2017

Convex-Set–Constrained Sparse Signal Recovery: Theory and Applications

Renliang Gu
Iowa State University

Convex-set–constrained sparse signal recovery: Theory and applications

by

Renliang Gu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Communications and Signal Processing)

Program of Study Committee:
Aleksandar Dogandžić, Major Professor
Aditya Ramamoorthy
Dan Nordman
Mingyi Hong
Zhengdao Wang

Iowa State University

Ames, Iowa

2017

DEDICATION

I would like to dedicate this thesis to my wife Yan and to my daughter Angela without whose support I would not have been able to complete this work. I would also like to thank my friends and family for their love and support during my Ph.D. career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS alternate convex search. 90

ADMM alternating direction method of multipliers. 51, 67, 71

AWGN additive white Gaussian noise. 15

BB Barzilai-Borwein. 9, 21, 91

BFGS Broyden-Fletcher-Goldfarb-Shanno. 90

BPDN basis pursuit denoising. 15, 37

CNDE Center for Nondestructive Evaluation. xii

CPU central processing unit. vii, 33, 38

CT computed tomography. 1, 3, 6, 72–74, 78, 79, 97, 100, 109, 118

DWT discrete wavelet transform. 6, 32, 33, 35, 36, 68–71, 73, 75, 88

FBP filtered backprojection. 32–34, 74, 98, 100–104, 107–109

FISTA fast iterative shrinkage-thresholding algorithm. iv, 9, 28, 34, 44

GFB generalized forward-backward. 9, 36–38

GLM generalized linear model. 1–3, 14, 15, 30, 32, 84, 86

GPU graphics processing unit. 73, 97

i.i.d. independent, identically distributed. 32, 35, 68

IRT Image Reconstruction Toolbox. 32, 33, 70

KL Kurdyka-Łojasiewicz. 3, 89, 95, 96, 109, 116, 117

L-BFGS-B limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints. 73, 90, 91, 118

ML maximum-likelihood. 101

MM majorization-minimization. 2

MRI magnetic resonance imaging. 6

NDE nondestructive evaluation. 73

NLL negative log-likelihood. 1, 3, 5–10, 13–15, 18, 23, 27, 28, 32, 34, 51, 52, 61, 64, 67, 68, 72, 73, 77, 84–91, 101, 115–117

NPG Nesterov's proximal-gradient. 72, 90, 92, 95, 101, 102, 108

NPG$_S$ Nesterov's proximal-gradient sparse. vii, 18, 35, 38

NSF National Science Foundation. xii

PDS primal-dual splitting. 9, 36, 38

PET positron emission tomography. 6, 30, 32, 35, 70

PG proximal-gradient. 1, 8, 10, 18–20, 22–27, 32, 34, 92

pmf probability mass function. 14

PNPG projected Nesterov's proximal-gradient. vii, 2, 5, 8, 10, 13, 15, 17, 18, 20, 23–26, 28, 29, 33–36, 38, 39, 67, 68, 118

POCS projections onto convex sets. 18

PPXA parallel proximal algorithm. 9

RSE relative square error. viii, 33, 38, 97, 101–104

SNR signal-to-noise ratio. 32, 69–71

SPECT single photon emission computed tomography. 6

SPIRAL sparse Poisson-intensity reconstruction algorithm. 7, 29, 30, 32–34, 36–38

TFOCS templates for first-order conic solvers. 10, 18, 27, 29

TV total-variation. vii, 2, 6, 18, 31–33, 56, 59, 60, 65, 68–72, 75, 88, 90, 92, 118

VMILA variable metric inexact line-search algorithm. 22, 32, 33, 35

ACKNOWLEDGEMENTS

I would like to thank Dr. Aleksandar Dogandžić, for his excellent guidance, persistent encouragement and invaluable inspirations. Without his support and selfless help, I would not be able to accomplish the work and writing of this dissertation.

My gratitude also goes to Dr. Joe Gray, Center for Nondestructive Evaluation (CNDE), for introducing me to the X-ray CT in the real world, and Dr. Zhengdao Wang for the encouragement at the lowest point in my graduate career. I deeply appreciate discussions with and constructive suggestions from Drs. Mingyi Hong, Dan Nordman, and Aditya Ramamoorthy. The list to thank extends to many others, including Dr. Kun Qiu, for guidance early in my studies; Drs. Paul Sacks and Eric Weber, for answering my harmonic-analysis questions; Drs. Yurii Nesterov and Antonin Chambolle, for their pioneering and inspirational work; and the anonymous reviewers, whose review comments pushed us to improve.

Finally, I am grateful to the support by the National Science Foundation (NSF) under Grant CCF-1421480 and the NSF/IU program, CNDE, ISU.

ABSTRACT

Convex-set constrained sparse signal reconstruction facilitates flexible measurement model and accurate recovery. The objective function that we wish to minimize is a sum of a convex differentiable data-fidelity (negative log-likelihood (NLL)) term and a convex regularization term. We apply sparse signal regularization where the signal belongs to a closed convex set within the closure of the domain of the NLL. Signal sparsity is imposed using the ℓ_1-norm penalty on the signal's linear transform coefficients.

First, we present a projected Nesterov's proximal-gradient (PNPG) approach that employs a projected Nesterov's acceleration step with restart and a duality-based inner iteration to compute the proximal mapping. We propose an *adaptive* step-size selection scheme to obtain a good local majorizing function of the NLL and reduce the time spent backtracking. We present an integrated derivation of the momentum acceleration and proofs of $\mathcal{O}(k^{-2})$ objective function convergence rate and convergence of the iterates, which account for adaptive step size, inexactness of the iterative proximal mapping, and the convex-set constraint. The tuning of PNPG is largely application independent. Tomographic and compressed-sensing reconstruction experiments with Poisson generalized linear and Gaussian linear measurement models demonstrate the performance of the proposed approach.

We then address the problem of upper-bounding the regularization constant for the convex-set–constrained sparse signal recovery problem behind the PNPG framework. This bound defines the maximum influence the regularization term has to the signal recovery. We formulate an optimization problem for finding these bounds when the regularization term can be globally minimized and develop an alternating direction method of multipliers (ADMM) type method for their computation. Simulation examples show that the derived and empirical bounds match.

Finally, we show application of the PNPG framework to X-ray computed tomography (CT) and outline a method for sparse image reconstruction from Poisson-distributed polychromatic X-ray CT measurements under the blind scenario where the material of the inspected object and the incident energy spectrum are unknown. To obtain a parsimonious mean measurement-model parameterization, we first rewrite the measurement equation by changing the integral variable from photon energy to mass attenuation, which allows us to combine the variations brought by the unknown incident spectrum and mass attenuation into a single unknown mass-attenuation spectrum function; the resulting measurement equation has the Laplace integral form. We apply a block coordinate-descent algorithm that alternates between an NPG image reconstruction step and a limited-memory BFGS with box constraints (L-BFGS-B) iteration for updating mass-attenuation spectrum parameters. Our NPG-BFGS algorithm is the first physical-model based image reconstruction method for simultaneous blind sparse image reconstruction and mass-attenuation spectrum estimation from polychromatic measurements. Real X-ray CT reconstruction examples demonstrate the performance of the proposed blind scheme.

CHAPTER 1. INTRODUCTION

This dissertation is inspired and developed while pursuing better quality of image reconstructions from real polychromatic X-ray CT data. The sparse signal recovery problem under constraints, and development of nonlinear physical-model–based reconstruction algorithms are important for the broad signal processing community. We now motivate the topics of this thesis and highlight our contribution.

1.1 Momentum-Accelerated Sparse Signal Recovery With Signal Constraints

We developed computationally efficient algorithms for minimization of general nonlinear NLL functions with sparse signal regularizations and convex-set constraints on the signal. Our algorithms can solve inverse problems under various statistical models, including the popular Gaussian and Poisson generalized linear models (GLMs). We incorporated the developed algorithm into our polychromatic X-ray CT reconstruction framework outlined in Section 1.3 by using it to estimate the underlying density map.

In Chapter 2, we present an integrated derivation of the momentum acceleration and proofs of $\mathcal{O}(k^{-2})$ objective function convergence rate and convergence of the iterates, which account for adaptive step size, inexactness of the iterative proximal mapping, and the convex-set constraint. These results are the first for an accelerated proximal-gradient (PG) method *with step-size adaptation* (and, therefore, adjustment to the local curvature of the NLL) that

- establish convergence of the iterates and

- incorporate inexact proximal operators into objective function convergence rate and convergence of the iterates analyses.

Thanks to its matrix-free nature, our proposed approach can fit large-scale sparse GLMs [BvG11]. Indeed, we demonstrate in Chapter 4 and [GD16b] that our PNPG algorithm exhibits state-of-the-art convergence speed under the large-scale scenario.

We demonstrated the power of combining step-size adaptation and momentum acceleration. In this process, we violated the conventional textbook definitions of majorizing functions that require global, rather than local, majorization. We believe that this will motivate further research on developing local majorizing functions in the general majorization-minimization (MM) algorithmic framework.

1.2 Upper-Bounding the Regularization Constant for Convex Sparse Signal Reconstruction

Selecting the range of regularization constants u is important in application areas that employ regularized statistical inference. Our upper bounds on u can be used to construct accurate prior distributions for the regularization constant and to design continuation procedures that gradually decrease a regularization constant from a large starting point down to the desired value, which improves numerical stability and convergence speed of the resulting minimization algorithm by taking advantage of the fact that minimization algorithms for penalized regularization schemes converge faster for "smoother" problems with larger u. The work described in Chapter 3 is the first to upper-bound the regularization constant for total-variation (TV) regularizations.

1.3 Polychromatic X-ray Source Modeling

X-ray sources are polychromatic. Ignoring this fact when performing reconstruction leads to artifacts, such as cupping and streaking, in reconstructed images. We proposed a new model pa-

rameterization that allows for blind correction of these artifacts and then developed reconstruction algorithms based on this parameterization. Here, blind correction means that we do not know

(i) incident spectrum (which is a characteristic of the X-ray machine) and

(ii) mass attenuation (inspected material).

Non-blind scenario. If the mass-attenuation spectrum is known (incident spectrum and mass attenuation in (i) and (ii) are known), we have established conditions under which the NLLs are *convex functions* of the density map assuming polychromatic measurements and Poisson (Chapter 4) or lognormal noise models [GD15b]. Considering the importance and accumulated knowledge in X-ray CT, establishing convexity of the NLLs associated with polychromatic X-ray CT is important; we proved these results *thanks to the Laplace-transform formulation of the noiseless single-material measurements and GLMs that follow from this formulation.*

Blind scenario. Accurately characterizing the incident spectrum of the X-ray machine and the mass-attenuation function of the inspected material is not easy or may not be possible, which justifies the blind scenario. Our reconstruction algorithm in Chapter 4 is the first physical-model–based method for simultaneous blind sparse image reconstruction and mass-attenuation spectrum estimation from polychromatic measurements. This algorithm

i) matches or outperforms non-blind linearization methods that assume perfect knowledge of the X-ray source and material properties.

We have identified and quantified inherent limitations of the blind model, such as the shift ambiguity of the mass-attenuation spectrum. Important results accomplished in this task are:

ii) conditions for biconvexity of the corresponding NLL with respect to the density map and mass-attenuation spectrum parameters (thanks to the Laplace-transform and GLM formulations of the measurement models),

iii) establishment of the Kurdyka-Łojasiewicz (KL) property of the underlying objective function (penalized NLL).

Use of parsimonious physics-based models and ambiguity, convexity, and convergence analyses differentiate our work from the existing efforts on blind beam-hardening correction [VVD+11; JBS15], which have an *ad hoc* flavor.

The underlying optimization problem for performing our blind sparse signal reconstruction in Chapter 4 is *biconvex* with respect to the density map and mass-attenuation spectrum parameters; solving and analyzing bi- and multiconvex problems is of great general interest in optimization theory, see [XY13] and references therein. It would be of interest to prove the conjecture that the general version of i) is a consequence of the separability in the measurement model.

CHAPTER 2. PROJECTED NESTEROV'S PROXIMAL-GRADIENT ALGORITHM FOR SPARSE SIGNAL RECOVERY

A paper published in *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3510–3525, 2017.

Renliang Gu and Aleksandar Dogandžić

Abstract

We develop a PNPG approach for sparse signal reconstruction that combines adaptive step size with Nesterov's momentum acceleration. The objective function that we wish to minimize is the sum of a convex differentiable data-fidelity (NLL) term and a convex regularization term. We apply sparse signal regularization where the signal belongs to a closed convex set within the closure of the domain of the NLL; the convex-set constraint facilitates flexible NLL domains and accurate signal recovery. Signal sparsity is imposed using the ℓ_1-norm penalty on the signal's linear transform coefficients. The PNPG approach employs a projected Nesterov's acceleration step with restart and a duality-based inner iteration to compute the proximal mapping. We propose an *adaptive* step-size selection scheme to obtain a good local majorizing function of the NLL and reduce the time spent backtracking. Thanks to step-size adaptation, PNPG converges faster than the methods that do not adjust to the local curvature of the NLL. We present an integrated derivation of the momentum acceleration and proofs of $\mathcal{O}(k^{-2})$ objective function convergence rate and convergence of the iterates, which account for adaptive step size, inexactness of the iterative proximal mapping, and the convex-set constraint. The tuning of PNPG is largely application independent. Tomographic and compressed-sensing reconstruction experiments with Poisson generalized linear and Gaussian linear measurement models demonstrate the performance of the proposed approach.

2.1 Introduction

Most natural signals are well described by only a few significant coefficients in an appropriate transform domain, with the number of significant coefficients much smaller than the signal size. Therefore, for a vector $\boldsymbol{x} \in \mathbb{R}^p$ that represents the signal and an appropriate *sparsifying dictionary* matrix Ψ, $\Psi^H \boldsymbol{x}$ is a signal transform-coefficient vector with most elements having negligible magnitudes. Real-valued $\Psi \in \mathbb{R}^{p \times p'}$ can accommodate discrete wavelet transform (DWT) or gradient-map sparsity with anisotropic TV sparsifying transform (with $\Psi = [\Psi_\mathrm{v} \ \Psi_\mathrm{h}]$); a complex-valued $\Psi = \Psi_\mathrm{v} + \mathrm{j}\Psi_\mathrm{h} \in \mathbb{C}^{p \times p'}$ can accommodate gradient-map sparsity and the 2D isotropic TV sparsifying transform; here $\Psi_\mathrm{v}, \Psi_\mathrm{h} \in \mathbb{R}^{p \times p'}$ are the vertical and horizontal difference matrices similar to those in [BV16, Sec. 15.3.3]. The idea behind compressed sensing [CT06] is to *sense* the significant components of $\Psi^H \boldsymbol{x}$ using a small number of measurements; here, "^{H}" denotes the conjugate transpose.

We use the NLL (data-fidelity) function $\mathcal{L}(\boldsymbol{x})$ to describe the noisy measurement process. Consider signals $\boldsymbol{x}$ that belong to a closed convex set $\boldsymbol{C}$ and assume

$$\boldsymbol{C} \subseteq \mathrm{cl}(\mathrm{dom}\, \mathcal{L}) \tag{2.1}$$

which ensures that $\mathcal{L}(\boldsymbol{x})$ is computable for all $\boldsymbol{x} \in \mathrm{int}\, \boldsymbol{C}$. If $\boldsymbol{C} \setminus \mathrm{dom}\, \mathcal{L}$ is not empty, then $\mathcal{L}(\boldsymbol{x})$ is not computable in it, which needs special attention; see Section 2.3. The nonnegative signal scenario with

$$\boldsymbol{C} = \mathbb{R}^p_+ \tag{2.2}$$

is of significant practical interest and applicable to X-ray CT, single photon emission computed tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI), where the pixel values correspond to inherently nonnegative density or concentration maps [PL15]. Harmany, Marcia, and Willett consider such a nonnegative sparse signal model and de-

velop in [HMW12] and [HTWM10] a convex-relaxation sparse Poisson-intensity reconstruction algorithm (SPIRAL) and a linearly constrained gradient projection method for Poisson and Gaussian linear measurements, respectively. In addition to signal nonnegativity, other convex-set constraints have been considered in the literature: prescribed value in the Fourier domain; box, geometric, and total-energy constraints; intersections of these sets [YW82]; and unit simplex [HR12].

We adopt the analysis regularization framework and minimize

$$f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + ur(\boldsymbol{x}) \tag{2.3a}$$

with respect to the signal $\boldsymbol{x}$, where $\mathcal{L}(\boldsymbol{x})$ is a differentiable convex NLL and

$$r(\boldsymbol{x}) = I_C(\boldsymbol{x}) + \rho(\boldsymbol{x}) \tag{2.3b}$$

is a convex regularization term that imposes convex-set constraint on $\boldsymbol{x}$, $\boldsymbol{x} \in C$, and sparsity of an appropriate transformed $\boldsymbol{x}$ through the convex penalty $\rho(\boldsymbol{x})$ [HMW12; GD15c; GD16b; BPR16; DFS12]. Here, $u > 0$ is a scalar tuning constant that quantifies the weight of the regularization term, and $I_C(\boldsymbol{x}) \triangleq \begin{cases} 0, & \boldsymbol{x} \in C \\ +\infty, & \text{otherwise} \end{cases}$ is the indicator function. The penalty $\rho(\boldsymbol{x})$ is often selected as the ℓ_1-norm of the signal transform-coefficient vector [DFS12]:

$$\rho(\boldsymbol{x}) = \|\Psi^H \boldsymbol{x}\|_1. \tag{2.4}$$

Define the proximal operator for a function $r(\boldsymbol{x})$ scaled by $\lambda > 0$ at argument $\boldsymbol{a} \in \mathbb{R}^p$:

$$\operatorname{prox}_{\lambda r} \boldsymbol{a} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^p} \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{a}\|_2^2 + \lambda r(\boldsymbol{x}). \tag{2.5}$$

In this chapter (see also [GD15c; GD16b]), we develop a PNPG method whose momentum acceleration accommodates adaptive step-size selection and convex-set constraint on the signal $\boldsymbol{x}$.

Computing the proximal operator with respect to $r(\boldsymbol{x})$ in (2.3b) needs iteration and is therefore inexact [DFS09; SRB11; VSBV13]. We establish conditions for the $\mathcal{O}(k^{-2})$ convergence rate of the objective function as well as the convergence of PNPG iterates. These results are the first for an accelerated PG method *with step-size adaptation* (and, therefore, adjustment to the local curvature of the NLL) that

- establish convergence of the iterates (Theorem 2.2) and
- incorporate inexact proximal operators into objective function convergence rate and convergence of the iterates analyses (Theorems 2.1 and 2.2).

We modify the original Nesterov's acceleration [Nes83; BT09a] so that we can establish these results when the step size is adaptive and adjusts to the local curvature of the NLL. (Local-curvature adjustments of the NLL by step-size adaptation have also been used in other algorithms under different contexts in [Nes13; BCG11; BLPP16]; see also the following and discussion in Section 2.4.1.1.) Our integration of the adaptive step size and convex-set constraint extends the application of the Nesterov-type acceleration to more general measurement models than those used previously, such as the Poisson compressed-sensing scenario described in Section 2.2.1. Furthermore, a convex-set constraint can bring significant improvement to signal reconstructions compared with imposing signal sparsity only, as illustrated in Section 2.5.2. See Section 2.4.1 for further discussion of $\mathcal{O}(k^{-2})$ acceleration approaches [AT06; BCG11; BPR16; BN16; BT09a].

Optimization problems (2.3a) with composite penalty-term structure in (2.3b) have been considered in [DFS09; CPP09; HMW12; AABM12], which use PG (forward-backward)–type methods with nested inner iterations. The general optimization approach in these references is close to ours. Unlike PNPG, these methods approximate the NLLs whose gradients are not Lipschitz continuous and [DFS09; CPP09; HMW12] do not have fast $\mathcal{O}(k^{-2})$ convergence-rate guarantees; [CPP09] observes the benefits of larger step sizes and step-size adaptation. The nested forward-backward splitting iteration in [AABM12] applies fast iterative shrinkage-thresholding algorithm (FISTA) [BT09b] in both the outer and inner loops using duality to formulate the inner iteration;

however, it does not employ step-size adaptation or analyze effects of inexact proximal-mapping computations. References [DFS12; RFP13; PCP11; CP11a; Con13; Vũ13; AABM12] describe splitting schemes to minimize (2.3a), where [DFS12; PCP11] are inspired by the parallel proximal algorithm (PPXA) [CP11b]. Some splitting schemes, e.g., [DFS12; PCP11], apply proximal operations on individual summands $\mathcal{L}(\boldsymbol{x})$, $u\rho(\boldsymbol{x})$, and $I_C(\boldsymbol{x})$, which is useful if all individual proximal operators are easy to compute. Both [DFS12] and generalized forward-backward (GFB) splitting [RFP13] require inner iterations to solve $\mathrm{prox}_{\lambda\rho}\,\boldsymbol{a}$ for $\rho(\boldsymbol{x})$ in (2.4) in the general case where the sparsifying matrix Ψ is not orthogonal. Reference [AABM12] applies the primal-dual approach by Chambolle and Pock [CP11a], which allows solving its Poisson reconstruction problems without approximating the NLL: (2.3a) is split into $\mathcal{L}(\boldsymbol{x})$ and $r(\boldsymbol{x})$ and also into $\mathcal{L}(\boldsymbol{x}) + \rho(\boldsymbol{x})$ and $I_C(\boldsymbol{x})$, where the second approach (termed CP) does not require nested iterations. The primal-dual splitting (PDS) method in [Con13; Vũ13] does not require inner iterations for general $\mathcal{L}(\boldsymbol{x})$ and sparsifying matrix. GFB and PDS need Lipschitz-continuous gradient of $\mathcal{L}$ and the value of the Lipschitz constant is important for tuning their parameters. The convergence rate of both GFB and PDS methods can be upper-bounded by C/k where k is the number of iterations and the constant C is determined by values of the tuning proximal and relaxation constants [LFP16; Dav15]. In Section 2.5, we show the performances of CP, GFB, and PDS.

Variable-metric methods with *problem-specific* diagonal scaling matrices have been considered in [BLPP16; Sal16; BPR16]; [BLPP16] applies Barzilai-Borwein (BB) step size and an Armijo line search for the overrelaxation parameter. It accounts for inexact proximal operator and establishes convergence of iterates but *does not* employ acceleration or provide fast convergence-rate guarantees. [BLPP16] does not require the Lipschitz continuity of the gradient of the NLL in general, except for proving the convergence rate of the objective function. Salzo [Sal16] analyzes variable-metric algorithms without acceleration (of the type [BLPP16]) and relies on the uniform continuity of $\nabla\mathcal{L}$ for the convergence analysis of both objective function and iterates; however, [Sal16] does not account for inexact proximal operators. In practice, special care is needed in selecting a good scaling matrix, and no clear guidelines are given in [BLPP16; Sal16; BPR16]

for this selection. Setting the overrelaxation parameter in [Sal16] to unity leads to a variable-metric/scaling scheme with an adaptive step size; further, setting the scaling matrix to identity leads to a PG iteration with adaptive step size.

Similar to templates for first-order conic solvers (TFOCS) [BCG11], PNPG code is easy to maintain: for example, the proximal-mapping computation can be easily replaced as a module by the latest state-of-the-art solver. Furthermore, PNPG requires minimal *application-independent tuning*; indeed, we use the same set of tuning parameters in two different application examples. This is in contrast with the existing splitting methods, which require problem-dependent (NLL- and u-dependent) tuning, with convergence speed sensitive to the choice of tuning constants.

We review the notation: $\mathbf{0}$, $\mathbf{1}$, I, denoting the vectors of zeros and ones and identity matrix, respectively; "$\succeq$" is the elementwise version of "$\geq$"; "^{T}" and "^{H}" are transpose and conjugate transpose, respectively. For a vector $\boldsymbol{a} = (a_i)_{i=1}^N \in \mathbb{R}^N$, define the projection and soft-thresholding operators:

$$P_C(\boldsymbol{a}) = \arg\min_{\boldsymbol{x}\in C} \|\boldsymbol{x} - \boldsymbol{a}\|_2^2 \tag{2.6a}$$

$$[\mathcal{T}_\lambda(\boldsymbol{a})]_i = \operatorname{sgn}(a_i)\max\left(|a_i| - \lambda, 0\right) \tag{2.6b}$$

and the elementwise exponential function $[\exp_\circ \boldsymbol{a}]_i = \exp a_i$. The projection onto $\mathbb{R}_+^N$ and the proximal operator (2.5) for the ℓ_1-norm $\|\boldsymbol{x}\|_1$ can be computed in closed form:

$$\left[P_{\mathbb{R}_+^N}(\boldsymbol{a})\right]_i = \max(a_i, 0), \qquad \operatorname{prox}_{\lambda\|\cdot\|_1} \boldsymbol{a} = \mathcal{T}_\lambda(\boldsymbol{a}). \tag{2.6c}$$

2.1.1 Preliminary Results

Define the ε-subgradient [Ber15, Sec. 3.3] ($\varepsilon > 0$):

$$\partial_\varepsilon r(\boldsymbol{x}) \triangleq \left\{\boldsymbol{g} \in \mathbb{R}^p \mid r(\boldsymbol{z}) \geq r(\boldsymbol{x}) + (\boldsymbol{z} - \boldsymbol{x})^T \boldsymbol{g} - \varepsilon, \forall \boldsymbol{z} \in \mathbb{R}^p\right\} \tag{2.7}$$

and an *inexact proximal operator* [VSBV13]:

Definition 1. We say that $\boldsymbol{x}$ approximates $\operatorname{prox}_{ur}(\boldsymbol{a})$ with ε-precision, denoted

$$\boldsymbol{x} \approxeq_{\varepsilon} \operatorname{prox}_{ur} \boldsymbol{a} \tag{2.8}$$

if $(\boldsymbol{a}-\boldsymbol{x})/u \in \partial_{\frac{\varepsilon^2}{2u}} r(\boldsymbol{x})$.

Proposition 2.1. $\boldsymbol{x} \approxeq_{\varepsilon} \operatorname{prox}_{ur} \boldsymbol{a}$ *implies* $\|\boldsymbol{x} - \operatorname{prox}_{ur} \boldsymbol{a}\|_2 \leq \varepsilon$.

Proof: By Definition 1, the following holds for any $\boldsymbol{z}$:

$$ur(\boldsymbol{z}) \geq ur(\boldsymbol{x}) + (\boldsymbol{z}-\boldsymbol{x})^T(\boldsymbol{a}-\boldsymbol{x}) - 0.5\varepsilon^2 \tag{2.9a}$$

which is equivalent to

$$0.5\|\boldsymbol{z}-\boldsymbol{a}\|_2^2 + ur(\boldsymbol{z}) \geq 0.5\|\boldsymbol{x}-\boldsymbol{a}\|_2^2 + ur(\boldsymbol{x}) + 0.5\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 - 0.5\varepsilon^2. \tag{2.9b}$$

Since $\boldsymbol{z} = \operatorname{prox}_{ur} \boldsymbol{a}$ minimizes the left-hand side of (2.9b), substituting it into (2.9b) completes the proof. □

Now, we adapt the results in [BT09b, Sec. IV-A] and [AABM12, Sec. 5.2.4] to complex Ψ using the fact that, for complex $\boldsymbol{y}$ and $\boldsymbol{p}$, $\|\boldsymbol{y}\|_1 = \max_{\|\boldsymbol{p}\|_\infty \leq 1} \operatorname{Re}(\boldsymbol{p}^H \boldsymbol{y})$. The proximal operator (2.5) with $\rho(\boldsymbol{x})$ in (2.4) can be rewritten as

$$\operatorname{prox}_{\lambda r} \boldsymbol{a} = \hat{\boldsymbol{x}}(\hat{\boldsymbol{p}}) \tag{2.10a}$$

where $\hat{\boldsymbol{p}} \in \mathbb{C}^{p'}$ solves the *dual problem* [BT09b; AABM12]:

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p} \in H} \tfrac{1}{2}\|\mathcal{S}(\boldsymbol{p})\|_2^2 - \tfrac{1}{2}\|\mathcal{S}(\boldsymbol{p}) - \hat{\boldsymbol{x}}(\boldsymbol{p})\|_2^2 \tag{2.10b}$$

and

$$H \triangleq \{\boldsymbol{w} \in \mathbb{C}^{p'} \mid \|\boldsymbol{w}\|_\infty \leq 1\} \tag{2.10c}$$

$$\mathcal{S}(\boldsymbol{p}) \triangleq \boldsymbol{a} - \lambda \operatorname{Re}(\Psi \boldsymbol{p}) \tag{2.10d}$$

$$\hat{\boldsymbol{x}}(\boldsymbol{p}) \triangleq P_C\big(\mathcal{S}(\boldsymbol{p})\big) \in \mathbb{R}^p. \tag{2.10e}$$

When $\boldsymbol{p} \in H$, the objective function in (2.10b) is differentiable with respect to the real and imaginary parts of $\boldsymbol{p}$. When Ψ is real-valued, the optimal $\hat{\boldsymbol{p}}$ must be real-valued and hence (2.10b) reduces to optimization with respect to $\boldsymbol{p}$ over the unit hypercube.

The duality gap for the optimization problem (2.10b) is

$$G(\boldsymbol{p}) = \lambda\{\rho\big(\hat{\boldsymbol{x}}(\boldsymbol{p})\big) - \hat{\boldsymbol{x}}^T(\boldsymbol{p}) \operatorname{Re}(\Psi \boldsymbol{p})\} + I_H(\boldsymbol{p}). \tag{2.11}$$

To simplify the notation, we omit the dependence of $G(\boldsymbol{p}), \hat{\boldsymbol{x}}(\boldsymbol{p}), \mathcal{S}(\boldsymbol{p})$ and $\hat{\boldsymbol{p}}$ on $\boldsymbol{a}$ and λ. We will add the subscripts "$_{\boldsymbol{a},\lambda}$" to these quantities when we wish to emphasize their dependence on $\boldsymbol{a}$ and λ.

The following proposition extends the result in [VSBV13, Sec. 2.1] to accommodate the composite penalty (2.3b) that includes the indicator function $I_C(\boldsymbol{x})$; if $C = \mathbb{R}^p$, it reduces to [VSBV13, Prop. 2.3]. It can be used to guarantee the ε-precision of the proximal mapping in (2.8).

Proposition 2.2. *If the duality gap* (2.11) *satisfies* $G(\boldsymbol{p}) \leq \varepsilon^2/2$, *then*

$$\hat{\boldsymbol{x}}(\boldsymbol{p}) \approxeq_\varepsilon \operatorname{prox}_{\lambda r} \boldsymbol{a}. \tag{2.12}$$

Proof: Finite $G(\boldsymbol{p})$ implies $\boldsymbol{p} \in H$. Therefore, $0 \geq \boldsymbol{z}^T \operatorname{Re}(\Psi \boldsymbol{p}) - \rho(\boldsymbol{z})$ for all $\boldsymbol{z} \in \mathbb{R}^p$ (see also (2.4)) and thus

$$G(\boldsymbol{p})/\lambda \geq \rho\big(\hat{\boldsymbol{x}}(\boldsymbol{p})\big) + [\boldsymbol{z} - \hat{\boldsymbol{x}}(\boldsymbol{p})]^T \operatorname{Re}(\Psi \boldsymbol{p}) - \rho(\boldsymbol{z}). \tag{2.13}$$

Use the projection theorem [Ber15, Prop. 1.1.9 in App. B] to obtain

$$I_C(\boldsymbol{z}) \geq [\boldsymbol{z} - \widehat{\boldsymbol{x}}(\boldsymbol{p})]^T [\mathcal{S}(\boldsymbol{p}) - \widehat{\boldsymbol{x}}(\boldsymbol{p})]/\lambda. \tag{2.14}$$

Adding (2.13), (2.14), and $\varepsilon^2/(2\lambda) \geq G_\lambda(\boldsymbol{p})/\lambda$ and reorganizing yields

$$r(\boldsymbol{z}) \geq r\big(\widehat{\boldsymbol{x}}(\boldsymbol{p})\big) + [\boldsymbol{z} - \widehat{\boldsymbol{x}}(\boldsymbol{p})]^T [\boldsymbol{a} - \widehat{\boldsymbol{x}}(\boldsymbol{p})]/\lambda - \varepsilon^2/(2\lambda) \tag{2.15}$$

where we used (2.10d), (2.3b), and $I_C\big(\widehat{\boldsymbol{x}}(\boldsymbol{p})\big) = 0$. According to Definition 1, (2.12) and (2.15) are equivalent. □

We introduce representative NLL functions (Section 2.2), describe the proposed PNPG signal reconstruction algorithm (Section 2.3), establish its convergence properties (Section 2.4), present numerical examples (Section 2.5), and make concluding remarks (Section 2.6).

2.2 Probabilistic Measurement Models

For numerical stability, we normalize the likelihood function so that the corresponding NLL $\mathcal{L}(\boldsymbol{x})$ is lower-bounded by zero.

2.2.1 Poisson Generalized Linear Model

GLMs with Poisson observations are often adopted in astronomic, optical, hyperspectral, and tomographic imaging [PL15; HMW12; OF97] and are used to model event counts, e.g., numbers of particles hitting a detector. Assume that the measurements $\boldsymbol{y} = (y_n)_{n=1}^N \in \mathbb{N}_0^N$ are independent Poisson-distributed[1] with means $[\boldsymbol{\phi}(\boldsymbol{x})]_n$.

[1] Here, we use the extended Poisson probability mass function (pmf) $\mathrm{Poisson}(y \mid \mu) = (\mu^y / y!)\, \mathrm{e}^{-\mu}$ for all $\mu \geq 0$ by defining $0^0 = 1$ to accommodate the identity-link model.

Upon normalization, we obtain the generalized Kullback-Leibler divergence form of the NLL [ZBBR15]

$$\mathcal{L}(\boldsymbol{x}) = \mathbf{1}^T[\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{y}] + \sum_{n, y_n \neq 0} y_n \ln \frac{y_n}{[\boldsymbol{\phi}(\boldsymbol{x})]_n}. \tag{2.16a}$$

The NLL $\mathcal{L}(\boldsymbol{x}) : \mathbb{R}^p \mapsto \mathbb{R}_+$ is a convex function of the signal $\boldsymbol{x}$. Here, the relationship between the linear predictor $\Phi\boldsymbol{x}$ and the expected value $\boldsymbol{\phi}(\boldsymbol{x})$ of the measurements $\boldsymbol{y}$ is summarized by the link function $\boldsymbol{g}(\cdot) : \mathbb{R}^N \mapsto \mathbb{R}^N$ [MN89]:

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{g}^{-1}(\Phi\boldsymbol{x}). \tag{2.16b}$$

Note that $\mathrm{cl}(\mathrm{dom}\, \mathcal{L}) = \{\boldsymbol{x} \in \mathbb{R}^p \mid \boldsymbol{\phi}(\boldsymbol{x}) \succeq \mathbf{0}\}$.

Two typical link functions in the Poisson GLM are log (described in [GD15d, Sec. I-A2], see also [GD16a]) and identity:

$$\boldsymbol{g}(\boldsymbol{\mu}) = \boldsymbol{\mu} - \boldsymbol{b}, \qquad \boldsymbol{\phi}(\boldsymbol{x}) = \Phi\boldsymbol{x} + \boldsymbol{b} \tag{2.17}$$

used for modeling the photon count in optical imaging and radiation activity in emission tomography [PL15, Ch. 9.2], as well as for astronomical image deconvolution. Here, $\Phi \in \mathbb{R}_+^{N \times p}$ and $\boldsymbol{b} \in \mathbb{R}_+^{N \times 1}$ are the known sensing matrix and intercept term, respectively; the intercept $\boldsymbol{b}$ models background radiation and scattering estimated, e.g., by calibration before the measurements $\boldsymbol{y}$ have been collected. The nonnegative set $\boldsymbol{C}$ in (2.2) satisfies (2.1), where we have used the fact that the elements of Φ are nonnegative. If $\boldsymbol{b}$ has zero components, $\boldsymbol{C} \setminus \mathrm{dom}\, \mathcal{L}$ is *not empty* and the NLL does not have a Lipschitz-continuous gradient.

2.2.2 Linear Model with Gaussian Noise

The linear measurement model with zero-mean additive white Gaussian noise (AWGN) leads to the following scaled NLL:

$$\mathcal{L}(\boldsymbol{x}) = \tfrac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 \tag{2.18}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the measurement vector, and constant terms (not functions of $\boldsymbol{x}$) have been ignored. This NLL belongs to the Gaussian GLM with identity link without intercept: $\boldsymbol{g}(\boldsymbol{\mu}) = \boldsymbol{\mu}$. Here, $\mathrm{dom}\,\mathcal{L}(\boldsymbol{x}) = \mathbb{R}^p$, any closed convex C satisfies (2.1), and the set $C \setminus \mathrm{dom}\,\mathcal{L}$ is empty.

Minimization of the objective function (2.3a) with Gaussian NLL (2.18) and penalty (2.3b) with $\rho(\boldsymbol{x})$ in (2.4) is an *analysis basis pursuit denoising (BPDN) problem with a convex signal constraint*.

2.3 Reconstruction Algorithm

We propose a PNPG approach for minimizing (2.3a) that combines convex-set projection with Nesterov acceleration [Nes83; BT09a] and applies adaptive step size to adapt to the local curvature of the NLL and restart to ensure monotonicity of the resulting iteration. The pseudo code in Algorithm 1 summarizes our PNPG method.

