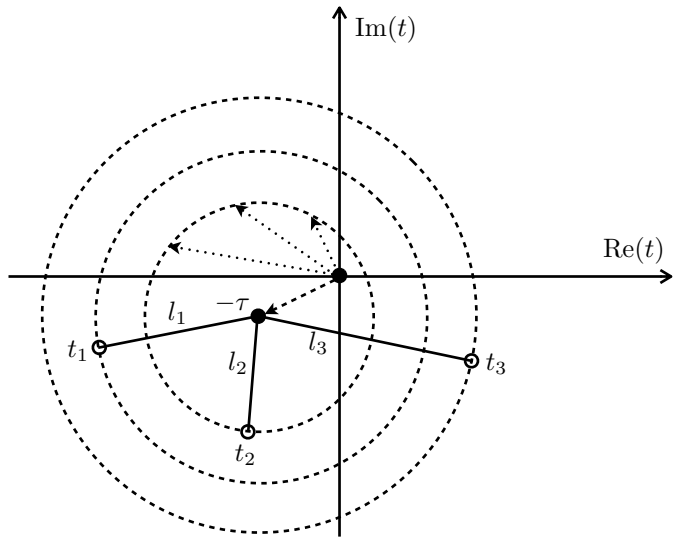


Figure 4.1: Geometrical interpretation of the pointwise nonlinearities involved in our holographic forward model.

The binary mask χ then predefines what samples are actually measured according to a measurement ratio K. Each sample $\gamma[\mathbf{k}]$ is thus counted as a measurement if and only if the value $\chi[\mathbf{k}]$ is unity for the same $\mathbf{k}$. The mask follows the pseudorandom structure of [21], and consists in binary values that are independent, identically distributed, and nonzero with probability K. Our overall holographic model is represented as a block diagram in Figure 4.2. The structure of this model is close to the one of our compressed-imaging scheme [11] that is described in Chapter 3. As shown in our experimental section, the generalized formulation that is proposed successfully handles cases where the amount of data available for reconstruction is substantially reduced. Note that downsampled acquisitions potentially allow for faster imaging when using cameras where pixel values are addressed in random access [74].

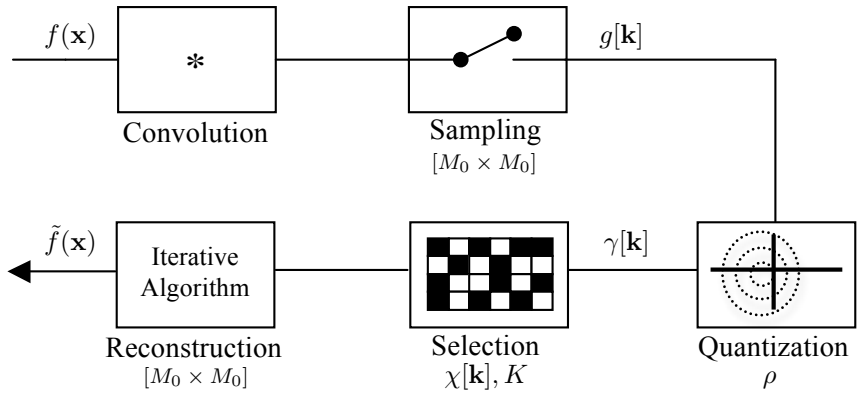


Figure 4.2: Proposed holographic model. The unknown complex field f is convolved with the spatial filter h and sampled with step Δ_{s} to obtain the sequence g. The latter is then modified pointwise according to (4.4). This yields the intensity measurements γ. Based on the measurements that are retained by the mask χ, our reconstruction algorithm finally produces an estimate $\tilde{f}$. The latter corresponds to a coefficient sequence $\tilde{c}$ whose resolution matches the one of the sensor array.

4.4 Reconstruction problem

According to (4.4), the goal of holographic reconstruction is to produce the most precise estimate of the object wave f at reconstruction distance z_{f}, given noisy and quantized versions of the available intensity samples γ. In this section, we propose an estimate $\tilde{f}$ of f that involves all the available measured data.

4.4.1 Discretization

In the context of our reconstruction problem, we assume that the estimated complex field $\tilde{f}$ is bandlimited in the frequency domain in accordance with the size Δ_{s} of the sensors. The signal $\tilde{f}$ thus admits a shift-invariant expansion of the form (1.3), where the coefficient-grid spacing Δ_{c} is equal to Δ_{s}. In mathematical terms,

$$\tilde{f}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \tilde{c}[\mathbf{k}]\varphi\left(\frac{\mathbf{x}}{\Delta_\mathrm{s}} - \mathbf{k}\right), \tag{4.6}$$

where $\tilde{c}$ is the equivalent coefficient sequence, and where φ is a separable generating kernel composed of two normalized sinc functions [116]. The mathematical definition of this kernel is $\varphi(\mathbf{x}) = \prod_{i=1}^{2} \sin(\pi x_i)/(\pi x_i)$. Note that the properties of the sinc function imply that the coefficients $\tilde{c}$ are ideal samples of $\tilde{f}$, *i.e.*, that

$$\tilde{c}[\mathbf{k}] = \tilde{f}(\mathbf{x})|_{\mathbf{x}=\mathbf{k}\Delta_\mathrm{s}}. \tag{4.7}$$

The above bandlimitedness assumption implies that the measurements $\tilde{g}$ related to $\tilde{f}$ according to our forward model are expressed as

$$\tilde{g}[\mathbf{k}] = (\tilde{c} * b)[\mathbf{k}], \tag{4.8}$$

where b is a sequence corresponding to the filter h in spatially discretized form. The type of discretization that is employed for this filter is the same as in (4.7).

Note that, in matrix notation, the relationship between the coefficients $\tilde{c}$ and the samples of $\tilde{g}$ that are kept according to χ can be summarized by (1.5), the corresponding measurement matrix $\mathbf{A}$ being defined as

$$\mathbf{A} = \boldsymbol{\chi}\mathbf{B}. \tag{4.9}$$

The convolution matrix $\mathbf{B}$ is associated with the kernel b, and the downsampling matrix $\boldsymbol{\chi}$ is linked to the sequence χ. As in the previous chapters, the latter matrix corresponds to an identity matrix whose rows associated with the discarded measurements are suppressed [11].

4.4.2 Consistent formulation

Following a variational framework, we define the coefficients $\tilde{c}$ of the estimated complex field $\tilde{f}$ as the solution of an inverse problem. This solution must be matched to the available measurements, and has to be regularized so as to make the problem well posed.

According to (4.4), we define the class of suitable solutions as those yielding measurements that are compatible with the image data after reintroduction into our forward model. Specifically, given the samples γ acquired by the sensor array and

the pre-estimated reference-wave sequence $\tilde{\rho}$, one can express a so-called *consistency constraint* on the solution $\tilde{c}$ as

$$\mathcal{Q}(\tilde{g}[\mathbf{k}], w_\rho \tilde{\rho}[\mathbf{k}]) = \gamma[\mathbf{k}], \ \forall \mathbf{k} \text{ s.t. } \chi[\mathbf{k}] = 1, \tag{4.10}$$

the sequence $\tilde{g}$ being implicitly a function of the coefficients $\tilde{c}$ according to (4.8). Note that the actual reference amplitudes are determined by the factor w_ρ that is obtained during the reconstruction process. Due to the presence of noise, we relax the strict data-fidelity constraint of (4.10) as

$$\mathcal{D}(\tilde{c}, w_\rho) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi[\mathbf{k}] \left(\mathcal{Q}(\tilde{g}[\mathbf{k}], w_\rho \tilde{\rho}[\mathbf{k}]) - \gamma[\mathbf{k}]\right)^2 \leq \mathcal{K}_\mathcal{D}. \tag{4.11}$$

Besides $\tilde{c}$, the multiplicative weight w_ρ influences the value of $\mathcal{D}$ in (4.11); both quantities shall be alternately optimized in our algorithm, as described in Section 4.5. Finally, the positive scalar $\mathcal{K}_\mathcal{D}$ determines to what extent the reconstructed solution can depart from the available measurements, depending on the level of noise that affect them. Based on the sampled intensities, this constant can be deduced from the SNR of the acquisition device $\mathcal{K}_{\text{SNR}}$ through the relation

$$\mathcal{K}_\mathcal{D} = \exp\left(-\mathcal{K}_{\text{SNR}} \ln(10)/10\right) \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi[\mathbf{k}]\gamma[\mathbf{k}]^2, \tag{4.12}$$

as shown in Appendix 4.8.2.

Under the sole data-fidelity constraint of (4.11), the solution is underdetermined. In order to make the problem well posed, we thus regularize it by minimizing the total variation [32] of its phase and amplitude components. From a qualitative standpoint, our regularization approach is derived from the assumption that the unwrapped-phase and amplitude maps of the focused hologram are both well approximated by piecewise-constant functions as they are in-focus. This assumption shall prove fruitful for improving the reconstruction quality, as shown in the experimental section. Accordingly, we define our regularization term $\mathcal{R}$ as

$$\mathcal{R}(\tilde{c}) = \int_{\mathbb{R}^2} \|\nabla(\arg_{\text{u}} \tilde{f})(\mathbf{x})\| \mathrm{d}\mathbf{x} + \Lambda_{\text{A}} \int_{\mathbb{R}^2} \|\nabla(|\tilde{f}|)(\mathbf{x})\| \mathrm{d}\mathbf{x}, \tag{4.13}$$

where $\tilde{f}$ is implicitly determined from $\tilde{c}$ according to the expansion (4.6). The functions $\arg_{\text{u}} \tilde{f}$ and $|\tilde{f}|$ correspond to the unwrapped-phase and amplitude maps

of the solution, and the positive scalar weight Λ_{A} balances the influence of both total-variation integrals. In order to be computationally tractable, we replace the integrals in (4.13) by approximate sums with step size Δ_{s}. The corresponding approximate regularization term $\mathcal{R}^0$ reads

$$\mathcal{R}^0(\tilde{c}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Psi_{\mathcal{R}}\left(\|(\arg_{\mathrm{u}}\tilde{c} * \mathbf{r})[\mathbf{k}]\|\right) + \Lambda_{\mathrm{A}} \sum_{\mathbf{k}\in\mathbb{Z}^2} \Psi_{\mathcal{R}}\left(\|(|\tilde{c}| * \mathbf{r})[\mathbf{k}]\|\right), \tag{4.14}$$

where the potential function $\Psi_{\mathcal{R}}$ is defined as

$$\Psi_{\mathcal{R}}(t) = (t^2 + \epsilon)^{\frac{1}{2}}, \tag{4.15}$$

and where ϵ is an additive parameter that ensures the differentiability of both TV functionals [41]. The horizontal and vertical components of the gradient filter $\mathbf{r}$ are specified as the convolution masks $[0\ -1\ 1]$ and $[0\ -1\ 1]^{\mathrm{T}}$. Note that, according to (4.7), the corresponding filtering operations are each associated with only two distinct phase or amplitude values of $\tilde{f}$.