Define the quadratic approximation of the NLL $\mathcal{L}(\boldsymbol{x})$ as

$$Q_\beta(\boldsymbol{x} \mid \bar{\boldsymbol{x}}) = \mathcal{L}(\bar{\boldsymbol{x}}) + (\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla\mathcal{L}(\bar{\boldsymbol{x}}) + \frac{1}{2\beta}\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2 \tag{2.19}$$

Algorithm 1: PNPG iteration

Input: $\boldsymbol{x}^{(0)}$, u, γ, b, $\mathtt{m}$, $\mathtt{m}$, ξ, η, and threshold ϵ
Output: $\arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$
Initialization: $\boldsymbol{x}^{(-1)} \leftarrow \mathbf{0}$, $i \leftarrow 0$, $\kappa \leftarrow 0$, $\beta^{(1)}$ by the BB method
repeat
  $i \leftarrow i + 1$ and $\kappa \leftarrow \kappa + 1$
  while *true* **do** `// backtracking search`
    evaluate (2.20a) to (2.20d)
    if $\overline{\boldsymbol{x}}^{(i)} \notin \operatorname{dom} \mathcal{L}$ **then** `// domain restart`
      $\theta^{(i-1)} \leftarrow 1$ and continue
    solve the proximal mapping in (2.20e)
    if *majorization condition* (2.21) *holds or the number of backtrackings exceeds* t_{MAX} **then**
      break
    else
      if $\beta^{(i)} > \beta^{(i-1)}$ **then** `// increase` $\mathtt{m}$
        $\mathtt{m} \leftarrow \mathtt{m} + \mathtt{m}$
      $\beta^{(i)} \leftarrow \xi\beta^{(i)}$ and $\kappa \leftarrow 0$
  if $f(\boldsymbol{x}^{(i)}) > f(\boldsymbol{x}^{(i-1)})$ **then**
    if (2.21) *holds* **then**
      if $\overline{\boldsymbol{x}}^{(i)} \neq \boldsymbol{x}^{(i-1)}$ *and* $f(\boldsymbol{x}^{(i)}) \leq f(\overline{\boldsymbol{x}}^{(i)})$ **then** `// function restart`
        $\theta^{(i-1)} \leftarrow 1$, $i \leftarrow i - 1$, and continue
      if $\eta > \eta_{\text{MIN}}$ *and* $f(\boldsymbol{x}^{(i)}) > f(\overline{\boldsymbol{x}}^{(i)})$ **then** `// more accurate proximal`
        $\eta \leftarrow \eta/10$, $i \leftarrow i - 1$, and continue
    declare convergence
  if *convergence cond.* (2.23a) *holds with threshold* ϵ **then**
    declare convergence
  if $\kappa \geq \mathtt{m}$ **then** `// adapt step size`
    $\kappa \leftarrow 0$ and $\beta^{(i+1)} \leftarrow \beta^{(i)}/\xi$
  else
    $\beta^{(i+1)} \leftarrow \beta^{(i)}$
until *convergence declared or maximum number of iterations exceeded*

with step-size tuning constant $\beta > 0$. Iteration i of the PNPG method proceeds as follows:

$$B^{(i)} = \beta^{(i-1)}/\beta^{(i)} \tag{2.20a}$$

$$\theta^{(i)} = \begin{cases} 1, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases} \tag{2.20b}$$

$$\Theta^{(i)} = (\theta^{(i-1)} - 1)/\theta^{(i)} \tag{2.20c}$$

$$\bar{\boldsymbol{x}}^{(i)} = P_C\left(\boldsymbol{x}^{(i-1)} + \Theta^{(i)}(\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i-2)})\right) \tag{2.20d}$$

$$\boldsymbol{x}^{(i)} = \mathrm{prox}_{\beta^{(i)}\boldsymbol{ur}}\left(\bar{\boldsymbol{x}}^{(i)} - \beta^{(i)}\nabla\mathcal{L}(\bar{\boldsymbol{x}}^{(i)})\right) \tag{2.20e}$$

where $\beta^{(i)} > 0$ is an *adaptive step size* chosen to satisfy the *majorization condition*

$$\mathcal{L}(\boldsymbol{x}^{(i)}) \leq Q_{\beta^{(i)}}(\boldsymbol{x}^{(i)} \mid \bar{\boldsymbol{x}}^{(i)}) \tag{2.21}$$

using a simple adaptation scheme that aims at keeping $\beta^{(i)}$ as large as possible; see also Section 2.3.2 and Algorithm 1. Here,

$$\gamma \geq 2, \qquad b \in [0, 1/4] \tag{2.22}$$

in (2.20b) are *momentum tuning constants*. We will denote $\theta^{(i)}$ as $\theta^{(i)}_{\gamma,b}$ when we wish to emphasize its dependence on γ and b. We declare convergence when

$$\sqrt{\delta^{(i)}} \leq \epsilon \|\boldsymbol{x}^{(i)}\|_2 \tag{2.23a}$$

where $\epsilon > 0$ is the convergence threshold and $\delta^{(i)}$ is the local variation of signal iterates:

$$\delta^{(i)} \triangleq \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i-1)}\|_2^2. \tag{2.23b}$$

We need $B^{(i)}$ in (2.20a) to derive the theoretical guarantee for the convergence speed of the PNPG iteration and its sequence convergence. A similar idea for handling the increasing step size in its TFOCS framework is seen in [BCG11]. However, [BCG11] does not address this modification in detail or establish convergence properties of the corresponding method.

The acceleration step (2.20d) extrapolates the two latest iteration points in the direction of their difference $\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i-2)}$, followed by the projection onto the convex set C, which has also been proposed in our preliminary work [GD15c] and in the variable-metric/scaling method [BPR16]. For nonnegative C in (2.2), this projection has closed form; see (2.6c). If C is an intersection of convex sets with a simple individual projection operator for each, we can apply projections onto convex sets (POCS) [YW82].

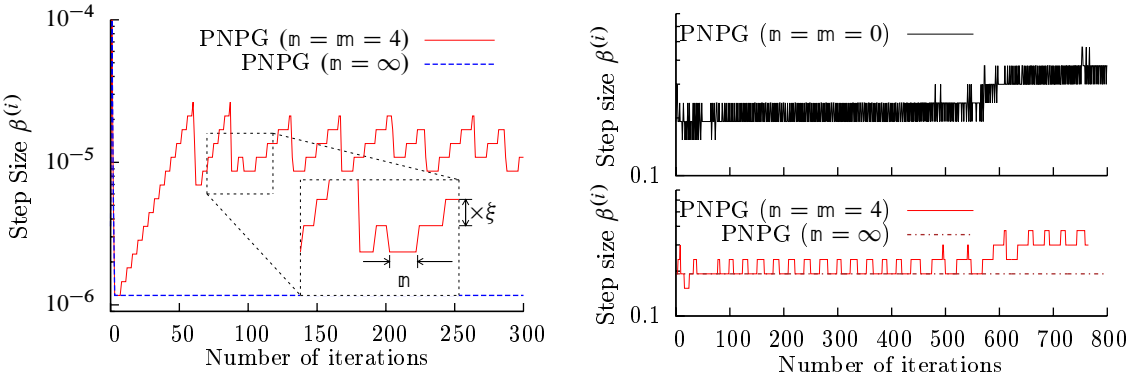
For $\rho(\boldsymbol{x})$ in (2.4), we compute the proximal mapping (2.20e) using the dual formulation in (2.10) and a simpler version of PNPG, Nesterov's projected-gradient algorithm, because the proximal step in this case reduces to projection onto H in (2.10c); for the TV penalty, this method is similar to the TV denoising scheme in [BT09b]. Because of its iterative nature, (2.20e) is *inexact*; this inexactness can be modeled as

$$\boldsymbol{x}^{(i)} \approx_{\varepsilon^{(i)}} \operatorname{prox}_{\beta^{(i)}\boldsymbol{ur}}\left(\bar{\boldsymbol{x}}^{(i)} - \beta^{(i)}\nabla\mathcal{L}(\bar{\boldsymbol{x}}^{(i)})\right) \tag{2.24}$$

where $\varepsilon^{(i)}$ quantifies the precision of the PG step in Iteration i.

If we remove the convex-set constraint by setting $C = \mathbb{R}^p$, iteration (2.20a)–(2.20e) reduces to the Nesterov's proximal-gradient iteration with adaptive step size that imposes signal sparsity *only* in the analysis form (termed NPG_S); see Section 2.5.2 for an illustrative comparison between NPG_S and PNPG.

We now extend [BT09a, Lemma 2.3] to the inexact proximal operation:



(a) Poisson model with identity link. (b) Gaussian linear model.

Figure 2.1: Step sizes $\beta^{(i)}$ as functions of the number of iterations for Poisson and Gaussian linear models.

Lemma 2.1. *Assume convex and differentiable NLL* $\mathcal{L}(\boldsymbol{x})$ *and convex* $\rho(\boldsymbol{x})$*, and consider an inexact PG step* (2.24) *with step size* $\beta^{(i)}$ *that satisfies the majorization condition* (2.21)*. Then,*

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^{(i)}) \geq \frac{1}{2\beta^{(i)}}\left[\|\boldsymbol{x}^{(i)} - \boldsymbol{x}\|_2^2 - \|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{x}\|_2^2 - (\varepsilon^{(i)})^2\right] \tag{2.25}$$

for all $i \geq 1$ *and any* $\boldsymbol{x} \in \mathbb{R}^p$.

Proof: See Appendix 2.A. □

Lemma 2.1 is general and algorithm independent, because $\bar{\boldsymbol{x}}^{(i)}$ can be any value in $\operatorname{dom}\mathcal{L}$ and we have used only the fact that step size $\beta^{(i)}$ satisfies the majorization condition (2.21), rather than depending on specific details of the step-size selection. We use this result to establish the monotonicity property in Proposition 2.3 and to derive and analyze our accelerated PG scheme.

2.3.1 Restart and Monotonicity

If $f(\bar{\boldsymbol{x}}^{(i)}) > f(\boldsymbol{x}^{(i)}) > f(\boldsymbol{x}^{(i-1)})$ or $\bar{\boldsymbol{x}}^{(i)} \in C \setminus \operatorname{dom}\mathcal{L}$, set

$$\theta^{(i-1)} = 1 \qquad \text{(restart)}, \tag{2.26}$$

re-evaluate (2.20b)–(2.20e), and refer to this action as *function restart* [OC15] or *domain restart*, respectively; see Algorithm 1. Function and domain restarts ensure that the PNPG iteration is monotonic and $\bar{\boldsymbol{x}}^{(i)}$ remains within $\mathrm{dom}\, f$ as long as the projected initial value is within $\mathrm{dom}\, f$: $f(P_C(\boldsymbol{x}^{(0)})) < +\infty$. In this chapter, we employ PNPG iteration *with* restart, unless specified otherwise (e.g., in Theorems 2.1 and 2.2 in Section 2.4).

Proposition 2.3 (Monotonicity)**.** *The inexact PG step* (2.24) *is monotonic:*

$$f(\boldsymbol{x}^{(i)}) \leq f(\bar{\boldsymbol{x}}^{(i)}) \tag{2.27a}$$

if it is sufficiently accurate such that

$$\varepsilon^{(i)} \leq \|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^{(i)}\|_2. \tag{2.27b}$$

Hence, the PNPG iteration with restart and inexact PG steps (2.24) *is non-increasing:*

$$f(\boldsymbol{x}^{(i)}) \leq f(\boldsymbol{x}^{(i-1)}) \tag{2.28}$$

if (2.27b) *holds for all* i.

Proof: (2.27a) is straightforward by plugging $\boldsymbol{x} = \bar{\boldsymbol{x}}^{(i)}$ and (2.27b) into (2.25).

If there is no restart in Iteration i, the objective function has not increased. If there is a restart, $\theta^{(i-1)} = 1$, (2.20d) simplifies to $\bar{\boldsymbol{x}}^{(i)} = P_C(\boldsymbol{x}^{(i-1)}) = \boldsymbol{x}^{(i-1)}$, and monotonicity follows due to $\bar{\boldsymbol{x}}^{(i)} = \boldsymbol{x}^{(i-1)}$. □

To establish the monotonicity in Proposition 2.3, the step size $\beta^{(i)}$ need satisfy only the majorization condition (2.21).

2.3.2 Adaptive Step Size

Define the *step-size adaptation parameter*

$$\xi \in (0, 1). \tag{2.29}$$

We propose the following adaptive scheme for selecting $\beta^{(i)}$:

i) • if there have been no step-size backtracking events or increase attempts for $\mathtt{m}$ consecutive iterations ($i - \mathtt{m}$ to $i - 1$), start with a larger step size:

$$\bar{\beta}^{(i)} = \beta^{(i-1)}/\xi \qquad \text{(increase attempt)}; \tag{2.30a}$$

• otherwise start with

$$\bar{\beta}^{(i)} = \beta^{(i-1)}; \tag{2.30b}$$

ii) (backtracking search) select

$$\beta^{(i)} = \xi^{t_i} \bar{\beta}^{(i)} \tag{2.30c}$$

where $0 \leq t_i \leq t_{\text{MAX}}$ is the smallest integer such that (2.30c) satisfies (2.21); *backtracking event* corresponds to $t_i > 0$.

iii) if $\max(\beta^{(i)}, \beta^{(i-1)}) < \bar{\beta}^{(i)}$, increase $\mathtt{m}$ by a nonnegative integer $\mathtt{m}$:

$$\mathtt{m} \leftarrow \mathtt{m} + \mathtt{m}. \tag{2.30d}$$

We select the initial step size $\bar{\beta}^{(1)}$ using the BB method [BB88]. If there has been an attempt to change the step size in any of the previous $\mathtt{m}$ consecutive iterations, we start the backtracking

search ii) with the step size from the latest completed iteration. Consequently, $\beta^{(i)}$ will be approximately piecewise constant as a function of the iteration index i; see Fig. 2.1, which shows the evolutions of $\beta^{(i)}$ for measurements following the Poisson generalized linear and Gaussian linear models corresponding to Figs. 2.4a and 2.6b in Sections 2.5.1 and 2.5.2. To reduce sensitivity to the choice of the tuning constant $\mathtt{m}$, we increase its value by $\mathtt{m}$ if there is a failed attempt to increase the step size in Iteration i; i.e., $\overline{\beta}^{(i)} > \beta^{(i-1)}$ and $\beta^{(i)} < \overline{\beta}^{(i)}$.

The adaptive step-size strategy keeps $\beta^{(i)}$ as large as possible subject to (2.21), which is important not only because the signal iterate may reach regions of $\mathcal{L}(\boldsymbol{x})$ with different local Lipschitz constants, but also because of the varying curvature of $\mathcal{L}(\boldsymbol{x})$ in different updating directions. For example, a (backtracking-only) PG-type algorithm with non-adaptive step size would fail or converge very slowly if the local Lipschitz constant of $\nabla\mathcal{L}(\boldsymbol{x})$ decreases as the algorithm iterates, because the step size will not adjust and track this decrease; see also Section 2.5, which demonstrates the benefits of step-size adaptation.

Setting $\mathtt{m} = +\infty$ corresponds to step-size backtracking only. A step-size adaptation scheme with $\mathtt{m} = \mathtt{m} = 0$ initializes the step-size search aggressively, with an increase attempt (2.30a) in each iteration.

2.3.3 Inner-Iteration Warm Start and Convergence Criteria

For $\rho(\boldsymbol{x})$ in (2.4), the inner iteration solves (2.20e) using the dual problem (2.10b). Denote by $\boldsymbol{p}^{(i,j)}$ the iterates of the dual variable $\boldsymbol{p}$ in the jth inner iteration step within Iteration i; this inner iteration solves (2.20e) using (2.10b). The initial $\boldsymbol{p}^{(i,0)}$ is the latest $\boldsymbol{p}$ from Iteration $i-1$, which is referred to in [VSBV13] as the *warm restart*. (The variable metric inexact line-search algorithm (VMILA) [BLPP16] also uses warm restart.)

We consider two convergence criteria. The first tracks local variation of the signal iterates (2.23b):

$$\|\boldsymbol{x}^{(i,j)} - \boldsymbol{x}^{(i,j-1)}\| \leq \eta\sqrt{\delta^{(i-1)}} \tag{2.31a}$$

where η is a tuning constant.

The second *duality-gap–based* criterion relies on the result in Proposition 2.2 to guarantee that $(\theta^{(k)}\varepsilon^{(k)})^2$ decreases at a rate of $\mathcal{O}(k^{-q})$ within each iteration segment without restart; this guarantee allows us to control the decrease of the convergence-rate upper bound in Section 2.4. Denote by ι_i the iteration index of the latest restart prior to (and excluding) Iteration i ($i \geq 1$); set its initial value $\iota_1 = 0$. We select the duality-gap based inner-iteration convergence criterion as (see also (2.10e) and (2.11))

$$\frac{G^{(i,j)}}{\beta^{(i)}u\,\rho\big(\widehat{\boldsymbol{x}}^{(i,j)}\big)} \leq \frac{\eta}{(i-\iota_i)^q(\theta^{(i)})^2} \tag{2.31b}$$

where η is a tuning constant and q is the *accuracy rate* [VSBV13]. Here, $G^{(i,j)}$ and $\widehat{\boldsymbol{x}}^{(i,j)}$ are the duality gap $G_{\boldsymbol{a},\lambda}(\boldsymbol{p}^{(i,j)})$ and $\widehat{\boldsymbol{x}}_{\boldsymbol{a},\lambda}(\boldsymbol{p}^{(i,j)})$ in (2.11) and (2.10e) (respectively) for $\boldsymbol{a} = \overline{\boldsymbol{x}}^{(i)} - \beta^{(i)}\nabla\mathcal{L}(\overline{\boldsymbol{x}}^{(i)})$ and $\lambda = \beta^{(i)}u$. Without restart (i.e., $\iota_i \equiv 0$) and step-size adaptation, (2.31b) reduces to the inner-iteration convergence criterion in [VSBV13, Sec. 6.1].

Adjusting η. We use η in (2.31a)–(2.31b) to trade off accuracy and speed of the inner iterations. If $f(\boldsymbol{x}^{(i)}) > \max\big(f(\boldsymbol{x}^{(i-1)}), f(\overline{\boldsymbol{x}}^{(i)})\big)$ indicating that the monotonicity condition (2.27b) does not hold, we decrease η by an order of magnitude (10 times) and re-evaluate (2.20a)–(2.20e).

2.4 Convergence Analysis

We now bound the convergence rate of the PNPG method without restart.

Theorem 2.1 (Convergence of the Objective Function)**.** *Assume that the NLL $\mathcal{L}(\boldsymbol{x})$ is convex and differentiable, $\rho(\boldsymbol{x})$ is convex, the closed convex set $\boldsymbol{C}$ satisfies*

$$\boldsymbol{C} \subseteq \operatorname{dom}\mathcal{L} \tag{2.32}$$

(implying no need for domain restart), and the momentum tuning constants are within the range (2.22)*. Consider the PNPG iteration without restart with the inexact PG step* (2.24) *in place of*

(2.20e). *The convergence rate of the PNPG iteration is bounded as follows: for* $k \geq 1$,

$$\Delta^{(k)} \leq \frac{\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star\|_2^2 + \mathcal{E}^{(k)}}{2\beta^{(k)}(\theta^{(k)})^2} \tag{2.33a}$$

$$\leq \gamma^2 \frac{\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star\|_2^2 + \mathcal{E}^{(k)}}{2\left(\sqrt{\beta^{(1)}} + \sum_{i=1}^k \sqrt{\beta^{(i)}}\right)^2} \tag{2.33b}$$

where $\boldsymbol{x}^\star$ *is a minimum point of* $f(\boldsymbol{x})$ *and*

$$\Delta^{(k)} \triangleq f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^\star) \tag{2.34a}$$

$$\mathcal{E}^{(k)} \triangleq \sum_{i=1}^k (\theta^{(i)}\varepsilon^{(i)})^2 \tag{2.34b}$$

are the centered objective function and the cumulative error term, which accounts for the inexact PG steps, respectively.

Proof: See Appendix 2.A for the proof of (2.33a); then, to obtain (2.33b), use

$$\theta^{(k)}\sqrt{\beta^{(k)}} \geq \frac{1}{\gamma}\sqrt{\beta^{(k)}} + \theta^{(k-1)}\sqrt{\beta^{(k-1)}} \tag{2.35a}$$

$$\geq \frac{1}{\gamma}\sum_{i=2}^k \sqrt{\beta^{(i)}} + \theta^{(1)}\sqrt{\beta^{(1)}} \tag{2.35b}$$

for all $k > 1$, where (2.35a) follows from the definitions of $B^{(k)}$ and $\theta^{(k)}$ in (2.20a) and (2.20b), and (2.35b) follows by repeated application of the inequality (2.35a) with k replaced by $k-1, k-2, \ldots, 2$. □

Theorem 2.1 shows that better initialization, smaller proximal-mapping approximation error, and larger step sizes $(\beta^{(i)})_{i=1}^k$ help lower the convergence-rate upper bounds in (2.33). This result motivates our step-size adaptation with the goal of maintaining large $(\beta^{(i)})_{i=1}^k$; see Section 2.3.2. To derive this theorem, we have used only the fact that the step size $\beta^{(i)}$ satisfies the majorization condition (2.21), rather than taking advantage of specific details of the step-size selection.

To minimize the upper bound in (2.33a), we can select $\theta^{(i)}$ to satisfy (2.63b) with equality, which corresponds to $\theta_{2,1/4}^{(i)}$ in (2.20b), on the boundary of the feasible region in (2.22). By (2.35a),

$\sqrt{\beta^{(k)}}\theta^{(k)}$ and the denominator of the bound in (2.33a) are strictly increasing sequences. The upper bound in (2.33b) is not a function of b and is minimized with respect to γ for $\gamma = 2$, given the fixed step sizes $(\beta^{(i)})_{i=0}^{+\infty}$.

Corollary 2.1. *Under the assumptions of Theorem 2.1, the convergence of PNPG iteration* $\boldsymbol{x}^{(k)}$ *without restart is bounded as follows:*

$$\Delta^{(k)} \leq \gamma^2 \frac{\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star\|_2^2 + \mathcal{E}^{(k)}}{2(k+1)^2 \beta_{\min}} \tag{2.36a}$$

for $k \geq 1$, *provided that*

$$\beta_{\min} \triangleq \min_{k=1}^{+\infty} \beta^{(k)} > 0. \tag{2.36b}$$

Proof: Use (2.33b) and the fact that $\sqrt{\beta^{(1)}} + \sum_{i=1}^{k} \sqrt{\beta^{(i)}} \geq (k+1)\sqrt{\beta_{\min}}$. □

The assumption (2.36b) is implied by, and weaker than, the Lipschitz continuity of $\nabla \mathcal{L}(\boldsymbol{x})$; indeed, $\beta_{\min} > \xi/L$ if $\nabla \mathcal{L}(\boldsymbol{x})$ has a Lipschitz constant L; see also (2.29).

According to Corollary 2.1, the PNPG iteration attains the $\mathcal{O}(k^{-2})$ convergence rate as long as the step size $\beta^{(i)}$ is bounded away from zero (see (2.36b)) and the cumulative error term (2.34b) converges:

$$\mathcal{E}^{(+\infty)} \triangleq \lim_{k \to +\infty} \mathcal{E}^{(k)} < +\infty \tag{2.37}$$

which requires that $(\theta^{(k)}\varepsilon^{(k)})^2$ decreases at a rate of $\mathcal{O}(k^{-q})$ with $q > 1$. This condition, also key for establishing convergence of iterates in Theorem 2.2, motivates us to use decreasing convergence criteria (2.31a)–(2.31b) for the inner proximal-mapping iterations, where (2.31b) guarantees (2.37) upon choosing an appropriate q.

We now contrast our result in Theorem 2.1 with existing work on accommodating inexact proximal mappings in PG schemes. By recursively generating a function sequence that approximates the objective function, [VSBV13] gives an asymptotic analysis of the effect of $\varepsilon^{(i)}$ on the conver-

gence rate of accelerated PG methods with inexact proximal mapping. However, no explicit upper bound is provided for $\Delta^{(k)}$. Schmidt, Roux, and Bach [SRB11] provide convergence-rate analysis and an upper bound on $\Delta^{(k)}$, but their analysis does not apply here because it relies on fixed step-size assumption, uses different form of acceleration [SRB11, Prop. 2], and has no convex-set constraint. Bonettini, Loris, Porta, and Prato [BLPP16] analyze the inexactness of proximal mapping but for proximal variable-metric/scaling methods with $\mathcal{O}(k^{-1})$ convergence rate for the objective function.

We now establish convergence of the PNPG iterates.

Theorem 2.2 (Convergence of Iterates)**.** *Assume that*

1) *the conditions of Theorem 2.1 hold;*

2) $\mathcal{E}^{(+\infty)}$ *exists:* (2.37) *holds;*

3) *the momentum tuning constants* (γ, b) *satisfy*

$$\gamma > 2, \qquad b \in [0, 1/\gamma^2]; \tag{2.38}$$

4) *the step-size sequence* $(\beta^{(i)})_{i=1}^{+\infty}$ *is bounded within the range* $[\beta_{\min}, \beta_{\max}]$*, for* $\beta_{\min} > 0$*.*

Consider the PNPG iteration without restart with the inexact PG step (2.24) *in place of* (2.20e)*. Then, the sequence of PNPG iterates* $\boldsymbol{x}^{(i)}$ *converges to a minimizer of* $f(\boldsymbol{x})$*.*

Proof: See Appendix 2.B. □

Observe that Assumption 3 requires a narrower range of (γ, b) than (2.22): indeed (2.38) is a strict subset of (2.22). The intuition is to leave a sufficient gap between the two sides of (2.63a) so that their difference becomes a quantity that is roughly proportional to the growth of $\theta^{(i)}$, which is important for proving the convergence of signal iterates [CD15]. Although the momentum term (2.20b) with $\gamma = 2$ is optimal in terms of minimizing the upper bound on the convergence rate (see Theorem 2.1), it appears difficult or impossible to prove convergence of the signal iterates $\boldsymbol{x}^{(i)}$ for

this choice of γ because, in this case, the gap between the two sides of (2.63a) is upper-bounded by a constant.

Iterate convergence results in [BPR16; AD15; CD15] apply to momentum-accelerated methods that require non-increasing step-size sequences and do not adjust to the local curvature of the NLL. Aujol and Dossal [AD15] analyze both the convergence of the objective function and the iterates with inexact proximal operator for $B^{(1)} = 1$ and $\mathtt{m} = \infty$, i.e., with decreasing step size only, and for a different (less aggressive) $\theta^{(i)}$ than ours in (2.20b). Bonettini, Porta, and Ruggiero use the ideas from [CD15] to establish convergence of iterates for their variable-metric/scaling approach in [BPR16], but this analysis does not account for inexact proximal steps.

2.4.1 $\mathcal{O}(k^{-2})$ Convergence Acceleration Approaches

A few variants that accelerate the PG method achieve the $\mathcal{O}(k^{-2})$ convergence rate [BCG11, Sec. 5.2]. One competitor proposed by Auslender and Teboulle in [AT06, Sec. 5] and restated in [BCG11] where it was referred to as AT, replaces (2.20d)–(2.20e) with

$$\bar{\boldsymbol{x}}^{(i)} = \Big(1 - \frac{1}{\theta^{(i)}}\Big)\boldsymbol{x}^{(i-1)} + \frac{1}{\theta^{(i)}}\tilde{\boldsymbol{x}}^{(i-1)} \tag{2.39a}$$

$$\tilde{\boldsymbol{x}}^{(i)} = \operatorname{prox}_{\theta^{(i)}\beta^{(i)}\boldsymbol{ur}}\big(\tilde{\boldsymbol{x}}^{(i-1)} - \theta^{(i)}\beta^{(i)}\nabla\mathcal{L}(\bar{\boldsymbol{x}}^{(i)})\big) \tag{2.39b}$$

$$\boldsymbol{x}^{(i)} = \Big(1 - \frac{1}{\theta^{(i)}}\Big)\boldsymbol{x}^{(i-1)} + \frac{1}{\theta^{(i)}}\tilde{\boldsymbol{x}}^{(i)} \tag{2.39c}$$

where $\theta^{(i)} = \theta^{(i)}_{2,1/4}$ in (2.20b). Here, $\beta^{(i)}$ in the TFOCS implementation [BCG11] is selected using the aggressive search with $\mathtt{m} = \mathtt{m} = 0$.

All intermediate signals in (2.39a)–(2.39c) belong to $\boldsymbol{C}$ and do not require projections onto $\boldsymbol{C}$. However, as $\theta^{(i)}$ increases with i, step (2.39b) becomes unstable, especially when an iterative solver is needed for its proximal operation. To stabilize its convergence, AT relies on periodic restart by resetting $\theta^{(i)}$ using (2.26) [BCG11]. However, the period of restart is a tuning parameter that is not easy to select. For a linear Gaussian model, this period varies according to the condition

number of the sensing matrix Φ [BCG11], which is generally unavailable and not easy to compute for large-scale problems. For other models, there are no guidelines how to select the restart period.

In Section 2.5, we show that AT converges slowly compared with PNPG, which justifies the use of projection onto C in (2.20d) and (2.20d)–(2.20e) instead of (2.39a)–(2.39c). PNPG usually runs uninterrupted (without restart) over long stretches and benefits from Nesterov's acceleration within these stretches, which may explain its better convergence properties compared with AT. PNPG may also be less sensitive than AT to proximal-step inaccuracies; we have established convergence-rate bounds for PNPG that account for inexact proximal steps (see (2.33) and (2.36a)), whereas AT does not yet have such bounds, to the best of our knowledge.

2.4.1.1 Relationship with FISTA

The PNPG method is a generalized FISTA [BT09a] that accommodates convex constraints, more general NLLs,[2] and (increasing) adaptive step size. In contrast with PNPG, FISTA has a non-increasing step size $\beta^{(i)}$, which allows for setting $B^{(i)} = 1$ in (2.20b) for all i (see Appendix 2.A.2); upon setting $(\gamma, b) = (2, 1/4)$, this choice yields the standard FISTA (and Nesterov's [Nes83]) update. Convergence of signal iterates has not been established for FISTA with $(\gamma, b) = (2, 1/4)$ [DDD16]. Theorem 2.2 comes close to this goal because it establishes convergence of iterates of PNPG and corresponding FISTA for (γ, b) *arbitrarily close* to $(2, 1/4)$.

The method in [BPR16] is a variable-metric/scaling version of FISTA with projection of the extrapolation step to account for the convex constraints. [BN16] analyzes a version of FISTA where the (decreasing) step size is adjusted using a condition in [Tse00] different from the majorization condition (2.21), and establishes objective-function convergence under the assumption that the step size is lower bounded. As FISTA, [BPR16] and [BN16] *do not* adapt the step size and hence do not adjust to the local curvature of the NLL.

[2] FISTA has been developed for the linear Gaussian model in Section 2.2.2.

2.5 Numerical Examples

We now evaluate our proposed algorithm by means of numerical simulations. We adopt the nonnegative $\boldsymbol{C}$ in (2.2) and the ℓ_1-norm sparsifying penalty in (2.4). The PNPG iterations with the local-variation and duality-gap inner convergence criteria (2.31a) and (2.31b) are labeled PNPG and PNPG_d, respectively.

All iterative methods that we compare use the convergence criterion (2.23a) with

$$\epsilon = 10^{-9} \tag{2.40}$$

and have the maximum number of iterations $I_{\max} = 10^4$. In the presented examples, PNPG uses momentum tuning constants $(\gamma, b) = (2, 1/4)$ and adaptive step-size parameters $\mathbb{n} = 4$ (unless specified otherwise), $\mathbb{m} = \mathbb{n}$, $\xi = 0.8$, inner-iteration convergence constants $\eta = 10^{-2}$ and $(\eta, q) = (1, 1.0001)$ for PNPG and PNPG_d (respectively), and maximum number of inner iterations $J_{\max} = 1000$. Here, PNPG_d uses $q = 1.0001$ with goal to guarantee (2.37).

We apply the AT method (2.39) implemented in the TFOCS package [BCG11] with adaptive step size and a periodic restart every 200 iterations (tuned for its best performance) and our proximal mapping. Our inner convergence criteria (2.31b) cannot be implemented in the TFOCS package (i.e., it require editing its code). Hence, we select the proximal mapping that has a relative-error inner convergence criterion

$$\|\boldsymbol{p}^{(i,j)} - \boldsymbol{p}^{(i,j-1)}\|_2 \leq \epsilon' \|\boldsymbol{p}^{(i,j)}\|_2, \tag{2.41a}$$

where $\boldsymbol{p}^{(i,j)}$ is the dual variable employed by the inner iterations. This relative-error inner convergence criterion is easy to incorporate into the TFOCS software package [BCG11] and is already used by the SPIRAL package; see [Har]. Here, we select

$$\epsilon' = 10^{-6} \tag{2.41b}$$

for both AT and SPIRAL and set their maximum number of inner iterations to 100.

We apply the CP approach based on [AABM12, Sec. 7.5]:

$$\boldsymbol{z} \leftarrow \mathrm{prox}_{\sigma_1 F^*}\left(\boldsymbol{z} + \sigma_1 \Phi \bar{\boldsymbol{x}}\right) \tag{2.42a}$$

$$\boldsymbol{p} \leftarrow P_{uH}(\boldsymbol{p} + \sigma_2 \Psi^H \bar{\boldsymbol{x}}) \tag{2.42b}$$

$$\bar{\boldsymbol{x}} \leftarrow 2P_C\left(\boldsymbol{x} - \tau\left(\Phi^T \boldsymbol{z} + \mathrm{Re}(\Psi \boldsymbol{p})\right)\right) - \boldsymbol{x} \tag{2.42c}$$

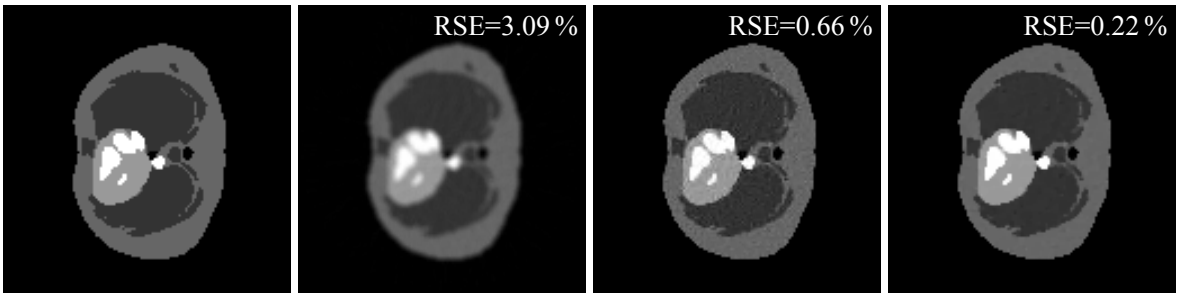
$$\boldsymbol{x} \leftarrow (\bar{\boldsymbol{x}} + \boldsymbol{x})/2 \tag{2.42d}$$

obtained by splitting the objective function (2.3a) into the sum of $F(\Phi \boldsymbol{x}) + u\|\Psi^H \boldsymbol{x}\|_1$ and $I_C(\boldsymbol{x})$, where the first summand is a convex lower semicontinuous function of $[\Phi^H \ \Psi]^H \boldsymbol{x}$ and $F(\Phi \boldsymbol{x}) = \mathcal{L}(\boldsymbol{x})$. In our examples in Sections 2.5.1 and 2.5.2, $F(\boldsymbol{y})$ and its convex conjugate $F^*(\boldsymbol{z})$ have analytical proximal operators [AABM12, Sec. 7.4]; hence, CP *does not* require an inner iteration. In the original CP algorithm in [CP11a], Chambolle and Pock select $\sigma_1 = \sigma_2 = \sigma$. However, when the difference between $\|\Phi\|_2$ and $\|\Psi\|_2$ becomes large, it is hard to find tuning constants of the form $(\sigma_1, \sigma_2, \tau) = (\sigma, \sigma, \tau)$ that ensure fast convergence of the CP algorithm, which is why we do not impose $\sigma_1 = \sigma_2 = \sigma$ here. Another version of CP can be obtained by associating the regularization parameter u with Ψ instead of the ℓ_1-norm function, which leads to replacing uH with H and Ψ, Ψ^H with $u\Psi, u\Psi^H$, respectively, in (2.42). In this chapter, we adopt the version of CP in (2.42).

All the numerical examples were performed on a Linux workstation with an Intel Xeon CPU E31245 (3.30 GHz) and 8 GB memory. The operating system is Ubuntu 14.04 LTS (64-bit). The Matlab implementation of the proposed algorithms and numerical examples is available [Gu].

2.5.1 PET Image Reconstruction from Poisson Measurements

In this example, we adopt the Poisson GLM (2.16a) with identity link in (2.17). Consider PET reconstruction of the 128×128 concentration map $\boldsymbol{x}$ in Fig. 2.2a, which represents simulated radiotracer activity in the human chest. Assume that the corresponding 128×128 attenuation map $\boldsymbol{\kappa}$ is known, which is needed to model the attenuation of the gamma rays [OF97] and compute the



(a) radio-isotope concentration (b) FBP (c) DWT (d) TV

Figure 2.2: (a) True emission image and (b)–(d) the reconstructions of the emission concentration map.

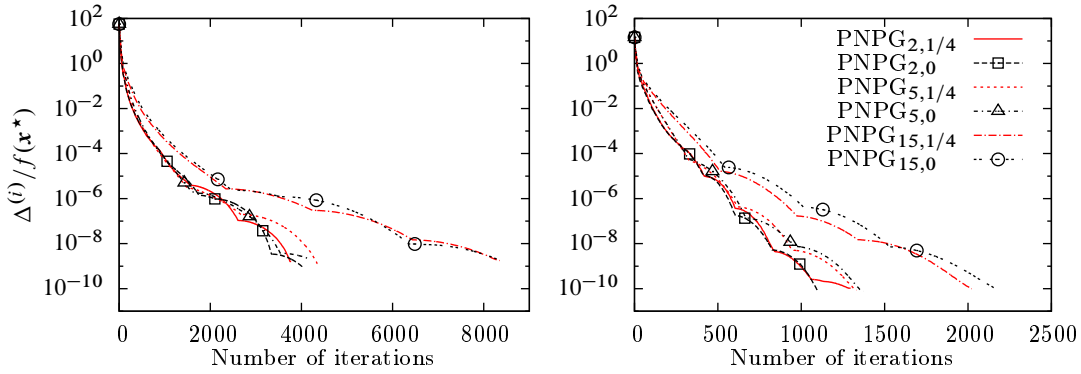
sensing matrix Φ in this application. We collect the photons from 90 equally spaced directions over 180°, with 128 radial samples in each direction. Here, we adopt the parallel strip-integral matrix Γ [Fes09, Ch. 25.2] and use its implementation in the Image Reconstruction Toolbox (IRT) [Fes16] with sensing matrix

$$\Phi = w \operatorname{diag}\big(\exp_{\circ}(-\Gamma \boldsymbol{\kappa} + \boldsymbol{c})\big)\Gamma \tag{2.43}$$

where $\boldsymbol{c}$ is a known vector generated using a zero-mean independent, identically distributed (i.i.d.) Gaussian sequence with variance 0.3 to model the detector efficiency variation; $w > 0$ is a known scaling constant controlling the expected total number of detected photons due to electron-positron annihilation; and $\mathbf{1}^T \operatorname{E}(\boldsymbol{y} - \boldsymbol{b}) = \mathbf{1}^T \Phi \boldsymbol{x}$, which is a signal-to-noise ratio (SNR) measure. Assume that the background radiation, scattering effect, and accidental coincidence combined lead to a known (generally nonzero) intercept term $\boldsymbol{b}$ in the Poisson GLM (2.17). The elements of the intercept term have been set to a constant equal to 10 % of the sample mean of $\Phi \boldsymbol{x}_{\text{true}}$: $\boldsymbol{b} = (\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}})/(10N)\mathbf{1}$.

The above model, choices of parameters in the PET system setup, and concentration map have been adopted from IRT [Fes16, `emission/em_test_setup.m`].

Here, we consider the DWT and isotropic TV sparsifying transforms. We use the 2-D Haar DWT with 6 decomposition levels and a full circular mask [DGQ11] to construct a sparsifying



(a) DWT, $\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}} = 10^8$, $a = 0.5$ (b) TV, $\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}} = 10^7$, $a = 0$

Figure 2.3: Normalized centered objectives as functions of the number of iterations for (a) DWT and (b) TV regularizations.