Following the above definitions, the solution of our reconstruction problem is expressed as

$$\arg\min_{\tilde{c}} \mathcal{R}^0(\tilde{c}) \ \text{ s.t. } \ \tilde{\mathcal{D}}(\tilde{c}) \leq \mathcal{K}_{\mathcal{D}}, \tag{4.16}$$

where

$$\tilde{\mathcal{D}}(\tilde{c}) = \min_{w_\rho} \mathcal{D}(\tilde{c}, w_\rho). \tag{4.17}$$

Given the definition of $\mathcal{Q}$ in (4.3), the expression of $\mathcal{D}(\tilde{c}, w_\rho)$ in (4.11) contains two imbricated squaring operations. This implies that the data term $\tilde{\mathcal{D}}(\tilde{c})$ is a non-convex functional of the field coefficients $\tilde{c}$. The problem of (4.16) can thus have several local optima; its solution is non-unique in general. Note that our reconstruction algorithm proposed in Section 4.5 does not require any phase unwrapping to regularize the solution.

4.5 Reconstruction algorithm

We solve the problem of (4.16) based on a gradient-descent approach that we combine with a specifically devised line-search strategy. Specifically, we propose to

1) Initialize the solution as $\tilde{\mathbf{c}} = \mathbf{0}$
2) Find the optimal factor $\tilde{w}_\rho = \arg\min_{w_\rho} \mathcal{D}(\tilde{\mathbf{c}}, w_\rho)$ to determine $\tilde{\mathcal{D}}(\tilde{\mathbf{c}})$
if $\tilde{\mathcal{D}}(\tilde{\mathbf{c}}) > \mathcal{K}_\mathcal{D}$ **then**
  a) Compute the data-term gradient $\nabla\tilde{\mathcal{D}}(\tilde{\mathbf{c}})$
  b) Compute the step $\tilde{\omega} \in \mathbb{R}_+$ s.t. $\tilde{\mathcal{D}}(\tilde{\mathbf{c}} - \tilde{\omega}\nabla\tilde{\mathcal{D}}(\tilde{\mathbf{c}})) = \mathcal{K}_\mathcal{D}$ if it exists; optimize $\tilde{\omega} = \arg\min_\omega \tilde{\mathcal{D}}(\tilde{\mathbf{c}} - \omega\nabla\tilde{\mathcal{D}}(\tilde{\mathbf{c}}))$ otherwise
  c) Update the solution as $\tilde{\mathbf{c}} \leftarrow \tilde{\mathbf{c}} - \tilde{\omega}\nabla\tilde{\mathcal{D}}(\tilde{\mathbf{c}})$
end
if $\tilde{\mathcal{D}}(\tilde{\mathbf{c}}) \leq \mathcal{K}_\mathcal{D}$ **then**
  a) Compute the regularization gradient $\nabla\mathcal{R}^0(\tilde{\mathbf{c}})$
  b) Update the solution as $\tilde{\mathbf{c}} \leftarrow \tilde{\mathbf{c}} - \omega_{\text{reg}}\nabla\mathcal{R}^0(\tilde{\mathbf{c}})$
end
3) Return to Step 2 until N_i iterations are reached

Algorithm 4.1: Proposed holographic-reconstruction algorithm.

apply the iterative procedure detailed in matrix notation in Algorithm 4.1.

The positive scalar ω_{reg} is a step parameter that determines the strength of the regularization flow. The corresponding regularization gradient $\nabla\mathcal{R}^0$ and the data-term gradient $\nabla\tilde{\mathcal{D}}$ are complex-valued with $\nabla\cdot(\tilde{\mathbf{c}}) = \partial\cdot/\partial\,\text{Re}(\tilde{\mathbf{c}}) + \text{j}(\partial\cdot/\partial\,\text{Im}(\tilde{\mathbf{c}}))$. The explicit forms of the update terms are then obtainable from (4.14) using differential calculus. Our regularization technique does not require any explicit phase unwrapping, as explained in Appendix 4.8.3. Finally, determining either of the two scalars $\tilde{w}_\rho$ and $\tilde{\omega}$ in Step 2 amounts to solving an elementary line-search problem. This issue is addressed in more detail in Appendix 4.8.4. Note that the algorithm that we propose is closely related to incremental gradient methods as defined in [117]. From that perspective, the way our algorithm enforces the data-fidelity constraint in (4.16) may be interpreted as a projection operation.

Since the problem of (4.16) is non-convex, the solution obtained with our algorithm potentially depends on the initial estimate. In practice, however, no significant changes were observed, hence our generic choice of zero initialization in Step 1. While rigorous convergence analysis is nontrivial in non-convex settings [118], experimental evidence suggests that our algorithm converges to a fixed point. Indeed, the solution and the corresponding value of the regularization functional in (4.16) have been observed to stabilize as the iterations proceed. Each iteration of our

algorithm is comparable to the standard technique in terms of computational cost. In the framework of our experiments, the computational time associated with one given holographic-reconstruction task on a Mac OS X machine with a Quad-Core Xeon 2 $\times$ 2.8 GHz is of the order of several tens of minutes.

4.6 Experiments

Based on the above developments, we conduct experiments in simulated and practical off-axis configurations. The holographic reconstructions are performed with our method as well as with the conventional approach consisting in demodulation followed by back-propagation. Besides the standard reconstruction paradigm, we investigate cases where the amount of available measurements is reduced through random subsampling. In the sequel, a first set of experiments on simulated holograms evaluates the algorithms quantitatively in each case, using the oracle complex fields for comparison. The experiments that follow allow to determine the relevance and the competitiveness of our approach when real data are involved.

The standard and proposed algorithms are implemented in MATLAB, using the angular-spectrum approximation to model light propagation [75]. The reconstructed objects are shown with the same resolution and field of view as those of the measured holograms. In practical configurations, an additive linear-phase field can appear in the reconstruction when tilt coming from sample positioning occurs between acquisition of the RCH and the one of the object. This additive field is removed to simplify the visual interpretation of the results. Regarding the boundary conditions that are used, our algorithm is set to regularize the reconstructions on a slightly larger field of view than the one displayed. As shown in the sequel, this allows to improve the quality of reconstruction at the boundaries compared to the standard technique that uses apodization.

4.6.1 Synthetic data

In this set of experiments, the original complex fields are available, their amplitudes and phases at the focus plane being defined from the 2D spatial maps of Table 4.1. The *Pentagon* picture originates from the BM3D database at `http://www.cs.tut.fi/~foi/GCF-BM3D/` and the others from the USC-SIPI database at `http://sipi.usc.edu/database/`. After normalization of its spatial-domain amplitude

			Amplitude & Phase SNR (ANR & PNR)							
Original Object			$K = 1$				$K = 1/2$			
#	Amplit.	Phases	Standard		Variational		Standard		Variational	
			ANR	PNR	ANR	PNR	ANR	PNR	ANR	PNR
1	*Pentagon*	*Man*	21.46	21.25	**24.20**	**22.38**	6.60	1.55	**23.47**	**21.84**
2	*Pentagon*	-	24.08	-	**24.53**	-	6.64	-	**23.91**	-
3	*Airplane*	*Airplane*	**18.29**	17.71	18.07	**17.75**	4.61	−3.77	**17.77**	**17.37**
4	*Airport*	*Airport*	19.14	18.66	**20.23**	**20.17**	5.87	0.46	**19.59**	**19.75**
5	-	*Man*	-	20.74	-	**22.98**	-	2.51	-	**22.54**

Table 4.1: Reconstruction quality in synthetic experiments.

and phase values to $[0, 1]$ and $[0, \pi]$, respectively, each complex field is used to generate a distinct 512×512 hologram according to our forward model. These holograms were first obtained from larger objects and then restricted to a limited field of view—centered and with 512×512 samples—in analogy with a physical setup. The acquisition parameters that are used for hologram generation correspond to an off-axis configuration, choosing the wavelength as corresponding to the He-Ne laser line, *i.e.*, $\lambda_0 = 633 \cdot 10^{-9}$, and choosing $\Delta_\mathrm{s} = 6.45 \cdot 10^{-6}$, $M_0 = 1{,}024$, $\mathrm{NA} = 0.25$, $n = 1$, $O = 10$, $z_\mathrm{f} = -4 \cdot 10^{-2}$, and $\tilde{\rho}[\mathbf{k}] = \exp(2\pi\,\mathrm{j}(k_1 + k_2)/5)$. The oracle maps mentioned in Table 4.1 are shown along with the corresponding holograms in Figure 4.3. Note that the scales at which our object maps, holograms, and reconstructions are displayed are normalized in each figure for convenience.

Given the synthetic holograms and their corresponding parameters, our goal is to determine how accurately the oracle complex fields can be reconstructed by the standard and proposed methods. For each hologram, we consider the classical setting where all samples are available for reconstruction, as well as the setting $K = 1/2$ where only 50% of the data are kept through random downsampling. As reconstruction parameters, our algorithm uses $\epsilon = 10^{-10}$, $\omega_\mathrm{reg} = 1.5 \cdot 10^{-2}$, $\mathcal{K}_\mathrm{SNR} = 35$, and $N_\mathrm{i} = 2{,}000$ in all synthetic experiments. The regularization weight is set to $\Lambda_\mathrm{A} = 0.5$. In order to neglect the possible influence of boundary artifacts, the SNR values are evaluated on fields of view that are centered and whose size is reduced by 200 pixels in each dimension compared to those of the corresponding holograms.

The numerical results that are obtained are reported in Tables 4.1 and 4.2.

			FOV Ratio		
Original Object			**Standard**	**Variational**	
#	**Amplit.**	**Phases**	$K = 1$	$K = 1$	$K = 1/2$
1	*Pentagon*	*Man*	0.84	**0.88**	**0.87**
2	*Pentagon*	-	0.84	**0.86**	**0.86**
3	*Airplane*	*Airplane*	0.84	**0.95**	**0.95**
4	*Airport*	*Airport*	0.84	**0.86**	**0.86**
5	-	*Man*	0.84	**0.90**	**0.89**

Table 4.2: Effective field of view in synthetic experiments.

As an example, the reconstruction is shown for the phase-only hologram #5 in Figure 4.4; the phase values $[0, \pi]$ are mapped to the grayscale range [black, white]. In the classical paradigm where all samples are available, the values of Table 4.1 indicate an overall improvement in terms of reconstruction quality when using our approach. Our visual results also demonstrate that less artifacts are produced by the proposed method at the object boundaries. Furthermore, besides backward compatibility with the standard approach, our reconstruction algorithm is able to successfully recover object profiles when the amount of sampled data is reduced. In particular, the corresponding SNR values of Table 4.1 remain stable as compared to the fully sampled configurations. By contrast, the SNR values that correspond to the standard reconstruction approach decrease dramatically when downsampling is used. The strong aliasing effects that are due to downsampling can indeed not be handled properly in that case since no regularization is used.

In Table 4.2, the proposed *FOV ratio* measures the relative area of the solution where the reconstruction quality remains highest with respect to the oracle, considering the downsampled case only for our method. This area is determined through local MSE computations performed at every distance from the boundaries of the given object. It is obtained as the centered region where this local MSE does not exceed twice the global MSE. The FOV results further document the improvement of our method over the standard one. Our algorithm yields data-dependent values because its regularization strategy is itself data-adaptive.

At this stage, these observations confirm the relevance of our approach for holographic reconstruction in both classical and downsampled configurations. They remain to be corroborated in the real-data experiments that follow.

Hologram		Acquisition					Algorithm
Name	**Type**	λ_0	M_0	NA	O	z_f	SNR
Neuron	*Phase-only*	$680 \cdot 10^{-9}$	1024	0.45	20	$-3 \cdot 10^{-2}$	25.0 dB
Epithelial	*Phase-only*	$680 \cdot 10^{-9}$	512	0.25	10	$-4 \cdot 10^{-2}$	32.0 dB
USAF 5-4	*Mixed*	$661 \cdot 10^{-9}$	512	0.4	20	$-5 \cdot 10^{-2}$	30.0 dB
USAF 9-8	*Mixed*	$661 \cdot 10^{-9}$	512	0.4	20	$-6 \cdot 10^{-2}$	18.5 dB

Table 4.3: Object-dependent parameters used for optical acquisition.

4.6.2 Real data

In this second experimental part, we consider holograms that are acquired practically from distinct physical objects. The 1,024 × 1,024 hologram *Neuron* and the 512 × 512 hologram *Epithelial* consist in acquisitions of neural and epithelial cell samples, respectively. The 512 × 512 holograms *USAF 5-4* and *USAF 9-8* correspond to USAF targets with mixed phase and amplitude information. As in the synthetic case of Section 4.6.1, the acquisition settings that have been used in the optical setup are off-axis. The parameters that are common to all acquisitions are $\Delta_\mathrm{s} = 6.45 \cdot 10^{-6}$ and $n = 1$. The other scalar parameters are reported in Table 4.3, while the approach followed to generate the reference waves is described in [114].

The proposed algorithm is parameterized as above, except $\omega_\mathrm{reg} = 3 \cdot 10^{-4}$, $\Lambda_\mathrm{A} = 1$ for the hologram *USAF 9-8*, and except $\mathcal{K}_\mathrm{SNR}$. In this practical setting, the latter quantity depends nontrivially on the acquisition conditions and is determined experimentally. The SNR values specific to each hologram are thus reported in Table 4.3 along with the acquisition parameters. Moreover, since the amplitude of the reference wave tends to vary slowly in space due to the non-ideality of the holographic device, we propose to determine w_ρ in (4.11) in a spatially adaptive manner. Specifically, we perform one distinct optimization for each 8×8-pixel block of the solution. Note that each optimized value of w_ρ is still estimated based on the principles of Appendix 4.8.4, although the sums involved in the computations are now restricted to the corresponding block.

The reconstructions from *Neuron*, *Epithelial*, and *USAF 5-4* are shown in Figures 4.5, 4.6, 4.7, and 4.8. While the standard technique only handles the fully sampled scenario, distinct downsampling factors $K = 1, 1/2, 1/4$ are considered with our algorithm. When all samples are available, the visual results show simi-

lar advantages of our algorithm over the classical method as in the synthetic case. For instance, the field of view of the reconstructed objects is extended when using our iterative approach, unlike in the standard approach where the signal must be attenuated at borders in order to properly handle the boundary conditions. These results thus demonstrate that the boundary issues arising in digital propagation can be suitably addressed when using an implicit formulation of the reconstruction problem such as (4.16). In this set of real experiments, our method is also observed to reduce the amount of noise in the recovered phase and amplitude profiles. In particular, the continuity of their background is improved. When decreasing K, the reconstruction quality remains acceptable even though some fine details are inevitably lost due to the associated decrease in resolution. This decrease is associated to small patterns of "pointillism" type appearing in the holograms.

The contrast and the sharpness of object features are also well preserved when using our algorithm. In particular, the results on *USAF 5-4* shown in Figures 4.7 and 4.8 demonstrate that sharp transitions and large regular structures are reconstructed accurately. The resolution performance of the standard and proposed methods can be compared in Figure 4.9 where the reconstruction from *USAF 9-8* is shown in the fully sampled setting. In this example, the resolution capability of both methods is similar. While the use of TV regularization produces a slight smearing effect in some places, the level of noise is strongly reduced in the reconstruction obtained with our method. This set of practical experiments thus corroborates the observations made in the synthetic case, and validates our approach as applied to real off-axis holograms subject to noise.

4.7 Conclusions

We have devised an algorithm for off-axis holographic reconstruction that is based on a consistent problem formulation. Based on suitable regularity assumptions, our technique has allowed to reconstruct complex-valued object profiles satisfactorily, including in the case where the amount of measured samples is decreased. Compared to the standard technique, the proposed method avoids the presence of boundary artifacts in the solutions, and reduces the perceived level of noise in practical configurations. From a general perspective, the obtained results further illustrate the interest of inverse-problem approaches in holography. In future research, distinct regularization strategies could be investigated for reconstruction.

This includes alternate variational terms that would take correlations between phase and amplitude components into account, and so-called *total cyclic variation* [119].

4.8 Appendix

4.8.1 Forward-model operators

The effect of the operator $\mathcal{H}$ can be described as a convolution with a specific filter. The latter is defined in the Fourier domain as the amplitude-only function

$$\begin{cases} 1, & \|\boldsymbol{\omega}\| \leq \Delta_\omega/2 \\ 0, & \text{otherwise.} \end{cases} \tag{4.18}$$

We remind that this frequency limitation fundamentally originates from the numerical aperture of the objective.

The propagation operator $\mathcal{W}$ is also associated with a continuous-domain filter. Given the distance z_f, it is defined in the Fourier domain as the phase-only function

$$\exp\left(-2\pi\,\mathrm{j}\,z_\mathrm{f}\left(1-{\lambda_0}^2\|\boldsymbol{\omega}\|^2\right)^{1/2}/\lambda_0\right). \tag{4.19}$$

This propagation model is based on the angular-spectrum method [75].

4.8.2 Data-fidelity constant

As mentioned in Section 4.4, we constrain suitable estimates of the unknown complex field to be consistent with the measured intensity samples. According to our forward model, a given solution $\tilde{c}$ should thus yield the measurements

$$\tilde{\gamma}[\mathbf{k}] = \mathcal{Q}(\tilde{g}[\mathbf{k}], w_\rho\tilde{\rho}[\mathbf{k}]) \tag{4.20}$$

that are close to the known samples $\gamma[\mathbf{k}]$, $\forall\mathbf{k}$ s.t. $\chi[\mathbf{k}]=1$. Assuming that the noise level affecting the samples of γ corresponds to a SNR of $\mathcal{K}_\mathrm{SNR}$ in dB, we constrain the similarity between $\tilde{\gamma}$ and γ to match this value as a lower limit. Taking only the non-masked samples into account, we obtain

$$\mathcal{K}_\mathrm{SNR} \leq 10\log_{10}\left(\frac{\sum_{\mathbf{k}\in\mathbb{Z}^2}\chi[\mathbf{k}]\gamma[\mathbf{k}]^2}{\sum_{\mathbf{k}\in\mathbb{Z}^2}\chi[\mathbf{k}](\tilde{\gamma}[\mathbf{k}]-\gamma[\mathbf{k}])^2}\right), \tag{4.21}$$

which expands as

$$\sum_{\mathbf{k}\in\mathbb{Z}^2} \chi[\mathbf{k}](\tilde{\gamma}[\mathbf{k}] - \gamma[\mathbf{k}])^2 \leq \exp\left(-\ln(10)\mathcal{K}_{\mathrm{SNR}}/10\right) \sum_{\mathbf{k}\in\mathbb{Z}^2} \chi[\mathbf{k}]\gamma[\mathbf{k}]^2. \tag{4.22}$$

This expression is equivalent to (4.11). The left-hand side of (4.22) corresponds to the data term $\mathcal{D}(\tilde{c}, w_\rho)$, while its right-hand side is the definition of the data-fidelity constant $\mathcal{K}_\mathcal{D}$ in (4.12). Note that this derived result is specific to the quadratic form of (4.11) that is matched to a Gaussian noise model. Another noise model (*e.g.*, Poisson) would correspond to a distinct formula for $\mathcal{K}_\mathcal{D}$.