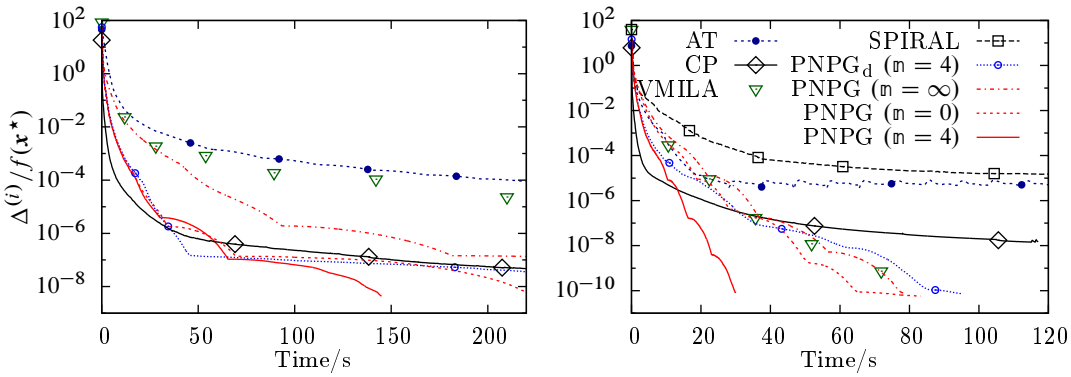
dictionary matrix $\Psi \in \mathbb{R}^{12\,449 \times 14\,056}$ with orthonormal rows, i.e., $\Psi\Psi^T = I$, which allows efficient inner iteration. We compare the filtered backprojection (FBP) [OF97] and PG methods that aim at minimizing (2.3) with nonnegative $\boldsymbol{C}$ in (2.2) and DWT and TV sparsifying transforms.

We implemented SPIRAL with TV penalty using the centered NLL term (2.16a), which improves the numerical stability compared with the original code in [Har]. We do not compare with SPIRAL that uses DWT penalty because its inner iteration for the proximal step requires a square orthogonal Ψ (see [HMW12, Sec. II-B]), which is not the case here. We also compare with VMILA [BLPP16; BLPP] with both DWT and TV penalties and its default tuning constants, which yield good performance; hence VMILA is insensitive to tuning.

In this example, we adopt the following form of the regularization constant u:

$$u = 10^a, \tag{2.44}$$

vary a in the range $[-6, 3]$ with a grid size of 0.5, and search for the reconstructions with the best average relative square error (RSE) performance; here, $\text{RSE} = \|\hat{\boldsymbol{x}} - \boldsymbol{x}_{\text{true}}\|_2^2 / \|\boldsymbol{x}_{\text{true}}\|_2^2$, where $\boldsymbol{x}_{\text{true}}$ and $\hat{\boldsymbol{x}}$ are the true and reconstructed signals, respectively. All iterative methods were initialized by FBP reconstructions implemented by IRT [Fes16].



(a) DWT, $\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}} = 10^8, a = 0.5$ (b) TV, $\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}} = 10^7, a = 0$

Figure 2.4: Normalized centered objectives as functions of the CPU time for (a) DWT and (b) TV regularizations.

Figs. 2.2b-2.2d show reconstructions and corresponding RSEs for one random realization of the noise and detector variation $\boldsymbol{c}$, with the expected total annihilation photon count (SNR) equal to 10^8; the optimal a is 0.5. All sparse reconstruction methods (PNPG, AT, CP, SPIRAL, and VMILA) perform similarly as long as they employ the same penalty: the TV sparsity penalty is superior to its DWT counterpart; see [GD16b, Fig. 6] which shows average RSEs of different methods as functions of $\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}}$.

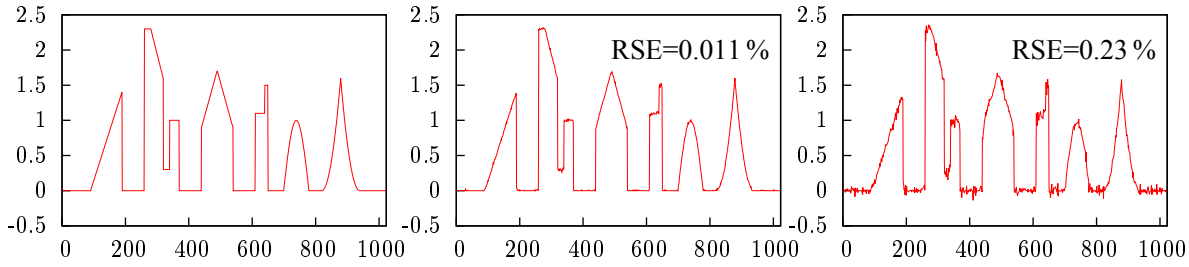
Figs. 2.3 and 2.4 show the normalized centered objectives $\Delta^{(i)}/f(\boldsymbol{x}^\star)$ as functions of the number of iterations and CPU time for the DWT and TV signal sparsity regularizations and two random realizations of the noise and detector variation with different total expected photon counts. The legends in Figs. 2.3b and 2.4b apply to Figs. 2.3a and 2.4a as well. Fig. 2.3 examines the convergence of PNPG as a function of the momentum tuning constants (γ, b) in (2.22), using $\gamma \in \{2, 5, 15\}$ and $b \in \{0, 1/4\}$. For small $\gamma \leq 5$, there is no significant difference between different selections and no choice is uniformly the best, consistent with [CD15] which considers only $b = 0$ and non-adaptive step size. As we increase γ further ($\gamma = 15$), we observe slower convergence. In the remainder of this section, we use $(\gamma, b) = (2, 1/4)$.

To illustrate the benefits of step-size adaptation, we present in Fig. 2.4 the performance of PNPG ($\mathbb{m} = \infty$), which does not adapt to the local curvature of the NLL and has monotonically

non-increasing step sizes $\beta^{(i)}$, similar to FISTA. PNPG ($\mathtt{m} = 4$) outperforms PNPG ($\mathtt{m} = \infty$) because it uses step-size adaptation; see also Fig. 2.1a, which corresponds to Fig. 2.4a and shows that the step size of PNPG ($\mathtt{m} = 4$) is consistently larger than that of PNPG ($\mathtt{m} = \infty$). Initializing PNPG iterations by a vector close to $\mathbf{0}$ (rather than with FBP) will lead to an even larger difference in convergence speed between PNPG ($\mathtt{m} = \infty$) and PNPG ($\mathtt{m} = 4$). The advantage of PNPG ($\mathtt{m} = 4$) over the aggressive PNPG ($\mathtt{m} = 0$) scheme is due to the *patient* nature of its step-size adaptation, which leads to a better local majorization function of the NLL and reduces time spent backtracking. Indeed, if we do not account for the time spent on each iteration and only compare the objectives as functions of the iteration index, then PNPG ($\mathtt{m} = 4$) and PNPG ($\mathtt{m} = 0$) perform similarly; see [GD15c, Fig. 4]. Although PNPG ($\mathtt{m} = 0$) and AT have the same step-size selection strategy and $\mathcal{O}(k^{-2})$ convergence-rate guarantees, PNPG ($\mathtt{m} = 0$) converges faster; both schemes are further outperformed by PNPG ($\mathtt{m} = 4$). Fig. 2.4b shows that SPIRAL, which does not employ PG step acceleration, takes at least three times longer than PNPG ($\mathtt{m} = 4$) to reach the same objective function.

In Fig. 2.4b, AT and SPIRAL reach the performance floor due to their fixed inner convergence criterion in (2.41a); we observe a performance floor for AT in Fig. 2.4a as well. Reducing ϵ' in (2.41b) will lower this floor, at the cost of slowing down the two algorithms. This result justifies our convex-set projection in (2.20d) for the Nesterov's acceleration step, shows the superiority of (2.20d) over AT's acceleration in (2.39a) and (2.39c), and is consistent with the results in Section 2.5.2.

PNPG_d ($\mathtt{m} = 4$) employs the duality-gap–based inner convergence criterion (2.31b) with $q = 1.0001$. Since the goal of using (2.31b) is to guarantee (2.37), this inner criterion is more stringent and leads to slower overall performance of PNPG_d ($\mathtt{m} = 4$) compared to PNPG ($\mathtt{m} = 4$). Indeed, if we do not account for the time spent on each iteration and only compare the objectives as functions of the iteration index, then PNPG_d ($\mathtt{m} = 4$) and PNPG ($\mathtt{m} = 4$) perform similarly with the former slightly better.



(a) true signal (b) PNPG (c) NPG_S

Figure 2.5: Nonnegative skyline signal and its PNPG and NPG_S reconstructions for $N/p = 0.34$.

The CP method uses the following tuning constants carefully selected for this particular problem: $(\sigma_1, \sigma_2, \tau) = (10^{-6}, 1, 10^{-3})$ and $(\sigma_1, \sigma_2, \tau) = (10^{-4}, 1, 10^{-2})$ for the DWT and TV penalties, respectively. CP is sensitive to tuning and a different selection of $(\sigma_1, \sigma_2, \tau)$ can significantly slow down its convergence. Initially, CP converges quickly and then slows down as it approaches the optimum.

Considering its $\mathcal{O}(k^{-1})$ theoretical convergence rate, VMILA performs quite well, thanks to its use of the variable-metric/scaling approach.

2.5.2 Skyline Signal Reconstruction from Linear Measurements

We adopt the DWT sparsifying transform and linear measurement model with Gaussian noise in Section 2.2.2 where each column of the sensing matrix Φ are i.i.d. and drawn from the uniform distribution on unit sphere. Due to the widespread use of this measurement model, we can compare a wider range of methods than in the Poisson PET example in Section 2.5.1.

We have designed a "skyline" signal of length $p = 1024$ by overlapping magnified and shifted triangle, rectangle, sinusoid, and parabola functions; see Fig. 2.5a. We generate the noiseless measurements using $\boldsymbol{y} = \Phi \boldsymbol{x}_{\text{true}}$. The DWT matrix Ψ is constructed using the Daubechies-4 wavelet with three decomposition levels whose approximation by the 5 % largest-magnitude wavelet coefficients achieves $\text{RSE} = 98\,\%$. We compare:

- AT, PNPG, and PNPG_d;

- CP with the parameters $\sigma_2 = \sigma_1$ [CP11a] and $\tau = 1$ with σ_1 tuned separately for best performance in each experiment;[3]
- linearly constrained gradient projection method [HTWM10], part of the SPIRAL toolbox [Har] and labeled SPIRAL herein;
- the GFB method [RFP13] (see (2.4)):

$$\boldsymbol{z}_1 \leftarrow \boldsymbol{z}_1 + \lambda\big[\mathrm{prox}_{(\gamma u/w)\rho}(2\boldsymbol{x} - \boldsymbol{z}_1 - \gamma\nabla\mathcal{L}(\boldsymbol{x})) - \boldsymbol{x}\big] \tag{2.45a}$$

$$\boldsymbol{z}_2 \leftarrow \boldsymbol{z}_2 + \lambda[P_C(2\boldsymbol{x} - \boldsymbol{z}_2 - \gamma\nabla\mathcal{L}(\boldsymbol{x})) - \boldsymbol{x}] \tag{2.45b}$$

$$\boldsymbol{x} \leftarrow w\boldsymbol{z}_1 + (1-w)\boldsymbol{z}_2 \tag{2.45c}$$

  with $\gamma = 1.9/\|\Phi\|_2^2$, $\lambda = 1$, and $w = 0.5$ tuned for best performance; and
- the PDS method [Con13]:

$$\bar{\boldsymbol{z}} \leftarrow P_{[-u,u]^p}(\boldsymbol{z} + \sigma\Psi^T\boldsymbol{x}) \tag{2.46a}$$

$$\bar{\boldsymbol{x}} \leftarrow P_C\big(\boldsymbol{x} - \tau\nabla\mathcal{L}(\boldsymbol{x}) - \tau\Psi(2\bar{\boldsymbol{z}} - \boldsymbol{z})\big) \tag{2.46b}$$

$$\boldsymbol{z} \leftarrow \boldsymbol{z} + \gamma(\bar{\boldsymbol{z}} - \boldsymbol{z}) \tag{2.46c}$$

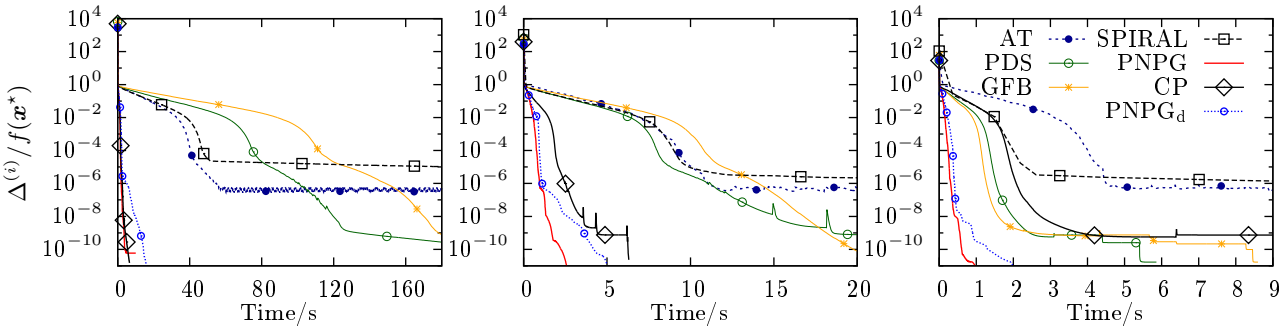
$$\boldsymbol{x} \leftarrow \boldsymbol{x} + \gamma(\bar{\boldsymbol{x}} - \boldsymbol{x}) \tag{2.46d}$$

  where we choose $\tau = 1/(\sigma + \|\Phi\|_2^2/2)$ and $\gamma = 2 - 0.5\|\Phi\|_2^2(\tau^{-1} - \sigma)^{-1}$ with σ tuned for best performance,[4]

all of which aim to solve the generalized analysis BPDN problem with a convex signal constraint. Here, $p' = p$, $\Psi \in \mathbb{R}^{p\times p}$ is an orthogonal matrix ($\Psi\Psi^T = \Psi^T\Psi = I$), and $\mathrm{prox}_{\lambda\rho}\,\boldsymbol{a} = \Psi\,\mathcal{T}_\lambda(\Psi^T\boldsymbol{a})$ has a closed-form solution (see (2.6c)), which simplifies the implementation of the GFB method ((2.45a), in particular); see the discussion in Section 2.1. The other tuning options for SPIRAL, and AT are kept to their default values, unless specified otherwise.

[3] We select $\sigma_1 = \sigma_2$ as in [CP11a] because $\|\Phi\|_2$ and $\|\Psi\|_2$ have approximately the same scale in this example.

[4] These choices of τ and γ are at the boundary of the convergence region in [Con13, Th. 3.1]. We have searched for τ and γ inside this convergence region as well, but found that the boundary choices that we select are the best, or close to the best.



(a) $a = -5$ (b) $a = -4$ (c) $a = -3$

Figure 2.6: Normalized centered objectives as functions of CPU time for normalized numbers of measurements $N/p = 0.34$ and different regularization constants a.

We initialize the iterative methods by the approximate minimum-norm estimate: $\boldsymbol{x}^{(0)} = \Phi^T[\mathrm{E}(\Phi\Phi^T)]^{-1}\boldsymbol{y} = N\Phi^T\boldsymbol{y}/p$ and select the regularization parameter u as

$$u = 10^a U, \qquad U \triangleq \|\Psi^T \nabla\mathcal{L}(\mathbf{0})\|_\infty \tag{2.47}$$

where a is an integer selected from the interval $[-9, -1]$ and U is an upper bound on u of interest. Indeed, the minimum point $\boldsymbol{x}^\star$ reduces to $\mathbf{0}$ if $u \geq U$ [GD15d, Sec. II-D].

As before, PNPG ($\mathtt{m} = 4$) and PNPG ($\mathtt{m} = 0$) converge at similar rates as functions of the number of iterations. However, due to the excessive attempts to increase the step size at every iteration, PNPG ($\mathtt{m} = 0$) spends more time backtracking and converges at a slower rate as a function of CPU time compared with PNPG ($\mathtt{m} = 4$); see also Fig. 2.1b which corresponds to Fig. 2.6b and shows the step sizes as functions of the number of iterations for $a = -4$ and $N/p = 0.34$. Hence, we present only the performances of PNPG with $\mathtt{m} = 4$ in this section.

Fig. 2.5 shows the advantage brought by the convex-set nonnegativity signal constraints (2.2). Figs. 2.5b and 2.5c present the PNPG ($a = -5$) and NPG$_\mathrm{S}$ ($a = -4$) reconstructions from one realization of the linear measurements with $N/p = 0.34$ and a tuned for the best RSE performance. Recall that NPG$_\mathrm{S}$ imposes signal sparsity only. Here, imposing signal nonnegativity significantly improves the overall reconstruction and *does not* simply rectify the signal values close to zero.

Fig. 2.6 presents the normalized centered objectives $\Delta^{(i)}/f(\boldsymbol{x}^\star)$ as functions of CPU time for a random realization of the sensing matrix Φ with normalized numbers of measurements $N/p = 0.34$ and several different regularization constants a. (For the GFB method, we compute the normalized centered objectives using $P_C(\boldsymbol{x}^{(i)})$ instead of $\boldsymbol{x}^{(i)}$ in (2.34a) because its $\boldsymbol{x}^{(i)}$ may be outside $\boldsymbol{C}$.) The legend in Fig. 2.6c applies also to Figs. 2.6a and 2.6b. To achieve good performance, CP and PDS need to be manually tuned for each a. CP and PDS have optimal $\sigma_1 = \sigma_2$ equal to $0.01, 0.1, 1$ and $0.0026, 0.026, 2.6$ for $a = -5, -4, -3$, respectively.

PNPG and PNPG_d have the steepest descent rate, followed by PNPG_d. AT and SPIRAL reach the performance floor around the relative precision of 10^{-6} due to their fixed inner convergence criterion in (2.41a). The GFB and primal-dual methods, PDS and CP, are sensitive to the selection of the tuning constants. After a careful selection of the tuning constants, CP performs exceptionally well in Figs. 2.6a and 2.6b. The performance of GFB is affected significantly by the value of the regularization parameter a.

2.6 Conclusion

We developed a fast algorithm for reconstructing sparse signals that belong to a closed convex set by employing a projected proximal-gradient scheme with Nesterov's acceleration, restart, and adaptive step size. We applied the PNPG method to construct one of the first Nesterov-accelerated proximal-gradient reconstruction algorithm for Poisson compressed sensing. We presented integrated derivation of the proposed algorithm and convergence-rate upper bound that accounts for inexactness of the proximal operator and also proved convergence of iterates. Our PNPG approach is computationally efficient compared with other state-of-the-art methods.

Appendices

2.A Derivation of Acceleration (2.20a)–(2.20d) and Proofs of Lemma 2.1 and Theorem 2.1

We first prove Lemma 2.1 and then derive the acceleration (2.20a)–(2.20d) and prove Theorem 2.1.

Proof of Lemma 2.1: According to Definition 1 and (2.24),

$$ur(\boldsymbol{x}) \geq ur(\boldsymbol{x}^{(i)}) + (\boldsymbol{x} - \boldsymbol{x}^{(i)})^T \Big[\frac{\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^{(i)}}{\beta^{(i)}} - \nabla \mathcal{L}(\bar{\boldsymbol{x}}^{(i)}) \Big] - \frac{(\varepsilon^{(i)})^2}{2\beta^{(i)}} \tag{2.48a}$$

for any $\boldsymbol{x} \in \mathbb{R}^p$. Moreover, due to the convexity of $\mathcal{L}(\boldsymbol{x})$,

$$\mathcal{L}(\boldsymbol{x}) \geq \mathcal{L}(\bar{\boldsymbol{x}}^{(i)}) + (\boldsymbol{x} - \bar{\boldsymbol{x}}^{(i)})^T \nabla \mathcal{L}(\bar{\boldsymbol{x}}^{(i)}). \tag{2.48b}$$

Summing (2.48a), (2.48b), and (2.21) completes the proof. □

The following result from [Ber15, Prop. 3.2.1 in Sec. 3.2] states that the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be reduced by projecting them onto a closed convex set $\boldsymbol{C}$.

Lemma 2.2 (Projection theorem)**.** *The projection mapping onto a nonempty closed convex set* $\boldsymbol{C} \subseteq \mathbb{R}^p$ *is nonexpansive*

$$\| P_C(\boldsymbol{x}) - P_C(\boldsymbol{y}) \|_2^2 \leq \| \boldsymbol{x} - \boldsymbol{y} \|_2^2 \tag{2.49}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$.

We now derive the projected Nesterov's acceleration step (2.20b)–(2.20d) with the goal of selecting the $\bar{\boldsymbol{x}}^{(i)}$ in the proximal step (2.20e) that achieves the convergence rate of $\mathcal{O}(k^{-2})$. This derivation and convergence-rate proof are inspired by—but are more general than—[BT09a]. We

start from (2.25) with $\boldsymbol{x}$ replaced by $\boldsymbol{x} = \boldsymbol{x}^\star$ and $\boldsymbol{x} = \boldsymbol{x}^{(i-1)}$,

$$-\Delta^{(i)} \geq \frac{\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^\star\|_2^2 - \|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^\star\|_2^2 - (\varepsilon^{(i)})^2}{2\beta^{(i)}} \tag{2.50a}$$

$$\Delta^{(i-1)} - \Delta^{(i)} \geq \frac{\delta^{(i)} - \|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^{(i-1)}\|_2^2 - (\varepsilon^{(i)})^2}{2\beta^{(i)}} \tag{2.50b}$$

and design two coefficient sequences $a^{(i)} > 0$ and $b^{(i)} > 0$ that multiply (2.50a) and (2.50b), respectively, which ultimately leads to (2.20a)–(2.20d) and the convergence-rate guarantee in (2.33a).

Consider sequences $a^{(i)} > 0$ and $b^{(i)} > 0$. Multiply them by (2.50a) and (2.50b), respectively, add the resulting expressions, and multiply by $\beta^{(i)}$ to obtain

$$\begin{aligned}
&-2\beta^{(i)}c^{(i)}\Delta^{(i)} + 2\beta^{(i)}b^{(i)}\Delta^{(i-1)} \\
&\geq \frac{1}{c^{(i)}}\left\|c^{(i)}\boldsymbol{x}^{(i)} - b^{(i)}\boldsymbol{x}^{(i-1)} - a^{(i)}\boldsymbol{x}^\star\right\|_2^2 \\
&\qquad - \frac{1}{c^{(i)}}\|c^{(i)}\bar{\boldsymbol{x}}^{(i)} - b^{(i)}\boldsymbol{x}^{(i-1)} - a^{(i)}\boldsymbol{x}^\star\|_2^2 - c^{(i)}(\varepsilon^{(i)})^2 \\
&= c^{(i)}\left[t^{(i)} - \bar{t}^{(i)} - (\varepsilon^{(i)})^2\right]
\end{aligned} \tag{2.51}$$

where

$$c^{(i)} \triangleq a^{(i)} + b^{(i)} \tag{2.52a}$$

$$t^{(i)} \triangleq \|\boldsymbol{x}^{(i)} - \boldsymbol{z}^{(i)}\|_2^2, \qquad \bar{t}^{(i)} \triangleq \|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{z}^{(i)}\|_2^2 \tag{2.52b}$$

$$\boldsymbol{z}^{(i)} \triangleq \frac{b^{(i)}}{c^{(i)}}\boldsymbol{x}^{(i-1)} + \frac{a^{(i)}}{c^{(i)}}\boldsymbol{x}^\star. \tag{2.52c}$$

We arranged (2.51) using completion of squares so that the first two summands are similar (but with opposite signs), with the goal of facilitating cancellations as we sum over i. Since we have

control over the sequences $a^{(i)}$ and $b^{(i)}$, we impose the following boundary conditions for $i \geq 1$:

$$c^{(i-1)}t^{(i-1)} \geq c^{(i)}\bar{t}^{(i)} \tag{2.53a}$$

$$\pi^{(i)} \geq 0 \tag{2.53b}$$

where

$$\pi^{(i)} \triangleq \beta^{(i)}c^{(i)} - \beta^{(i+1)}b^{(i+1)}. \tag{2.54}$$

Now, apply the inequality (2.53a) to the right-hand side of (2.51):

$$-2\beta^{(i)}c^{(i)}\Delta^{(i)} + 2\beta^{(i)}b^{(i)}\Delta^{(i-1)} \geq c^{(i)}t^{(i)} - c^{(i-1)}t^{(i-1)} - c^{(i)}(\varepsilon^{(i)})^2 \tag{2.55a}$$

and sum (2.55a) over $i = 1, 2, \ldots, k$, which leads to summand cancellations and

$$-2\beta^{(k)}c^{(k)}\Delta^{(k)} + 2\beta^{(1)}b^{(1)}\Delta^{(0)} - 2\sum_{i=1}^{k-1}\pi^{(i)}\Delta^{(i)} \geq -c^{(0)}t^{(0)} - \sum_{i=1}^{k}c^{(i)}(\varepsilon^{(i)})^2 \tag{2.55b}$$

where (2.55b) follows by discarding a nonnegative term $c^{(k)}t^{(k)}$.

Now, due to $\pi^{(i)}\Delta^{(i)} \geq 0$ (see (2.34a) and (2.53b)), the inequality (2.55b) leads to

$$\Delta^{(k)} \leq \frac{2\beta^{(1)}b^{(1)}\Delta^{(0)} + c^{(0)}t^{(0)} + \sum_{i=1}^{k}c^{(i)}(\varepsilon^{(i)})^2}{2\beta^{(k)}c^{(k)}} \tag{2.56}$$

with simple upper bound on the right-hand side, thanks to summand cancellations facilitated by the assumptions (2.53).

As long as $\beta^{(k)}c^{(k)}$ grows at a rate of k^2 and the inexactness of the proximal mappings leads to bounded $\sum_{i=1}^{k}c^{(i)}(\varepsilon^{(i)})^2$, the centered objective function $\Delta^{(k)}$ can achieve the desired bound decrease rate of $1/k^2$.

In the following section, we show how to satisfy (2.53a), which will lead to the projected momentum acceleration step (2.20d). We approach the constraints (2.53a) by first aiming to meet them with equality, which is possible in the absence of the convex-set constraint ($C = \mathbb{R}^p$). We then use the nonexpansiveness of the convex-set projection to construct $a^{(i)}$ and $b^{(i)}$ that satisfy (2.53a) with inequality in the general case where the convex-set constraint is present. Finally, we show how to satisfy (2.53b), which will allow us to construct the recursive update of $\theta^{(i)}$ in (2.20b) and verify the allowed range of momentum tuning constants in (2.22).

2.A.1 Satisfying Conditions (2.53)

2.A.1.1 Imposing equality in (2.53a)

(2.53a) holds with equality for all i and any $\boldsymbol{x}^\star$ when we choose $\overline{\boldsymbol{x}}^{(i)} = \widehat{\boldsymbol{x}}^{(i)}$ that satisfies

$$\sqrt{c^{(i-1)}}(\boldsymbol{x}^{(i-1)} - \boldsymbol{z}^{(i-1)}) = \sqrt{c^{(i)}}(\widehat{\boldsymbol{x}}^{(i)} - \boldsymbol{z}^{(i)}). \tag{2.57}$$

Now, (2.57) requires equal coefficients multiplying $\boldsymbol{x}^\star$ on both sides; thus $a^{(i)}/\sqrt{c^{(i)}} = 1/\mathrm{w}$ for all i, where $\mathrm{w} > 0$ is a constant (not a function of i), which implies $c^{(i)} = \mathrm{w}^2(a^{(i)})^2$ and $b^{(i)} = \mathrm{w}^2(a^{(i)})^2 - a^{(i)}$; see also (2.52a). Upon defining

$$\theta^{(i)} \triangleq \mathrm{w}^2 a^{(i)} \tag{2.58a}$$

we have

$$\mathrm{w}^2 c^{(i)} = (\theta^{(i)})^2; \qquad \mathrm{w}^2 b^{(i)} = (\theta^{(i)})^2 - \theta^{(i)}. \tag{2.58b}$$

Plug (2.58) into (2.57) and reorganize to obtain the following form of momentum acceleration:

$$\widehat{\boldsymbol{x}}^{(i)} = \boldsymbol{x}^{(i-1)} + \Theta^{(i)}(\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i-2)}). \tag{2.59}$$

Although $\bar{\boldsymbol{x}}^{(i)} = \hat{\boldsymbol{x}}^{(i)}$ satisfies (2.53a), it is not guaranteed to be within $\mathrm{dom}\,\mathcal{L}$; consequently, the proximal-mapping step for this selection *may not* be computable.

2.A.1.2 Selecting $\bar{\boldsymbol{x}}^{(i)} \in C$ that satisfies (2.53a)

We now seek $\bar{\boldsymbol{x}}^{(i)}$ within C that satisfies the inequality (2.53a). Since $\boldsymbol{x}^{(i-1)}$ and $\boldsymbol{x}^\star$ are in C, $\boldsymbol{z}^{(i)} \in C$ by the convexity of C; see (2.52c). According to Lemma 2.2, projecting (2.59) onto C preserves or reduces the distance between points. Therefore,

$$\bar{\boldsymbol{x}}^{(i)} = P_C(\hat{\boldsymbol{x}}^{(i)}) \tag{2.60}$$

belongs to C *and* satisfies the condition (2.53a):

$$c^{(i-1)}t^{(i-1)} = c^{(i)}\|\hat{\boldsymbol{x}}^{(i)} - \boldsymbol{z}^{(i)}\|_2^2 \tag{2.61a}$$

$$\geq c^{(i)}\|\bar{\boldsymbol{x}}^{(i)} - \boldsymbol{z}^{(i)}\|_2^2 = c^{(i)}\bar{t}^{(i)} \tag{2.61b}$$

where (2.61a) and (2.61b) follow from (2.57) and by using Lemma 2.2, respectively; see also (2.52b).

Without loss of generality, set $\mathrm{w} = 1$ and rewrite and modify (2.54), (2.52b), and (2.55b) using (2.58) to obtain

$$\pi^{(i)} = \beta^{(i)}(\theta^{(i)})^2 - \beta^{(i+1)}\theta^{(i+1)}(\theta^{(i+1)} - 1), \quad i \geq 1 \tag{2.62a}$$

$$(\theta^{(i)})^2 t^{(i)} = \left\|\theta^{(i)}\boldsymbol{x}^{(i)} - (\theta^{(i)} - 1)\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^\star\right\|_2^2 \tag{2.62b}$$

$$\sum_{i=1}^{k-1} \pi^{(i)}\Delta^{(i)} \leq \frac{1}{2}\Big[(\theta^{(0)})^2 t^{(0)} + \sum_{i=1}^{k}(\theta^{(i)}\varepsilon^{(i)})^2\Big] \tag{2.62c}$$

where (2.62c) is obtained by discarding the negative term $-2\beta^{(k)}(\theta^{(k)})^2\Delta^{(k)}$ and the zero term $\beta^{(1)}\theta^{(1)}(\theta^{(1)} - 1)\Delta^{(0)}$ (because $\theta^{(1)} = 1$) on the left-hand side of (2.55b). Now, (2.33a) follows from (2.56) by using $\theta^{(0)} = \theta^{(1)} = 1$ (see (2.20b)), (2.58), and (2.62b) with $i = 0$.

2.A.1.3 Satisfying (2.53b)

By substituting (2.62a) into (2.53b), we obtain the conditions

$$\beta^{(i-1)}(\theta^{(i-1)})^2 \geq \beta^{(i)}\big[(\theta^{(i)})^2 - \theta^{(i)}\big] \tag{2.63a}$$

and interpret $(\pi^{(i)})_{i=1}^{+\infty}$ as the sequence of gaps between the two sides of (2.63a); (2.63a) implies

$$\theta^{(i)} \leq 1/2 + \sqrt{1/4 + B^{(i)}(\theta^{(i-1)})^2}. \tag{2.63b}$$

Comparing (2.20b) with (2.63b) justifies the constraints in (2.22).

2.A.2 Connection to Convergence-Rate Analysis of FISTA in [BT09a]

If the step-size sequence $(\beta^{(i)})$ is non-increasing (e.g., in the backtracking-only scenario with $\mathbb{m} = +\infty$), (2.20b) with $B^{(i)} = 1$ also satisfies the inequality (2.63b). In this case, (2.33a) still holds but (2.33b) does not because (2.35) no longer holds. However, because $B^{(i)} = 1$, we have $\theta^{(k)} \geq (k+1)/\gamma$ and

$$\Delta^{(k)} \leq \gamma^2 \frac{\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star\|_2^2 + \mathcal{E}^{(k)}}{2\beta^{(k)}(k+1)^2} \tag{2.64}$$

which generalizes [BT09a, Th. 4.4] to include the inexactness of the proximal operator and the convex-set projection.

2.B Convergence of Iterates

To prove convergence of iterates, we need to show that the centered objective function $\Delta^{(k)}$ decreases faster than the right-hand side of (2.33b). We introduce Lemmas 2.3 and 2.4 and then use them to prove Theorem 2.2. Throughout this Appendix, we assume that Assumption 1 of

Theorem 2.2 holds, which justifies (2.50) and (2.63) as well as results from Appendix 2.A that we use in the proofs.

Lemma 2.3. *Under Assumptions 1–3 of Theorem 2.2,*

$$\sum_{i=1}^{+\infty}(2\theta^{(i)}-1)\delta^{(i)} < +\infty. \tag{2.65}$$

Proof: By letting $k \to +\infty$ in (2.62c) and using (2.37), we obtain

$$\sum_{i=1}^{+\infty}\pi^{(i)}\Delta^{(i)} < +\infty. \tag{2.66}$$

For $i \geq 1$, rewrite (2.62a) using $\theta^{(i)}$ expressed in terms of $\theta^{(i+1)}$ (based on (2.20b)):

$$\begin{aligned}\pi^{(i)} &= \frac{\beta^{(i+1)}}{\gamma}\left[(\gamma-2)\theta^{(i+1)} + (1-b\gamma^2)/\gamma\right] \\ &\geq \frac{\gamma-2}{\gamma}\beta^{(i+1)}\theta^{(i+1)}\end{aligned} \tag{2.67}$$

where the inequality in (2.67) is due to $b\gamma^2 - 1 < 0$; see Assumption 3. Apply nonexpansiveness of the projection operator to (2.50b) and use (2.59) to obtain

$$2\beta^{(i)}(\Delta^{(i-1)} - \Delta^{(i)}) \geq \delta^{(i)} - (\Theta^{(i)})^2\delta^{(i-1)} - (\varepsilon^{(i)})^2; \tag{2.68}$$

then multiply both sides of (2.68) by $(\theta^{(i)})^2$, sum over $i = 1, 2, \ldots, k$ and reorganize:

$$\begin{aligned}\sum_{i=1}^{k-1}(2\theta^{(i)}-1)\delta^{(i)} &\leq (\theta^{(0)}-1)^2\delta^{(0)} - (\theta^{(k)})^2\delta^{(k)} + \mathcal{E}^{(k)} \\ &\quad + 2\beta^{(1)}\Delta^{(0)} - 2\beta^{(k)}(\theta^{(k)})^2\Delta^{(k)} + 2\sum_{i=1}^{k-1}\varrho^{(i)}\Delta^{(i)} \qquad (2.69a) \\ &\leq \mathcal{E}^{(k)} + 2\beta^{(1)}\Delta^{(0)} + \frac{4}{\gamma-2}\sum_{i=1}^{k-1}\pi^{(i)}\Delta^{(i)} \qquad (2.69b)\end{aligned}$$

where (see (2.62a))

$$\varrho^{(i)} \triangleq \beta^{(i+1)}(\theta^{(i+1)})^2 - \beta^{(i)}(\theta^{(i)})^2 \tag{2.69c}$$

$$= \beta^{(i+1)}\theta^{(i+1)} - \pi^{(i)}, \tag{2.69d}$$

and we drop the zero term $(\theta^{(0)} - 1)^2\delta^{(0)}$ and the negative term $-(\theta^{(k)})^2\delta^{(k)} - 2\beta^{(k)}(\theta^{(k)})^2\Delta^{(k)}$ from (2.69a) and use the fact that $\varrho^{(i)} \leq [2/(\gamma - 2)]\pi^{(i)}$ implied by (2.67) to obtain (2.69b). Finally, let $k \to +\infty$ and use (2.37) and (2.66) to conclude (2.65). □

Lemma 2.4. *For $j \geq 3$,*

$$\Pi_j \triangleq \sum_{k=j}^{+\infty} \prod_{\ell=j}^{k} \Theta^{(\ell)} \leq \gamma\theta^{(j-1)} - 1. \tag{2.70}$$

Proof: For $j \geq 3$,

$$\frac{1}{\sqrt{\beta^{(k-1)}\theta^{(k-1)}\theta^{(k)}}} \leq \frac{\gamma}{\sqrt{\beta^{(k-1)}\theta^{(k-1)}}} - \frac{\gamma}{\sqrt{\beta^{(k)}\theta^{(k)}}} \tag{2.71a}$$

$$\leq \frac{\gamma}{\sqrt{\beta^{(k-2)}\theta^{(k-2)}}} - \frac{\gamma}{\sqrt{\beta^{(k)}\theta^{(k)}}} \tag{2.71b}$$

where we obtain the inequality (2.71a) by combining the terms on the right-hand size and using (2.35a) and (2.71b) holds because $\sqrt{\beta^{(k)}}\theta^{(k)}$ is an increasing sequence (see Section 2.4). Now,

$$\Pi_j \leq \sum_{k=j}^{+\infty} \prod_{\ell=j}^{k} \frac{\beta^{(\ell-2)}(\theta^{(\ell-2)})^2}{\beta^{(\ell-1)}\theta^{(\ell-1)}\theta^{(\ell)}} = \sum_{k=j}^{+\infty} \frac{\beta^{(j-2)}(\theta^{(j-2)})^2\theta^{(j-1)}}{\beta^{(k-1)}(\theta^{(k-1)})^2\theta^{(k)}} \tag{2.72a}$$

$$\leq \frac{\gamma\beta^{(j-2)}(\theta^{(j-2)})^2\theta^{(j-1)}}{\sqrt{\beta^{(j-2)}}\theta^{(j-2)}\sqrt{\beta^{(j-1)}}\theta^{(j-1)}} = \gamma\sqrt{B^{(j-1)}}\theta^{(j-2)} \tag{2.72b}$$

where (2.72a) follows by using (2.20c), (2.63a) with $i = \ell - 1$, and fraction-term cancellation; (2.72b) is obtained by substituting (2.71b) into (2.72a) and canceling summation terms. (2.72b) implies (2.70) by using (2.35a) with $k = j - 1$. □

Define

$$\lambda^{(i)} \triangleq \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^\star\|_2^2, \qquad \Lambda^{(i)} \triangleq \lambda^{(i)} - \lambda^{(i-1)}. \tag{2.73}$$

Since $f(\boldsymbol{x}^{(i)})$ converges to $f(\boldsymbol{x}^\star) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$ as the iteration index i grows and $\boldsymbol{x}^\star$ is a minimizer, it is sufficient to prove the convergence of $\lambda^{(i)}$; see [CD15, Th. 4.1].