4.8.3 Spatial derivatives of unwrapped phase

The phase derivative can be estimated independently of wrapping [120]. More specifically, since our finite-difference filters only take two distinct phase values into account at each time, only phase differences are relevant. Assuming that the phase jumps between two adjacent coefficient samples do not exceed π, and defining $v_i(\tilde{c}) = ((\arg \tilde{c} * r_i)[\mathbf{k}] \bmod 2\pi)$ at a given position $\mathbf{k}$, we thus have the relation

$$(\arg_{\mathrm{u}} \tilde{c} * r_i)[\mathbf{k}] = \begin{cases} v_i(\tilde{c}), & v_i(\tilde{c}) < \pi \\ v_i(\tilde{c}) - 2\pi, & \text{otherwise.} \end{cases} \tag{4.23}$$

The above assumption on maximal phase jumps corresponds to the required properties for proper phase sampling [120]. In settings such as [121] where this assumption is violated, our method may still be applied, provided that only the solution amplitudes are regularized. While preliminary investigations indicate satisfactory results on objects whose phases are random and uniformly distributed in $[-\pi, \pi[$, this issue remains to be addressed in further research.

4.8.4 Line search

Each of the optimization problems that arise in Step 2 of Algorithm 4.1 when determining $\tilde{w}_\rho$ and $\tilde{\omega}$ can be recast as the resolution of a simple polynomial equation. According to the form of the data term that is specified in (4.11), these line-search problems both comply with the generic formulation

$$P_0(t) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \left(|s_1[\mathbf{k}] + ts_2[\mathbf{k}]|^2 - s_3[\mathbf{k}]\right)^2, \tag{4.24}$$

where s_i are known constant sequences (s_3 being real-valued and nonnegative), and where t is the scalar variable of interest to optimize. Expanding the terms of (4.24), and factoring out t, we obtain the simplified expression

$$P_0(t) = a_4t^4 + a_3t^3 + a_2t^2 + a_1t + a_0, \tag{4.25}$$

where

$$\begin{aligned}
a_4 &= \sum_{\mathbf{k}\in\mathbb{Z}^2} |s_2[\mathbf{k}]|^4, \\
a_3 &= \sum_{\mathbf{k}\in\mathbb{Z}^2} 4\,\mathrm{Re}(s_1[\mathbf{k}]s_2^*[\mathbf{k}])|s_2[\mathbf{k}]|^2, \\
a_2 &= \sum_{\mathbf{k}\in\mathbb{Z}^2} 2|s_2[\mathbf{k}]|^2(|s_1[\mathbf{k}]|^2 - s_3[\mathbf{k}]) + 4\left(\mathrm{Re}(s_1[\mathbf{k}]s_2^*[\mathbf{k}])\right)^2, \\
a_1 &= \sum_{\mathbf{k}\in\mathbb{Z}^2} 4\,\mathrm{Re}(s_1[\mathbf{k}]s_2^*[\mathbf{k}])(|s_1[\mathbf{k}]|^2 - s_3[\mathbf{k}]), \\
a_0 &= \sum_{\mathbf{k}\in\mathbb{Z}^2} (|s_1[\mathbf{k}]|^2 - s_3[\mathbf{k}])^2.
\end{aligned} \tag{4.26}$$

Given their definition, the coefficients a_i are real scalars that only depend on the constant sequences s_i. Depending on the optimization problem, $P_0(t)$ has to be either equalized to $\mathcal{K}_\mathcal{D}$ or minimized. In the former case, the solution—if it exists—corresponds to the smallest real and nonnegative root of the fourth-degree polynomial

$$P_1(t) = a_4t^4 + a_3t^3 + a_2t^2 + a_1t + (a_0 - \mathcal{K}_\mathcal{D}). \tag{4.27}$$

In the latter case, the argument minimizing P_0 must cancel its first derivative. Accordingly, the solution is the real and nonnegative root of the polynomial

$$P_2(t) = 4a_4t^3 + 3a_3t^2 + 2a_2t + a_1 \tag{4.28}$$

that yields the smallest value of P_0.

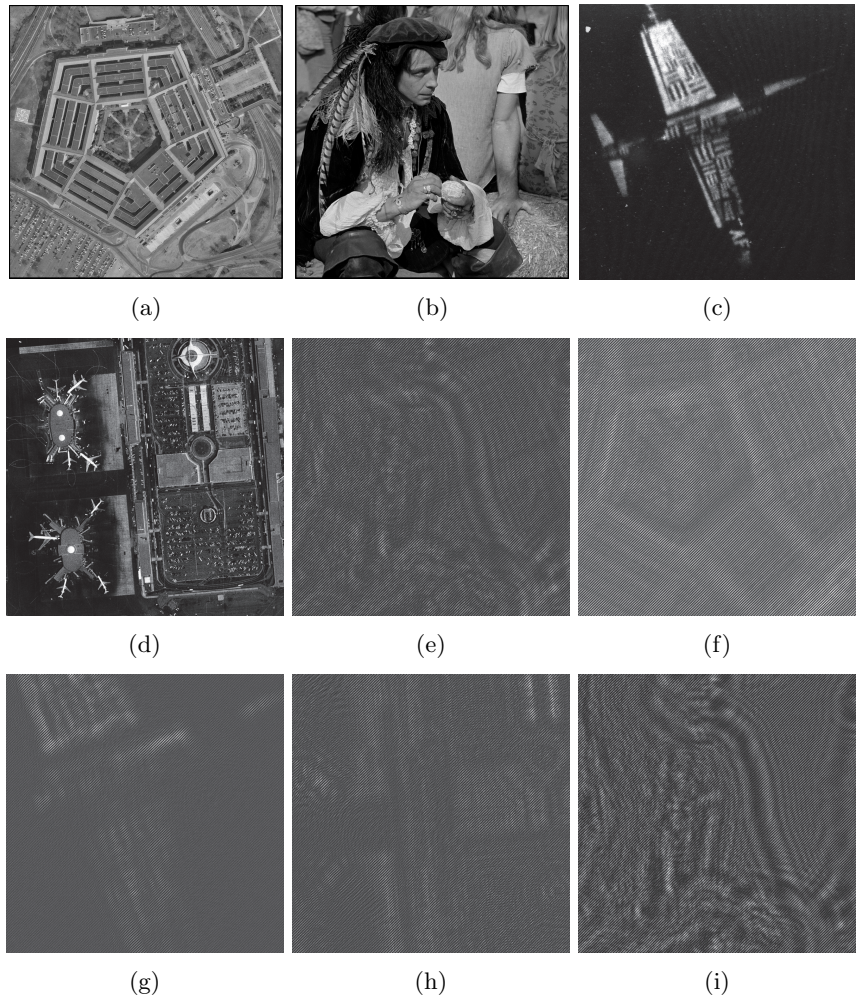


Figure 4.3: Full-sized oracle maps and corresponding hologram acquisitions used in the synthetic experiments. (a)–(d) *Pentagon*, *Man*, *Airplane*, and *Airport* maps defining the objects of Table 4.1 (e)–(i) Intensity holograms of size 512×512 associated with the objects no. 1, 2, 3, 4, and 5.

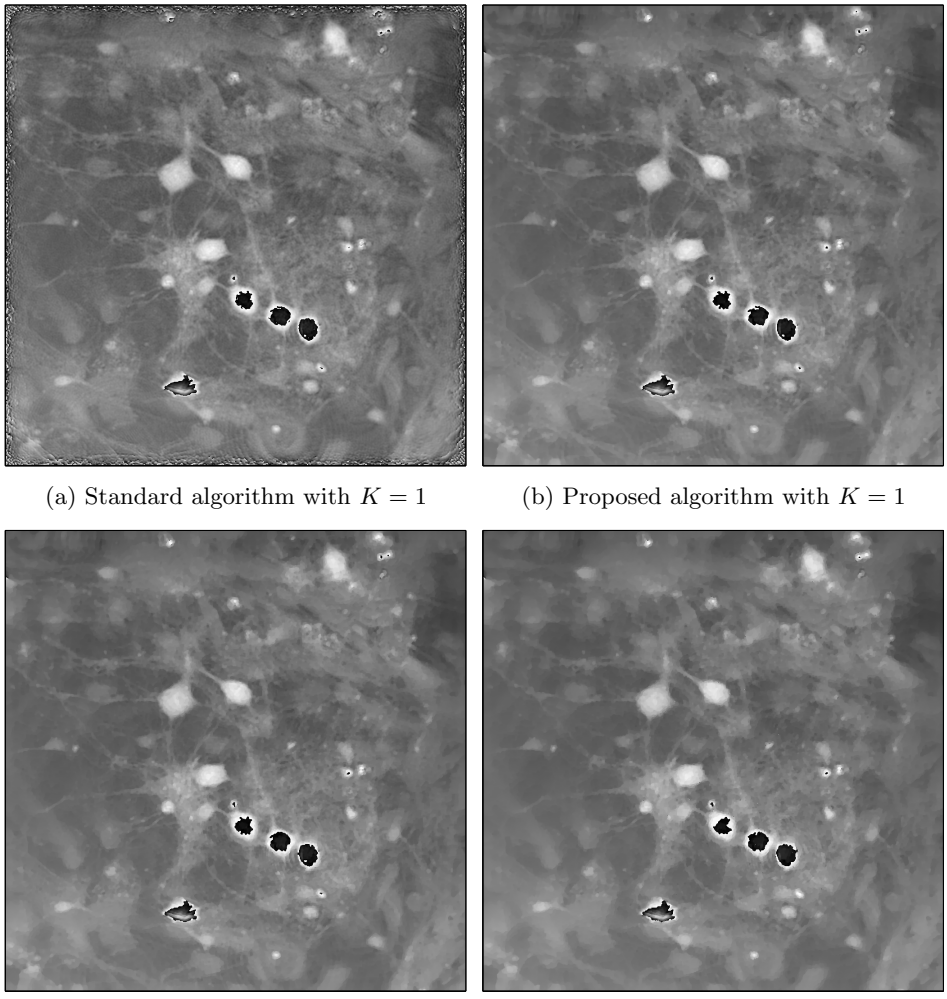
(a) Standard algorithm with $K = 1$

(b) Proposed algorithm with $K = 1$

(c) Standard algorithm with $K = 1/2$

(d) Proposed algorithm with $K = 1/2$

Figure 4.4: Reconstruction of the phase-only *Man* object from the synthetic hologram #5 without downsampling ($K = 1$) and with downsampling ($K = 1/2$).



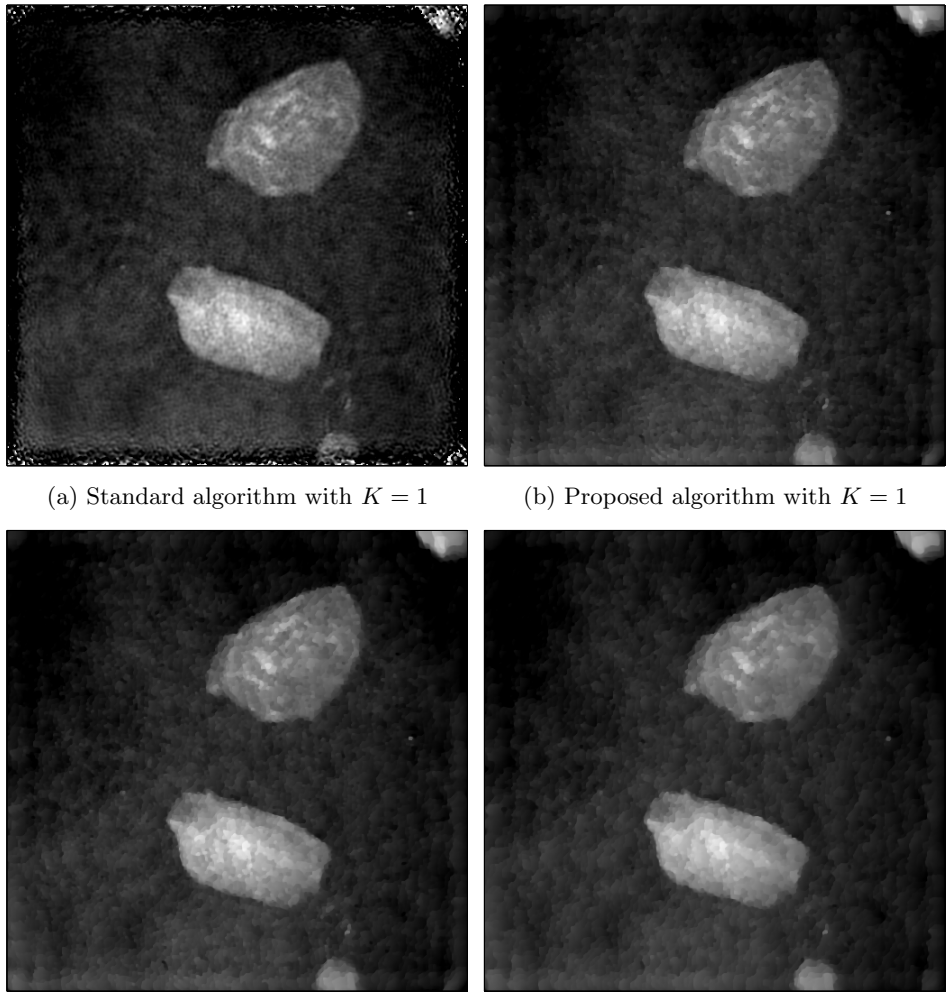
(a) Standard algorithm with $K = 1$

(b) Proposed algorithm with $K = 1$

(c) Proposed algorithm with $K = 1/2$

(d) Proposed algorithm with $K = 1/4$

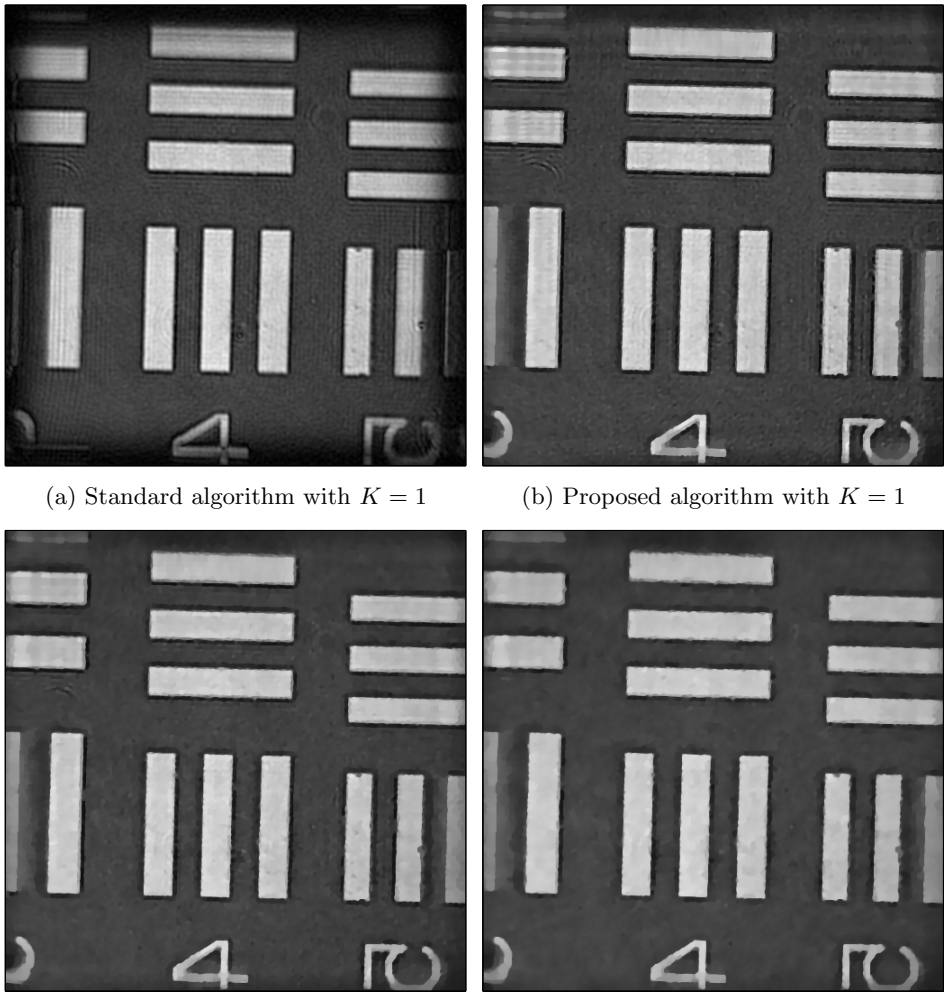
Figure 4.5: Reconstruction from the phase-only hologram *Neuron*.



(a) Standard algorithm with $K = 1$ (b) Proposed algorithm with $K = 1$

(c) Proposed algorithm with $K = 1/2$ (d) Proposed algorithm with $K = 1/4$

Figure 4.6: Reconstruction from the phase-only hologram *Epithelial*.