Proof of Theorem 2.2: Use (2.50a) and $\Delta^{(i)} \geq 0$ to obtain

$$0 \geq \lambda^{(i)} - \|\overline{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^\star\|_2^2 - (\varepsilon^{(i)})^2. \tag{2.74}$$

Now,

$$\begin{aligned} \|\overline{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^\star\|_2^2 \leq \|\widehat{\boldsymbol{x}}^{(i)} - \boldsymbol{x}^\star\|_2^2 &= \lambda^{(i-1)} + (\Theta^{(i)})^2 \delta^{(i-1)} \\ &\quad + 2\Theta^{(i)} (\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^\star)^T (\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i-2)}) && \text{(2.75a)} \\ &\leq \lambda^{(i-1)} + (\Theta^{(i)})^2 \delta^{(i-1)} + \Theta^{(i)} (\Lambda^{(i-1)} + \delta^{(i-1)}) && \text{(2.75b)} \end{aligned}$$

where (2.75a) and (2.75b) follow by using the nonexpansiveness of the projection operator (see also (2.59)) and the identity

$$2(\boldsymbol{a} - \boldsymbol{b})^T (\boldsymbol{a} - \boldsymbol{c}) = \|\boldsymbol{a} - \boldsymbol{b}\|_2^2 + \|\boldsymbol{a} - \boldsymbol{c}\|_2^2 - \|\boldsymbol{b} - \boldsymbol{c}\|_2^2 \tag{2.76}$$

respectively. Combine the inequalities (2.75b) and (2.74) to get

$$\begin{aligned} \Lambda^{(i)} &\leq \Theta^{(i)} \big[\Lambda^{(i-1)} + (\Theta^{(i)} + 1) \delta^{(i-1)} \big] + (\varepsilon^{(i)})^2 && \text{(2.77a)} \\ &\leq \Theta^{(i)} (\Lambda^{(i-1)} + 2\delta^{(i-1)}/\xi) + (\varepsilon^{(i)})^2 && \text{(2.77b)} \end{aligned}$$

where (2.77b) is due to $1 < 1/\xi$ (see (2.29)) and the following:

$$\Theta^{(i)} < \frac{\theta^{(i-1)}}{\theta^{(i)}} = \frac{\sqrt{\beta^{(i-1)}}\theta^{(i-1)}\sqrt{\beta^{(i)}}}{\sqrt{\beta^{(i)}}\theta^{(i)}\sqrt{\beta^{(i-1)}}} \tag{2.78a}$$

$$< \frac{\sqrt{\beta^{(i)}}}{\sqrt{\beta^{(i-1)}}} \leq \frac{1}{\sqrt{\xi}} < \frac{1}{\xi} \tag{2.78b}$$

where we have used (2.20c), the fact that $\sqrt{\beta^{(i)}}\theta^{(i)}$ is an increasing sequence, $\beta^{(i)}/\beta^{(i-1)} \geq 1/\xi$ (see Section 2.3.2), and (2.29).

According to (2.35b) and the fact that the sequence $(\beta^{(i)})$ is bounded (by Assumption 4), there exists an integer J such that

$$\theta^{(j-1)} \geq 2, \qquad \Theta^{(j)} \geq \frac{1}{\theta^{(j)}} > 0 \tag{2.79}$$

for all $j \geq J$, where the second inequality follows from the first and the definition of $\Theta^{(j)}$; see (2.20c). Then

$$\begin{aligned}\Omega^{(i)} &\triangleq \max(0, \Lambda^{(i)}) \\ &\leq \Theta^{(i)}\left[\Omega^{(i-1)} + \frac{2\delta^{(i-1)}}{\xi} + \frac{(\varepsilon^{(i)})^2}{\Theta^{(i)}}\right] &\text{(2.80a)} \\ &\leq \sum_{j=J}^{i}\left[\frac{2\delta^{(j-1)}}{\xi} + \frac{(\varepsilon^{(j)})^2}{\Theta^{(j)}}\right]\prod_{\ell=j}^{i}\Theta^{(\ell)} + \Omega^{(J-1)}\prod_{\ell=J}^{i}\Theta^{(\ell)} &\text{(2.80b)}\end{aligned}$$

for $i \geq J$, where the inequality in (2.80a) follows by combining the inequalities (2.77b) and $\Omega^{(i-1)} \geq \Lambda^{(i-1)}$, and (2.80b) follows by recursively applying inequality (2.80a) with i replaced by $i-1, i-2, \ldots, J$. Now, sum the inequalities (2.80b) over $i = J, J+1, \ldots, +\infty$ and exchange the order of summation over i and j on the right-hand side (see also (2.70)):

$$\sum_{i=J}^{+\infty}\Omega^{(i)} \leq \sum_{j=J}^{+\infty}\Pi_j\left[\frac{2\delta^{(j-1)}}{\xi} + \frac{(\varepsilon^{(j)})^2}{\Theta^{(j)}}\right] + \Pi_J\Omega^{(J-1)}. \tag{2.81}$$

For $j \geq J \geq 3$,

$$\gamma(2\theta^{(j-1)} - 1) - \Pi_j \geq \gamma(\theta^{(j-1)} - 1) + 1 > 0 \tag{2.82a}$$

$$2\gamma(\theta^{(j-1)} - 1) - \Pi_j \geq \gamma(\theta^{(j-1)} - 2) + 1 > 0 \tag{2.82b}$$

where the first and second inequalities in (2.82) follow by applying Lemma 2.4 and (2.79), respectively; consequently,

$$\sum_{j=J}^{+\infty} \Pi_j \delta^{(j-1)} \leq \gamma \sum_{j=J}^{+\infty} (2\theta^{(j)} - 1)\delta^{(j)} < +\infty \tag{2.83a}$$

$$\sum_{j=J}^{+\infty} \Pi_j \frac{(\varepsilon^{(j)})^2}{\Theta^{(j)}} \leq 2\gamma \sum_{j=J}^{+\infty} (\varepsilon^{(j)})^2 \frac{\theta^{(j-1)} - 1}{\Theta^{(j)}} \tag{2.83b}$$

$$= 2\gamma \sum_{j=J}^{+\infty} (\varepsilon^{(j)})^2 \theta^{(j)} \tag{2.83c}$$

$$\leq 2\gamma \sum_{j=J}^{+\infty} (\theta^{(j)} \varepsilon^{(j)})^2 \tag{2.83d}$$

where (2.83a) follows from (2.82a) and Lemma 2.3 (for the second inequality) and (2.83b) follows by using (2.82b); (2.83c) and (2.83d) are due to (2.20c) and (2.79), respectively. Combine (2.83a) and (2.83d) with (2.81) to conclude that

$$\sum_{i=1}^{+\infty} \Omega^{(i)} < +\infty. \tag{2.84}$$

The remainder of the proof uses the technique employed by Chambolle and Dossal to conclude the proof of [CD15, Th. 4.1, p. 978], which we repeat for completeness. Define $X^{(i)} \triangleq \lambda^{(i)} - \sum_{j=1}^{i} \Omega^{(j)}$, which is lower bounded because $\lambda^{(i)}$ and $\sum_{j=1}^{i} \Omega^{(j)}$ are lower and upper bounded, respectively; see (2.73) and (2.84). Furthermore, $(X^{(i)})$ is an non-increasing sequence:

$$X^{(i+1)} = \lambda^{(i+1)} - \Omega^{(i+1)} - \sum_{j=1}^{i} \Omega^{(j)} \leq X^{(i)}, \tag{2.85}$$

where we used the fact that $\Omega^{(i+1)} \geq \Lambda^{(i+1)} = \lambda^{(i+1)} - \lambda^{(i)}$. Hence, $(X^{(i)})$ converges as $i \to +\infty$. Since $\sum_{j=1}^{i} \Omega^{(j)}$ converges, $(\lambda^{(i)})$ also converges. □

CHAPTER 3. UPPER-BOUNDING THE REGULARIZATION CONSTANT FOR CONVEX SPARSE SIGNAL RECONSTRUCTION

Submitted for publication.

Renliang Gu and Aleksandar Dogandžić

Abstract

Consider reconstructing a signal $\boldsymbol{x}$ by minimizing a weighted sum of a convex differentiable NLL (data-fidelity) term and a convex regularization term that imposes a convex-set constraint on $\boldsymbol{x}$ and enforces its sparsity using ℓ_1-norm analysis regularization. We compute upper bounds on the regularization tuning constant beyond which the regularization term overwhelmingly dominates the NLL term so that the set of minimum points of the objective function does not change. Necessary and sufficient conditions for irrelevance of sparse signal regularization and a condition for the existence of finite upper bounds are established. We formulate an optimization problem for finding these bounds when the regularization term can be globally minimized by a feasible $\boldsymbol{x}$ and also develop an ADMM type method for their computation. Simulation examples show that the derived and empirical bounds match.

3.1 Introduction

Selection of the regularization tuning constant $u > 0$ in convex Tikhonov-type [TA77] penalized NLL minimization

$$f_u(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + u\,r(\boldsymbol{x}) \tag{3.1}$$

is a challenging problem critical for obtaining accurate estimates of the signal $\boldsymbol{x}$ [Vog02, Ch. 7]. Too little regularization leads to unstable reconstructions with large noise and artifacts due to, for example, aliasing. With too much regularization, the reconstructions are too smooth and often degenerate to constant signals. Finding bounds on the regularization constant u or finding conditions for the irrelevance of signal regularization has received little attention. In this chapter, we determine upper bounds on u beyond which the regularization term $r(\boldsymbol{x})$ overwhelmingly dominates the NLL term $\mathcal{L}(\boldsymbol{x})$ in (3.1) so that the minima of the objective function $f_u(\boldsymbol{x})$ *do not* change. For a linear measurement model with white Gaussian noise and ℓ_1-norm regularization, a closed-form expression for such a bound is determined in [KKL+07, eq. (4)]; see also Example 3.4. The obtained bounds can be used to design continuation procedures [HYZ08; WNF09] that gradually decrease u from a large starting point down to the desired value, which improves the numerical stability and convergence speed of the resulting minimization algorithm by taking advantage of the fact that penalized NLL schemes converge faster for smoother problems with larger u [AG03]. In some scenarios, users can monitor the reconstructions as u decreases and terminate when the result is satisfactory.

Consider a convex NLL $\mathcal{L}(\boldsymbol{x})$ and a regularization term

$$r(\boldsymbol{x}) = \mathbb{I}_C(\boldsymbol{x}) + \|\Psi^H \boldsymbol{x}\|_1 \tag{3.2}$$

that imposes a convex-set constraint on $\boldsymbol{x}$, $\boldsymbol{x} \in C \subseteq \mathbb{R}^p$, and sparsity of an appropriate linearly transformed $\boldsymbol{x}$, where $\Psi \in \mathbb{C}^{p \times p'}$ is a known *sparsifying dictionary* matrix. Assume that the NLL

$\mathcal{L}(\boldsymbol{x})$ is differentiable and lower bounded within the closed convex set C, and satisfies

$$\operatorname{dom} \mathcal{L}(\boldsymbol{x}) \supseteq C \tag{3.3}$$

which ensures that $\mathcal{L}(\boldsymbol{x})$ is computable for all $\boldsymbol{x} \in C$. Define the convex sets of solutions to $\min_{\boldsymbol{x}} f_u(\boldsymbol{x})$, $\min_{\boldsymbol{x}} r(\boldsymbol{x})$, and $\min_{\boldsymbol{x} \in Q} \mathcal{L}(\boldsymbol{x})$:[1]

$$\mathcal{X}_u \triangleq \left\{\boldsymbol{x} \;\middle|\; f_u(\boldsymbol{x}) = \min_{\boldsymbol{\chi}} f_u(\boldsymbol{\chi})\right\} \tag{3.4a}$$

$$\begin{aligned} Q &\triangleq \left\{\boldsymbol{x} \mid r(\boldsymbol{x}) = \min_{\boldsymbol{\chi}} r(\boldsymbol{\chi})\right\} \\ &= \left\{\boldsymbol{x} \in C \;\middle|\; \|\Psi^H \boldsymbol{x}\|_1 \leq \min_{\boldsymbol{\chi} \in C} \|\Psi^H \boldsymbol{\chi}\|_1\right\} \end{aligned} \tag{3.4b}$$

$$\mathcal{X}^\diamond \triangleq \left\{\boldsymbol{x} \in Q \mid \mathcal{L}(\boldsymbol{x}) \leq \min_{\boldsymbol{\chi} \in Q} \mathcal{L}(\boldsymbol{\chi})\right\} \neq \emptyset \tag{3.4c}$$

where the existence of $\mathcal{X}^\diamond$ is ensured by the assumption that $\mathcal{L}(\boldsymbol{x})$ is lower bounded in C.

We review the notation: "$*$", "^{T}", "^{H}", "$^+$", $\|\cdot\|_p$, $|\cdot|$, $\otimes$, "$\succeq$", "$\preceq$", I_N, $\mathbf{1}_{N\times 1}$, and $\mathbf{0}_{N\times 1}$ denote complex conjugation, transpose, Hermitian transpose, Moore-Penrose matrix inverse, ℓ_p-norm over the complex vector space $\mathbb{C}^N$ defined by $\|\boldsymbol{z}\|_p^p = \sum_{i=1}^N |z_i|^p$ for $\boldsymbol{z} = (z_i) \in \mathbb{C}^N$, absolute value, Kronecker product, elementwise versions of "$\geq$" and "$\leq$", the identity matrix of size N and the $N \times 1$ vectors of ones and zeros, respectively (replaced by I, $\mathbf{1}$, and $\mathbf{0}$ when the dimensions can be inferred). $\mathbb{I}_C(\boldsymbol{a}) = \begin{cases} 0, & \boldsymbol{a} \in C \\ +\infty, & \text{otherwise} \end{cases}$, $P_C(\boldsymbol{a}) = \arg\min_{\boldsymbol{x} \in C} \|\boldsymbol{x} - \boldsymbol{a}\|_2^2$, and $\exp_\circ \boldsymbol{a}$ denote the indicator function, projection onto C, and the elementwise exponential function: $[\exp_\circ \boldsymbol{a}]_i = \exp a_i$.

Denote by $\mathcal{N}(A)$ and $\mathcal{R}(A)$ the null space and range (column space) of a matrix A. These vector spaces are real or complex depending on whether A is a real- or complex-valued matrix. For a set S of complex vectors of size p, define $\operatorname{Re} S \triangleq \left\{\boldsymbol{s} \in \mathbb{R}^p \mid \boldsymbol{s} + \mathrm{j}\boldsymbol{t} \in S \text{ for some } \boldsymbol{t} \in \mathbb{R}^p\right\}$

[1] The use of "$\leq$" in the definitions of Q and $\mathcal{X}^\diamond$ in (3.4b) and (3.4c) makes it easier to identify both as convex sets.

and $S \cap \mathbb{R}^p \triangleq \{\boldsymbol{s} \in \mathbb{R}^p \mid \boldsymbol{s} + \mathrm{j}\boldsymbol{0} \in S\}$, where $\mathrm{j} = \sqrt{-1}$. For $A \in \mathbb{C}^{M \times N}$,

$$\mathcal{N}(A^H) \cap \mathbb{R}^M = \mathcal{N}(\underline{A}^T), \quad \mathrm{Re}\big(\mathcal{R}(A)\big) = \mathcal{R}(\underline{A}) \tag{3.5}$$

are the *real* null space and range of $\underline{A}^T$ and $\underline{A}$, respectively, where

$$\underline{A} \triangleq \big[\mathrm{Re}\, A \;\; \mathrm{Im}\, A\big] \in \mathbb{R}^{M \times 2N}. \tag{3.6}$$

If $\underline{A}$ in (3.6) has full row rank, we can define

$$A^{\ddagger} \triangleq A^H [\mathrm{Re}(AA^H)]^{-1} \tag{3.7}$$

which reduces to A^+ for real-valued A. The following are equivalent: $\mathrm{Re}(\mathcal{R}(\Psi)) = \mathbb{R}^p, \mathcal{N}(\Psi^H) \cap \mathbb{R}^p = \{\boldsymbol{0}\}$, and $d = p$, where

$$d \triangleq \dim(\mathrm{Re}(\mathcal{R}(\Psi))) \leq \min(p, 2p'). \tag{3.8}$$

We can decompose Ψ as

$$\Psi = FZ \tag{3.9}$$

where $F \in \mathbb{R}^{p \times d}$ and $Z \in \mathbb{C}^{d \times p'}$ with $\mathrm{rank}\, F = d$ and $\mathrm{rank}\, \underline{Z} = d$; $\underline{Z} = \big[\mathrm{Re}\, Z \;\; \mathrm{Im}\, Z\big] \in \mathbb{R}^{d \times 2p'}$, consistent with the notation in (3.6). Here, $\mathcal{R}(F)$ denotes the real range of the real-valued matrix F. Clearly, $d \geq 1$ is of interest; otherwise $\Psi = 0$. Observe that (see (3.7))

$$\mathrm{Re}(\Psi Z^{\ddagger}) = F \tag{3.10a}$$

$$\mathcal{R}(F) = \mathrm{Re}(\mathcal{R}(\Psi)). \tag{3.10b}$$

The subdifferential of the indicator function $N_C(\boldsymbol{x}) = \partial \mathbb{I}_C(\boldsymbol{x})$ is the *normal cone to* C *at* $\boldsymbol{x}$ [Ber09, Sec. 5.4] and, by the definition of a cone, satisfies

$$N_C(\boldsymbol{x}) = a N_C(\boldsymbol{x}), \qquad \text{for any } a > 0. \tag{3.11}$$

Define

$$G(s) \triangleq \begin{cases} \{s/|s|\}, & s \neq 0 \\ \{w \in \mathbb{C} \mid |w| \leq 1\}, & s = 0 \end{cases} \tag{3.12}$$

and its elementwise extension $G(\boldsymbol{s})$ for vector arguments $\boldsymbol{s}$, which can be interpreted as twice the Wirtinger subdifferential of $\|\boldsymbol{s}\|_1$ with respect to $\boldsymbol{s}$ [BST12]. Note that $\boldsymbol{s}^H G(\boldsymbol{s}) = \{\|\boldsymbol{s}\|_1\}$, and, when $\boldsymbol{s}$ is a real vector, $\mathrm{Re}(G(\boldsymbol{s}))$ is the subdifferential of $\|\boldsymbol{s}\|_1$ with respect to $\boldsymbol{s}$ [KR15, Sec. 11.3.4].

Lemma 3.1. *For* $\Psi \in \mathbb{C}^{p \times p'}$ *and* $\boldsymbol{x} \in \mathbb{R}^p$*, the subdifferential of* $\|\Psi^H \boldsymbol{x}\|_1$ *with respect to* $\boldsymbol{x}$ *is*

$$\partial_{\boldsymbol{x}} \|\Psi^H \boldsymbol{x}\|_1 = \mathrm{Re}\big(\Psi G(\Psi^H \boldsymbol{x})\big). \tag{3.13}$$

Proof: (3.13) follows from

$$\partial_{\boldsymbol{x}} |\boldsymbol{\psi}_j^H \boldsymbol{x}| = \mathrm{Re}\big(\boldsymbol{\psi}_j G(\boldsymbol{\psi}_j^H \boldsymbol{x})\big) \tag{3.14}$$

where $\boldsymbol{\psi}_j$ is the jth column of Ψ. We obtain (3.14) by replacing the linear transform matrix in [WYYZ08, Prop. 2.1] with $\begin{bmatrix} \mathrm{Re}\,\boldsymbol{\psi}_j & \mathrm{Im}\,\boldsymbol{\psi}_j \end{bmatrix}^T$. □

We now use Lemma 3.1 to formulate the necessary and sufficient conditions for $\boldsymbol{x} \in \mathcal{X}_u$:

$$\mathbf{0} \in u\,\mathrm{Re}\big(\Psi G(\Psi^H \boldsymbol{x})\big) + \nabla \mathcal{L}(\boldsymbol{x}) + N_C(\boldsymbol{x}) \tag{3.15a}$$

and $\boldsymbol{x} \in Q$:

$$\mathbf{0} \in \mathrm{Re}\big(\Psi G(\Psi^H \boldsymbol{x})\big) + N_C(\boldsymbol{x}) \tag{3.15b}$$

respectively.

When the signal vector $\boldsymbol{x} = \mathrm{vec}\, X$ corresponds to an image $X \in \mathbb{R}^{J\times K}$, its isotropic and anisotropic TV regularizations correspond to [CP16, Sec. 2.1]

$$\Psi = \Psi_{\mathrm{v}} + \mathrm{j}\Psi_{\mathrm{h}} \in \mathbb{C}^{JK\times JK} \qquad \text{(isotropic)} \tag{3.16a}$$

$$\Psi = \big[\Psi_{\mathrm{v}}\ \Psi_{\mathrm{h}}\big] \in \mathbb{R}^{JK\times 2JK} \qquad \text{(anisotropic)} \tag{3.16b}$$

respectively, where $\Psi_{\mathrm{v}} = I_K \otimes D^T(J)$ and $\Psi_{\mathrm{h}} = D^T(K) \otimes I_J$ are the vertical and horizontal difference matrices (similar to those in [BV16, Sec. 15.3.3]), and

$$D(L) \triangleq \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{L\times L} \tag{3.17}$$

obtained by appending an all-zero row from below to the $(L-1)\times L$ upper-trapezoidal matrix with first row $\big[1, -1, 0, \ldots, 0\big]$; note that $D(1) = 0$. Here, $d = JK - 1$ and

$$\mathcal{N}(\Psi^H) = \mathcal{R}(\mathbf{1}) \tag{3.18}$$

for both the isotropic and anisotropic TV regularizations.

The scenario where

$$\mathcal{N}(\Psi^H) \cap C \neq \emptyset \tag{3.19}$$

holds is of practical interest: then $Q = \mathcal{N}(\Psi^H) \cap C$ and $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ globally minimize the regularization term: $r(\boldsymbol{x}^\diamond) = 0$. If (3.19) holds and $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$, then $G(\Psi^H \boldsymbol{x}^\diamond) = H$, where

$$H \triangleq \{\boldsymbol{w} \in \mathbb{C}^{p' \times 1} \mid \|\boldsymbol{w}\|_\infty \leq 1\}. \tag{3.20}$$

If, in addition to (3.19),

- $d = p$, then $\mathcal{X}^\diamond = Q = \{\mathbf{0}\}$;
- $\mathcal{N}(\Psi^H) \cap \mathbb{R}^p = \mathcal{R}(\mathbf{1})$, then $Q = \mathcal{R}(\mathbf{1}) \cap C$ and $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ are constant signals of the form $\boldsymbol{x}^\diamond = \mathbf{1}x_0^\diamond$, $x_0^\diamond \in \mathbb{R}$.

In Section 3.2, we define and explain an upper bound U on useful regularization constants u and establish conditions under which signal sparsity regularization is *irrelevant* and finite U *does not exist*. We then present an optimization problem for finding U when (3.19) holds (Section 3.3), develop a general numerical method for computing bounds U (Section 3.4), present numerical examples (Section 3.5), and make concluding remarks (Section 3.6).

3.2 Upper Bound Definition and Properties

Define

$$U \triangleq \inf\{u \geq 0 \,|\, \mathcal{X}_u \cap Q \neq \emptyset\}. \tag{3.21}$$

If $\mathcal{X}_u \cap Q = \emptyset$ for all u, then finite U does not exist, which we denote by $U = +\infty$.

We now show that, if $u \geq U$, then the set of minimum points $\mathcal{X}_u$ of the objective function does not change.

Remark 3.1. (a) For any u, $\mathcal{X}_u \cap Q = \mathcal{X}^\diamond$ if and only if $\mathcal{X}_u \cap Q \neq \emptyset$.

(b) Assuming $\mathcal{X}_U \cap Q \neq \emptyset$ for some $U \geq 0$, $\mathcal{X}_u = \mathcal{X}^\diamond$ for $u > U$.

Proof: We first prove (a). Necessity follows by the existence of $\mathcal{X}^\diamond$; see (3.4c). We argue sufficiency by contradiction. Consider any $\boldsymbol{x}_u \in \mathcal{X}_u \cap Q$; i.e., $\boldsymbol{x}_u$ minimizes both $f_u(\boldsymbol{x})$ and $r(\boldsymbol{x})$. If $\boldsymbol{x}_u \notin \mathcal{X}^\diamond$, there exists a $\boldsymbol{y} \in \mathcal{X}^\diamond$ with $\mathcal{L}(\boldsymbol{y}) < \mathcal{L}(\boldsymbol{x}_u)$ that, by the definition of $\mathcal{X}^\diamond$, also minimizes $r(\boldsymbol{x})$. Therefore, $f_u(\boldsymbol{y}) = \mathcal{L}(\boldsymbol{y}) + ur(\boldsymbol{y}) < f_u(\boldsymbol{x}_u)$, which contradicts the assumption $\boldsymbol{x}_u \in \mathcal{X}_u$. Therefore, $\mathcal{X}_u \cap Q \subseteq \mathcal{X}^\diamond$. If there exists a $\boldsymbol{z} \in \mathcal{X}^\diamond \subseteq Q$ such that $\boldsymbol{z} \notin \mathcal{X}_u$, then $f_u(\boldsymbol{z}) > f_u(\boldsymbol{x}_u)$ which, since both $\boldsymbol{z}$ and $\boldsymbol{x}_u$ are in Q, implies that $\mathcal{L}(\boldsymbol{z}) > \mathcal{L}(\boldsymbol{x}_u)$ and contradicts the definition of $\mathcal{X}^\diamond$. Therefore, $\mathcal{X}^\diamond \subseteq \mathcal{X}_u$.

We now prove (b). By (a), $\mathcal{X}_U \cap Q = \mathcal{X}^\diamond$, which confirms (b) for $u = U$. Consider now $u > U$, a $\boldsymbol{y} \in \mathcal{X}_U \cap Q = \mathcal{X}^\diamond$, and any $\boldsymbol{x} \in \mathcal{X}_u$. Then,

$$\mathcal{L}(\boldsymbol{x}) + Ur(\boldsymbol{x}) \geq \mathcal{L}(\boldsymbol{y}) + Ur(\boldsymbol{y}) \tag{3.22a}$$

$$\mathcal{L}(\boldsymbol{y}) + ur(\boldsymbol{y}) \geq \mathcal{L}(\boldsymbol{x}) + ur(\boldsymbol{x}). \tag{3.22b}$$

By summing the two inequalities in (3.22) and rearranging, we obtain $r(\boldsymbol{y}) \geq r(\boldsymbol{x})$. Since $\boldsymbol{y} \in Q$, $\boldsymbol{x}$ is also in Q; i.e., $\mathcal{X}_u \subseteq Q$, which implies $\mathcal{X}_u = \mathcal{X}^\diamond$ by (a). □

As u increases, $\mathcal{X}_u$ moves gradually towards Q and, according to the definition (3.21), $\mathcal{X}_u$ and Q do not intersect when $u < U$. Once $u = U$, the intersection of the two sets is $\mathcal{X}^\diamond$, and, by Remark 3.1(b), $\mathcal{X}_u = \mathcal{X}^\diamond$ for all $u > U$.

3.2.1 Irrelevant Signal Sparsity Regularization

Remark 3.2. The following claims are equivalent:

(a) $\mathcal{X}^\diamond \cap \mathcal{X}_0 \neq \emptyset$; i.e., there exists an $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ such that

$$\mathbf{0} \in \nabla \mathcal{L}(\boldsymbol{x}^\diamond) + N_C(\boldsymbol{x}^\diamond); \tag{3.23}$$

(b) $\mathcal{X}^\diamond \subseteq \mathcal{X}_0$; and

(c) $U = 0$; i.e., $\mathcal{X}_0 \cap Q \neq \emptyset$.

Proof: (c) follows from (a) because $\mathcal{X}^\diamond \subseteq Q$. (b) follows from (c) by applying Remark 3.1(a) to obtain $\mathcal{X}_0 \cap Q = \mathcal{X}^\diamond$, which implies (b). Finally, (b) implies (a). □

Having $\nabla\mathcal{L}(\boldsymbol{x}^\diamond) = \mathbf{0}$ for at least one $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ implies (3.23) and is therefore a stronger condition than (3.23).

Example 3.1. Consider $\mathcal{L}(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$ and $C = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x} - \mathbf{1}_{2\times 1}\|_2 \leq 1\}$. (Here, $\mathcal{L}(\boldsymbol{x})$ could correspond to the Gaussian measurement model with measurements equal to zero.) Since C is a circle within $\mathbb{R}^2_+$, the objective functions for the identity ($\Psi = I_2$) and 1D TV sparsifying transforms are

$$f_u(\boldsymbol{x}) = x_1^2 + x_2^2 + u(x_1 + x_2) + \mathbb{I}_C(\boldsymbol{x}), \quad \text{(identity)} \tag{3.24a}$$

$$f_u(\boldsymbol{x}) = x_1^2 + x_2^2 + u|x_1 - x_2| + \mathbb{I}_C(\boldsymbol{x}), \quad \text{(1D TV)} \tag{3.24b}$$

respectively, where $\mathcal{X}_u = \mathcal{X}^\diamond = Q = \{\boldsymbol{x}^\diamond\}$ and $\boldsymbol{x}^\diamond = \big(1 - \sqrt{2}/2\big)\mathbf{1}$. Here, $\nabla\mathcal{L}(\boldsymbol{x}^\diamond) = (2 - \sqrt{2})\mathbf{1}_{2\times 1}$ and $N_C(\boldsymbol{x}^\diamond) = \{a\mathbf{1} \mid a \leq 0\}$, which confirms that (3.23) holds.

3.2.2 Condition for Infinite U and Guarantees for Finite U

Remark 3.3. If there exists $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ such that

$$[\nabla\mathcal{L}(\boldsymbol{x}^\diamond) + N_C(\boldsymbol{x}^\diamond)] \cap \mathrm{Re}(\mathcal{R}(\Psi)) = \emptyset. \tag{3.25}$$

then $U = +\infty$. When (3.19) holds, the reverse is also true with a stronger claim: $U = +\infty$ implies (3.25) for all $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$.

Proof: First, we prove sufficiency by contradiction. If a finite U exists, then $\mathcal{X}^\diamond \subseteq \mathcal{X}_u$ for all $u \geq U$. Therefore, (3.15a) holds with $\boldsymbol{x}$ being any $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$, which contradicts (3.25).

In the case where (3.19) holds, we prove the necessity by contradiction. If (3.25) does not hold for all $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$, there exist $\boldsymbol{t} \in N_C(\boldsymbol{x}^\diamond)$ and $\boldsymbol{w} \in \mathbb{C}^{p'}$ such that

$$\mathbf{0} = \nabla \mathcal{L}(\boldsymbol{x}^\diamond) + \mathrm{Re}(\Psi \boldsymbol{w}) + \boldsymbol{t}. \tag{3.26}$$

Since (3.19) holds, $\Psi^H \boldsymbol{x}^\diamond = \mathbf{0}$ and $G(\Psi^H \boldsymbol{x}^\diamond) = H$; see (3.20). When $u \geq \|\boldsymbol{w}\|_\infty$, $\boldsymbol{w} \in uH$ and $\mathrm{Re}(\Psi \boldsymbol{w}) \in u\,\mathrm{Re}\big(\Psi G(\Psi^H \boldsymbol{x}^\diamond)\big)$. Therefore, (3.15a) holds at $\boldsymbol{x} = \boldsymbol{x}^\diamond$ for all $u \geq \|\boldsymbol{w}\|_\infty$, which contradicts $U = +\infty$. □

Example 3.2. Consider $\mathcal{L}(\boldsymbol{x}) = x_1 + \mathbb{I}_{\mathbb{R}_+}(x_1)$, $\Psi = I_2$, and $C = \big\{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x} - \mathbf{1}_{2\times 1}\|_2 \leq 1\big\}$. (Here, $\mathcal{L}(\boldsymbol{x})$ could correspond to the $\mathrm{Poisson}(x_1)$ measurement model with measurement equal to zero.) Since C is a circle within $\mathbb{R}^2_+$, the objective function is

$$f_u(\boldsymbol{x}) = (1+u)x_1 + ux_2 + \mathbb{I}_C(\boldsymbol{x}) \tag{3.27}$$

with $\mathcal{X}_u = \{\boldsymbol{x}_u\}$, $\mathcal{X}^\diamond = Q = \{\boldsymbol{x}^\diamond\}$, and

$$\boldsymbol{x}_u = \mathbf{1}_{2\times 1} - \frac{1}{\sqrt{2 + 2/u + 1/u^2}} \begin{bmatrix} 1 + 1/u \\ 1 \end{bmatrix} \tag{3.28a}$$

$$\boldsymbol{x}^\diamond = \big(1 - \sqrt{2}/2\big)\mathbf{1}_{2\times 1} \tag{3.28b}$$

which implies $U = +\infty$, consistent with the observation that $\mathcal{X}_u \cap Q = \emptyset$. Here, (3.19) is not satisfied: (3.25) is only a sufficient condition for $U = +\infty$ and does not hold in this example.

Example 3.3. Consider $\mathcal{L}(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$, 1D TV sparsifying transform with $\Psi = D^T(2)$, and $C = \big\{\boldsymbol{x} \in \mathbb{R}^2 \mid \big\|\boldsymbol{x} - \big[2,\ 0\big]^T\big\|_2^2 \leq 2\big\}$. Since C is a circle with $x_1 - x_2 \geq 0$, the objective function is

$$f_u(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 + u|x_1 - x_2| + \mathbb{I}_C(\boldsymbol{x}) \tag{3.29a}$$

$$= \|\boldsymbol{x} - \tfrac{1}{2}[u\ -u]^T\|_2^2 - u^2/2 + \mathbb{I}_C(\boldsymbol{x}) \tag{3.29b}$$

with $\mathcal{X}_u = \left\{\left[2-(1+4/u)/q(u),\ 1/q(u)\right]^T\right\}$, $q(u) \triangleq \sqrt{1+4/u+8/u^2}$, and $\mathcal{X}^\diamond = Q = \{\mathbf{1}_{2\times 1}\}$, which implies $U = +\infty$. Since (3.19) holds in this example, (3.25) is necessary and sufficient for $U = +\infty$. Since $-\mathbf{1}^T\nabla\mathcal{L}(\boldsymbol{x}^\diamond) = -4$ and $N_C(\boldsymbol{x}^\diamond) = \{(-a,a)^T \mid a \geq 0\}$, (3.25) holds.

3.2.2.1 Two cases of finite U

If $d = p$ and (3.19) holds, then U must be finite: in this case, condition (3.25) in Remark 3.3 cannot hold, which is easy to confirm by substituting $\mathrm{Re}(\mathcal{R}(\Psi)) = \mathbb{R}^p$ into (3.25).

U must also be finite if

$$\mathcal{X}^\diamond \cap \mathrm{int}\, \boldsymbol{C} \neq \emptyset. \tag{3.30}$$

Indeed, (3.30) implies (3.19) and that for $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond \cap \mathrm{int}\, \boldsymbol{C}$,

$$N_C(\boldsymbol{x}^\diamond) = \{\mathbf{0}\} \tag{3.31a}$$

$$\nabla\mathcal{L}(\boldsymbol{x}^\diamond) \in \mathrm{Re}(\mathcal{R}(\Psi)) \tag{3.31b}$$

and hence (3.25) cannot hold upon substituting (3.31a) and (3.31b). Here, (3.31b) follows from $\mathbf{0} \in \nabla\mathcal{L}(\boldsymbol{x}^\diamond) + N_Q(\boldsymbol{x}^\diamond)$, the condition for optimality of the optimization problem $\min_{\boldsymbol{x}\in Q}\mathcal{L}(\boldsymbol{x})$ that defines $\mathcal{X}^\diamond$, by using the fact that $N_Q(\boldsymbol{x}^\diamond) = \mathrm{Re}(\mathcal{R}(\Psi))$ when $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond \cap \mathrm{int}\, C$.

If (3.30) holds then, by Remark 3.2, $U = 0$ if and only if $\nabla\mathcal{L}(\boldsymbol{x}^\diamond) = \mathbf{0}$.

3.3 Bounds When (3.19) Holds

We now present an optimization problem for finding U when (3.19) holds.

Theorem 3.1. *Assume that* (3.19) *holds and that the convex NLL* $\mathcal{L}(\boldsymbol{x})$ *is differentiable within* $\mathcal{X}^\diamond$. *Consider the following optimization problem:*

$$(P_0): \quad U_0(\boldsymbol{x}^\diamond) = \min_{\boldsymbol{a}\in\mathbb{R}^p,\, \boldsymbol{t}\in\mathbb{C}^{p'}} \|\boldsymbol{p}(\boldsymbol{x}^\diamond, \boldsymbol{a}, \boldsymbol{t})\|_\infty \tag{3.32a}$$

$$\text{subject to} \quad \boldsymbol{a} \in N_C(\boldsymbol{x}^\diamond) \tag{3.32b}$$

$$\nabla\mathcal{L}(\boldsymbol{x}^\diamond) + \boldsymbol{a} \in \mathcal{R}(F) \tag{3.32c}$$

with

$$\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{t}) \triangleq \boldsymbol{t} + Z^\ddagger\big\{F^+[\nabla\mathcal{L}(\boldsymbol{x}) + \boldsymbol{a}] - \mathrm{Re}(Z\boldsymbol{t})\big\}. \tag{3.33}$$

Then, $U_0(\boldsymbol{x}^\diamond) = U$ *for all* $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ *and* U *in* (3.21).

Here, $U = +\infty$ if and only if the constraints in (3.32b) and (3.32c) cannot be satisfied for any $\boldsymbol{a}$, which is equivalent to $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ satisfying (3.25) in Remark 3.3.

Proof: Observe that $G(\Psi^H\boldsymbol{x}^\diamond) = H$ for all $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ and

$$\mathrm{Re}\big(\Psi\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{t})\big) = \nabla\mathcal{L}(\boldsymbol{x}) + \boldsymbol{a}. \tag{3.34}$$

due to (3.19) and (3.10a), respectively.

We first prove that $\mathcal{X}^\diamond \subseteq \mathcal{X}_u$ if $u \geq U_0(\boldsymbol{x}^\diamond)$. Consider *any* $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ and denote by $(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{t}})$ a pair $(\boldsymbol{a}, \boldsymbol{t})$ that solves the minimization problem (P_0). Since $u \geq U_0(\boldsymbol{x}^\diamond)$, there exists an $\tilde{\boldsymbol{h}} \in H$ such that $\boldsymbol{p}(\boldsymbol{x}^\diamond, \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{t}}) + u\tilde{\boldsymbol{h}} = \boldsymbol{0}$. Using (3.34), we obtain

$$\boldsymbol{0} = \mathrm{Re}\{\Psi[\boldsymbol{p}(\boldsymbol{x}^\diamond, \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{t}}) + u\tilde{\boldsymbol{h}}]\} = u\,\mathrm{Re}(\Psi\tilde{\boldsymbol{h}}) + \nabla\mathcal{L}(\boldsymbol{x}^\diamond) + \tilde{\boldsymbol{a}} \tag{3.35}$$

which implies $\boldsymbol{x}^\diamond \in \mathcal{X}_u$ according to (3.15a).

Second, we prove that if $u < U_0(\boldsymbol{x}^\diamond)$ for any $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$, then $\mathcal{X}^\diamond \cap \mathcal{X}_u = \emptyset$. We employ proof by contradiction. Suppose $\mathcal{X}^\diamond \cap \mathcal{X}_u \neq \emptyset$; then, there exists an $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond \cap \mathcal{X}_u$. According

to (3.15a), there exist an $\check{\boldsymbol{h}} \in H$ and an $\check{\boldsymbol{a}} \in N_C(\boldsymbol{x}^\diamond)$ such that $\mathbf{0} = u\,\mathrm{Re}(\Psi\check{\boldsymbol{h}}) + \nabla\mathcal{L}(\boldsymbol{x}^\diamond) + \check{\boldsymbol{a}}$. Using (3.34), we have

$$\mathbf{0} = \mathrm{Re}\big(\Psi[u\check{\boldsymbol{h}} + \boldsymbol{p}(\boldsymbol{x}^\diamond, \check{\boldsymbol{a}}, -u\check{\boldsymbol{h}})]\big). \tag{3.36}$$

Note that

$$u\check{\boldsymbol{h}} + \boldsymbol{p}(\boldsymbol{x}^\diamond, \check{\boldsymbol{a}}, -u\check{\boldsymbol{h}}) = Z^\ddagger\big\{F^+[\nabla\mathcal{L}(\boldsymbol{x}^\diamond) + \check{\boldsymbol{a}}] + u\,\mathrm{Re}(Z\check{\boldsymbol{h}})\big\}. \tag{3.37}$$

Inserting (3.37) into (3.36) and using (3.10a) and the fact that F has full column rank leads to $\mathbf{0} = F^+[\nabla\mathcal{L}(\boldsymbol{x}^\diamond) + \check{\boldsymbol{a}}] + u\,\mathrm{Re}(Z\check{\boldsymbol{h}})$; thus

$$\mathbf{0} = u\check{\boldsymbol{h}} + \boldsymbol{p}(\boldsymbol{x}^\diamond, \check{\boldsymbol{a}}, -u\check{\boldsymbol{h}}). \tag{3.38}$$

Now, rearrange and use the fact that $\|\check{\boldsymbol{h}}\|_\infty \leq 1$ (see (3.20)) to obtain

$$\|\boldsymbol{p}(\boldsymbol{x}^\diamond, \check{\boldsymbol{a}}, -u\check{\boldsymbol{h}})\|_\infty = u\|-\check{\boldsymbol{h}}\|_\infty \leq u < U_0(\boldsymbol{x}^\diamond) \tag{3.39}$$

which contradicts (3.32), where $U_0(\boldsymbol{x}^\diamond)$ is the minimum.

Finally, we prove by contradiction that $U_0(\boldsymbol{x}^\diamond)$ is invariant within $\mathcal{X}^\diamond$ if $\mathcal{X}^\diamond$ has more than one element. Assume that there exist $\boldsymbol{x}_1^\diamond, \boldsymbol{x}_2^\diamond \in \mathcal{X}^\diamond$ and u such that $U_0(\boldsymbol{x}_1^\diamond) \leq u < U_0(\boldsymbol{x}_2^\diamond)$. We obtain contradictory results: $\boldsymbol{x}_1^\diamond \in \mathcal{X}_u$ and $\mathcal{X}^\diamond \cap \mathcal{X}_u \neq \emptyset$ because $u \geq U_0(\boldsymbol{x}_1^\diamond)$ and $u < U_0(\boldsymbol{x}_2^\diamond)$, respectively. Therefore, $U = U_0(\boldsymbol{x}^\diamond)$ is invarant to $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$.

The constraints on $\boldsymbol{a}$ in (3.32b) and (3.32c) are equivalent to stating that (3.25) does not hold for any $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$; see also (3.10b). If an $\boldsymbol{a}$ does not exist that satisfies these constraints, (3.25) holds and $U = +\infty$ according to Remark 3.3. □

We make a few observations: (P_0) is a linear programming problem with linear constraints and can be solved using CVX [GB14] and Matlab's optimization toolbox upon identifying $N_C(\boldsymbol{x}^\diamond)$

and $\mathcal{R}(F)$ in (3.32b) and (3.32c), respectively. Theorem 3.1 requires differentiability of the NLL only at $\boldsymbol{x} = \boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$. If Ψ is real, then Z is real as well, the optimal $\boldsymbol{t}$ in (P_0) has zero imaginary component and the corresponding simplified version of Theorem 3.1 follows and requires optimization in (P_0) with respect to real-valued $\boldsymbol{t} \in \mathbb{R}^{p'}$.

If Ψ is real and $d = p'$, then we can select $Z = I$, which leads to $Z^\ddagger = I$ and cancellation of the variable $\boldsymbol{t}$ in (3.32a) and simplification of (P_0).

We now specialize Theorem 3.1 to two cases with finite U.

Corollary 3.1 ($d = p$)**.** *If $d = p$ and if* (3.19) *holds, then U in* (3.21) *can be computed as*

$$U = \min_{\boldsymbol{a} \in N_C(\boldsymbol{0}),\, \boldsymbol{t} \in \mathbb{C}^{p'}} \left\| \boldsymbol{t} + \Psi^\ddagger [\nabla \mathcal{L}(\boldsymbol{0}) + \boldsymbol{a} - \mathrm{Re}(\Psi \boldsymbol{t})] \right\|_\infty . \tag{3.40}$$

Proof: Theorem 3.1 applies, $\mathcal{X}^\diamond = \{\boldsymbol{0}\}$, and U must be finite. Setting $F = I$ in (3.32) leads to (3.40). □

If $C = \mathbb{R}^p_+$, then $N_C(\boldsymbol{0}) = \mathbb{R}^p_-$ and the condition $\boldsymbol{a} \in N_C(\boldsymbol{0})$ reduces to $\boldsymbol{a} \preceq \boldsymbol{0}$.

Corollary 3.2 ($\mathcal{X}^\diamond \cap \operatorname{int} C \neq \emptyset$)**.** *If* (3.30) *holds, then U in* (3.21) *can be computed as*

$$U = \min_{\boldsymbol{t} \in \mathbb{C}^d} \left\| \boldsymbol{t} + Z^\ddagger [F^+ \nabla \mathcal{L}(\boldsymbol{x}^\diamond) - \mathrm{Re}(Z \boldsymbol{t})] \right\|_\infty \tag{3.41}$$

with any $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond \cap \operatorname{int} \boldsymbol{C}$.

Proof: Thanks to (3.30), (3.19) and (3.31a)–(3.31b) are satisfied, Theorem 3.1 applies, U must be finite, and $\boldsymbol{a} = \boldsymbol{0}$ (by (3.31a)). By using these facts, we simplify (3.32) to obtain (3.41). □

If $d = p$ and $\boldsymbol{0} \in \operatorname{int} C$, then both Corollaries 3.1 and 3.2 apply and the upper bound U can be obtained by setting $\boldsymbol{a} = \boldsymbol{0}$ and $N_C(\boldsymbol{0}) = \{\boldsymbol{0}\}$ in (3.40) or by setting $\boldsymbol{x}^\diamond = \boldsymbol{0}$ and $F = I$ in (3.41).

Example 3.4. Consider a real invertible $\Psi \in \mathbb{R}^{p \times p}$.

(a) If $C = \mathbb{R}^p_+$, Corollary 3.1 applies and (3.40) becomes

$$U = \min_{\boldsymbol{a} \preceq \mathbf{0}} \|\Psi^{-1}[\nabla \mathcal{L}(\mathbf{0}) + \boldsymbol{a}]\|_\infty. \tag{3.42a}$$

In this case, $U = 0$ and signal sparsity regularization is irrelevant if $\nabla \mathcal{L}(\mathbf{0}) \succeq \mathbf{0}$, which follows by inspection from (3.42a), as well as from (3.23) in Remark 3.2. If $\Psi = I$, (3.42a) further reduces to $U = -\min\big(0, \min_i [\nabla \mathcal{L}(\mathbf{0})]_i\big)$.

(b) If $\mathbf{0} \in \operatorname{int} C$, Corollaries 3.1 and 3.2 apply and the bound U simplifies to

$$U = \|\Psi^{-1} \nabla \mathcal{L}(\mathbf{0})\|_\infty. \tag{3.42b}$$

For $\Psi = I$ and a linear measurement model with white Gaussian noise, (3.42b) reduces to the expressions in [KKL+07, eq. (4)] and [WNF09, Sec. III], used in [WNF09] to design its continuation scheme; [KKL+07] and [WNF09] also assume $C = \mathbb{R}^p$.

Example 3.5 (One-dimensional TV regularization). Consider 1D TV regularization with $\Psi = D^T(p) \in \mathbb{R}^{p \times p}$ obtained by setting $K = 1$, $J = p$ in (3.16a); note that $d = p - 1$. Consider a constant signal $\boldsymbol{x}^\diamond = \mathbf{1} x_0^\diamond \in \mathcal{X}^\diamond$. Then Theorem 3.1 applies and yields

$$U = \min_{\boldsymbol{a} \in N_C(\mathbf{1} x_0^\diamond)} \max_{1 \le j < p} \left| \sum_{i=1}^{j} \big[\nabla \mathcal{L}(\mathbf{1} x_0^\diamond) + \boldsymbol{a}\big]_i \right| \tag{3.43a}$$

where we have used the factorization (3.9) with F obtained by the block partitioning $\Psi = \big[F \ \mathbf{0}_{p \times 1}\big]$, $Z = \big[I_{p-1} \ \mathbf{0}_{(p-1) \times 1}\big]$, and the fact that F^+ is equal to the $(p-1) \times p$ lower-triangular matrix of ones. When (3.30) holds, $\mathbf{1} x_0^\diamond \in \mathcal{X}^\diamond \cap \operatorname{int} C$, Corollary 3.2 applies, $\boldsymbol{a} = \mathbf{0}$ (see (3.31a)), and (3.43a) reduces to:

$$U = \max_{1 \le j < p} \left| \sum_{i=1}^{j} \big[\nabla \mathcal{L}(\mathbf{1} x_0^\diamond)\big]_i \right|. \tag{3.43b}$$

The bounds obtained by solving (P_0) are often simple but restricted to the scenario where (3.19) holds. In the following section, we remove assumption (3.19) and develop a general numerical method for finding U in (3.21).

3.4 ADMM Algorithm for Computing U

We focus on the nontrivial scenario where (3.23) does not hold and assume $u > 0$. We also assume that an $\boldsymbol{x}^\diamond \in \mathcal{X}^\diamond$ is available, which will be sufficient to obtain the U in (3.21). We use the duality of norms [BV04, App. A.1.6]:

$$\|\Psi^H \boldsymbol{x}\|_1 = \max_{\|\boldsymbol{w}\|_\infty \le 1} \mathrm{Re}(\boldsymbol{w}^H \Psi^H \boldsymbol{x}) \tag{3.44}$$

to rewrite the minimization of (3.1) as the following min-max problem (see also (3.20)):

$$\min_{\boldsymbol{x}} \max_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{x}) + u\,\mathrm{Re}(\boldsymbol{w}^H \Psi^H \boldsymbol{x}) + \mathbb{I}_C(\boldsymbol{x}) - \mathbb{I}_H(\boldsymbol{w}). \tag{3.45}$$

Since the objective function in (3.45) is convex with respect to $\boldsymbol{x}$ and concave with respect to $\boldsymbol{w}$, the optimal $(\boldsymbol{x}, \boldsymbol{w}) = (\boldsymbol{x}_u, \boldsymbol{w}_u)$ is at the saddle point of (3.45) and satisfies

$$\mathbf{0} \in \nabla \mathcal{L}(\boldsymbol{x}_u) + u\,\mathrm{Re}(\Psi \boldsymbol{w}_u) + N_C(\boldsymbol{x}_u) \tag{3.46a}$$

$$\boldsymbol{w}_u \in G(\Psi^H \boldsymbol{x}_u). \tag{3.46b}$$

Now, select U as the smallest u for which (3.46a)–(3.46b) hold with $\boldsymbol{x}_u = \boldsymbol{x}^\diamond$:

$$U = \frac{1}{v^\diamond} \|\nabla \mathcal{L}(\boldsymbol{x}^\diamond)\|_2 \tag{3.47}$$

where $(v^\diamond, \boldsymbol{w}^\diamond, \boldsymbol{t}^\diamond)$ is the solution to the following constrained linear programming problem:

$$(\mathrm{P}_1): \quad \underset{v,\boldsymbol{w},\boldsymbol{t}}{\text{minimize}} \quad -v + \mathbb{I}_{G(\Psi^H \boldsymbol{x}^\diamond)}(\boldsymbol{w}) + \mathbb{I}_{N_C(\boldsymbol{x}^\diamond)}(\boldsymbol{t}) \tag{3.48a}$$

$$\text{subject to} \quad v\boldsymbol{g} + \mathrm{Re}(\Psi \boldsymbol{w}) + \boldsymbol{t} = \boldsymbol{0} \tag{3.48b}$$

obtained from (3.46a)–(3.46b) with $\boldsymbol{x}_u$ and $\boldsymbol{w}_u$ replaced by $\boldsymbol{x}^\diamond$ and $\boldsymbol{w}$. Here,

$$\boldsymbol{g} \triangleq \nabla \mathcal{L}(\boldsymbol{x}^\diamond) \big/ \|\nabla \mathcal{L}(\boldsymbol{x}^\diamond)\|_2 \tag{3.49}$$

is the normalized gradient (for numerical stability) of the NLL at $\boldsymbol{x}^\diamond$; $\nabla \mathcal{L}(\boldsymbol{x}^\diamond) \neq \boldsymbol{0}$ because (3.23) does not hold. Due to (3.15b), $v = 0$ is a feasible point that satisfies the constraints (3.48b), which implies that $v^\diamond \geq 0$. When (3.25) holds, v has to be zero, implying $U = +\infty$.

To solve (P_1) and find $v^\diamond$, we apply an iterative algorithm based on ADMM [BPC+11; HL13]

$$\boldsymbol{w}^{(i+1)} = \underset{\boldsymbol{w} \in G(\Psi^H \boldsymbol{x}^\diamond)}{\arg\min} \|v^{(i)}\boldsymbol{g} + \mathrm{Re}(\Psi \boldsymbol{w}) + \boldsymbol{t}^{(i)} + \boldsymbol{z}^{(i)}\|_2^2 \tag{3.50a}$$

$$v^{(i+1)} = \rho - \boldsymbol{g}^T \left[\mathrm{Re}(\Psi \boldsymbol{w}^{(i+1)}) + \boldsymbol{t}^{(i)} + \boldsymbol{z}^{(i)}\right] \tag{3.50b}$$

$$\boldsymbol{t}^{(i+1)} = P_{N_C(\boldsymbol{x}^\diamond)}\left(-v^{(i+1)}\boldsymbol{g} - \mathrm{Re}(\Psi \boldsymbol{w}^{(i+1)}) - \boldsymbol{z}^{(i)}\right) \tag{3.50c}$$

$$\boldsymbol{z}^{(i+1)} = \boldsymbol{z}^{(i)} + \mathrm{Re}(\Psi \boldsymbol{w}^{(i+1)}) + v^{(i+1)}\boldsymbol{g} + \boldsymbol{t}^{(i+1)} \tag{3.50d}$$

where $\rho > 0$ is a tuning parameter for the ADMM iteration and we solve (3.50a) using the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm with box constraints [BLNZ95] and PNPG algorithm [GD16b] for real and complex Ψ, respectively. We initialize the iteration (3.50) with $v^{(0)} = 1$, $\boldsymbol{t}^{(0)} = \boldsymbol{0}$, $\boldsymbol{z}^{(0)} = \boldsymbol{0}$, and $\rho = 1$, where ρ is adaptively adjusted thereafter using the scheme in [BPC+11, Sec. 3.4.1].

In special cases, (3.50) simplifies. If (3.19) holds, then $\Psi^H \boldsymbol{x}^\diamond = \boldsymbol{0}$ and the constraint in (3.50a) simplifies to $\|\boldsymbol{w}\|_\infty \leq 1$; see (3.20). If $\mathrm{Re}(\Psi\Psi^H) = cI$, $c > 0$, and $\Psi \in \mathbb{R}^{p \times p}$ or

Table 3.1: Theoretical and empirical bounds U for the linear Gaussian model.

	$C = \mathbb{R}^p_+$, DWT		$C = \mathbb{R}^p$, DWT		$C = \mathbb{R}^p_+$, TV		$C = \mathbb{R}^p$, TV	
SNR/dB	theoretical	empirical	theoretical	empirical	theoretical	empirical	theoretical	empirical
30	8.87	8.87	9.43	9.43	101.55	101.54		
20	8.91	8.91	9.47	9.47	100.21	100.21		
10	9.03	9.03	9.59	9.59	96.47	96.47		
0	9.43	9.43	9.98	9.98	87.49	87.49	same as $C = \mathbb{R}^p_+$, TV	
−10	11.88	11.89	14.03	14.02	152.07	152.07		
−20	27.77	27.78	43.28	43.28	361.56	361.56		
−30	88.78	88.82	139.67	139.66	1024.04	1024.04		
−30	77.29	77.31	123.91	123.90	683.43	683.43	909.50	909.48

Table 3.2: Theoretical and empirical bounds U for the PET example.

	DWT		Anisotropic TV		Isotropic TV	
$\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}}$	theoretical	empirical	theoretical	empirical	theoretical	empirical
10^1	9.660×10^{-1}	9.662×10^{-1}	7.550×10^{-2}	7.544×10^{-2}	7.971×10^{-2}	7.937×10^{-2}
10^3	1.155×10^{2}	1.156×10^{2}	4.154×10^{0}	4.153×10^{0}	4.888×10^{0}	4.877×10^{0}
10^5	1.153×10^{4}	1.153×10^{4}	3.951×10^{2}	3.950×10^{2}	4.666×10^{2}	4.656×10^{2}
10^7	1.145×10^{6}	1.145×10^{6}	3.947×10^{4}	3.946×10^{4}	4.661×10^{4}	4.651×10^{4}
10^9	1.153×10^{8}	1.154×10^{8}	3.950×10^{6}	3.949×10^{6}	4.665×10^{6}	4.654×10^{6}

$\Psi \in \mathbb{C}^{p \times p/2}$, (3.50a) has the following analytical solution:

$$\boldsymbol{w}^{(i+1)} = P_{G(\Psi^H \boldsymbol{x}^\diamond)}\Big(-\frac{1}{c}\Psi^H\big(v^{(i)}\boldsymbol{g} + \boldsymbol{t}^{(i)} + \boldsymbol{z}^{(i)}\big)\Big). \tag{3.51}$$

When (3.30) holds, (3.50c) reduces to $\boldsymbol{t}^{(i)} = \mathbf{0}$ for all i, thanks to (3.31a).

When Ψ is real, the constraints imposed by $\mathbb{I}_{G(\Psi^H \boldsymbol{x}^\diamond)}(\boldsymbol{w})$ become linear and (P_1) becomes a linear programming problem with linear constraints.

3.5 Numerical Examples

Matlab implementations of the presented examples are available at `https://github.com/isucsp/pnpg/tree/master/uBoundEx`. In all numerical examples, the empirical upper bounds U were obtained by a grid search over u with $\mathcal{X}_u = \{\boldsymbol{x}_u\}$ obtained using the PNPG method [GD16b].

3.5.1 Signal reconstruction for Gaussian linear model

We adopt the linear measurement model with white Gaussian noise and scaled NLL $\mathcal{L}(\boldsymbol{x}) = 0.5\|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2$, where the elements of the sensing matrix $\Phi \in \mathbb{R}^{N\times p}$ are i.i.d. and drawn from the uniform distribution on a unit sphere. We reconstruct the nonnegative "skyline" signal $\boldsymbol{x}_{\text{true}} \in \mathbb{R}^{1024\times 1}$ in Section 2.5.2 from noisy linear measurements $\boldsymbol{y}$ using the DWT and 1D TV regularizations, where the DWT matrix Ψ is orthogonal ($\Psi\Psi^T = \Psi^T\Psi = I$), constructed using the Daubechies-4 wavelet with three decomposition levels. Define the SNR as

$$\text{SNR (dB)} = 10\log_{10}\frac{\|\Phi\boldsymbol{x}_{\text{true}}\|_2^2}{N\sigma^2} \tag{3.52}$$

where σ^2 is the variance of the Gaussian noise added to $\Phi\boldsymbol{x}_{\text{true}}$ to create the noisy measurement vector $\boldsymbol{y}$.

For $C = \mathbb{R}_+^p$ and $C = \mathbb{R}^p$ with DWT regularization, $\mathcal{X}^\diamond = \{\mathbf{0}\}$ and Example 3.4 applies and yields the upper bounds (3.42a) and (3.42b), respectively.

For TV regularization, we apply the result in Example 3.5. For $C = \mathbb{R}^p$ and $C = \mathbb{R}_+^p$, we have $\mathcal{X}^\diamond = \{\mathbf{1}x_0\}$ and $\mathcal{X}^\diamond = \{\mathbf{1}\max(x_0, 0)\}$, respectively, where

$$x_0 \triangleq \arg\min_{x\in\mathbb{R}} \mathcal{L}(\mathbf{1}x) = \mathbf{1}^T\Phi^T\boldsymbol{y}/\|\Phi\mathbf{1}\|_2^2. \tag{3.53}$$

If $\mathbf{1}x_0 \in \operatorname{int} C$, which holds when $C = \mathbb{R}^p$ or when $C = \mathbb{R}_+^p$ and $x_0 > 0$, then the bound U is given by (3.43b). For $C = \mathbb{R}_+^p$ and if $x_0 \le 0$, then $\mathcal{X}^\diamond = \{\mathbf{0}\}$ and (3.43a) applies. In this case, $U = 0$ if $[\nabla\mathcal{L}(\mathbf{0})]_i \ge 0$ for $i = 1, \ldots, p-1$, which occurs only when $[\nabla\mathcal{L}(\mathbf{0})]_i = 0$ for all i.

Table 3.1 shows the theoretical and empirical bounds for DWT and TV regularizations and $C = \mathbb{R}_+^p$ and $C = \mathbb{R}^p$; we decrease the SNR from 30 dB to −30 dB with independent noise realizations for different SNRs. The theoretical bounds in Sections 3.3 and 3.4 coincide. For DWT regularization, $\mathcal{X}^\diamond$ is the same for both convex sets C and thus the upper bound U for $C = \mathbb{R}_+^p$ is always smaller than its counterpart for $C = \mathbb{R}^p$, thanks to being optimized over variable $\boldsymbol{a}$ in

(3.42a). For TV regularization, when $x_0 > 0$, the upper bounds U coincide for both C because, in this case, $\mathcal{X}^\diamond$ is the same for both C and $\mathcal{X}^\diamond \in \operatorname{int} C$. In the last row of Table 3.1 we show the case where $x_0 \leq 0$; then, $\mathcal{X}^\diamond$ differs for the two convex sets C, and the upper bound U for $C = \mathbb{R}^p_+$ is smaller than its counterpart for $C = \mathbb{R}^p$, thanks to being optimized over variable $\boldsymbol{a}$ in (3.43a): compare (3.43a) with (3.43b).

3.5.2 PET image reconstruction from Poisson measurements

Consider PET reconstruction of the 128×128 concentration map $\boldsymbol{x}_{\text{true}}$ in Figure 2.2a, which represents simulated radiotracer activity in a human chest, from independent noisy Poisson-distributed measurements $\boldsymbol{y} = (y_n)$ with means $[\Phi \boldsymbol{x}_{\text{true}} + \boldsymbol{b}]_n$. The choices of parameters in the PET system setup and concentration map $\boldsymbol{x}_{\text{true}}$ have been taken from the IRT [Fes16, `emission/em_test_setup.m`]. Here,

$$\mathcal{L}(\boldsymbol{x}) = \mathbf{1}^T(\Phi \boldsymbol{x} + \boldsymbol{b} - \boldsymbol{y}) + \sum_{n, y_n \neq 0} y_n \ln \frac{y_n}{[\Phi \boldsymbol{x} + \boldsymbol{b}]_n} \tag{3.54a}$$

and

$$\Phi = w \operatorname{diag}\big(\exp_\circ(-S\boldsymbol{\kappa} + \boldsymbol{c})\big) S \in \mathbb{R}^{N \times p}_+ \tag{3.54b}$$

is the known sensing matrix; $\boldsymbol{\kappa}$ is the density map needed to model the attenuation of the gamma rays [OF97]; $\boldsymbol{b} = (b_i)$ is the known intercept term accounting for background radiation, scattering effect, and accidental coincidence;[2] $\boldsymbol{c}$ is a known vector that models the detector efficiency variation; and $w > 0$ is a known scaling constant, which we use to control the expected total number of detected photons due to electron-positron annihilation, $\mathbf{1}^T \operatorname{E}(\boldsymbol{y} - \boldsymbol{b}) = \mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}}$, an SNR measure. We collect the photons from 90 equally spaced directions over 180°, with 128 radial samples

[2] The elements of the intercept term have been set to a constant equal to 10% of the sample mean of $\Phi \boldsymbol{x}_{\text{true}}$: $\boldsymbol{b} = [\mathbf{1}^T \Phi \boldsymbol{x}_{\text{true}}/(10N)]\mathbf{1}$.

at each direction. Here, we adopt the parallel strip-integral matrix S [Fes09, Ch. 25.2] and use its implementation in the IRT [Fes16].

We now consider the nonnegative convex set $C = \mathbb{R}_+^p$, which ensures that (3.3) holds, and 2D isotropic and anisotropic TV and DWT regularizations, where the 2D DWT matrix Ψ is constructed using the Daubechies-6 wavelet with six decomposition levels.

For TV regularizations, $\mathcal{X}^\diamond = \{\mathbf{1}\max(0, x_0)\}$, where $x_0 = \arg\min_{x\in\mathbb{R}} \mathcal{L}(\mathbf{1}x)$, computed using the bisection method that finds the zero of $\partial\mathcal{L}(\mathbf{1}x)/\partial x$, which is an increasing function of $x \in \mathbb{R}_+$. Here, no search for x_0 is needed when $\partial\mathcal{L}(\mathbf{1}x)/\partial x|_{x=0} > 0$, because in this case $x_0 < 0$.

We computed the theoretical bounds using the ADMM-type algorithm in Section 3.4.

Table 3.2 shows the theoretical and empirical bounds for DWT and TV regularizations and the SNR $\mathbf{1}^T\Phi\boldsymbol{x}_{\text{true}}$ varying from 10^1 to 10^9, with independent measurement realizations for different SNRs.

Denote the isotropic and anisotropic 2D TV bounds by U_{iso} and U_{ani}, respectively. Then, it is easy to show that when (3.19) holds, $U_{\text{ani}} \le U_{\text{iso}} \le \sqrt{2}U_{\text{ani}}$, which follows by using the inequalities $\sqrt{2}\sqrt{a^2+b^2} \ge |a|+|b| \ge \sqrt{a^2+b^2}$ and is confirmed in Table 3.2.

3.6 Concluding Remarks

We derived upper bounds on the regularization constant for convex sparse signal reconstruction and presented for the first time such bounds for total-variation regularization. The developed bounds can be used to construct accurate prior distributions for the regularization constant and to design continuation procedures. Future work will include obtaining simple expressions for upper bounds U for isotropic 2D TV regularization, based on Theorem 3.1. It would be also of interest to compute corresponding bounds for low-rank matrix models with nuclear-norm regularization.

CHAPTER 4. BLIND X-RAY CT IMAGE RECONSTRUCTION FROM POLYCHROMATIC POISSON MEASUREMENTS

A paper published in *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 150–165, 2016.

Renliang Gu and Aleksandar Dogandžić

Abstract

We develop a framework for reconstructing images that are sparse in an appropriate transform domain from polychromatic CT measurements under the blind scenario where the material of the inspected object and incident-energy spectrum are unknown. Assuming that the object that we wish to reconstruct consists of a single material, we obtain a parsimonious measurement-model parameterization by changing the integral variable from photon energy to mass attenuation, which allows us to combine the variations brought by the unknown incident spectrum and mass attenuation into a single unknown *mass-attenuation spectrum* function; the resulting measurement equation has the Laplace-integral form. The mass-attenuation spectrum is then expanded into basis functions using B-splines of order one. We consider a Poisson noise model and establish conditions for biconvexity of the corresponding NLL function with respect to the density-map and mass-attenuation spectrum parameters. We derive a block-coordinate descent algorithm for constrained minimization of a penalized NLL objective function, where penalty terms ensure nonnegativity of the mass-attenuation spline coefficients and nonnegativity and gradient-map sparsity of the density-map image, imposed using a convex TV norm; the resulting objective function is biconvex. This algorithm alternates between a Nesterov's proximal-gradient (NPG) step and a

limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints (L-BFGS-B) iteration for updating the image and mass-attenuation spectrum parameters, respectively. We prove the Kurdyka-Łojasiewicz property of the objective function, which is important for establishing local convergence of block-coordinate descent schemes in biconvex optimization problems. Our framework applies to other NLLs and signal-sparsity penalties, such as lognormal NLL and ℓ_1 norm of 2D DWT image coefficients. Numerical experiments with simulated and real X-ray CT data demonstrate the performance of the proposed scheme.

4.1 Introduction

X-ray CT measurement systems are important in modern nondestructive evaluation (NDE) and medical diagnostics. The past decades have seen great progress in CT hardware and (reconstruction) software development. CT sees into the interior of the inspected object and gives 2D and 3D reconstruction at a high resolution. It is a fast, high-resolution method that can distinguish density differences as small as 1 %. As it shows the finest interior detail, it has been one of the most important techniques in medical diagnosis, material analysis and characterization, and NDE [DTK12; WYD08].

Because of the importance of the technique in these application areas, improving reconstruction accuracy and speed of data collection in these systems could have a significant impact on these broad areas. Thanks to recent computational and theoretical advances, such as graphics processing units (GPUs) and sparse signal reconstruction theory and methods, it is now possible to design iterative reconstruction methods that incorporate accurate nonlinear physical models into sparse signal reconstructions from significantly undersampled measurements.

Due to the polychromatic nature of the X-ray source and the fact that mass attenuation generally decreases as a function of photon energy, the center of the spectrum shifts to higher energy as X-rays traverse the object, an effect known as “hardening” [KS88]. This effect destroys the linearity between the attenuation coefficient and the logarithm of the noiseless measurements.

Therefore, linear reconstructions such as FBP exhibit beam-hardening artifacts, e.g., cupping and streaking [Hsi09, Ch. 7.6], which limit the quantitative analysis of the reconstruction. In medical CT applications, severe artifacts can look similar to certain pathologies and further mislead the diagnosis [Hsi09, Sec. 7.6.2]. Fulfilling the promise of compressed sensing and sparse signal reconstruction in X-ray CT depends on accounting for the polychromatic measurements, in addition to other effects such as ring artifacts, metal artifacts in medical applications, X-ray scatter, and detector crosstalk and afterglow [NDF+13; BK04]. It is not clear how aliasing and beam-hardening artifacts interact, and our experience is that we cannot achieve great undersampling when applying sparse linear reconstruction to polychromatic measurements. Indeed, the error caused by the model mismatch may well be larger than the aliasing error that we wish to correct using sparse signal reconstruction.

Beam-hardening correction methods can be categorized into pre-filtering, linearization, dual-energy, and post-reconstruction approaches [KKF08]. Reconstruction methods have recently been developed in [EF02; EF03; VVD+11] that aim to optimize nonlinear objective functions based on the underlying physical model; [EF02; EF03] assume known incident polychromatic source spectrum and imaged materials, whereas [VVD+11] considers a blind scenario for a lognormal measurement model with *unknown* incident spectrum and imaged materials, but employs a photon-energy discretization [GD13, eq. (2)], [Hsi09, Sec. 8.4] with an excessive number of parameters (which leads to permutation and scaling ambiguities; see [GD13] for details) and suffers from numerical instability [GD15b]. The methods in [VVD+11] do not impose sparsity of the reconstructed density-map image, only its nonnegativity, and they have been tested in [VVD+11] using real and noiseless simulated data.

It is often expensive to determine the X-ray spectrum and the materials of the object. X-ray spectrum measurements based on semiconductor detectors are usually distorted by charge trapping, escape events, and other effects [RPP+09], and the corresponding correction requires a highly collimated beam and special procedures [LRG+14]. Even after measuring the spectrum, it is not feasible to scan different objects with fixed scanning configurations, e.g., X-ray tube voltage, cur-

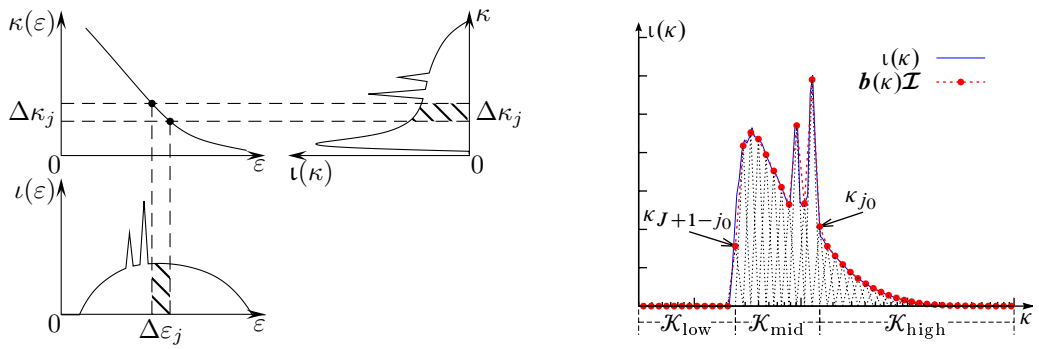


Figure 4.1: (a) Mass-attenuation spectrum $\iota(\kappa)$ obtained by combining the mass attenuation $\kappa(\varepsilon)$ and incident spectrum $\iota(\varepsilon)$ and (b) its B1-spline expansion, with κ-axis in log scale.

rent, prefiltrations, and scanning time. Knowing the mass-attenuation function can be challenging as well when the inspected material is unknown or the inspected object is made of a compound or a mixture with an unknown percentage of each constituent.

In this chapter (see also [GD13; GD15a; GD15b]), we adopt the nonlinear measurement scenario resulting from the polychromatic X-ray source and formulate a parsimonious measurement-model parameterization by exploiting the relationship between the *mass-attenuation coefficients*, *X-ray photon energy*, and *incident spectrum*; see Fig. 4.1a. This simplified model allows *blind* density-map reconstruction and estimation of the composite *mass-attenuation spectrum* $\iota(\kappa)$ in the case for which both the mass attenuation and incident spectrum are unknown. We develop a blind sparse density-map reconstruction scheme from measurements corrupted by Poisson noise, where the signal sparsity in the density-map domain is enforced using a TV norm penalty. The Poisson noise model is appropriate for measurements from photon-counting detectors and a good approximation for the more precise compound Poisson distribution for measurements from energy-integrating detectors [XT14; LWW07].

Although we focus on Poisson noise and gradient-map image sparsity in this chapter, our framework is general and easy to adapt to, for example, lognormal noise and image sparsity in a 2D DWT domain; see [GD15b; GD15a].

We introduce the notation: I_N, $\mathbf{1}_{N\times 1}$, and $\mathbf{0}_{N\times 1}$ are the identity matrix of size N and the $N\times 1$ vectors of ones and zeros, respectively (replaced by I, $\mathbf{1}$, and $\mathbf{0}$ when the dimensions can be inferred easily); $|\cdot|$, $\|\cdot\|_p$, and "^{T}" are the absolute value, ℓ_p norm, and transpose, respectively. Denote by $\lceil x \rceil$ the smallest integer larger than or equal to $x \in \mathbb{R}$. For a vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T \in \mathbb{R}^p$, define the nonnegativity indicator function

$$\mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha}) \triangleq \begin{cases} 0, & \boldsymbol{\alpha} \succeq \mathbf{0} \\ +\infty, & \text{otherwise} \end{cases} \tag{4.1}$$

where "$\succeq$" and "$\succ$"are the elementwise versions of "$\geq$" and "$>$", respectively. Furthermore, $\boldsymbol{a}^{\mathrm{L}}(s) \triangleq \int \boldsymbol{a}(\kappa)\mathrm{e}^{-s\kappa}\,\mathrm{d}\kappa$ is the *Laplace transform* of a vector function $\boldsymbol{a}(\kappa)$ and

$$\big((-\kappa)^m \boldsymbol{a}\big)^{\mathrm{L}}(s) = \int (-\kappa)^m \boldsymbol{a}(\kappa)\mathrm{e}^{-s\kappa}\,\mathrm{d}\kappa = \frac{\mathrm{d}^m \boldsymbol{a}^{\mathrm{L}}(s)}{\mathrm{d}s^m} \tag{4.2}$$

is the mth derivative of $\boldsymbol{a}^{\mathrm{L}}(s)$. Define also the set of nonnegative real numbers as $\mathbb{R}_+ = [0, +\infty)$, the elementwise logarithm $\ln_\circ \boldsymbol{x} = [\ln x_1, \ldots, \ln x_N]^T$ where $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$, and Laplace transforms $\boldsymbol{a}^{\mathrm{L}}_\circ(\boldsymbol{s}) = \big(\boldsymbol{a}^{\mathrm{L}}(s_n)\big)_{n=1}^N$ and $(\kappa\boldsymbol{a})^{\mathrm{L}}_\circ(\boldsymbol{s}) = \big((\kappa\boldsymbol{a})^{\mathrm{L}}(s_n)\big)_{n=1}^N$ obtained by stacking $\boldsymbol{a}^{\mathrm{L}}(s_n)$ and $(\kappa\boldsymbol{a})^{\mathrm{L}}(s_n)$ columnwise, where $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^T$. We define the proximal operator for function $r(\boldsymbol{\alpha})$ scaled by λ [PB13]:

$$\mathrm{prox}_{\lambda r}\,\boldsymbol{a} = \arg\min_{\boldsymbol{\alpha}} \tfrac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{a}\|_2^2 + \lambda r(\boldsymbol{\alpha}). \tag{4.3}$$

Finally, $\mathrm{supp}(\iota(\cdot))$ is the support set of a function $\iota(\cdot)$, $\mathrm{dom}(f) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < +\infty\}$ is the domain of function $f(\cdot)$, and $\mathrm{diag}(\boldsymbol{x})$ is the diagonal matrix with diagonal elements defined by the corresponding elements of vector $\boldsymbol{x}$.