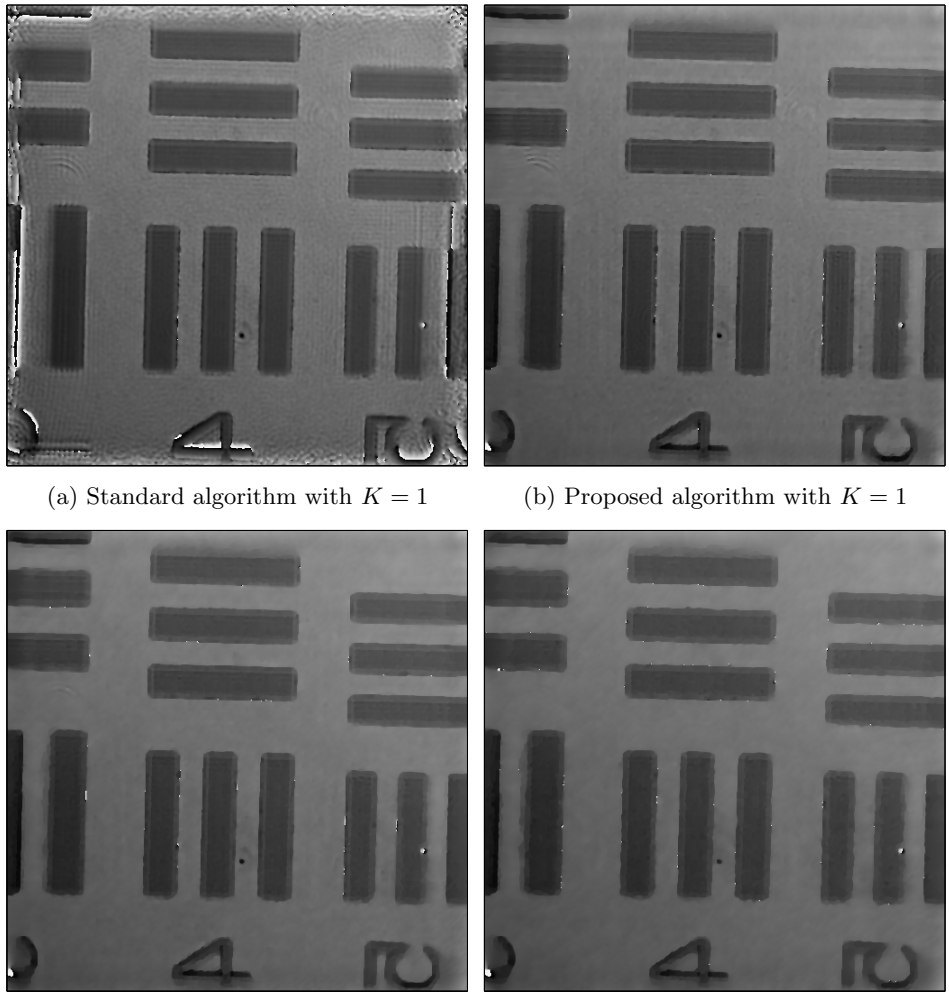
(a) Standard algorithm with $K = 1$

(b) Proposed algorithm with $K = 1$

(c) Proposed algorithm with $K = 1/2$

(d) Proposed algorithm with $K = 1/4$

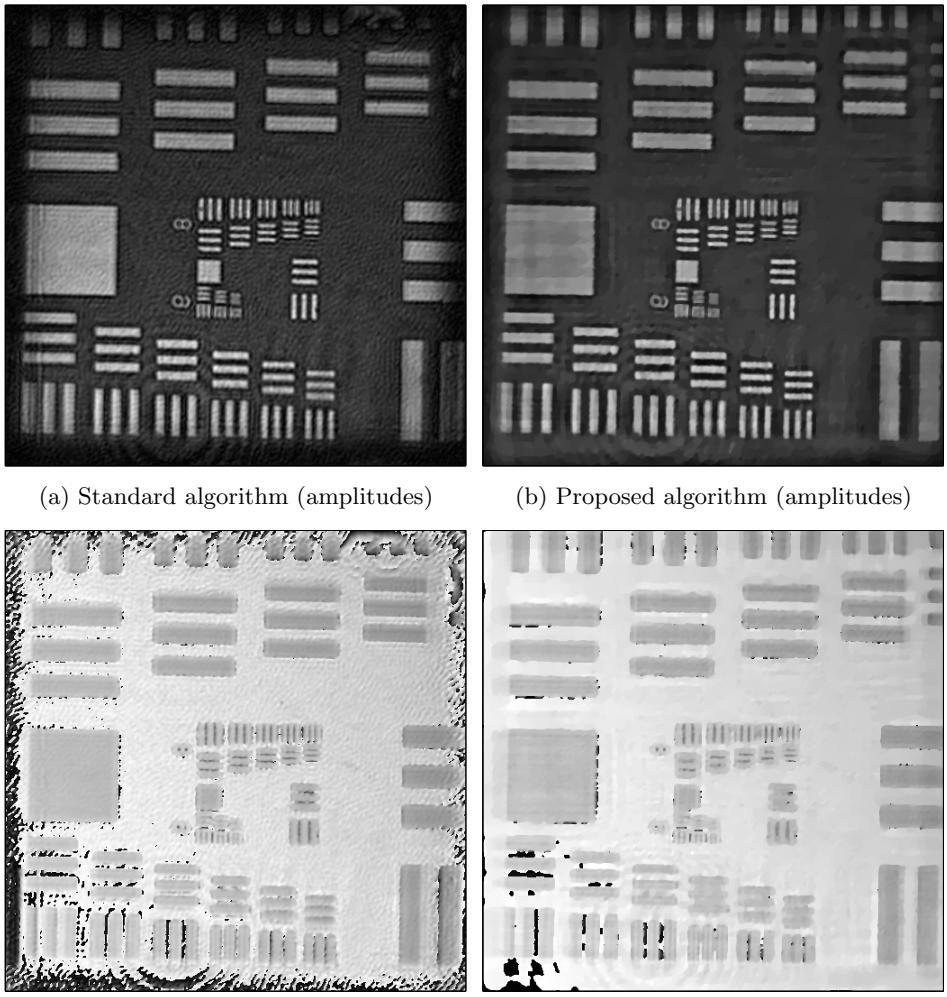
Figure 4.7: Reconstructed amplitudes from the hologram *USAF 5-4*.



(a) Standard algorithm with $K = 1$ (b) Proposed algorithm with $K = 1$

(c) Proposed algorithm with $K = 1/2$ (d) Proposed algorithm with $K = 1/4$

Figure 4.8: Reconstructed phases from the hologram *USAF 5-4*.



(a) Standard algorithm (amplitudes)

(b) Proposed algorithm (amplitudes)

(c) Standard algorithm (phases)

(d) Proposed algorithm (phases)

Figure 4.9: Reconstructed amplitudes and phases from the fully sampled hologram *USAF 9-8*.

Chapter 5

Conclusions

In this thesis, we investigated how the use of numerical techniques allows to modify the way visual optical data are acquired in several settings. In particular, we proposed specifically tailored reconstruction algorithms so as to enable imaging devices to be reconfigured or redesigned to operate with a lesser amount of data. As discussed in Chapter 1, imaging strategies that involve iterative reconstruction algorithms constitute a relatively new field of research given the required computational power. From the latter perspective, the resolution of non-convex problems in imaging is especially challenging. In the case of digital holographic reconstruction, our contribution in Chapter 4 illustrates the potential benefits of approaches that are based on inverse-problem formulations.

In Chapter 2, we proposed an efficient interpolation method based on edge-enhancing diffusion that is able to reconstruct image features accurately from only 2% of the sampled data. Subsequently, in Chapter 3, we proposed to reconstruct images from optically compressed measurements in binarized form, achieving satisfactory results with overall compression factors of 64. In both of these reconstruction problems, the presence of linear filtering effects in the forward model has been observed to improve the overall reconstruction quality. The differences are most dramatic in the case of our binary-measurement model where the preliminary step of optical filtering is necessary to retrieve grayscale information. Finally, in Chapter 4, we developed an iterative algorithm that handles phase and amplitude profile reconstruction from single downsampled intensity holograms. On the one

hand, the downsampling factors of 2 and 4 that were used in our experiments on holographic data are relatively moderate compared to what is achieved in a related work on digital holographic reconstruction [111]. On the other hand, our method has the advantage to directly handle intensity measurements according to our forward model (4.4). Note that, while the Fresnel operator used in [111] has been shown theoretically to be suitable for reconstruction in downsampled regimes, making precise statements in the case of our own reconstruction method [15] is not straightforward. Indeed, unlike in [111], our forward model is not linear and is thus incompatible with the conventional compressed-sensing framework that is proposed and studied in the literature. Better theoretical understanding might arise in the context of non-linear extensions of the compressed-sensing theory, such as the quadratic framework that is formalized and analyzed in [122].

5.1 Future work

The acquisition and reconstruction strategies that we proposed in this thesis allow to improve the performance and increase the robustness of imaging in several settings compared to the state of the art. However, for the problems that we considered, the precise influence of the forward model on the quality of reconstruction is still not well understood theoretically, up to our knowledge. Having a suitable understanding of this influence would allow to better optimize the design of the optical acquisition devices, knowing that, in each case, the amount of freedom in the optimization can be increased according to what is taken into account in the reconstruction algorithm. In the context of digital holographic reconstruction, for instance, and in the case of the algorithm proposed in Chapter 4, the reference wave is not constrained a priori to be composed of one single frequency, unlike in reconstruction techniques that are based on demodulation. Accordingly, if proven more suitable for the quality of reconstruction and technically feasible, holographic acquisition devices could be modified to generate customized reference waves; the reconstruction from single holograms obtained in an on-axis configuration could also be investigated for the same reason.

As mentioned in Section 3.7.7, the practical realization of the binary-compressed-imaging model that we proposed remains to be addressed in further research. An important challenge is the fact that, due to the non-negativity of the optical filters (3.11) in our incoherent-light setting, the light intensities acquired by the sensor

array have a very low contrast. In accordance with the finite-differentiation modality described in Section 3.7.4, this issue could be addressed using sensors acting as binary comparators [11]. Note that the electronics of such comparators should provide a sufficiently high common-mode rejection ratio in order to yield accurate measurements.

Other potential innovations lie in the development of new classes of reconstruction algorithms. In that regard, in settings where $\mathbf{A}$ corresponds to a Gaussian-type random mixing matrix and where $\mathcal{Q}$ corresponds to a scalar quantizer, problems of the form (1.6) have been successfully solved using *generalized approximate message passing* (GAMP) [123, 124]. The latter approach relies on a Bayesian formulation of the reconstruction problem, and yields very promising results in terms of reconstruction quality and computational efficiency. It is able to provide approximate minimum-mean-square-error estimates based on some given statistical priors on the solution. Related works have also studied the applicability of GAMP to other types of measurement matrices, such as convolution matrices [125], as well as to nonlinearities that are similar to the ones involved in our holographic forward model [126]. Accordingly, reconstruction problems such as the ones in Chapters 3 and 4 may be solved based on specific variants of GAMP in the future.

Bibliography

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, January 1998.

[2] P. Duhamel and M. Vetterli, "Fast Fourier transforms: A tutorial review and a state of the art," *Signal Processing*, vol. 19, no. 4, pp. 259–299, April 1990.

[3] T. Geva, "Magnetic resonance imaging: Historical perspective," *Journal of Cardiovascular Magnetic Resonance*, vol. 8, no. 4, pp. 573–580, April 2006.

[4] A. Kumar, D. Welti, and R. R. Ernst, "NMR Fourier zeugmatography," *Journal of Magnetic Resonance*, vol. 213, no. 2, pp. 495–509, April 2011.

[5] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Academic Press, 2nd edition, 2010.

[6] D. S. C. Biggs and M. Andrews, "Acceleration of iterative image restoration algorithms," *Applied Optics*, vol. 36, no. 8, pp. 1766–1775, March 1997.

[7] M. Uecker, Z. Shuo, D. Voit, A. Karaus, K.-D. Merboldt, and J. Frahm, "Real-time MRI at a resolution of 20 ms," *NMR in Biomedicine*, vol. 23, no. 8, pp. 986–994, October 2010.

[8] A. Bourquard and M. Unser, "Anisotropic interpolation of sparse generalized image samples," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 459–472, February 2013.

[9] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2915–2925, November 1994.

[10] A. Bourquard, F. Aguet, and M. Unser, "Optical imaging using binary sensors," *Optics Express*, vol. 18, no. 5, pp. 4876–4888, March 2010.

[11] A. Bourquard and M. Unser, "Binary compressed imaging," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1042–1055, March 2013.

[12] J. Romberg, "Sensing by random convolution," in *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, December 2007, pp. 137–140.

[13] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *42nd Annual Conference on Information Sciences and Systems*, March 2008, pp. 16–21.

[14] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embedding of sparse vectors," arXiv:submit/0417664 [cs.IT], February 2012.

[15] A. Bourquard, N. Pavillon, E. Bostan, C. Depeursinge, and M. Unser, "A practical inverse-problem approach to digital holographic reconstruction," *Optics Express*, vol. 21, no. 3, pp. 3417–3433, February 2013.

[16] M. K. Whitaker, *Estimating Signal Features from Noisy Images with Stochastic Backgrounds*, ProQuest, 2008.

[17] A. Stern and B. Javidi, "Random projections imaging with extended space-bandwidth product," *Journal of Display Technology*, vol. 3, no. 3, pp. 315–320, September 2007.

[18] A. Stern, Y. Rivenson, and B. Javidi, "Optically compressed image sensing using random aperture coding," in *Proceedings of the International Society for Optical Engineering*, 2008, vol. 6975, pp. 69750D–1–10.

[19] A. Bourquard, P. Thévenaz, K. Balać, and M. Unser, "Consistent and regularized magnification of images," in *Proceedings of the IEEE International Conference on Image Processing*, October 2008.

[20] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, December 2006.

[21] M. Marim, E. Angelini, J.-C. Olivo-Marin, and M. Atlan, "Off-axis compressed holographic microscopy in low-light conditions," *Optics Letters*, vol. 36, no. 1, pp. 79–81, January 2011.