4.1.1 Polychromatic X-ray CT Model

We review the standard noiseless polychromatic X-ray CT measurement model.

Assume that the incident intensity $\mathcal{I}^{\text{in}}$ of a polychromatic X-ray source spreads along photon energy ε following the density $\iota(\varepsilon) \geq 0$:

$$\mathcal{I}^{\text{in}} = \int \iota(\varepsilon)\, \mathrm{d}\varepsilon; \tag{4.4a}$$

see Fig. 4.1a, which shows a typical $\iota(\varepsilon)$. The noiseless measurement collected by an energy-integrating detector upon traversing a straight line $\ell = \ell(x, y)$ in a Cartesian coordinate system has the superposition-integral form [KS88, Ch. 4.1], [NDF+13, Sec. 6]:

$$\begin{aligned} \mathcal{I}^{\text{out}} &= \int \iota(\varepsilon) \exp\Big[-\int_{\ell} \mu(x, y, \varepsilon)\, \mathrm{d}\ell\Big]\, \mathrm{d}\varepsilon \\ &= \int \iota(\varepsilon) \exp\Big[-\kappa(\varepsilon) \int_{\ell} \alpha(x, y)\, \mathrm{d}\ell\Big]\, \mathrm{d}\varepsilon, \end{aligned} \tag{4.4b}$$

where we model the attenuation coefficients $\mu(x, y, \varepsilon)$ of the inspected object consisting of a *single* material using the following *separable form* [NDF+13, Sec. 6]:

$$\mu(x, y, \varepsilon) = \kappa(\varepsilon)\alpha(x, y). \tag{4.5}$$

Here, $\kappa(\varepsilon) > 0$ is the mass-attenuation coefficient of the material, a function of the photon energy ε (illustrated in Fig. 4.1a), and $\alpha(x, y) \geq 0$ is the density-map of the object. For a monochromatic source at photon energy ε, $\ln[\mathcal{I}^{\text{in}}(\varepsilon)/\mathcal{I}^{\text{out}}(\varepsilon)]$ is a linear function of $\alpha(x, y)$, which is a basis for traditional linear reconstruction. However, X-rays generated by vacuum tubes are not monochromatic [KS88; Hsi09], and we cannot transform the underlying noiseless measurements to a linear model unless we know perfectly the incident energy spectrum $\iota(\varepsilon)$ and mass attenuation of the inspected material $\kappa(\varepsilon)$.

In Section 4.2, we introduce our parsimonious parameterization of the measurement model (4.4b) tailored for signal reconstruction. In Section 4.3, we define the parameters to be estimated and discuss their identifiability. Section 4.4 presents the measurement model and establishes bi-convexity of the underlying NLL function with respect to the density-map and mass-attenuation

parameters. Section 4.5 introduces the penalized NLL function that incorporates the parameter constraints, establishes its properties, and describes a block coordinate-descent algorithm for its minimization. In Section 4.6, we show the performance of the proposed method using simulated and real X-ray CT data. Concluding remarks are given in Section 4.7.

4.2 Mass-Attenuation Parameterization

Since the mass attenuation $\kappa(\varepsilon)$ and incident spectrum density $\iota(\varepsilon)$ are both functions of ε (see Fig. 4.1a), we combine the variations of these two functions and write (4.4a) and (4.4b) as integrals of κ rather than ε, seeking to represent our model using two functions $\mathfrak{l}(\kappa)$ (defined below) and $\alpha(x, y)$ instead of three ($\iota(\varepsilon)$, $\kappa(\varepsilon)$, and $\alpha(x, y)$); see also [GD13]. Hence, we rewrite (4.4a) and (4.4b) as (see Appendix 4.A)

$$\mathcal{I}^{\text{in}} = \mathfrak{l}^{\text{L}}(0) \tag{4.6a}$$

$$\mathcal{I}^{\text{out}} = \mathfrak{l}^{\text{L}}\left(\int_{\ell} \alpha(x, y)\, \mathrm{d}\ell\right), \tag{4.6b}$$

where $\mathfrak{l}^{\text{L}}(s) = \int \mathfrak{l}(\kappa)\mathrm{e}^{-s\kappa}\, \mathrm{d}\kappa$ is the Laplace transform of the mass-attenuation spectrum $\mathfrak{l}(\kappa)$, which represents the density of the incident X-ray energy at attenuation κ; here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex. For invertible $\kappa(\varepsilon)$ with differentiable inverse function $\varepsilon(\kappa)$,

$$\mathfrak{l}(\kappa) \triangleq \iota(\varepsilon(\kappa))|\varepsilon'(\kappa)| \geq 0 \tag{4.7}$$

with $\varepsilon'(\kappa) = \mathrm{d}\varepsilon(\kappa)/\,\mathrm{d}\kappa$. In Fig. 4.1a, the area $\iota(\varepsilon_j)\Delta\varepsilon_j$ depicting the X-ray energy within the $\Delta\varepsilon_j$ slot is the same as area $\mathfrak{l}(\kappa_j)\Delta\kappa_j$, the amount of X-ray energy attenuated within the corresponding $\Delta\kappa_j$ slot. In Appendix 4.A, we generalize (4.7) to non-invertible $\kappa(\varepsilon)$ with K-edges.

The mass-attenuation spectrum $\mathfrak{l}(\kappa)$ is nonnegative for all κ; see (4.7) and its generalization (4.39) in Appendix 4.A. Due to its nonnegative support and range, $\mathfrak{l}^{\text{L}}(s)$ is a decreasing function

of s. Here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex. The function $(\iota^{\mathrm{L}})^{-1}$ maps the noiseless measurement $\mathcal{I}^{\mathrm{out}}$ in (4.6), which is a nonlinear function of the density-map $\alpha(x, y)$, into a noiseless linear "measurement" $\int_\ell \alpha(x, y)\,\mathrm{d}\ell$. The $(\iota^{\mathrm{L}})^{-1} \circ \exp(-\cdot)$ mapping corresponds to the *linearization function* in [Her79] (where it was defined through (4.4b) rather than the mass-attenuation spectrum) and converts $-\ln \mathcal{I}^{\mathrm{out}}$ into a noiseless linear "measurement" $\int_\ell \alpha(x, y)\,\mathrm{d}\ell$.

The mass-attenuation spectrum depends on the measurement system (through the incident energy spectrum) and inspected object (through the mass attenuation of the inspected material). In the blind scenario with unknown inspected material and incident signal spectrum, parameterization (4.6) allows us to estimate two functions: $\iota(\kappa)$ and $\alpha(x, y)$ rather than three: $\iota(\varepsilon), \kappa(\varepsilon)$, and $\alpha(x, y)$. This blind scenario is the focus of this chapter.

4.3 Discrete Parameter Definition and Ambiguity

We first define the discrete density map and mass-attenuation spectrum parameters and then discuss their identifiability.

4.3.1 Density-Map Discretization and Mass-Attenuation Spectrum Basis-Function Expansion

Upon spatial-domain discretization into p pixels, approximate the integral $\int_\ell \alpha(x, y)\,\mathrm{d}\ell$ with $\boldsymbol{\phi}^T\boldsymbol{\alpha}$:

$$\int_\ell \alpha(x, y)\,\mathrm{d}\ell = \boldsymbol{\phi}^T\boldsymbol{\alpha}, \tag{4.8}$$

where $\boldsymbol{\alpha} \succeq \mathbf{0}$ is a $p \times 1$ vector representing the 2D image that we wish to reconstruct and $\boldsymbol{\phi} \succeq \mathbf{0}$ is a $p \times 1$ vector of known weights quantifying how much each element of $\boldsymbol{\alpha}$ contributes to the X-ray attenuation on the straight-line path ℓ. An X-ray CT scan consists of hundreds of projections with the beam intensity measured by thousands of detectors for each projection. Denote by N the total number of measurements from all projections collected at the detector array. For the nth

measurement, define its discretized line integral as $\boldsymbol{\phi}_n^T \boldsymbol{\alpha}$. Stacking all N such integrals into a vector yields $\Phi\boldsymbol{\alpha}$, where

$$\Phi = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_N \end{bmatrix}^T \in \mathbb{R}^{N \times p} \tag{4.9}$$

is the *projection matrix*, also known as the Radon transform matrix in a parallel-beam X-ray tomographic imaging system. We call the corresponding transformation, $\Phi\boldsymbol{\alpha}$, the *monochromatic projection* of $\boldsymbol{\alpha}$.

Approximate $\iota(\kappa)$ with a linear combination of J ($J \ll N$) basis functions:

$$\iota(\kappa) = \boldsymbol{b}(\kappa)\boldsymbol{\mathcal{I}}, \tag{4.10a}$$

where

$$\boldsymbol{\mathcal{I}} \triangleq [\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_J]^T \succeq \mathbf{0} \tag{4.10b}$$

is the $J \times 1$ vector of corresponding basis-function coefficients, and the $1 \times J$ row-vector function

$$\boldsymbol{b}(\kappa) \triangleq \left[b_1(\kappa), b_2(\kappa), \ldots, b_J(\kappa)\right] \tag{4.11}$$

consists of B-splines [Sch07] of order one (termed B1 splines, illustrated in Fig. 4.1b). In this case, the decomposition (4.10a) yields nonnegative elements of the spline coefficients $\boldsymbol{\mathcal{I}}$ (based on (4.7)) and thus allows us to impose the physically meaningful nonnegativity constraint (4.10b) when estimating $\boldsymbol{\mathcal{I}}$. Substituting (4.8) and (4.10a) into (4.6a)–(4.6b) for each of the N measurements yields the following expressions for the incident energy and the $N \times 1$ vector of noiseless

measurements:

$$\mathcal{I}^{\text{in}}(\boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}(0)\boldsymbol{\mathcal{I}} \tag{4.12a}$$

$$\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})\boldsymbol{\mathcal{I}} \tag{4.12b}$$

where, following the notation introduced in Section 4.1, $\boldsymbol{b}^{\text{L}}_{\circ}(\boldsymbol{s}) = \left(\boldsymbol{b}^{\text{L}}(s_n)\right)_{n=1}^{N}$ is an *output basis-function matrix* obtained by stacking the $1 \times J$ vectors $\boldsymbol{b}^{\text{L}}(s_n)$ columnwise, and $\boldsymbol{s} = \Phi\boldsymbol{\alpha}$ is the monochromatic projection. Since the Laplace transform of (4.11) (see also (4.13b)) can be computed analytically, $\boldsymbol{b}^{\text{L}}(s)$ has a closed-form expression.

4.3.1.1 Spline selection

We select the spline knots from a growing geometric series $(\kappa_j)_{j=0}^{J+1}$ with $\kappa_0 > 0$:

$$\kappa_j = q^j \kappa_0 \tag{4.13a}$$

and common ratio $q > 1$, which yields the B1 splines

$$b_j(\kappa) = \begin{cases} \dfrac{\kappa - \kappa_{j-1}}{\kappa_j - \kappa_{j-1}}, & \kappa_{j-1} \le \kappa < \kappa_j \\ \dfrac{-\kappa + \kappa_{j+1}}{\kappa_{j+1} - \kappa_j}, & \kappa_j \le \kappa < \kappa_{j+1} \\ 0, & \text{otherwise} \end{cases} \tag{4.13b}$$

that satisfy the q-scaling property:

$$b_j(\kappa) = b_{j+1}(q\kappa) \tag{4.13c}$$

see also Fig. 4.1b. The geometric-series knots (4.13a) appear uniformly spaced in Fig. 4.1b because the κ-axis in this figure is shown in the log scale. When computing $b_j^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})$, larger j implies exponentially smaller $\mathrm{e}^{-\boldsymbol{\phi}_n^T\boldsymbol{\alpha}\kappa}$ terms within the integral range $[\kappa_{j-1}, \kappa_{j+1})$. The geometric-series knot

selection (4.13a) compensates for larger j with a geometrically wider integral range $[\kappa_{j-1}, \kappa_{j+1})$, which results in a more effective approximation of (4.6). In particular, this knot selection leads to $\left(b_j^{\mathrm{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right)_{j=1}^{J}$ with similar values for different values of j, which allows us to balance the weight of each $(\mathcal{I}_j)_{j=1}^{J}$ in $\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}$. Furthermore, the geometric-series knots (4.13a) span a range from κ_0 to κ_{J+1}, which can be made wide with a moderate number of knots J.

The common ratio q determines the resolution of the B1-spline approximation. Here, we select q and J so that the range of κ spanning the mass-attenuation spectrum is constant:

$$\frac{\kappa_{J+1}}{\kappa_0} = q^{J+1} = \text{const.} \tag{4.13d}$$

In summary, the following three tuning constants define our B1-spline basis functions $\boldsymbol{b}(\kappa)$:

$$(q, \kappa_0, J). \tag{4.13e}$$

4.3.2 Density-Map and Mass-Attenuation Spectrum Ambiguities

By noting (4.13c) and the κ-scaling property of the Laplace transform,

$$b_j(q\kappa) \overset{\mathcal{L}}{\rightarrow} \frac{1}{q} b_j^{\mathrm{L}}\left(\frac{s}{q}\right), \qquad q > 0 \tag{4.14}$$

we conclude that selecting basis functions $\left[b_0(\kappa), b_1(\kappa), \ldots, b_{J-1}(\kappa)\right]$ that are q times narrower than those in $\boldsymbol{b}(\kappa)$ and density-map and spectral parameters q times larger than $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$: $q\boldsymbol{\alpha}$ and $q\boldsymbol{\mathcal{I}}$, yields the same mean output photon energy. Consequently,

$$\boldsymbol{\mathcal{I}}^{\text{out}}\left(\boldsymbol{\alpha}, [0, \mathcal{I}_2, \ldots, \mathcal{I}_J]^T\right) = \boldsymbol{\mathcal{I}}^{\text{out}}\left(q\boldsymbol{\alpha}, q[\mathcal{I}_2, \ldots, \mathcal{I}_J, 0]^T\right). \tag{4.15}$$

We refer to this property as the *shift ambiguity* of the mass-attenuation spectrum, which allows us to rearrange leading or trailing zeros in the mass-attenuation coefficient vector $\boldsymbol{\mathcal{I}}$ and position the central nonzero part of $\boldsymbol{\mathcal{I}}$.

4.3.3 Rank of $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$ and Selection of the Number of Splines J

If $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$ does not have full column rank, then $\boldsymbol{\mathcal{I}}$ is not identifiable even if $\boldsymbol{\alpha}$ is known; see (4.12b). The estimation of $\boldsymbol{\mathcal{I}}$ may be numerically unstable if $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$ is poorly conditioned and has small minimum singular values. We can think of the noiseless X-ray CT measurements as $\boldsymbol{b}^{\mathrm{L}}(s)\boldsymbol{\mathcal{I}}$ sampled at different $s = \boldsymbol{\phi}_n^T\boldsymbol{\alpha} \in \left[0, \max_n(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right]$. The following remark implies that if we could collect all $s \in [0, a]$, $a > 0$ (denoted $\boldsymbol{s}$), the corresponding $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\boldsymbol{s})$ would be a full-rank matrix.

Remark 4.1. $\boldsymbol{\mathcal{J}} = \mathbf{0}_{J\times 1}$ is necessary for $\boldsymbol{b}^{\mathrm{L}}(s)\boldsymbol{\mathcal{J}} = 0$ over the range $s \in [0, a]$, where $\boldsymbol{\mathcal{J}} \in \mathbb{R}^J$ and $a > 0$.

Proof: See [GD15b, Sec. 4.3.3]. □

If our data collection system can sample over $\left[0, \max_n(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right]$ sufficiently densely, we expect $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$ to have full column rank.

As the number of splines J increases for fixed support $[\kappa_0, \kappa_{J+1}]$ (see (4.13d)), we achieve better resolution of the mass-attenuation spectrum, but $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$ becomes poorly conditioned with its smallest singular values approaching zero. To estimate this spectrum well, we should choose a J that provides both good resolution *and* sufficiently large smallest singular value of $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$. Fortunately, we focus on the reconstruction of $\boldsymbol{\alpha}$, which is affected by $\boldsymbol{\mathcal{I}}$ only through the function $\boldsymbol{b}^{\mathrm{L}}(s)\boldsymbol{\mathcal{I}}$, and $\boldsymbol{b}^{\mathrm{L}}(s)\boldsymbol{\mathcal{I}}$ is stable as we increase J. Indeed, we observe that when we choose a J significantly larger than the rank of $\boldsymbol{b}_{\circ}^{\mathrm{L}}(\Phi\boldsymbol{\alpha})$, the estimation of $\boldsymbol{\alpha}$ will be good and $\boldsymbol{b}^{\mathrm{L}}(s)\boldsymbol{\mathcal{I}}$ stable, even though the estimation of $\boldsymbol{\mathcal{I}}$ is poor due to its non-identifiability. The increase of J will also increase the computational complexity of signal reconstruction under the blind scenario for which the mass-attenuation spectrum is unknown.

4.4 Measurement Model and Its Properties

For an $N \times 1$ vector $\boldsymbol{\mathcal{E}}$ of independent Poisson measurements, the NLL in the form of generalized Kullback-Leibler divergence [ZBBR15] is (see also (4.12b))

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) = \mathbf{1}^T \left[\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) - \boldsymbol{\mathcal{E}}\right] - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{\mathcal{I}_n^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})}{\mathcal{E}_n}. \tag{4.16}$$

In the following, we express the NLL (4.16) as a function of $\boldsymbol{\alpha}$ with $\boldsymbol{\mathcal{I}}$ fixed and vice versa, and derive conditions for its convexity under the two scenarios. These results will then be used to establish biconvexity conditions for this NLL.

NLL of $\boldsymbol{\alpha}$. Recall (4.10a) and define

$$\iota_{\circ}^{\text{L}}(\Phi\boldsymbol{\alpha}) = \boldsymbol{b}_{\circ}^{\text{L}}(\Phi\boldsymbol{\alpha})\boldsymbol{\mathcal{I}} \tag{4.17}$$

obtained by stacking $\left(\iota^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right)_{n=1}^N$ columnwise. The NLL of $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}}$ is

$$\mathcal{L}_{\iota}(\boldsymbol{\alpha}) = \mathbf{1}^T\left[\iota_{\circ}^{\text{L}}(\Phi\boldsymbol{\alpha}) - \boldsymbol{\mathcal{E}}\right] - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{\iota^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})}{\mathcal{E}_n}, \tag{4.18}$$

which corresponds to the Poisson GLM with design matrix Φ and link function equal to the inverse of $\iota^{\text{L}}(\cdot)$. See [MN89] for an introduction to GLMs.

To establish convexity of the NLL (4.18), we enforce monotonicity of the mass-attenuation spectrum $\iota(\kappa)$ in low- and high-κ regions and also assume that the mid-κ region has higher spectrum than the low-κ region. Note that we do not require here that $\iota(\kappa)$ satisfy the basis-function expansion (4.10a); however, (4.10a) will be needed to establish the biconvexity of the NLL in (4.16). Hence, we define the three κ regions using the spline parameters (4.13e) as well as an additional integer constant

$$j_0 \geq \lceil (J+1)/2 \rceil. \tag{4.19}$$

In particular, κ_{J+1-j_0} and κ_{j_0} partition the range $[\kappa_0, \kappa_{J+1}]$ into the low-, mid-, and high-κ regions: $\mathcal{K}_{\text{low}}$, $\mathcal{K}_{\text{mid}}$, and $\mathcal{K}_{\text{high}}$, respectively, see Fig. 4.1b.

Assumption 4.1. *The mass-attenuation spectrum satisfies*

$$\mathcal{A} = \Big\{ \iota : \ [\kappa_0, \kappa_{J+1}] \to \mathbb{R}_+ \ \Big| \ \iota \textit{ non-decreasing in } \mathcal{K}_{low}, \textit{ non-increasing in } \mathcal{K}_{high}, \textit{ and } \iota(\kappa) \geq \iota\big(\kappa_{J+1-j_0}\big) \ \forall \kappa \in \mathcal{K}_{mid} \Big\}. \tag{4.20a}$$

If the basis-function expansion (4.10a) *holds,* (4.20a) *reduces to*

$$\mathcal{A} = \Big\{ \boldsymbol{\mathcal{I}} \in \mathbb{R}_+^J \ \Big| \ \mathcal{I}_1 \leq \mathcal{I}_2 \leq \cdots \leq \mathcal{I}_{J+1-j_0}, \ \mathcal{I}_{j_0} \geq \cdots \geq \mathcal{I}_J, \textit{ and } \mathcal{I}_j \geq \mathcal{I}_{J+1-j_0}, \ \forall\, j \in [J+1-j_0, j_0] \Big\}. \tag{4.20b}$$

Here, the monotonic low- and high-κ regions each contain $J - j_0$ knots, whereas the central region contains $2j_0 - J$ knots in the B1-spline representation.

In practice, the X-ray spectrum $\iota(\varepsilon)$ starts at the lowest effective energy that can penetrate the object, vanishes at the tube voltage (the highest photon energy), and has a region in the center higher than the two ends; see Fig. 4.1a. When the support of $\iota(\varepsilon)$ is free of K-edges (see the discussion in Appendix 4.A), the mass-attenuation coefficient $\kappa(\varepsilon)$ is a monotonic function of ε; thus $\iota(\kappa)$ as a function of κ has similar shape as $\iota(\varepsilon)$ as a function of ε, which justifies Assumption 4.1. If a K-edge is present within the support of $\iota(\varepsilon)$, it is difficult to infer the shape of $\iota(\kappa)$. In most cases, Assumption 4.1 holds.

For the approximation of $\iota(\kappa)$ using a B1-spline basis expansion, as long as $[\kappa_0, \kappa_{J+1}]$ is sufficiently large to cover the range of $\kappa(\varepsilon)$ with $\varepsilon \in \text{supp}(\iota(\varepsilon))$, we can always meet Assumption 4.1 by the appropriate selection of j_0.

Multiple different $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ share the same noiseless output $\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ and thus the same NLL; see Section 4.3.2. In particular, equivalent $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ can be constructed by left- or right-shifting the

mass attenuation spectrum and properly rescaling it and the density-map; see (4.15). Selecting a fixed j_0 in (4.19) can exclude all these equivalent values except the one in which the mass-attenuation spectrum satisfies (4.20a) and where the biconvexity of the NLL can be established.

Lemma 4.1. *Provided that Assumption 4.1 holds, the Poisson NLL $\mathcal{L}_\iota(\boldsymbol{\alpha})$ is a convex function of $\boldsymbol{\alpha}$ over the following region:*

$$\left\{\boldsymbol{\alpha} \,\middle|\, \iota_\circ^{\mathrm{L}}(\Phi\boldsymbol{\alpha}) \succeq (1-V)\boldsymbol{\mathcal{E}},\ \boldsymbol{\alpha} \in \mathbb{R}_+^p\right\} \tag{4.21a}$$

where

$$V \triangleq \frac{2q^{j_0}}{q^{2j_0}+1}. \tag{4.21b}$$

Proof: See Appendix 4.B. □

Note that (4.21a) is only a subset of the region where $\mathcal{L}_\iota(\boldsymbol{\alpha})$ is convex and that Lemma 4.1 does not assume a basis-function expansion of the mass-attenuation spectrum, only that it satisfies (4.20a).

The condition in (4.21a) corresponds to lower-bounding $\mathcal{I}_n^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})\big/\mathcal{E}_n$ by $1-V$ for all n. The constant V is a function of q^{j_0}, which is the ratio of the point where $\iota(\kappa)$ starts to be monotonically decreasing to the point where the support of $\iota(\kappa)$ starts; see Fig. 4.1b.

NLL of $\boldsymbol{\mathcal{I}}$. The NLL of $\boldsymbol{\mathcal{I}}$ for fixed $\boldsymbol{\alpha}$ reduces to a Poisson GLM with design matrix

$$A = \boldsymbol{b}_\circ^{\mathrm{L}}(\Phi\boldsymbol{\alpha}) \tag{4.22a}$$

all of whose elements are positive, and the identity link function:

$$\mathcal{L}_A(\boldsymbol{\mathcal{I}}) = \mathbf{1}^T(A\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{E}}) - \sum_{n,\mathcal{E}_n\neq 0} \mathcal{E}_n \ln \frac{[A\boldsymbol{\mathcal{I}}]_n}{\mathcal{E}_n}. \tag{4.22b}$$

We now prove the convexity of $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$.

Lemma 4.2. *The NLL* $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ *in* (4.22b) *is a convex function of* $\boldsymbol{\mathcal{I}}$ *for all* $\boldsymbol{\mathcal{I}} \in \mathbb{R}_+^J$.

Proof: The Hessian of the NLL in (4.22b)

$$\frac{\partial^2 \mathcal{L}_A(\boldsymbol{\mathcal{I}})}{\partial \boldsymbol{\mathcal{I}} \partial \boldsymbol{\mathcal{I}}^T} = A^T \operatorname{diag}(\boldsymbol{\mathcal{E}}) \operatorname{diag}^{-2}(A\boldsymbol{\mathcal{I}}) A \tag{4.23}$$

is positive semidefinite. Thus, $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ is convex on $\mathbb{R}_+^J$. □

The Hessian expression in (4.23) implies that $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ in (4.22b) is strongly convex if the design matrix A has full rank. Combining the convexity results in Lemmas 4.1 and 4.2 yields the biconvexity region for the NLL $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ in (4.16).

Theorem 4.1 (Biconvexity of the NLL)**.** *Suppose that Assumption 4.1 in* (4.20b) *holds. Then, the Poisson NLL* (4.16) *is biconvex [GPK07] with respect to* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\mathcal{I}}$ *in the following set:*

$$\mathcal{P} = \left\{ (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \,\middle|\, \boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \succeq (1 - V)\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{I}} \in \mathcal{A}, \boldsymbol{\alpha} \in \mathbb{R}_+^p \right\}, \tag{4.24}$$

which bounds $\mathcal{I}_n^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \big/ \mathcal{E}_n$ *from below by* $1 - V$ *for all* n*; see also* (4.21b).

Proof: We first show the convexity of $\mathcal{P}$ with respect to each variable ($\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$) with the other fixed. We then show the convexity of the NLL (4.16) for each variable.

Region $\mathcal{A}$ in (4.20b) is a subspace, thus a convex set. Since $\boldsymbol{\mathcal{I}}^{\text{out}}$ in (4.12b) is a linear function of $\boldsymbol{\mathcal{I}}$, the inequalities comparing $\boldsymbol{\mathcal{I}}^{\text{out}}$ to constants specify a convex set. Therefore, $\mathcal{P}_{\boldsymbol{\alpha}} = \left\{ \boldsymbol{\mathcal{I}} \,|\, (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \in \mathcal{P} \right\}$ is convex for fixed $\boldsymbol{\alpha} \in \mathbb{R}_+^p$, for it is the intersection of the subspace $\mathcal{A}$ and a convex set via $\boldsymbol{\mathcal{I}}^{\text{out}}$. Since $b_j(\kappa) \geq 0$, $\left(b_j^{\text{L}}(s)\right)_{j=1}^J = \int_{\kappa_{j-1}}^{\kappa_{j+1}} b_j(\kappa) \mathrm{e}^{-s\kappa} \, \mathrm{d}\kappa$ are decreasing functions of s, which, together with the fact that $\boldsymbol{\mathcal{I}} \succeq \mathbf{0}$, implies that $\boldsymbol{b}^{\text{L}}(\boldsymbol{s})\boldsymbol{\mathcal{I}}$ is a decreasing function of $\boldsymbol{s}$. Since the linear transform $\Phi\boldsymbol{\alpha}$ preserves convexity, $\mathcal{P}_{\boldsymbol{\mathcal{I}}} = \left\{ \boldsymbol{\alpha} \,|\, (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \in \mathcal{P} \right\}$ is convex with respect to $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}} \in \mathcal{A}$. Therefore, $\mathcal{P}$ is biconvex with respect to $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\alpha}$.

Observe that $\mathcal{P}$ in (4.24) is the intersection of the regions specified by Assumption 4.1 and Lemmas 4.1 and 4.2. Thus, within $\mathcal{P}$, the Poisson NLL (4.16) is a convex function of $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}}$ and a convex function of $\boldsymbol{\mathcal{I}}$ for fixed $\boldsymbol{\alpha}$, respectively.

By combining the above region and function convexity results, we conclude that (4.16) is biconvex within $\mathcal{P}$. □

In [GD15b], we establish conditions for biconvexity of the NLL under the lognormal noise model.

4.5 Parameter Estimation

Our goal is to compute penalized maximum-likelihood estimates of the density-map and mass-attenuation spectrum parameters $(\boldsymbol{\alpha}, \mathcal{I})$ by solving the following minimization problem:

$$\min_{\boldsymbol{\alpha},\mathcal{I}} f(\boldsymbol{\alpha}, \mathcal{I}) \tag{4.25a}$$

where

$$f(\boldsymbol{\alpha}, \mathcal{I}) = \mathcal{L}(\boldsymbol{\alpha}, \mathcal{I}) + ur(\boldsymbol{\alpha}) + \mathbb{I}_{[0,+\infty)}(\mathcal{I}) \tag{4.25b}$$

$$r(\boldsymbol{\alpha}) = \sum_{i=1}^{p} \sqrt{\sum_{j \in \mathcal{N}_i} (\alpha_i - \alpha_j)^2} + \mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha}) \tag{4.25c}$$

are the penalized NLL objective function and the density-map regularization term that enforces nonnegativity and sparsity of the image $\boldsymbol{\alpha}$; $u > 0$ is a scalar tuning constant. We impose the nonnegativity of the mass-attenuation coefficients (4.10b) using the indicator-function term in (4.25b). In this chapter, we adopt the Poisson NLL (4.16) and impose gradient-map sparsity of the density-map image using the TV penalty. Here, $\mathcal{N}_i$ is the index set of neighbors of α_i, where the elements of $\boldsymbol{\alpha}$ are arranged to form a 2D image: Each set $\mathcal{N}_i$ consists of two pixels at most, with one on the top and the other on the right of the ith pixel, if possible [BT09b]. The optimization problem in (4.25b) is general and allows for different NLL and density-map regularization terms: [GD15a; GD15b] use lognormal NLL and the image-sparsity regularization term in the form of the ℓ_1 norm of DWT coefficients of $\boldsymbol{\alpha}$, which is a convex function of $\boldsymbol{\alpha}$.

4.5.1 Properties of the Objective Function $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$

Since $r(\boldsymbol{\alpha})$ in (4.25c) and $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\mathcal{I}})$ in (4.25b) are convex functions of $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$ for all $\boldsymbol{\alpha} \succeq \mathbf{0}$ and $\boldsymbol{\mathcal{I}} \succeq \mathbf{0}$, the following holds:

Corollary 4.1. *The objective* $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *in* (4.25b) *is biconvex with respect to* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\mathcal{I}}$ *under the conditions specified by Theorem 4.1.*

Although the NLL may have multiple local minima of the form $q^{\ell}\hat{\boldsymbol{\alpha}}$ with integer ℓ (see Section 4.3.2), those with large ℓ can be eliminated by the regularization penalty. We first examine the impact of the ambiguity on the scaling of the first derivative of the objective function $f(\boldsymbol{z}) \triangleq f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$, where $\boldsymbol{z} \triangleq (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$. From (4.15), we conclude that $\boldsymbol{z}_0 = \left(\boldsymbol{\alpha}, [0, \mathcal{I}_2, \ldots, \mathcal{I}_J]^T\right)$ and $\boldsymbol{z}_1 = \left(q\boldsymbol{\alpha}, q[\mathcal{I}_2, \ldots, \mathcal{I}_J, 0]^T\right)$ have the same noiseless output $\boldsymbol{\mathcal{I}}^{\text{out}}$ and thus the same NLL. Hence, the partial derivative of $\mathcal{L}(\boldsymbol{z}) \triangleq \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ over $\boldsymbol{\alpha}$ at $\boldsymbol{z}_1$ is $1/q$ times that at $\boldsymbol{z}_0$. Meanwhile, the subgradients of the regularization term at $\boldsymbol{z}_0$ and $\boldsymbol{z}_1$ with respect to $\boldsymbol{\alpha}$ are the same. So, for the same regularization u, it is easier for the penalty term to dominate the subgradient of $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ around $\boldsymbol{z}_1$ than $\boldsymbol{z}_0$. This is also experimentally confirmed: we see that, upon initialization $\boldsymbol{\alpha}^{(0)} = q^{\ell}\boldsymbol{\alpha}$ with some $\boldsymbol{\alpha}$ and large ℓ, the magnitude of the iterates $\boldsymbol{\alpha}^{(i)}$ reduces as the iteration proceeds.

We now show that the objective function (4.25b) satisfies the KL property [ABRS10], which is important for establishing local convergence of block-coordinate schemes in biconvex optimization problems. The KL property [ABRS10] regularizes the (sub)gradient of a function through its value at a certain point or over the whole domain and also ensures the steepness of the function around the optimum so that the length of the gradient trajectory is bounded.

Theorem 4.2 (KL Property)**.** *The objective function* $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *satisfies the KL property in any compact subset* $\mathbb{C} \subseteq \mathrm{dom}(f)$.

Proof: See Appendix 4.C. □

Note that all $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ that lead to positive noiseless measurements, i.e. $\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \succ \mathbf{0}$, are in the domain of f, which excludes the case $\boldsymbol{\mathcal{I}} = \mathbf{0}$ when no incident X-ray is applied; see also (4.12b).

4.5.2 Minimization Algorithm

The parameters that we wish to estimate are naturally divided into two blocks, $\boldsymbol{\alpha}$ and $\mathcal{I}$. The large size of $\boldsymbol{\alpha}$ prohibits effective second-order methods under sparsity regularization, whereas $\mathcal{I}$ has much smaller size and only nonnegative constraints, thus allowing for more sophisticated solvers, such as the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach [Thi89, Sec. 4.3.3.4] that we adopt here. In addition, the scaling difference between $\boldsymbol{\alpha}$ and $\mathcal{I}$ can be significant, so that the joint gradient method for $\boldsymbol{\alpha}$ and $\mathcal{I}$ together would converge slowly. Therefore, we adopt a block coordinate-descent algorithm to minimize $f(\boldsymbol{\alpha}, \mathcal{I})$ in (4.25b), where the NPG [Nes83; BT09a] and L-BFGS-B [BLNZ95] methods are employed to update estimates of the density-map and mass-attenuation spectrum parameters, respectively. The choice of block coordinate-descent optimization is also motivated by the related alternate convex search (ACS) and block coordinate-descent schemes in [GPK07] and [XY13], respectively, both with convergence guarantees under certain conditions.

We minimize the objective function (4.25b) by alternatively updating $\boldsymbol{\alpha}$ and $\mathcal{I}$ using Steps 1 and 2, respectively, where Iteration i proceeds as follows:

1) (NPG) Set the mass-attenuation spectrum $\iota(\kappa) = \boldsymbol{b}(\kappa)\mathcal{I}^{(i-1)}$, treat it as known[1], and descend the regularized NLL function $f(\boldsymbol{\alpha}, \mathcal{I}^{(i-1)}) = \mathcal{L}_\iota(\boldsymbol{\alpha}) + ur(\boldsymbol{\alpha})$ by applying an *NPG step* for $\boldsymbol{\alpha}$, which yields $\boldsymbol{\alpha}^{(i)}$:

$$\theta^{(i)} = \frac{1}{2}\left[1 + \sqrt{1 + 4\left(\theta^{(i-1)}\right)^2}\right] \tag{4.26a}$$

$$\overline{\boldsymbol{\alpha}}^{(i)} = \boldsymbol{\alpha}^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}}\left(\boldsymbol{\alpha}^{(i-1)} - \boldsymbol{\alpha}^{(i-2)}\right) \tag{4.26b}$$

$$\boldsymbol{\alpha}^{(i)} = \operatorname{prox}_{\beta^{(i)} ur}\left(\overline{\boldsymbol{\alpha}}^{(i)} - \beta^{(i)} \nabla \mathcal{L}_\iota\left(\overline{\boldsymbol{\alpha}}^{(i)}\right)\right) \tag{4.26c}$$

where the minimization (4.26c) is computed using an inner iteration that employs the TV-based denoising method in [BT09b, Sec. IV], and $\beta^{(i)} > 0$ is an adaptive step size chosen to satisfy

[1] This selection corresponds to $\mathcal{L}_\iota(\boldsymbol{\alpha}) = \mathcal{L}\left(\boldsymbol{\alpha}, \mathcal{I}^{(i-1)}\right)$; see also (4.18).

the *majorization condition*:

$$\mathcal{L}_\iota(\boldsymbol{\alpha}^{(i)}) \leq \mathcal{L}_\iota(\overline{\boldsymbol{\alpha}}^{(i)}) + (\boldsymbol{\alpha}^{(i)} - \overline{\boldsymbol{\alpha}}^{(i)})^T \nabla \mathcal{L}_\iota(\overline{\boldsymbol{\alpha}}^{(i)}) + \frac{1}{2\beta^{(i)}} \|\boldsymbol{\alpha}^{(i)} - \overline{\boldsymbol{\alpha}}^{(i)}\|_2^2 \qquad (4.26d)$$

using an adaptation scheme [GD15c] that aims at finding the largest $\beta^{(i)}$ that satisfies (4.26d):

i) • if there have been no step-size backtracking events or increase attempts for $\mathtt{m}$ consecutive iterations ($i - \mathtt{m}$ to $i - 1$), start with a larger step size $\overline{\beta}^{(i)} = \beta^{(i-1)}/\xi$ where $\xi \in (0, 1)$ is a *step-size adaptation parameter*;

• otherwise start with $\overline{\beta}^{(i)} = \beta^{(i-1)}$;

ii) (backtracking search) select

$$\beta^{(i)} = \xi^{t_i} \overline{\beta}^{(i)} \qquad (4.27a)$$

where $t_i \geq 0$ is the smallest integer such that (4.27a) satisfies the majorization condition (4.26d); *backtracking event* corresponds to $t_i > 0$.

We select the initial step size $\overline{\beta}^{(0)}$ using the BB method [BB88]. We also apply the *function restart* [OC15] to restore the monotonicity and improve convergence; see the following discussion.

2) (BFGS) Set the design matrix $A = \boldsymbol{b}_\circ^{\mathrm{L}}\big(\Phi\boldsymbol{\alpha}^{(i)}\big)$, treat it as known[2], and minimize the regularized NLL function $f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}\right)$ with respect to $\boldsymbol{\mathcal{I}}$; i.e.,

$$\boldsymbol{\mathcal{I}}^{(i)} = \arg\min_{\boldsymbol{\mathcal{I}} \geq \mathbf{0}} \mathcal{L}_A(\boldsymbol{\mathcal{I}}) \qquad (4.28)$$

using the inner L-BFGS-B iteration, initialized by $\boldsymbol{\mathcal{I}}^{(i-1)}$.

[2] This selection corresponds to $\mathcal{L}_A(\boldsymbol{\mathcal{I}}) = \mathcal{L}\big(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}\big)$; see also (4.22b).

Iterate between Steps 1 and 2 until the relative distance of consecutive iterates of the density map $\boldsymbol{\alpha}$ does not change significantly:

$$\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\|_2 < \epsilon \|\boldsymbol{\alpha}^{(i)}\|_2, \tag{4.29}$$

where $\epsilon > 0$ is the convergence threshold. The convergence criteria for the inner TV-denoising and L-BFGS-B iterations in Steps 1 and 2 are chosen to trade off the accuracy and speed of the inner iterations and provide sufficiently accurate solutions to (4.26c) and (4.28); see [GD15b, Sec. IV-B2] for details.

We refer to the iteration between Steps 1 and 2 as the *NPG-BFGS algorithm*: it is the first physical-model–based image reconstruction method for simultaneous *blind* sparse image reconstruction and mass-attenuation spectrum estimation from polychromatic measurements; see also our preliminary work in [GD13]. In [GD13], we approximated Laplace integrals with Riemann sums, used a smooth approximation of the nonnegativity penalties in (4.25c), and did not employ signal-sparsity regularization.

If the mass-attenuation spectrum $\iota(\kappa)$ is *known* and we iterate Step 1 only to estimate the density-map image $\boldsymbol{\alpha}$, we refer to this iteration as the *NPG algorithm (known $\iota(\kappa)$)*.

If we *do not* apply the Nesterov's acceleration (4.26a)–(4.26b) and use only the PG step (4.26c) to update the density-map iterates $\boldsymbol{\alpha}$, i.e., assign (4.31c) instead of (4.26b) *in every iteration*, we refer to the corresponding iteration as the *PG-BFGS algorithm*.

Scale-and-shift adjustment of the NPG-BFGS and PG-BFGS estimates. Denote by $\widehat{\boldsymbol{\mathcal{I}}}$ and $\widehat{\boldsymbol{\alpha}}$ the mass-attenuation spectrum parameter and density-map image estimates upon convergence of the NPG-BFGS iteration. To emphasize the dependence of the objective function (4.25b) on u, we denote it here by $f_u(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$. If the last element $\widehat{\mathcal{I}}_J$ of $\widehat{\boldsymbol{\mathcal{I}}}$ is zero, we can trivially improve this objective function by using the shift ambiguity: remove this zero element by circularly shifting $\widehat{\boldsymbol{\mathcal{I}}}$ and divide $\widehat{\boldsymbol{\mathcal{I}}}$ and $\widehat{\boldsymbol{\alpha}}$ by q; after this adjustment, we would need to continue the NPG-BFGS iteration and seek the new local minimum. However, we can avoid additional iteration and simply adjust

the regularization constant u *as well as* $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\mathcal{I}}$ by assigning new values to them: $(u, \widehat{\boldsymbol{\alpha}}, \widehat{\mathcal{I}}) \leftarrow \left(qu, \widehat{\boldsymbol{\alpha}}/q, [0, \widehat{\mathcal{I}}_1, \ldots, \widehat{\mathcal{I}}_{J-1}]^T/q\right)$. Apply this adjustment sequentially until the last element of the new $\widehat{\mathcal{I}}$ is nonzero, which yields a local minimum $\left(\widehat{\boldsymbol{\alpha}}, \widehat{\mathcal{I}}\right)$ of the new objective function $f_u(\boldsymbol{\alpha}, \mathcal{I})$ that is *not possible* to improve on by a simple shift adjustment. Our empirical experience is that scale-and-shift adjustment is either not needed (no zero elements at the end of $\widehat{\mathcal{I}}$) or minor (very few zero elements): it slightly changes the grid of u over which we search for the best reconstructions; see also Section 4.6 for discussion on selection of u. The key insight is that blind methods *cannot* estimate the magnitude level in density-map reconstructions; see also Fig. 4.4b in Section 4.6. Hence, difference in density-map magnitude level that these methods exhibit is *not significant*.

4.5.3 Function Restart and Monotonicity

If $f\left(\boldsymbol{\alpha}, \mathcal{I}^{(i-1)}\right)$ is a convex function of $\boldsymbol{\alpha}$, apply [BT09a, Lemma 2.3] to establish that the iterate $\boldsymbol{\alpha}^{(i)}$ attains lower (or equal) objective function than the intermediate signal $\overline{\boldsymbol{\alpha}}^{(i)}$

$$f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i-1)}\right) \leq f\left(\overline{\boldsymbol{\alpha}}^{(i)}, \mathcal{I}^{(i-1)}\right) - \frac{1}{2\beta^{(i)}} \left\|\boldsymbol{\alpha}^{(i)} - \overline{\boldsymbol{\alpha}}^{(i)}\right\|_2^2, \tag{4.30}$$

where we have used the fact that step size $\beta^{(i)}$ satisfies the majorization condition (4.26d). However, (4.30) *does not* guarantee monotonicity of Step 1. We apply the function restart [OC15] to ensure this monotonicity and improve convergence. In particular, we apply the function restart as follows: if monotonicity of Step 1 is violated in Iteration i, i.e., if

$$f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i-1)}\right) > f\left(\boldsymbol{\alpha}^{(i-1)}, \mathcal{I}^{(i-1)}\right) \qquad \text{(restart cond.)} \tag{4.31a}$$

set

$$\theta^{(i-1)} = 1 \tag{4.31b}$$

and *repeat* Step 1 using this selection. In this repeated step, the *momentum term* $\frac{\theta^{(i-1)}-1}{\theta^{(i)}}\left(\boldsymbol{\alpha}^{(i-1)} - \boldsymbol{\alpha}^{(i-2)}\right)$ in (4.26b) becomes zero, and

$$\bar{\boldsymbol{\alpha}}^{(i)} = \boldsymbol{\alpha}^{(i-1)} \tag{4.31c}$$

holds. Consequently, the new Step 1 is *monotonic*:

$$f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i-1)}\right) \leq f\left(\boldsymbol{\alpha}^{(i-1)}, \mathcal{I}^{(i-1)}\right), \tag{4.31d}$$

which follows by substituting (4.31c) into (4.30).

Once we can guarantee the monotonicity of Step 1 in every Iteration i, it is easy to establish the monotonicity of the entire NPG-BFGS iteration:

Remark 4.2 (Monotonicity)**.** Under condition (4.24) of Theorem 4.1, the NPG-BFGS iteration with function restart is monotonically non-increasing:

$$f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)}\right) \leq f\left(\boldsymbol{\alpha}^{(i-1)}, \mathcal{I}^{(i-1)}\right) \tag{4.32}$$

for all i.

Proof: Under condition (4.24), $f\left(\boldsymbol{\alpha}, \mathcal{I}\right)$ is a convex function of $\boldsymbol{\alpha}$. In this case, we have established that (4.31d) holds and Step 1 is monotonic. By Step 2, $f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i-1)}\right) \geq f\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)}\right)$ and (4.32) follows. □

Clearly, PG-BFGS and NPG (for known $\iota(\kappa)$) are monotonic as well under the convexity condition (4.24). To derive the monotonicity results, we have used only the fact that step size $\beta^{(i)}$ satisfies the majorization condition (4.26d), rather than using any specific details of the step-size selection.

Note that the conditions of Theorem 4.1 are only *sufficient* for establishing the convexity of $f\left(\boldsymbol{\alpha}, \mathcal{I}\right)$ as a function of $\boldsymbol{\alpha}$.

In the following, we show that our PG-BFGS algorithm converges to a critical point of the objective function; interestingly, this convergence analysis *does not* require convexity of the objective function with respect to $\boldsymbol{\alpha}$. Unfortunately, these theoretical convergence properties do not carry over to the NPG-BFGS iteration, which empirically outperforms the PG-BFGS method; see Figs. 4.5 and 4.10 in Section 4.6.

4.5.4 Convergence Analysis of the PG-BFGS Iteration

We analyze the convergence of the PG-BFGS iteration using arguments similar to those in [XY13]. Although NPG-BFGS converges faster than PG-BFGS empirically, it is not easy to analyze its convergence due to NPG's Nesterov's acceleration step and adaptive step size. In this section, we denote the sequence of PG-BFGS iterates by $\{(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)})\}_{i=0}^{\infty}$.

We have established the monotonicity of the PG-BFGS iteration for step sizes $\beta^{(i)}$ that satisfy the majorization condition, which includes the above step-size selection as well.

Since our $f(\boldsymbol{\alpha}, \mathcal{I})$ are lower bounded (which is easy to argue; see Appendix 4.C), the sequence $f(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)})$ converges. It is also easy to conclude that the sequence $a_i \triangleq \|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\|_2^2 / \beta^{(i)}$ is Cauchy by showing $\sum_{i=0}^{\infty} a_i < +\infty$ according to (4.30) when (4.31c) holds. Thus $\boldsymbol{\alpha}^{(i)}$ converges if $\{\beta^{(i)}\}_{i=1}^{\infty}$ is upper bounded.

A better result $\sum_{i=0}^{\infty} \|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i+1)}\|_2 < +\infty$ [XY13] can be established because $f(\boldsymbol{\alpha}, \mathcal{I})$ satisfies the KL property. This property has been first used in [ABRS10] to establish the critical-point convergence for an alternating proximal-minimization method, which is then extended in [XY13] to the more general block coordinate-descent method. Using the analysis in [ABRS10], [LZZ+15] shows the convergence of the alternating proximal-minimization algorithm by applying the KL property to a biconvex objective function.

Next, we make the following claim on the convergence of the PG-BFGS iteration.

Theorem 4.3. *Consider the sequence* $\{(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)})\}_{i=0}^{\infty}$ *of PG-BFGS iterates, with step size* $\beta^{(i)}$ *satisfying the majorization condition* (4.26d)*. Assume*

1) bounded step size: there exist positive $\beta_+ > \beta_- > 0$ *such that* $\beta^{(i)} \in [\beta_-, \beta_+]$ *for all* i*,*

2) $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *is a strongly convex function of* $\boldsymbol{\mathcal{I}}$*, and*

3) the gradient of $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *with respect to* $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *is Lipschitz continuous.*

Then $\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i)}\right)$ *converges to one of the critical points* $(\boldsymbol{\alpha}^\star, \boldsymbol{\mathcal{I}}^\star)$ *of* $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ *and*

$$\sum_{i=1}^{\infty} \left\| \boldsymbol{\alpha}^{(i+1)} - \boldsymbol{\alpha}^{(i)} \right\|_2 < +\infty, \qquad \sum_{i=1}^{\infty} \left\| \boldsymbol{\mathcal{I}}^{(i+1)} - \boldsymbol{\mathcal{I}}^{(i)} \right\|_2 < +\infty. \tag{4.33}$$

Proof: We apply [XY13, Lemma 2.6] to establish the convergence of $\left\{\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i)}\right)\right\}_{i=1}^{+\infty}$. Since $r(\boldsymbol{\alpha})$ in (4.25c) and $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\mathcal{I}})$ are lower-bounded, we need to prove only that (4.16) is lower-bounded. By using the fact that $\ln x \leq x - 1$, we have

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \geq 0. \tag{4.34}$$

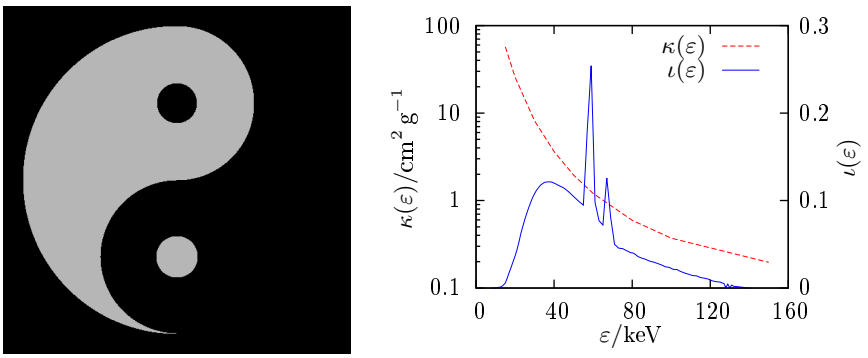
According to the assumption, $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ is strongly convex over $\boldsymbol{\mathcal{I}}$ and the step size $\beta^{(i)}$ is bounded. Hence, there exist constants $0 < \ell < L < +\infty$ such that

$$f\left(\boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\mathcal{I}}^{(i)}\right) - f\left(\boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\mathcal{I}}^{(i+1)}\right) \geq \frac{\ell}{2} \left\| \boldsymbol{\mathcal{I}}^{(i)} - \boldsymbol{\mathcal{I}}^{(i+1)} \right\|_2^2 \tag{4.35a}$$

$$L \geq \frac{1}{\beta^{(i)}} \geq \ell. \tag{4.35b}$$

In addition, $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ satisfies the KL property according to Theorem 4.2. We have now verified all conditions of [XY13, Lemma 2.6]. □

The conditions for strong convexity of $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ as a function of $\boldsymbol{\mathcal{I}}$ are discussed in Section 4.4; see also Section 4.3.3. The KL property can provide guarantees on the convergence rate under additional assumptions; see [ABRS10, Theorem 3.4]. The convergence properties of NPG-BFGS are of great interest because NPG-BFGS converges faster than PG-BFGS; establishing these properties is left as future work.



(a) (b)

Figure 4.2: (a) Density-map image used to generate the sinogram, and (b) mass attenuation and incident X-ray spectrum as functions of the photon energy ε.

4.6 Numerical Examples

We now evaluate the proposed algorithms using simulated and real-data examples.

We construct the fan-beam X-ray projection transform matrix Φ and its adjoint operator Φ^T directly on GPU with circular masks [DGQ11]; the multi-thread version on CPU is also available; see `https://github.com/isucsp/imgRecSrc`, which also contains Matlab implementation of the proposed algorithms.

4.6.1 Simulation Example

Consider the reconstruction of the 512×512 image in Fig. 4.2a of an iron object with density-map $\boldsymbol{\alpha}_{\text{true}}$. We generated a fan-beam polychromatic sinogram, with distance from the X-ray source to the rotation center equal to 2000 times the pixel size, using the interpolated mass attenuation $\kappa(\varepsilon)$ of iron [HS95] and the incident spectrum $\iota(\varepsilon)$ from tungsten anode X-ray tubes at 140 keV with 5 % relative voltage ripple [BS97]; see Fig. 4.2b. The mass-attenuation spectrum $\iota(\kappa)$ is constructed by combining $\kappa(\varepsilon)$ and $\iota(\varepsilon)$ and shown in Fig. 4.1b, see also Fig. 4.1a. Our simulated approximation of the noiseless measurements uses 130 equi-spaced discretization points over the range 20 keV to 140 keV. We simulated independent Poisson measurements $(\mathcal{E}_n)_{n=1}^N$ with means $(\mathrm{E}\,\mathcal{E}_n)_{n=1}^N = \boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$. We mimic real X-ray CT system calibration by scaling projection matrix

Φ and spectrum $\iota(\varepsilon)$ so that the maximum and minimum of the noiseless measurements $(\mathrm{E}\,\mathcal{E}_n)_{n=1}^N$ are 2^{16} and 20, respectively. Here, the scale of Φ corresponds to the real size that each image pixel represents, and the scale of $\iota(\varepsilon)$ corresponds to the current of the electrons hitting the tungsten anode as well as the overall scanning time.

Our goal is to reconstruct a 512×512 density-map using the measurements from an energy-integrating detector array of size 512 for each projection.

Since the true density-map is known, we adopt RSE as the main metric to assess the performance of the compared algorithms:

$$\mathrm{RSE}\{\widehat{\boldsymbol{\alpha}}\} = 1 - \left(\frac{\widehat{\boldsymbol{\alpha}}^T \boldsymbol{\alpha}_{\text{true}}}{\|\widehat{\boldsymbol{\alpha}}\|_2 \|\boldsymbol{\alpha}_{\text{true}}\|_2} \right)^2 \tag{4.36}$$

where $\boldsymbol{\alpha}_{\text{true}}$ and $\widehat{\boldsymbol{\alpha}}$ are the true and reconstructed signals, respectively. Note that (4.36) is invariant to scaling $\widehat{\boldsymbol{\alpha}}$ by a nonzero constant, which is needed because the magnitude level of $\boldsymbol{\alpha}$ is not identifiable due to the ambiguity of the density-map and mass-attenuation spectrum; see Section 4.3.2.

We compare

- the traditional FBP methods
  - without linearization [KS88, Ch. 3] (termed FBP) and
  - with linearization to correct for the polychromatic source [Her79] (linearized FBP)

  based on the 'data'

$$\boldsymbol{y} = -\ln_\circ \boldsymbol{\mathcal{E}} \qquad \text{(without linearization)} \tag{4.37a}$$

$$\boldsymbol{y} = \left(\iota^{\mathrm{L}}\right)_\circ^{-1}(\boldsymbol{\mathcal{E}}) \qquad \text{(with linearization)} \tag{4.37b}$$

respectively;

- linearized basis pursuit denoising (linearized BPDN), which applies the NPG approach to solve the analysis BPDN problem [BT09b]: $\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{\alpha}\|_2^2 + u' r(\boldsymbol{\alpha})$, where $\boldsymbol{y}$ are the linearized measurements in (4.37b) and the penalty $r(\boldsymbol{\alpha})$ has been defined in (4.25c);
- our
  - NPG-BFGS algorithm with the B1-spline tuning constants (4.13e) chosen to satisfy[3]

$$q^J = 10^3, \qquad \kappa_{\lceil 0.5(J+1)\rceil} = 1, \qquad J = 30 \tag{4.38}$$

  - NPG (known $\iota(\kappa)$) algorithm for estimating $\boldsymbol{\alpha}$

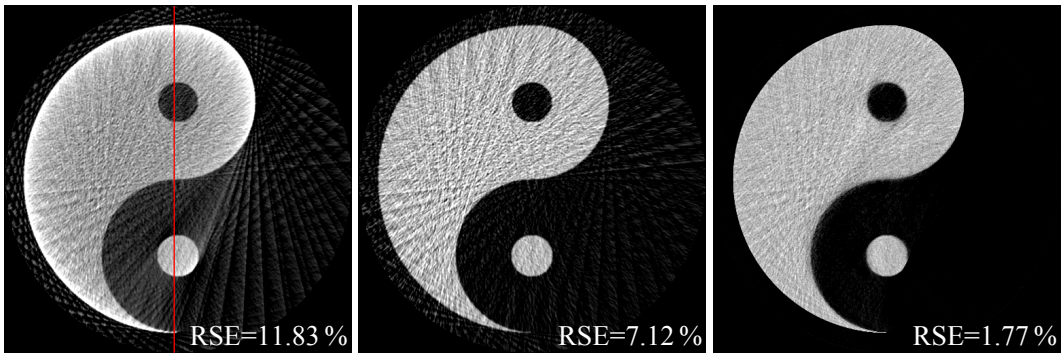
  with $\mathbb{m} = 4$; see Section 4.5.2.

The linearizing transform (4.37b) assumes knowledge of the mass-attenuation spectrum $\iota(\kappa)$ and, in the absence of noise, leads to the linear model $\boldsymbol{y} = \Phi\alpha$ under the general polychromatic-source scenario. In contrast, the standard logarithm transformation of the X-ray measurements (4.37a) *ignores* the hardening effect and can possibly lead to the linear model only for monochromatic X-ray sources. If the X-ray source is monochromatic, (4.37a) and (4.37b) *coincide* up to a known additive constant, and the two FBP methods are identical; in this case, linearized BPDN also coincides with the standard analysis BPDN approach applied to X-ray CT data.

For all methods that use sparsity and nonnegativity regularization (NPG-BFGS, NPG, and linearized BPDN), the regularization constants u and u' have been tuned manually for the best average RSE performance for each number of projections using a 9-point grid spanning 9 orders of magnitude.

All iterative algorithms employ the convergence criterion (4.29) with the threshold $\epsilon = 10^{-6}$ and the maximum number of iterations set to 4000. We initialize iterative reconstruction schemes with or without linearization using the corresponding FBP reconstructions; see also [GD15b, Sec. IV-B4] for details on NPG-BFGS initialization.

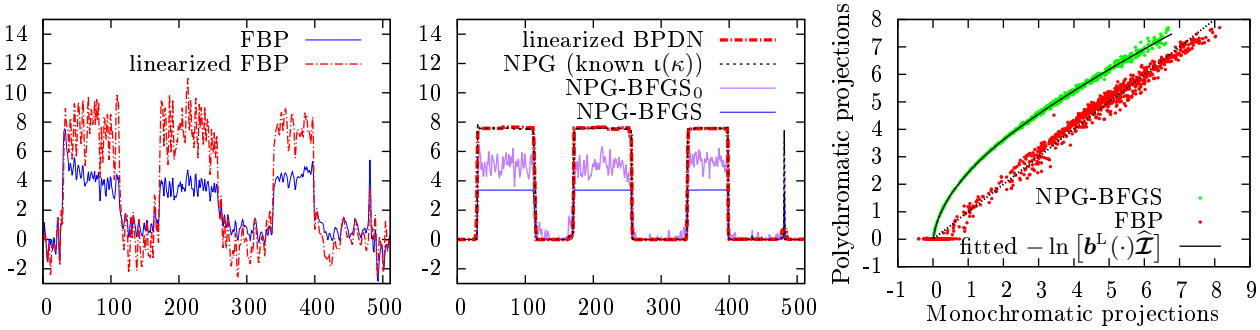
[3] This selection ensures sufficient coverage (three orders of magnitude) and resolution (30 basis functions) of the basis-function representation of the mass-attenuation spectrum and centers its support around 1.



(a) FBP (b) linearized FBP (c) NPG-BFGS$_0$

(d) linearized BPDN (e) NPG-BFGS and NPG (known $\iota(\kappa)$)

Figure 4.3: Reconstructions from 60 projections.



(a) 250th column (b) 250th column (c)

Figure 4.4: (a)–(b) Reconstruction profiles of different methods from 60 projections and (c) the polychromatic measurements as function of the monochromatic projections and corresponding fitted inverse linearization curves.

Here, the *non-blind* linearized FBP, NPG (known $\iota(\kappa)$), and linearized BPDN methods assume known $\iota(\kappa)$ (which requires knowledge of the incident spectrum of the X-ray machine and mass attenuation (material)), computed using (4.10a), with $\mathcal{I}$ equal to the exact sampled $\iota(\kappa)$ and $J = 100$ spline basis functions spanning three orders of magnitude.

Neither FBP nor NPG-BFGS assumes knowledge of the mass-attenuation spectrum $\iota(\kappa)$: FBP *ignores* the polychromatic-source effects whereas NPG-BFGS corrects *blindly* for these effects *without* knowledge of $\iota(\kappa)$.

Figs. 4.3 and 4.4 show the reconstructed density-map images and profiles of different methods from 60 equi-spaced fan-beam projections with spacing 6°, using one realization of noisy Poisson measurements. Fig. 4.5 shows the RSEs of several methods as functions of the iteration index i and demonstrates that RSE of NPG-BFGS decreases significantly faster with increasing i than the RSE of PG-BFGS; NPG-BFGS also converges faster than PG-BFGS. The FBP reconstruction in Fig. 4.3a is corrupted by both aliasing and beam-hardening (cupping and streaking) artifacts. Linearized FBP removes the beam-hardening artifacts but retains the aliasing artifacts and enhances noise due to the zero-forcing nature of linearization; see Fig. 4.3b. Linearized BPDN enforces the signal nonnegativity and sparsity constraints and achieves a smooth reconstruction in Fig. 4.3d with a 0.55 % RSE. Thanks to the superiority of the proposed model that accounts for both the polychromatic X-ray source and Poisson noise, NPG-BFGS and NPG achieve the best (and nearly the same) reconstructions; see Fig. 4.3e.

We also show in Fig. 4.3c the reconstruction by the NPG-BFGS method with very small u (labeled NPG-BFGS$_0$), which effectively removes the signal sparsity constraint and imposes only the signal nonnegativity constraint; consequently, Step 1 in NPG-BFGS$_0$ iteration has a closed form and reduces to simple nonnegativity thresholding. Hence, NPG-BFGS$_0$ is a maximum-likelihood (ML) approach that aims at minimizing the NLL (4.16) subject to the physical parameter constraints $\boldsymbol{\alpha} \succeq \mathbf{0}$ and $\mathcal{I} \succeq \mathbf{0}$. As NPG-BFGS$_0$ iterates, its RSE decreases, reaches a minimum, and then increases; see Fig. 4.5. This is a common behavior for unregularized ML image reconstruction approaches [LV89]. Fig. 4.3c shows this method's reconstruction at iteration step $i = 500$,

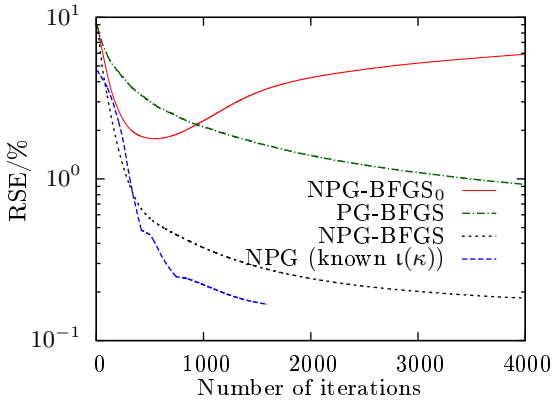


Figure 4.5: The RSEs as functions of the iteration index i.

which gives the best RSE; see also Fig. 4.5. Since it terminates early and has a simple Step 1, NPG-BFGS$_0$ running only 500 iterations is roughly 8 times faster than NPG-BFGS. The NPG-BFGS$_0$ method can be thought of as an improved version of [GD13], which also imposes only signal non-negativity. A comparison of NPG-BFGS$_0$ and NPG-BFGS shows the benefit of signal-sparsity regularization.

Figs. 4.4a and 4.4b show the reconstruction profiles of the 250th column, indicated by the red line in Fig. 4.3a. Note that the "tail" in the linearized BPDN reconstruction in Fig. 4.3d fades quickly and does not maintain the sharp end; see also the small bump in its corresponding profile in Fig. 4.4b. Recall that NPG-BFGS *cannot* identify the magnitude level of the density-map image $\boldsymbol{\alpha}$, which explains the corresponding magnitude discrepancy between NPG-BFGS, NPG-BFGS$_0$, and the non-blind methods in Fig. 4.4b. We have corrected this discrepancy manually in Fig. 4.3 because we wish to show visual quality and ability of different methods to remove artifacts and suppress noise, rather than the trivial difference in image contrast.

In Fig. 4.4c, we show the scatter plots with 1000 randomly selected points representing FBP and NPG-BFGS reconstructions from 60 fan-beam projections. Denote by $(\widehat{\boldsymbol{\alpha}}, \widehat{\mathcal{I}})$ the estimate of $(\boldsymbol{\alpha}, \mathcal{I})$ obtained upon convergence of the NPG-BFGS iteration. The y-coordinates in the scatter plots in Fig. 4.4c are the *noisy* measurements in log scale $-\ln \mathcal{E}_n$, and the corresponding x-coordinates are the monochromatic projections $\boldsymbol{\phi}_n^T \widehat{\boldsymbol{\alpha}}_{\text{FBP}}$ (red) and $\boldsymbol{\phi}_n^T \widehat{\boldsymbol{\alpha}}$ (green) of the estimated

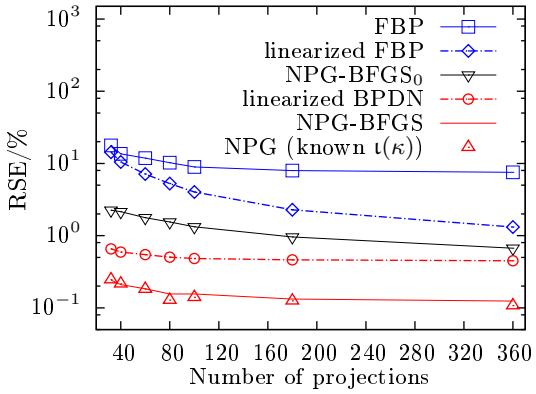


Figure 4.6: Average RSEs as functions of the number of projections.

density-maps. $-\ln\left[\boldsymbol{b}^{\mathrm{L}}(\cdot)\widehat{\boldsymbol{\mathcal{I}}}\right]$ is the inverse linearization function that maps monochromatic projections to fitted *noiseless* polychromatic projections $-\ln \mathcal{I}_n^{\mathrm{out}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\mathcal{I}}})$. The vertical-direction differences between the NPG-BFGS scatter plot and the corresponding linearization curve show goodness of fit between the measurements and our model.

Since FBP assumes a linear relation between $-\ln_{\circ} \boldsymbol{\mathcal{I}}^{\mathrm{out}}$ and $\Phi\boldsymbol{\alpha}$, its scatter plot (red) can be fitted by a straight line $y = x$, as shown in Fig. 4.4c. A few points in the FBP scatter plot with $\ln \mathcal{E}_n = 0$ and positive monochromatic projections indicate severe streaking artifacts. Observe relatively large residuals with bias, which remain even when more sophisticated linear models, e.g. iterative algorithms with sparsity and nonnegativity constraints, were adopted, thereby necessitating the need for accounting for the polychromatic source. The nonnegativity constraints on $\boldsymbol{\alpha}$ are particularly important for good estimation of $\boldsymbol{b}^{\mathrm{L}}(\cdot)\boldsymbol{\mathcal{I}}$.

Fig. 4.6 shows the average RSEs (over 5 Poisson noise realizations) of different methods as functions of the number of fan-beam projections in the range from 0° to 359°. Average RSEs of the methods that do not assume knowledge of the mass-attenuation spectrum $\iota(\kappa)$ are shown using solid lines; dashed lines represent non-blind methods that assume known mass-attenuation spectrum $\iota(\kappa)$. Red color represents methods that employ both signal-sparsity regularization and nonnegativity image constraints, black is for the method that employs the nonnegativity image

constraints only, and blue marks the methods that apply neither signal-sparsity regularization nor nonnegativity image constraints.

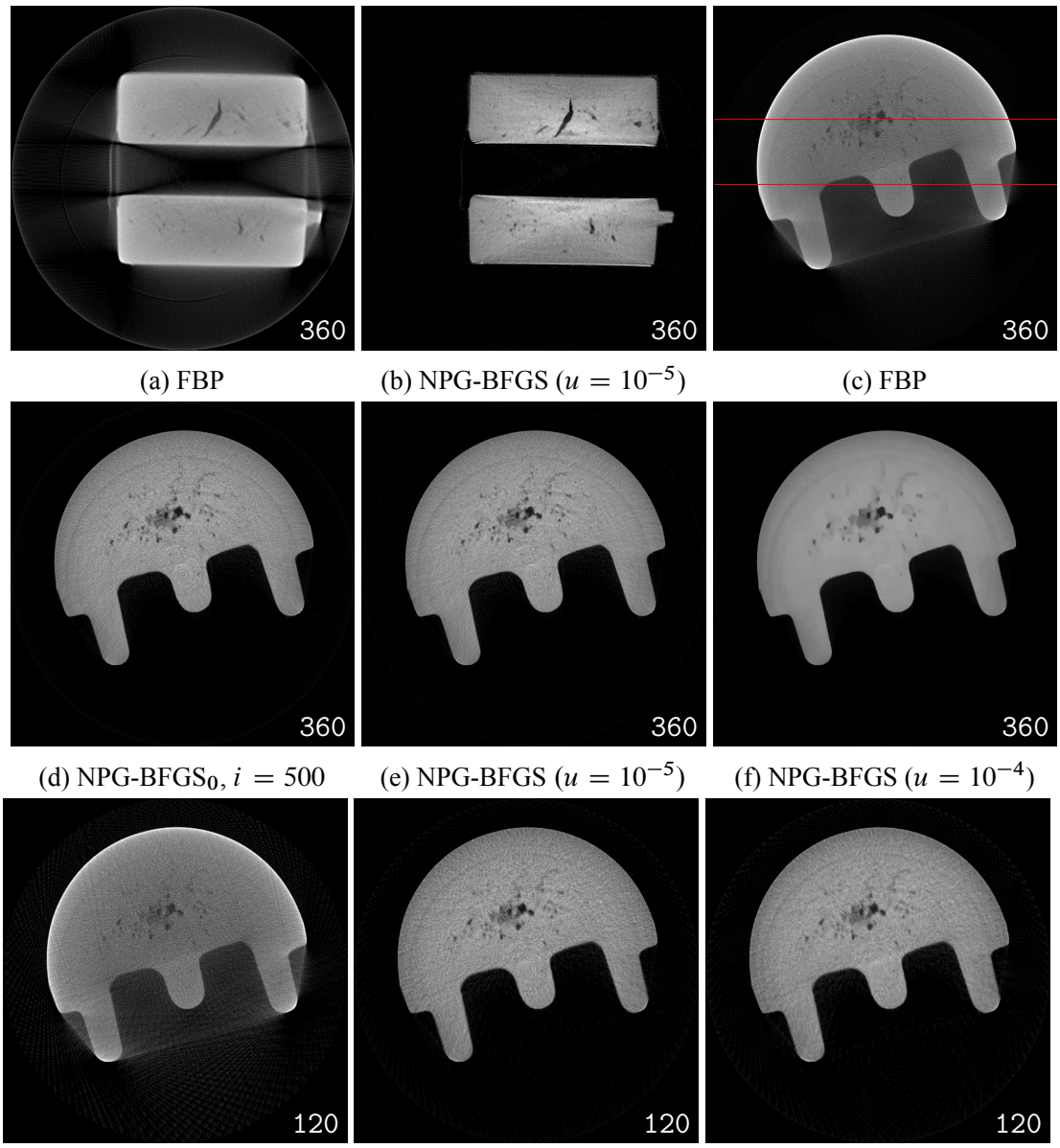
FBP ignores the polychromatic nature of the measurements; consequently, it performs poorly and does not improve as the number of projections increases. Linearized FBP, which assumes perfect knowledge of the mass-attenuation spectrum, performs much better than FBP, as shown in Fig. 4.6. Thanks to the signal nonnegativity and sparsity that it imposes, linearized BPDN achieves up to 20 times smaller RSEs compared with the linearized FBP. However, due to its zero-forcing nature, linearized BPDN enhances noise and breaches the Poisson measurement model, which explains its inferior performance compared with NPG (known $\iota(\kappa)$). Linearized BPDN exhibits a noise floor as the number of projections increases.

FBP ignores the polychromatic nature of the measurements; consequently, it performs poorly and does not improve as the number of projections increases. Linearized FBP, which assumes perfect knowledge of the mass-attenuation spectrum, performs much better than FBP, as shown in Fig. 4.6. Thanks to the signal nonnegativity and sparsity that it imposes, linearized BPDN achieves up to 20 times smaller RSEs compared with the linearized FBP. However, due to its zero-forcing nature, linearized BPDN enhances noise and breaches the Poisson measurement model, which explains its inferior performance compared with NPG (known $\iota(\kappa)$).

As expected, NPG (known $\iota(\kappa)$) performs slightly better than NPG-BFGS because it uses perfect knowledge of $\iota(\kappa)$. NPG (known $\iota(\kappa)$) and NPG-BFGS attain RSEs that are 24 % to 37 % of that achieved by linearized BPDN, which can be attributed to optimal statistical processing by these methods, in contrast with the suboptimal linearization. RSEs of NPG (known $\iota(\kappa)$) and NPG-BFGS reach a noise floor when the number of projections increases beyond 180. It is remarkable that the blind NPG-BFGS method effectively matches the performance of NPG (known $\iota(\kappa)$).

4.6.2 Real-Data Examples

We compare the NPG-BFGS and linear FBP methods by applying them to reconstruct two industrial objects containing defects, labeled `C-I` and `C-II`, from real fan-beam projections. Here,

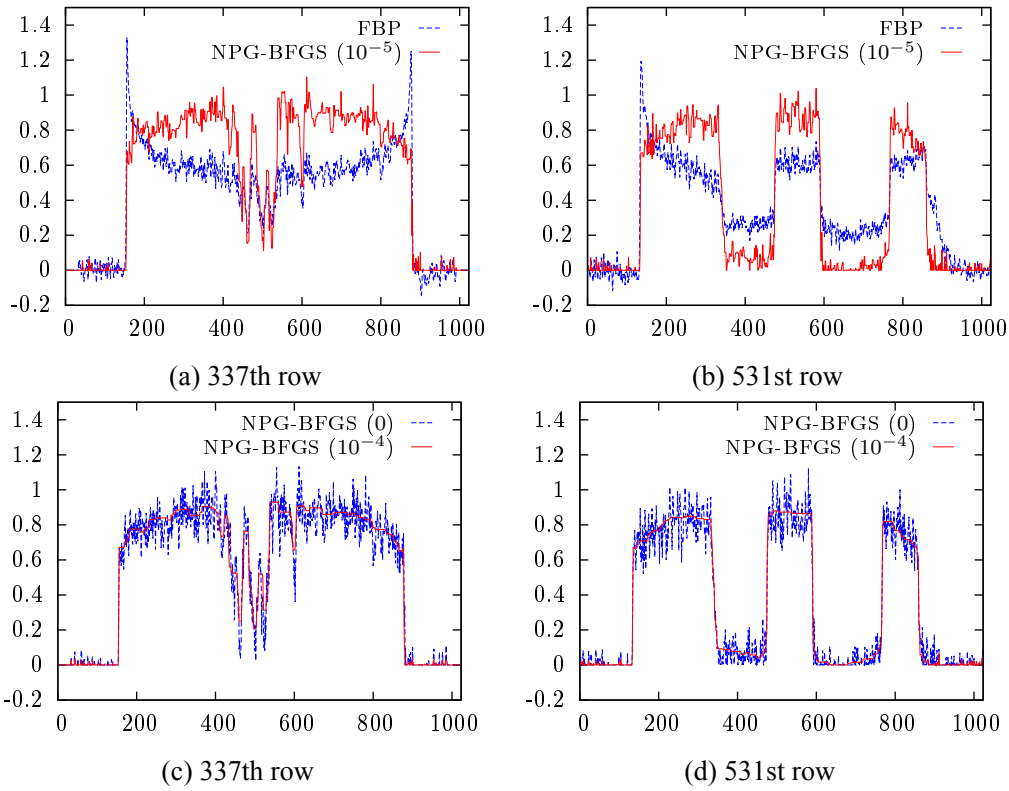


(a) FBP (b) NPG-BFGS ($u = 10^{-5}$) (c) FBP

(d) NPG-BFGS$_0$, $i = 500$ (e) NPG-BFGS ($u = 10^{-5}$) (f) NPG-BFGS ($u = 10^{-4}$)

(g) FBP (h) NPG-BFGS$_0$, $i = 500$ (i) NPG-BFGS ($u = 10^{-5}$)

Figure 4.7: Real X-ray CT reconstructions of objects `C-I` and `C-II` from (a)–(f) 360 and (g)–(h) 120 projections.



(a) 337th row (b) 531st row

(c) 337th row (d) 531st row

Figure 4.8: C-II object reconstruction profiles from 360 projections with (a)–(b) $u = 10^{-5}$ and (c)–(d) $u = 10^{-4}$ used by the NPG-BFGS method.