[22] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972.

[23] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IEEE*, vol. 72, no. 9, pp. 1192–1201, September 1984.

[24] P. Thevenaz, T. Blu, and M. Unser, "Interpolation revisited," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 739–758, July 2000.

[25] F. Malgouyres and F. Guichard, "Edge direction preserving image zooming: A mathematical and numerical analysis," *SIAM Journal on Numerical Analysis*, vol. 39, no. 1, pp. 1–37, 2001.

[26] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1647–1659, October 2005.

[27] A. Hirabayashi and M. Unser, "Consistent sampling and signal recovery," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4104–4115, August 2007.

[28] A. Roussos and P. Maragos, "Reversible interpolation of vectorial images by an anisotropic diffusion-projection PDE," *International Journal on Computer Vision*, vol. 84, no. 2, pp. 130–145, August 2009.

[29] P. Getreuer, "Contour stencils: Total variation along curves for adaptive image interpolation," *SIAM Journal on Imaging Sciences*, vol. 4, no. 3, pp. 954–979, September 2011.

[30] M. Unser, "Multigrid adaptive image processing," in *Proceedings of the IEEE International Conference on Image Processing*, October 1995, vol. 1, pp. 49–52.

[31] M. Arigovindan, M. Sühling, P. Hunziker, and M. Unser, "Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 450–460, April 2005.

[32] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, November 1992.

[33] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "Global solutions of variational models with convex regularization," *SIAM Journal on Imaging Sciences*, vol. 3, no. 4, pp. 1122–1145, December 2010.

[34] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.

[35] I. Galic, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, "Image compression with anisotropic diffusion," *Journal of Mathematical Imaging and Vision*, vol. 31, no. 2-3, pp. 255–269, July 2008.

[36] G. Facciolo, P. Arias, V. Caselles, and G. Sapiro, "Exemplar-based interpolation of sparsely sampled images," in *Lecture Notes in Computer Science*, Bonn, Germany, August 24-27, 2009, pp. 331–344.

[37] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving PDEs," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 65–82, June 2006.

[38] F. Bornemann and T. März, "Fast image inpainting based on coherence transport," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 259–278, July 2007.

[39] J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner, Stuttgart, Germany, 1998.

[40] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," in *SIAM Journal on Scientific Computing*, Breckenridge, CO, USA, April 5-9, 1994, vol. 17, pp. 227–238.

[41] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM Journal on Numerical Analysis*, vol. 36, no. 2, pp. 354–367, February 1999.

[42] A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," *Measurement Science & Technology*, vol. 18, no. 1, pp. 87–95, January 2007.

[43] S. Grewenig, J. Weickert, and A. Bruhn, "From box filtering to fast explicit diffusion," in *Proceedings of the 32nd DAGM Symposium in Pattern Recognition*, September 2010, pp. 533–542.

[44] A. Roussos and P. Maragos, "Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals," in *Proceedings of the IEEE International Conference on Image Processing*, September 2010, pp. 4141–4144.

[45] D. Geman and Y. Chengda, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, July 1995.

[46] M. Nikolova and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1623–1627, June 2007.

[47] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, November 1999.

[48] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, June 2005.

[49] D. Tschumperlé and R. Deriche, "Anisotropic diffusion partial differential equations for multichannel image regularization: Framework and applications," *Advances in Imaging and Electron Physics*, vol. 145, no. suppl., pp. 149–209, March 2007.

[50] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 231–241, February 2001.

[51] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Processing (Netherlands)*, vol. 89, no. 9, pp. 1683–93, September 2009.

[52] F. H. Harlow and J. E. Welch, "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface," *Physics of Fluids*, vol. 8, no. 12, pp. 2182–2189, December 1965.

[53] D.R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, February 2004.

[54] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM Journal of Scientific Computing*, vol. 27, no. 3, pp. 937–966, July 2006.

[55] I. Galic, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, "Towards PDE-based image compression," in *Proceedings of the 3rd international conference on Variational, Geometric, and Level Set Methods in Computer Vision*, October 2005, pp. 37–48.

[56] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 506–517, April 2005.

[57] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 310–318, March 1998.

[58] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, July 1990.

[59] A. Selinger, R. C. Nelson, and A. C. Nelson, "Using directional variance to extract curves in images, thus improving object recognition in clutter," Tech. Rep. UR CSD / TR712, Department of Computer Science, University of Rochester, 1999.

[60] M. Grasmair and F. Lenzen, "Anisotropic total variation filtering," *Applied Mathematics and Optimization*, vol. 62, no. 3, pp. 323–339, December 2010.

[61] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, SIAM, 2000.

[62] D. M. Young, "Iterative methods for solving partial difference equations of elliptic type," *Transactions of the American Mathematical Society*, vol. 76, no. 1, pp. 92–111, January 1954.

[63] J. Gilles and Y. Meyer, "Properties of $BV - G$ structures+textures decomposition models. Application to road detection in satellite images," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2793–2800, November 2012.

[64] S. Arseneau and J. R. Cooperstock, "An improved representation of junctions through asymmetric tensor diffusion," in *Proceedings of the 2nd International Symposium on Advances in Visual Computing, Part I*, November 2006, pp. 363–372.

[65] P. Getreuer, "Roussos-Maragos tensor-driven diffusion for image interpolation," *Image Processing On Line*, September 2011, DOI:`http://dx.doi.org/10.5201/ipol.2011.g_rmdi`.

[66] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, September 2010.

[67] M. F. Duarte, M. A. Davenport, D. Takbar, J. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling: Building simpler, smaller, and less-expensive digital cameras," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[68] R. F. Marcia and R. M. Willett, "Compressive coded aperture superresolution image reconstruction," in *IEEE International Conference on Acoustic, Speech and Signal Processes*, March-April 2008, pp. 833–836.

[69] F. Sebert, Y. M. Zou, and L. Ying, "Toeplitz block matrices in compressed sensing and their applications in imaging," in *Proceedings of the 5th International Conference on Information Technology and Application in Biomedicine*, May 2008, pp. 47–50.

[70] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, February 2009.

[71] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," arXiv:1109.4299v4 [cs.IT], October 2011.

[72] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, June 2007.

[73] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, December 2005.

[74] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectronics Journal*, vol. 37, no. 5, pp. 433–451, May 2006.

[75] J. W. Goodman, *Introduction to Fourier Optics*, McGraw Hills Companies, Inc., second edition, 1996.

[76] J. N. Laska, Z. Wen, W. Yin, and R. G. Baraniuk, "Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5289–5301, November 2011.

[77] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," arXiv:1202.1212v1 [cs.IT], February 2012.

[78] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *IEEE Workshop on Statistical Signal Processing Proceedings*, August 2007, pp. 294–298.

[79] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proceedings of the 2nd Workshop on Signal Processing with Adaptive Sparse Representations*, April 2009.

[80] J. Romberg and R. Neelamani, "Sparse channel separation using random probes," *Inverse Problems*, vol. 26, no. 11, pp. 115015 (25 pages), November 2010.

[81] J. P. Slavinsky, J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "The compressive multiplexer for multi-channel compressive sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2011, pp. 3980–3983.

[82] Y. E. Nesterov, "A method of solving a convex programming problem with convergence speed $O(1/k^2)$," *Doklady Akademii Nauk SSSR*, vol. 27, no. 2, pp. 372–376, 1983.

[83] R. T. Rockafellar, *Convex Analysis (Princeton Mathematical Series)*, Princeton University Press, 1970.

[84] B. Schölkopf and A. J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, The MIT Press, December 2001.

[85] M. M. Marim, E. D. Angelini, and J.-C. Olivo-Marin, "A compressed sensing approach for biological microscopy image denoising," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, June–July 2009, pp. 1374–1377.

[86] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[87] A. Beck and M. Teboulle, "Smoothing and first order methods: A unified framework," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 557–580, June 2012.

[88] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, January 2011.

[89] S. McKee, M. F. Tomé, V. G. Ferreira, J. A. Cuminato, A. Castelo, F. S. Sousa, and N. Mangiavacchi, "The MAC method," *Computers and Fluids*, vol. 37, no. 8, pp. 907–930, September 2008.

[90] R. Pan and S. J. Reeves, "Efficient Huber-Markov edge-preserving image restoration," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3728–3735, December 2006.

[91] R. Chan, J. G. Nagy, and R. J. Plemmons, "FFT-based preconditioners for Toeplitz-block least squares problems," *SIAM Journal on Numerical Analysis*, vol. 30, no. 6, pp. 1740–1768, December 1993.

[92] K. Lange, *Numerical Analysis for Statisticians*, Springer, 2nd edition, 2010.

[93] G. E. P. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

[94] A. Schulz, L. Velho, and E. A. B. da Silva, "On the empirical rate-distortion performance of compressive sensing," in *Proceedings of the IEEE International Conference on Image Processing*, November 2009, pp. 3049–3052.

[95] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, March 1982.

[96] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. IT-6, no. 1, pp. 7–12, March 1960.

[97] L. Sbaiz, F. Yang, E. Charbon, S. Susstrunk, and M. Vetterli, "The Gigavision Camera," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 1093–1096.

[98] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[99] G. Cardano, *Artis Magnæ, Sive de Regulis Algebraicis Liber Unus*, Nuremberg, 1545.

[100] E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Applied Optics*, vol. 38, no. 34, pp. 6994–7001, December 1999.

[101] I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," *Optics Letters*, vol. 22, no. 16, pp. 1268–1270, August 1997.

[102] G. Popescu, T. Ikeda, R. Dasari, and M. Feld, "Diffraction phase microscopy for quantifying cell structure and dynamics," *Optics Letters*, vol. 31, no. 6, pp. 775–777, March 2006.

[103] Y. Awatsuji, T. Tahara, A. Kaneko, T. Koyama, K. Nishio, S. Ura, T. Kubota, and O. Matoba, "Parallel two-step phase-shifting digital holography," *Applied Optics*, vol. 47, no. 19, pp. D183–D189, July 2008.

[104] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 227–246, 1972.

[105] V. Katkovnik, A. Migukin, and J. Astola, "3D wave field reconstruction from intensity-only data: variational inverse imaging techniques," in *9th Euro-American Workshop on Information Optics*, July 2010.

[106] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and fienup variants: A view from convex optimization," *Journal of the Optical Society of America A*, vol. 19, no. 7, pp. 1334–1345, July 2002.

[107] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, August 2013.

[108] M. Cetin, W. C. Karl, and A. S. Willsky, "Edge-preserving image reconstruction for coherent imaging applications," in *Proceedings of the IEEE International Conference on Image Processing*, September 2002.

[109] X. Zhang and E. Y. Lam, "Edge-preserving sectional image reconstruction in optical scanning holography," *Journal of the Optical Society of America A*, vol. 27, no. 7, pp. 1630–1637, July 2010.

[110] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," *Optics Express*, vol. 17, no. 15, pp. 13040–13049, July 2009.

[111] M. M. Marim, M. Atlan, E. Angelini, and J. C. Olivo-Marin, "Compressed sensing with off-axis frequency-shifting holography," *Optics Letters*, vol. 35, no. 6, pp. 871–873, March 2010.

[112] Y. Rivenson, A. Stern, and B. Javidi, "Compressive Fresnel holography," *EEE/OSA Journal of Display Technology*, vol. 6, no. 10, pp. 506–509, October 2010.

[113] S. Sotthivirat and J. A. Fessler, "Penalized-likelihood image reconstruction for digital holography," *Journal of the Optical Society of America A*, vol. 21, no. 5, pp. 737–750, May 2004.

[114] N. Pavillon, C. S. Seelamantula, J. Kuhn, M. Unser, and C. Depeursinge, "Suppression of the zero-order term in off-axis digital holography through nonlinear filtering," *Applied Optics*, vol. 48, no. 34, pp. H186–H195, December 2009.

[115] T. Colomb, J. Kühn, F. Charrière, C. Depeursinge, P. Marquet, and N. Aspert, "Total aberrations compensation in digital holographic microscopy with a reference conjugated hologram," *Optics Express*, vol. 14, no. 10, pp. 4300–4306, May 2006.

[116] M. Unser, "Sampling—50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, April 2000.

[117] D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," Report LIDS 2848, Massachusetts Institute of Technology, December 2010.

[118] O. Scherzer, Ed., *Energy Minimization Methods*, Springer, 2011.

[119] E. Cremers, D. Strekalovskiy, "Total cyclic variation and generalizations," *Journal of Mathematical Imaging and Vision*, vol. 47, no. 3, pp. 258–277, November 2013.

[120] M. Guizar-Sicairos, A. Diaz, M. Holler, M. S. Lucas, A. Menzel, R. A. Wepf, and O. Bunk, "Phase tomography from X-ray coherent diffractive imaging projections," *Optics Express*, vol. 19, no. 22, pp. 21345–21357, October 2011.

[121] K. Choi, R. Horisaki, J. Hahn, S. Lim, D. L. Marks, T. J. Schulz, and D. J. Brady, "Compressive holography of diffuse objects," *Applied Optics*, vol. 49, no. 34, pp. H1–H10, December 2010.

[122] H. Ohlsson, A. Y. Yang, R. Dong, M. Verhaegen, and S. S. Sastry, "Quadratic basis pursuit," arXiv:1301.7002 [cs.IT], February 2013.

[123] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing dequantization with applications to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6270–6281, December 2012.

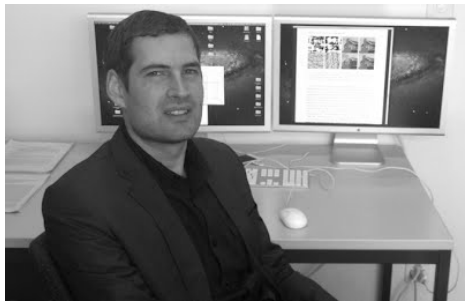
[124] U. S. Kamilov, A. Bourquard, A. Amini, and M. Unser, "One-bit measurements with adaptive thresholds," *IEEE Signal Processing Letters*, vol. 19, no. 10, pp. 607–610, October 2012.

[125] U. S. Kamilov, A. Bourquard, and M. Unser, "Sparse image deconvolution with message passing," accepted, July 2013.

[126] P Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, October 2012.

Index

Curriculum Vitæ



Aurélien Bourquard
Ecole polytechnique fédérale de Lausanne
CH-1015 Lausanne (VD), Switzerland

http://bigwww.epfl.ch/bourquard
aurelien.bourquard@epfl.ch
Tel: +41 (0)21 693 11 85

Education

Ph.D. in Electrical Engineering, EPFL 2008 - 2013
M.Sc. in Microengineering, EPFL 2006 - 2008
Total GPA: 5.87/6 (ranked first among 92 master students)
Degree completed with a Minor in Biomedical Engineering
B.Sc. in Microengineering, EPFL 2003 - 2006
Swiss Federal Maturity, Gymnase de Burier
Exchange second year at Schweizer Schule, Milano, Italy
Bilingual degree with excellence award

Experience

- Research in image processing, compressed sensing, computational optics
  6 *journal articles and* 9 *conference papers accepted as per September 2013*
- Teaching assistant in image-processing courses (exercises and laboratories)
- Reviewer for *IEEE International Symposium on Biomedical Imaging*,
  IEEE Trans. on Image Processing, IEEE Trans. on Medical Imaging,
  IEEE Trans. on Medical Imaging, Medical Image Analysis,
  Optics Express, IEEE Photonics Journal
- Collaboration on an industrial project with Essilor International S.A.

Extracurricular Activities

- Student Member of the EPFL School of Engineering Council (2005 - 2006)
- Class Representative of the EPFL Microengineering Section (2004 - 2007)

Languages

French:	Mother tongue
English:	Fluent (C2 level according to European CEFR scale)
German:	Intermediate (B2 level according to European CEFR scale)
Italien:	Intermediate
Spanish:	Intermediate

Skills

Soft: Teamwork and collaboration, transdisciplinarity
Technical: Matlab, C/C++, Java, ImageJ, COMSOL, ProEngineer, SVN

Publications

Journal Articles

- U.S. Kamilov, A. Bourquard, M. Unser, "Wavelet-Domain Approximate Message Passing for Bayesian Image Deconvolution," *submitted to IEEE Trans-*

actions on Image Processing.

- H. Kirshner, A. Bourquard, J.P. Ward, M. Porat, M. Unser, "Adaptive Image Resizing Based on Continuous-Domain Stochastic Modeling," *submitted to IEEE Transactions on Image Processing.*

- A. Bourquard, M. Unser, "Anisotropic Interpolation of Sparse, Generalized Image Samples," IEEE Transactions on Image Processing, vol. 22, no. 2, pp. 459-472, February 2013.

- A. Bourquard, M. Unser, "Binary Compressed Imaging," IEEE Transactions on Image Processing, vol. 22, no. 3, pp. 1042-1055, March 2013.

- A. Bourquard, N. Pavillon, E. Bostan, C. Depeursinge, M. Unser, "A Practical Inverse-Problem Approach to Digital Holographic Reconstruction," Optics Express, vol. 21, no. 3, pp. 3417-3433, February 11, 2013.

- U.S. Kamilov, A. Bourquard, A. Amini, M. Unser, "One-Bit Measurements with Adaptive Thresholds," IEEE Signal Processing Letters, vol. 19, no. 10, pp. 607-610, October 2012.

- S. Lefkimmiatis, A. Bourquard, M. Unser, "Hessian-Based Norm Regularization for Image Restoration with Biomedical Applications," IEEE Transactions on Image Processing, vol. 21, no. 3, pp. 983-995, March 2012.

- A. Bourquard, F. Aguet, M. Unser, "Optical Imaging Using Binary Sensors," Optics Express, vol. 18, no. 5, pp. 4876-4888, March 1, 2010.

Conference Papers

- U.S. Kamilov, A. Bourquard, M. Unser, "Sparse Image Deconvolution with Message Passing," Proceedings of Signal Processing with Adaptive Sparse Structured Representations (SPARS'13), Lausanne VD, Switzerland, July 8-11, 2013, in press.

- U.S. Kamilov, A. Bourquard, E. Bostan, M. Unser, "Autocalibrated Signal Reconstruction from Linear Measurements Using Adaptive GAMP," Proceedings of the Thirty-Eighth IEEE International Conference on Acoustics,

Speech, and Signal Processing (ICASSP'13), Vancouver BC, Canada, May 26-31, 2013, in press.

- H. Kirshner, A. Bourquard, J.P. Ward, M. Unser, "Linear Interpolation of Biomedical Images Using a Data-Adaptive Kernel," Proceedings of the Tenth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'13), San Francisco CA, USA, April 7-11, 2013, pp. 926-929.

- S. Lefkimmiatis, A. Bourquard, M. Unser, "Hessian-Based Regularization for 3-D Microscopy Image Restoration," Proceedings of the Ninth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12), Barcelona, Spain, May 2-5, 2012, pp. 1731-1734.

- R. Madani, A. Bourquard, M. Unser, "Image Segmentation with Background Correction Using a Multiplicative Smoothing-Spline Model," Proceedings of the Ninth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12), Barcelona, Spain, May 2-5, 2012, pp. 186-189.

- A. Bourquard, H. Kirshner, M. Unser, "Resolution-Invariant Separable ARMA Modeling of Images," Proceedings of the 2011 IEEE International Conference on Image Processing (ICIP'11), Brussels, Kingdom of Belgium, September 11-14, 2011, pp. 1873-1876.

- Z. Dogan, S. Lefkimmiatis, A. Bourquard, M. Unser, "A Second-Order Extension of TV Regularization for Image Deblurring," Proceedings of the 2011 IEEE International Conference on Image Processing (ICIP'11), Brussels, Kingdom of Belgium, September 11-14, 2011, pp. 713-716.

- A. Bourquard, P. Thévenaz, K. Balać, M. Unser, "Consistent and Regularized Magnification of Images," Proceedings of the 2008 IEEE International Conference on Image Processing (ICIP'08), San Diego CA, USA, October 12-15, 2008, pp. 325-328.

- D. Porto, A. Bourquard, Y. Perriard, "Genetic Algorithm Optimization for a Surgical Ultrasonic Transducer," Proceedings of the 2008 IEEE Ultrasonics Symposium (IUS'08), Beijing, China, November 2-5, 2008, pp. 1457-1460

Interests

Fitness and jogging, languages and travel, martial arts