NPG-BFGS achieves visually good reconstructions for $u = 10^{-5}$, presented in Fig. 4.7, where we also show its reconstruction for $u = 10^{-4}$.

The C-I data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 694 detectors, with X-ray source to rotation center distance equal to 3492 times the detector size. Figs. 4.7a and 4.7b show 512×512 density-map image reconstructions of object C-I using the FBP and NPG-BFGS methods, respectively. The linear FBP reconstruction, which does not account for the polychromatic nature of the X-ray source, suffers from severe streaking and cupping artifacts, whereas the NPG-BFGS reconstruction removes these artifacts by accounting for the polychromatic X-ray source.

The C-II data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 1380 detectors, with X-ray source to rotation center distance equal to 8696 times the detector size. Figs. 4.7c–4.7e show 1024×1024 density-map image reconstructions of

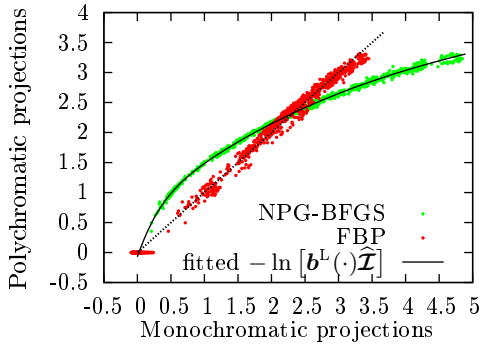


Figure 4.9: Polychromatic measurements as functions of monochromatic projections and corresponding inverse linearization curves.

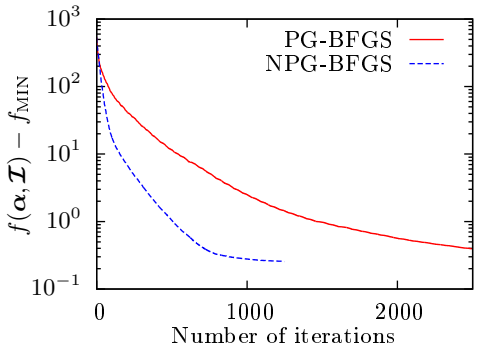


Figure 4.10: Centered objectives as functions of the iteration index i.

object `C-II` by the FBP, NPG-BFGS$_0$, and NPG-BFGS methods, respectively. The NPG-BFGS and NPG-BFGS$_0$ reconstructions do not have streaking and cupping artifacts exhibited by FBP. NPG-BFGS$_0$ terminates after 500 iterations and is 2 to 3 times faster than NPG-BFGS.

Figs. 4.7g–4.7i show the FBP, NPG-BFGS$_0$ (terminated at $i = 500$ iterations), and NPG-BFGS reconstructions from a downsampled `C-II` data set with 120 equi-spaced fan-beam projections with 3° separation. The FBP reconstruction in Fig. 4.7g exhibits both beam-hardening and aliasing artifacts. In contrast, the NPG-BFGS reconstruction in Fig. 4.7i does not have these artifacts because it accounts for the polychromatic X-ray source and employs signal-sparsity regularization in (4.25c). Indeed, if we reduce regularization constant u sufficiently, the aliasing effect will occur in the NPG-BFGS reconstruction in Fig. 4.7i as well. A comparison of NPG-BFGS$_0$ and NPG-BFGS shows the benefit of signal-sparsity regularization, particularly its ability to re-

duce noise. If we run NPG-BFGS$_0$ beyond $i = 500$ iterations, it will exhibit aliasing artifacts, in addition to noise.

Fig. 4.8 shows the reconstruction profiles of the 337th and 531th rows highlighted by the red horizontal lines across Figs. 4.7c and 4.7e. Noise in the NPG-BFGS reconstructions can be reduced by increasing regularization parameter u: Figs. 4.8c and 4.8d show the corresponding NPG-BFGS reconstruction profiles for $u = 10^{-4}$, which is 10 times that in Figs. 4.8a and 4.8b.

The NPG-BFGS reconstructions of `C-I` and `C-II` have higher contrast around the inner region where cracks reside, which may be due to the detector saturation that leads to measurement truncation, scattering, noise-model mismatch, or the bowtie filter applied to the X-ray source. This effect is visible in the `C-I` reconstruction in Fig. 4.7b and is barely visible in the `C-II` reconstruction in Fig. 4.7e, but it can be observed in the profiles in Fig. 4.8. We leave further verification of causes and potential correction of this problem to future work and note that this issue does not occur in the simulated-data examples that we constructed; see Section 4.6.1.

In Fig. 4.9, we show the scatter plots with 1000 randomly selected points representing FBP and NPG-BFGS reconstructions of the `C-II` object from 360 projections. A few points in the FBP scatter plot with $\ln \mathcal{E}_n = 0$ and positive monochromatic projections indicate severe streaking artifacts, which we also observed in the simulation example; see Fig. 4.4c.

We now illustrate the advantage of using Nesterov's acceleration in Step 1 of NPG-BFGS. Fig. 4.10 shows the centered objective $f(\boldsymbol{\alpha}, \mathcal{I}) - f_{\text{MIN}}$ with $u = 10^{-5}$ as a function of the iteration index i for the NPG-BFGS and PG-BFGS methods applied to the `C-II` reconstruction from 360 projections; here $f_{\text{MIN}} = \min_{\boldsymbol{x}} f(\boldsymbol{x})$. Thanks to the Nesterov's acceleration (4.26b), NPG-BFGS is 2 to 3 times faster than PG-BFGS.

4.7 Conclusion

We developed a model for single-material beam-hardening artifact correction that requires no more information than the conventional FBP method. The proposed model relies on separability

of the attenuation to combine the variations of the mass attenuation and X-ray spectrum into the *mass-attenuation spectrum*. We

- used this model to develop a framework for reconstructing density-map images that are sparse in an appropriate transform domain from polychromatic CT measurements under the *blind* scenario where the material of the inspected object and incident-energy spectrum are unknown,
- established the KL property and gave sufficient conditions for the biconvexity of the underlying objective function with respect to the density-map and mass-attenuation spectrum parameters under the Poisson measurement scenario, and
- developed a block-coordinate descent algorithm for constrained minimization of this objective function.

Numerical experiments on both simulated and real X-ray CT data were presented. Our *blind* method for sparse X-ray CT reconstruction matches or outperforms non-blind linearization methods that assume perfect knowledge of the X-ray source and material properties. Future work will include extending our parsimonious polychromatic measurement-model parameterization to multiple materials [ZTB+14] and developing corresponding reconstruction algorithms.

Appendices

4.A Mass-Attenuation Parameterization

All mass-attenuation functions $\kappa(\varepsilon)$ encountered in practice can be divided into piecewise-continuous segments, where each segment is a differentiable monotonically decreasing function of ε; see [HS95, Tables 3 and 4] and [Hud10, Sec. 2.3]. The points of discontinuity in $\kappa(\varepsilon)$ are referred to as K-edges and are caused by the interaction between photons and K shell electrons, which occurs only when ε reaches the binding energy of the K shell electron. One example in Fig. 4.11 is the mass attenuation coefficient curve of iron with a single K-edge at 7.11 keV.

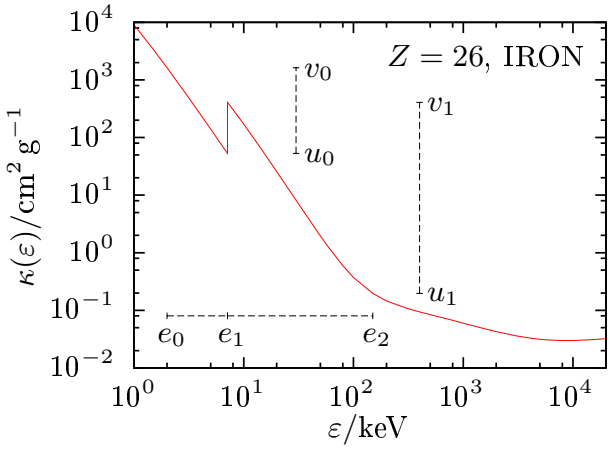


Figure 4.11: The mass attenuation coefficients κ of iron versus the photon energy ε with a K-edge at 7.11 keV.

We define the domain $\mathcal{E}$ of ε and partition it into $M+1$ intervals $\big((e_m, e_{m+1})\big)_{m=0}^{M}$ with $e_0 = \min(\mathcal{E})$ and $e_{M+1} = \max(\mathcal{E})$, such that in each interval $\kappa(\varepsilon)$ is invertible and differentiable. Here, $\mathcal{E}$ is the support set of the incident X-ray spectrum $\iota(\varepsilon)$ and $(e_m)_{m=1}^{M}$ are the M K-edges in $\mathcal{E}$. Taking Fig. 4.11 as an example, there is only one K-edge at e_1, given that the incident spectrum has its support as (e_0, e_2). The range and inverse of $\kappa(\varepsilon)$ within (e_m, e_{m+1}) are (u_m, v_m) and $\varepsilon_m(\kappa)$, respectively, with $u_m \triangleq \inf_{\varepsilon \nearrow e_{m+1}} \kappa(\varepsilon) < v_m \triangleq \sup_{\varepsilon \searrow e_m} \kappa(\varepsilon)$. Then, the noiseless measurement in (4.4b) can be written as

$$\mathcal{I}^{\text{out}} = \int \sum_{m=0}^{M} 1_{(u_m, v_m)}(\kappa)\iota(\varepsilon_m(\kappa)) \left|\varepsilon_m'(\kappa)\right| \mathrm{e}^{-\kappa \int \alpha(x,y)\,\mathrm{d}\ell}\,\mathrm{d}\kappa,$$

and (4.6b) and (4.6a) follow by noting that

$$\iota(\kappa) = \sum_{m=0}^{M} 1_{(u_m, v_m)}(\kappa)\iota(\varepsilon_m(\kappa)) \left|\varepsilon_m'(\kappa)\right| \geq 0 \tag{4.39}$$

and that $\mathcal{I}^{\text{out}}$ equals $\mathcal{I}^{\text{in}}$ when $\alpha(x, y) = 0$. Here, $1_{(u_m, v_m)}(\kappa)$ is an indicator function that takes value 1 when $\kappa \in (u_m, v_m)$ and 0 otherwise. Observe that (4.39) reduces to (4.7) when $M = 0$.

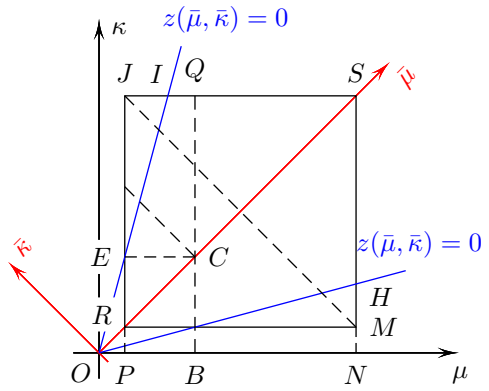


Figure 4.12: Integral region illustration.

4.B Proof of Lemma 4.1

We first introduce a lemma.

Lemma 4.3. *For $\iota(\kappa)$ that satisfy Assumption 4.1, the following holds:*

$$\begin{aligned} \mathrm{w} &\triangleq \iint \left[\mu\kappa - \frac{q^{j_0}}{(q^{j_0}+1)^2}(\mu+\kappa)^2\right]\iota(\kappa)\iota(\mu)h(\kappa+\mu)\,\mathrm{d}\mu\,\mathrm{d}\kappa \\ &\geq 0 \end{aligned} \tag{4.40}$$

for $q > 1$ and any nonnegative function $h : \mathbb{R} \to \mathbb{R}_+$.

Proof: In Fig. 4.12, the (μ, κ) coordinates of P, B and N are $(\kappa_0, 0)$, $(\kappa_{j_0}, 0)$ and $(\kappa_{J+1}, 0)$, respectively; the line OS is defined by $\kappa = \mu$.

Considering the finite support set of $\iota(\kappa)$, the effective integral range is $[\kappa_0, \kappa_{J+1}]^2$, which is the rectangle $RMSJ$ in Fig. 4.12. Using the symmetry between κ and μ in (4.40), we change the integral variables of (4.40) by rotating the coordinates by 90°:

$$\mu = \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}, \qquad \kappa = \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} \tag{4.41}$$

which yields

$$\mathrm{w} = \int_{\sqrt{2}\kappa_0}^{\sqrt{2}\kappa_{J+1}} \int_{0}^{g(\bar{\mu})} \bar{w}(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa}\; h\big(\sqrt{2}\bar{\mu}\big)\,\mathrm{d}\bar{\mu} \tag{4.42a}$$

where

$$\bar{w}(\bar{\mu},\bar{\kappa}) \triangleq z(\bar{\mu},\bar{\kappa})\iota\left(\frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}\right)\iota\left(\frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}\right) \tag{4.42b}$$

$$z(\bar{\mu},\bar{\kappa}) \triangleq \left(\frac{q^{j_0}-1}{q^{j_0}+1}\right)^2 \bar{\mu}^2 - \bar{\kappa}^2 \tag{4.42c}$$

$$g(\bar{\mu}) \triangleq \begin{cases} \bar{\mu} - \sqrt{2}\kappa_0, & \bar{\mu} \le \frac{1}{\sqrt{2}}(\kappa_0 + \kappa_{J+1}) \\ \sqrt{2}\kappa_{J+1} - \bar{\mu}, & \bar{\mu} > \frac{1}{\sqrt{2}}(\kappa_0 + \kappa_{J+1}) \end{cases} \tag{4.42d}$$

and (4.42a) follows because (4.42b) is even-symmetric with respect to $\bar{\kappa}$. Hence, the integration region is reduced to the triangle RSJ.

Note that $z(\bar{\mu},\bar{\kappa}) \ge 0$ in the cone between lines OH and OI, [both of which are specified by $z(\bar{\mu},\bar{\kappa}) = 0$], which implies that $\bar{w}(\bar{\mu},\bar{\kappa}) \ge 0$ within RCE and CSQ; hence, the integrals of $\bar{w}(\bar{\mu},\bar{\kappa})h\big(\sqrt{2}\bar{\mu}\big)$ over RCE and CSQ are nonnegative and, consequently,

$$\mathrm{w} \ge \iint_{\mathcal{R}} \bar{w}(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa}\; h\big(\sqrt{2}\bar{\mu}\big)\,\mathrm{d}\bar{\mu}. \tag{4.43}$$

Now

$$\mathcal{R} \triangleq \left\{ (\bar{\mu},\bar{\kappa}) \;\middle|\; \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}} \in [\kappa_0, \kappa_{j_0}],\ \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} \in [\kappa_{j_0}, \kappa_{J+1}] \right\} \tag{4.44}$$

is our new integration region, which is the rectangle $ECQJ$.

Next, we split the inner integral over $\bar{\kappa}$ on the right-hand side of (4.43) for fixed $\bar{\mu}$ into two regions: $z(\bar{\mu},\bar{\kappa}) \ge 0$ and $z(\bar{\mu},\bar{\kappa}) < 0$, i.e., trapezoid $ECQI$ and triangle EIJ, and prove that

the positive contribution of the integral over $ECQI$ is larger than the negative contribution of the integral over the EIJ.

The line OI is specified by $z(\bar{\mu},\bar{\kappa}) = 0$, and the (μ,κ)-coordinate of I in Fig. 4.12 is thus $(\kappa_{J+1-j_0}, \kappa_{J+1})$. Define

$$c \triangleq \frac{\sqrt{2}}{1+q^{j_0}} \tag{4.45}$$

and note that $ECQI \subseteq (\mathcal{K}_{\text{low}} \cup \mathcal{K}_{\text{mid}}) \times \mathcal{K}_{\text{high}}$ and $EIJ \subseteq \mathcal{K}_{\text{low}} \times \mathcal{K}_{\text{high}}$. We now use Assumption 4.1 to conclude that the following hold within $\mathcal{R}$:

- When $z(\bar{\mu},\bar{\kappa}) \geq 0$, i.e., in region $ECQI$,

$$\iota(\kappa)\,|_{\kappa=\frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}} \geq \iota\big(cq^{j_0}\bar{\mu}\big) \tag{4.46a}$$

$$\iota(\mu)\,|_{\mu=\frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}} \geq \iota\big(c\bar{\mu}\big) \tag{4.46b}$$

  where (4.46a) follows because $\kappa = \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}$ takes values between κ_{j_0} and $cq^{j_0}\bar{\mu} \in [\kappa_{j_0}, \kappa_{J+1}]$; i.e., $\kappa \in \mathcal{K}_{\text{high}}$ and $\iota(\kappa)$ decreases in $\mathcal{K}_{\text{high}}$. (4.46b) follows because $\mu = \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}$ takes values between $c\bar{\mu} \in [\kappa_0, \kappa_{J+1-j_0}]$ and κ_{j_0}; i.e., μ crosses $\mathcal{K}_{\text{low}}$ ($\iota(\kappa)$ increasing) and $\mathcal{K}_{\text{mid}}$ ($\iota(\kappa)$ high) regions. Here, $(c\bar{\mu}, cq^{j_0}\bar{\mu})$ is the (μ,κ)-coordinate of one point on line OI specified by $\bar{\mu}$ in $(\bar{\mu},\bar{\kappa})$-coordinate system.

- When $z(\bar{\mu},\bar{\kappa}) < 0$, i.e., in region EIJ,

$$\iota(\kappa)\,|_{\kappa=\frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}} < \iota\big(cq^{j_0}\bar{\mu}\big) \tag{4.46c}$$

$$\iota(\mu)\,|_{\mu=\frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}} < \iota\big(c\bar{\mu}\big) \tag{4.46d}$$

  where (4.46c) follows because $\kappa = \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} > cq^{j_0}\bar{\mu}$, i.e., $\kappa \in \mathcal{K}_{\text{high}}$, and (4.46d) follows because $\mu = \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}} < c\bar{\mu}$, i.e., $\mu \in \mathcal{K}_{\text{low}}$.

By combining (4.46) and (4.43), we have

$$\mathrm{w} \geq \int_{(\kappa_0 + \kappa_{j_0})/\sqrt{2}}^{(\kappa_{J+1} + \kappa_{J+1-j_0})/\sqrt{2}} \int_{\{\bar{\kappa} \,|\, (\bar{\mu},\bar{\kappa}) \in \mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\, \mathrm{d}\bar{\kappa} \bar{h}(\bar{\mu})\, \mathrm{d}\bar{\mu} \tag{4.47}$$

where $\bar{h}(\bar{\mu}) \triangleq \iota\big(cq^{j_0}\bar{\mu}\big)\iota\big(c\bar{\mu}\big)h\big(\sqrt{2}\bar{\mu}\big) \geq 0$. It is easy to verify that $\int_{\{\bar{\kappa} \,|\, (\bar{\mu},\bar{\kappa}) \in \mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\, \mathrm{d}\bar{\kappa}$ is an increasing function of $\bar{\mu}$ over the range of the outer integral $[(\kappa_0 + \kappa_{j_0})/\sqrt{2}, (\kappa_{J+1} + \kappa_{J+1-j_0})/\sqrt{2}]$, and, consequently,

$$\int_{\{\bar{\kappa} \,|\, (\bar{\mu},\bar{\kappa}) \in \mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\, \mathrm{d}\bar{\kappa} \geq 0, \tag{4.48}$$

where the equality is attained for $\bar{\mu} = (\kappa_0 + \kappa_{j_0})/\sqrt{2}$. Finally, (4.40) follows from (4.47) and (4.48). □

This proof of convexity of Lemma 4.3 is conservative as we loosen the positive integrals in regions RCE and CSQ by replacing them with zeros.

We now use Lemma 4.3 to prove the convexity of $\mathcal{L}_\iota(\boldsymbol{\alpha})$ in Lemma 4.1. Note that the mass-attenuation spectrum $\iota(\kappa)$ is considered known in Lemma 4.1. We define $\xi(\cdot) \triangleq \iota^{\mathrm{L}}(\cdot)$ and the corresponding first and second derivatives: $\dot{\xi}(s) = (-\kappa\iota)^{\mathrm{L}}(s)$ and $\ddot{\xi}(s) = \big(\kappa^2\iota\big)^{\mathrm{L}}(s)$. Observe that $\boldsymbol{\mathcal{I}}^{\mathrm{out}} = \big(\mathcal{I}_n^{\mathrm{out}}\big)_{n=1}^N = \xi_\circ(\Phi\boldsymbol{\alpha}) = \big(\xi(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\big)_{n=1}^N$. For notational simplicity, we omit the dependence of $\boldsymbol{\mathcal{I}}^{\mathrm{out}}$ on $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$ and use $\boldsymbol{\mathcal{I}}^{\mathrm{out}}$ and $\xi_\circ(\Phi\boldsymbol{\alpha})$ interchangeably.

We use the identities

$$\frac{\partial \xi_\circ(\Phi\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T} = \mathrm{diag}\left(\dot{\xi}_\circ(\Phi\boldsymbol{\alpha})\right)\Phi \tag{4.49a}$$

$$\frac{\partial \xi(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^T} = \ddot{\xi}\big(\boldsymbol{\phi}_n^T\boldsymbol{\alpha}\big)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \tag{4.49b}$$

to compute the gradient and Hessian of the Poisson NLL in (4.18):

$$\frac{\partial \mathcal{L}_\iota(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \Phi^T \operatorname{diag}\left(\dot{\xi}_\circ\left(\Phi\boldsymbol{\alpha}\right)\right)\left[\mathbf{1} - \operatorname{diag}^{-1}\left(\mathcal{I}^{\text{out}}\right)\boldsymbol{\mathcal{E}}\right] \tag{4.50a}$$

$$\frac{\partial \mathcal{L}_\iota(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^T} = \Phi^T \operatorname{diag}^{-2}\left(\mathcal{I}^{\text{out}}\right)\operatorname{diag}\left(\boldsymbol{\mathcal{E}}\right)\operatorname{diag}\left(\boldsymbol{x}\right)\Phi \tag{4.50b}$$

where the $N \times 1$ vector $\boldsymbol{x} = (x_n)_{n=1}^N$ is defined as

$$x_n = \dot{\xi}^2(s) + \ddot{\xi}(s)\xi(s)\left(\frac{\mathcal{I}_n^{\text{out}}}{\mathcal{E}_n} - 1\right)\Bigg|_{s=\boldsymbol{\phi}_n^T\boldsymbol{\alpha}}. \tag{4.50c}$$

Since $\mathcal{I}_n^{\text{out}} \geq (1-V)\mathcal{E}_n \geq 0$ according to (4.21a), we have

$$\frac{\mathcal{I}_n^{\text{out}}}{\mathcal{E}_n} - 1 \geq -V \tag{4.51}$$

and

$$x_n \geq \iint \left(\mu\kappa - \kappa^2 V\right)\iota(\mu)\iota(\kappa)\exp[-(\mu+\kappa)\boldsymbol{\phi}_n^T\boldsymbol{\alpha}]\,\mathrm{d}\kappa\,\mathrm{d}\mu \tag{4.52a}$$

$$= \iint \left(\mu\kappa - \frac{\mu^2+\kappa^2}{2}V\right)\iota(\mu)\iota(\kappa)\exp[-(\mu+\kappa)\boldsymbol{\phi}_n^T\boldsymbol{\alpha}]\,\mathrm{d}\kappa\,\mathrm{d}\mu \tag{4.52b}$$

$$\geq \frac{(q^{j_0}+1)^2}{q^{2j_0}+1}\mathrm{w} \geq 0 \tag{4.52c}$$

where (4.52a) follows by applying inequality (4.51) to (4.50c), using the Laplace-transform identity for derivatives (4.2), and combining the multiplication of the integrals; and (4.52b) is due to the symmetry with respect to μ and κ. Now, plug (4.21b) into (4.52b) and apply Lemma 4.3 with $h(\kappa) = \mathrm{e}^{-\kappa\boldsymbol{\phi}_n^T\boldsymbol{\alpha}}$ to conclude (4.52c). Therefore, the Hessian of $\mathcal{L}_\iota(\boldsymbol{\alpha})$ in (4.50b) is positive semidefinite.

4.C Proof of Theorem 4.2

According to [XY13], real-analytic and semialgebraic functions and their summations satisfy the KL property automatically. Therefore, the proof consists of showing the following two parts:

(a) the NLL in (4.16) is a real-analytic function of $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ on $\mathbb{C} \subseteq \mathrm{dom}(f)$ and

(b) both $r(\boldsymbol{\alpha})$ in (4.25c) and $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\mathcal{I}})$ are semialgebraic functions.

Real-analytic NLL. The NLL in (4.16) is in the form of weighted summations of terms $\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}$, $\ln\big[\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}\big]$, and $\ln^2\big[\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}\big]$ for $n = 1, 2, \ldots, N$. Weighted summation of real-analytic functions is real-analytic; hence, we need to prove that $\ell_1(t) = \boldsymbol{b}^{\mathrm{L}}\big(\boldsymbol{\phi}^T(\boldsymbol{\alpha}+t\boldsymbol{\gamma})\big)(\boldsymbol{\mathcal{I}}+t\boldsymbol{\mathcal{J}})$, $\ell_2(t) = \ln \ell_1(t)$, and $\ell_3(t) = \ell_2^2(t)$ are real-analytic functions. Since $\big(\ell_i(t)\big)_{i=1}^3$ are smooth, it is sufficient to prove that the mth derivatives, $\ell_i^{(m)}(t)$, are bounded for all m, $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$, $(\boldsymbol{\gamma}, \boldsymbol{\mathcal{J}})$, and t such that $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \boldsymbol{\mathcal{I}} + t\boldsymbol{\mathcal{J}}) \in \mathrm{dom}(f)$.

The mth derivative of $\ell_1(t)$ is

$$\begin{aligned}\ell_1^{(m)} = {} & (\boldsymbol{\phi}^T\boldsymbol{\gamma})^m\big((-\kappa)^m\boldsymbol{b}\big)^{\mathrm{L}}(\boldsymbol{\alpha} + t\boldsymbol{\gamma})(\boldsymbol{\mathcal{I}} + t\boldsymbol{\mathcal{J}}) \\ & + m(\boldsymbol{\phi}^T\boldsymbol{\gamma})^{m-1}\big((-\kappa)^{m-1}\boldsymbol{b}\big)^{\mathrm{L}}(\boldsymbol{\alpha} + t\boldsymbol{\gamma})\boldsymbol{\mathcal{J}} \end{aligned} \tag{4.53}$$

which is bounded for any $\boldsymbol{\alpha}$, $\boldsymbol{\mathcal{I}}$, $\boldsymbol{\gamma}$, $\boldsymbol{\mathcal{J}}$, and t such that $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \boldsymbol{\mathcal{I}} + t\boldsymbol{\mathcal{J}})$ is in one of compact subsets $\mathbb{C} \subseteq \mathrm{dom}(f)$.

For any compact set $\mathbb{C} \subseteq \mathrm{dom}(f)$, there exists $\epsilon > 0$ such that $\ell_1(t) \geq \epsilon$ for all $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \boldsymbol{\mathcal{I}} + t\boldsymbol{\mathcal{J}}) \in \mathbb{C}$. $\ln(\cdot)$ and $(\cdot)^2$ are analytic on $[\epsilon, +\infty)$. Since the compositions and products of analytic functions are analytic [KP02, Ch. 1.4], both $\ell_2(t)$ and $\ell_3(t)$ are analytic. Therefore, the NLL in (4.16) is analytic.

Semialgebraic regularization terms. According to [XY13],

i) the ℓ_2 norm $\|\cdot\|_2$ is semialgebraic,

ii) the indicator function $\mathbb{I}_{[0,+\infty)}(\cdot)$ is semialgebraic,

iii) finite sums and products of semialgebraic functions are semialgebraic, and

iv) the composition of semialgebraic functions are semialgebraic.

Therefore, $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha})$ and $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\mathcal{I}})$ are both semialgebraic. Since we can write

$$\sqrt{\sum_{j\in\mathcal{N}_i}(\alpha_i-\alpha_j)^2}=\|P_i\boldsymbol{\alpha}\|_2 \tag{4.54}$$

for some matrix P_i, using i), iii), and iv) leads to semialgebraic (4.54), thus semialgebraic $r(\boldsymbol{\alpha})$ in (4.25c). Finally, according to [XY13], the sum of real-analytic and semialgebraic functions satisfies the KL property. Therefore, $f(\boldsymbol{\alpha},\boldsymbol{\mathcal{I}})$ satisfies the KL property on a compact subset of $\mathrm{dom}\, f(\boldsymbol{\alpha},\boldsymbol{\mathcal{I}})$.

CHAPTER 5. CONCLUSION

We developed a fast framework for reconstructing signals that are sparse in a transform domain and belong to a closed convex set by employing a projected proximal-gradient scheme with Nesterov's acceleration, restart, and *adaptive* step size. This framework allows us to construct one of the first Nesterov-accelerated proximal-gradient reconstruction algorithm for Poisson compressed sensing. We derived convergence-rate upper-bound that accounts for inexactness of the proximal operator and proved the convergence of iterates. When compared with the state-of-the-art, our proposed PNPG approach is computationally efficient.

The regularization constant is a key parameter for achieving good reconstructions. We derived upper bounds on this constant for convex sparse signal reconstruction and presented for the first time such bounds for total-variation regularization. These bounds can be used to construct accurate prior distributions for the regularization constant and to design continuation procedures. The potential future work can be to obtain simpler expressions for upper bounds U for isotropic 2D TV regularization and low-rank matrix models with nuclear-norm regularization, based on Theorem 3.1.

Finally, we developed the first physical-model–based image reconstruction method for simultaneous *blind* sparse image reconstruction and mass-attenuation spectrum estimation from polychromatic measurements. We developed an algorithm that alternatively updates the attenuation map (using PNPG) and the mass attenuation spectrum (using L-BFGS-B) and successfully reconstructs attenuation map from real polychromatic X-ray CT measurements. An exciting direction for the future is to generalize our polychromatic signal model to handle multiple materials and develop corresponding reconstruction schemes for this scenario.

BIBLIOGRAPHY

[AABM12] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for Poisson intensity CBCT and PET," *Inverse Probl. Imag.*, vol. 6, no. 4, pp. 565–598, 2012 (cit. on pp. 8, 9, 11, 30).

[ABRS10] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010 (cit. on pp. 89, 95, 96).

[AD15] J.-F. Aujol and C. Dossal, "Stability of over-relaxations for the forward-backward algorithm, application to FISTA," *SIAM J. Optim.*, vol. 25, no. 4, pp. 2408–2433, 2015 (cit. on p. 27).

[AG03] E. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*. Philadelphia, PA: SIAM, 2003 (cit. on p. 52).

[AT06] A. Auslender and M. Teboulle, "Interior gradient and proximal methods for convex and conic optimization," *SIAM J. Optim.*, vol. 16, no. 3, pp. 697–725, 2006 (cit. on pp. 8, 27).

[BB88] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988 (cit. on pp. 21, 91).

[BCG11] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comp.*, vol. 3, no. 3,

pp. 165–218, 2011. [Online]. Available: `http://cvxr.com/tfocs` (cit. on pp. 8, 10, 18, 27–29).

[Ber09] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA: Athena Scientific, 2009 (cit. on p. 55).

[Ber15] D. P. Bertsekas, *Convex Optimization Algorithms*. Belmont, MA: Athena Scientific, 2015 (cit. on pp. 10, 13, 39).

[BK04] J. Barrett and N. Keat, "Artifacts in CT recognition and avoidance," *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004 (cit. on p. 74).

[BLNZ95] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995 (cit. on pp. 67, 90).

[BLPP] S. Bonettini, I. Loris, F. Porta, and M. Prato, *Variable metric inexact line–search algorithm (VMILA)*. [Online]. Available: `http://www.oasis.unimore.it/site/home/software.html` (visited on 01/03/2017) (cit. on p. 32).

[BLPP16] S. Bonettini, I. Loris, F. Porta, and M. Prato, "Variable metric inexact line-search-based methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 26, no. 2, pp. 891–921, 2016 (cit. on pp. 8, 9, 22, 26, 32).

[BN16] J. Y. Bello Cruz and T. T. A. Nghia, "On the convergence of the forward–backward splitting method with linesearches," *Optim. Method. Softw.*, vol. 31, no. 6, pp. 1209–1238, 2016 (cit. on pp. 8, 28).

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011 (cit. on p. 67).

[BPR16] S. Bonettini, F. Porta, and V. Ruggiero, "A variable metric forward-backward method with extrapolation," *SIAM J. Sci. Comput.*, vol. 38, no. 4, A2558–A2584, 2016 (cit. on pp. 7–9, 18, 27, 28).

[BS97] J. M. Boone and J. A. Seibert, "An accurate method for computer-generating tungsten anode X-ray spectra from 30 to 140 kV," *Med. Phys.*, vol. 24, no. 11, pp. 1661–1670, 1997 (cit. on p. 97).

[BST12] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients," vol. 23, no. 3, pp. 425–438, 2012 (cit. on p. 55).

[BT09a] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009 (cit. on pp. 8, 15, 18, 28, 39, 44, 90, 93).

[BT09b] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009 (cit. on pp. 8, 11, 18, 88, 90, 99).

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004 (cit. on p. 66).

[BV16] S. Boyd and L. Vandenberghe, *Vectors, Matrices, and Least Squares*, 2016. [Online]. Available: `http://stanford.edu/class/ee103/mma.pdf` (visited on 08/23/2016) (cit. on pp. 6, 56).

[BvG11] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York: Springer, 2011 (cit. on p. 2).

[CD15] A. Chambolle and C. Dossal, "On the convergence of the iterates of the 'fast iterative shrinkage/thresholding algorithm'," *J. Optim. Theory Appl.*, vol. 166, no. 3, pp. 968–982, 2015 (cit. on pp. 26, 27, 33, 47, 49).

[Con13] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013 (cit. on pp. 9, 36).

[CP11a] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011 (cit. on pp. 9, 30, 36).

[CP11b] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., vol. 49, New York: Springer, 2011, ch. 10, pp. 185–212 (cit. on p. 9).

[CP16] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numer.*, vol. 25, pp. 161–319, May 1, 2016 (cit. on p. 56).

[CPP09] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 730–762, 2009 (cit. on p. 8).

[CT06] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006 (cit. on p. 6).

[Dav15] D. Davis, "Convergence rate analysis of primal-dual splitting schemes," *SIAM J. Optim.*, vol. 25, no. 3, pp. 1912–1943, 2015 (cit. on p. 9).

[DDD16] I. Daubechies, M. Defrise, and C. De Mol, "Sparsity-enforcing regularisation and ISTA revisited," *Inverse Probl.*, vol. 32, no. 10, 104001, 2016 (cit. on p. 28).

[DFS09] F.-X. Dupé, J. M. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310–321, 2009 (cit. on p. 8).

[DFS12] F.-X. Dupé, M. J. Fadili, and J.-L. Starck, "Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors," *Stat. Methodol.*, vol. 9, no. 1-2, pp. 4–18, 2012 (cit. on pp. 7, 9).

[DGQ11] A. Dogandžić, R. Gu, and K. Qiu, "Mask iterative hard thresholding algorithms for sparse image reconstruction of objects with known contour," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2011, pp. 2111–2116 (cit. on pp. 31, 97).

[DTK12] W. Dewulf, Y. Tan, and K. Kiekens, "Sense and non-sense of beam hardening correction in CT metrology," *CIRP Ann.*, vol. 61, no. 1, pp. 495–498, 2012 (cit. on p. 73).

[EF02] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, 2002 (cit. on p. 74).

[EF03] I. A. Elbakri and J. A. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation," *Phys. Med. Biol.*, vol. 48, no. 15, pp. 2453–2477, 2003 (cit. on p. 74).

[Fes09] J. A. Fessler, *Image Reconstruction*, 2009. [Online]. Available: `http://web.eecs.umich.edu/~fessler/book/a-geom.pdf` (visited on 08/23/2016) (cit. on pp. 31, 71).

[Fes16] J. A. Fessler, *Image reconstruction toolbox*, 2016. [Online]. Available: `http://www.eecs.umich.edu/~fessler/code` (visited on 08/23/2016) (cit. on pp. 31, 32, 70, 71).

[GB14] M. Grant and S. Boyd. (Mar. 2014). CVX: matlab software for disciplined convex programming, version 2.1, [Online]. Available: `http://cvxr.com/cvx` (cit. on p. 63).

[GD13] R. Gu and A. Dogandžić, "Beam hardening correction via mass attenuation discretization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, Canada, May 2013, pp. 1085–1089 (cit. on pp. 74, 75, 78, 92, 102).

[GD15a] R. Gu and A. Dogandžić, "Polychromatic sparse image reconstruction and mass attenuation spectrum estimation via B-spline basis function expansion," in *Rev. Prog. Quant. Nondestr. Eval.*, D. E. Chimenti and L. J. Bond, Eds., ser. AIP Conf. Proc. Vol. 34 1650, Melville, NY, 2015, pp. 1707–1716 (cit. on pp. 75, 88).

[GD15b] R. Gu and A. Dogandžić. (Sep. 2015). Polychromatic X-ray CT image reconstruction and mass-attenuation spectrum estimation. arXiv: 1509.02193 [stat.ME] (cit. on pp. 3, 74, 75, 83, 88, 92, 99).

[GD15c] R. Gu and A. Dogandžić, "Projected Nesterov's proximal-gradient signal recovery from compressive Poisson measurements," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2015, pp. 1490–1495 (cit. on pp. 7, 18, 34, 91).

[GD15d] R. Gu and A. Dogandžić. (Mar. 2015). Reconstruction of nonnegative sparse signals using accelerated proximal-gradient algorithms. version 3. arXiv: 1502.02613v3 [stat.CO] (cit. on pp. 14, 37).

[GD16a] R. Gu and A. Dogandžić, "Blind X-ray CT image reconstruction from polychromatic Poisson measurements," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 150–165, 2016 (cit. on p. 14).

[GD16b] R. Gu and A. Dogandžić. (Oct. 2016). Projected Nesterov's proximal-gradient algorithm for sparse signal reconstruction with a convex constraint. version 6. arXiv: 1502.02613 [stat.CO] (cit. on pp. 2, 7, 33, 67, 68).

[GPK07] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007 (cit. on pp. 87, 90).

[Gu] R. Gu, *Projected Nesterov's proximal-gradient algorithm source code*. [Online]. Available: `https://github.com/isucsp/pnpg` (visited on 01/18/2017) (cit. on p. 30).

[Har] Z. Harmany, *The sparse Poisson intensity reconstruction algorithms (SPIRAL) toolbox*. [Online]. Available: `http://drz.ac/code/spiraltap` (visited on 09/03/2016) (cit. on pp. 29, 32, 36).

[Her79] G. T. Herman, "Correction for beam hardening in computed tomography," *Phys. Med. Biol.*, vol. 24, no. 1, pp. 81–106, 1979 (cit. on pp. 79, 98).

[HL13] M. Hong and Z.-Q. Luo. (Mar. 2013). On the linear convergence of the alternating direction method of multipliers. arXiv: `1208.3922 [math.OC]` (cit. on p. 67).

[HMW12] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012 (cit. on pp. 6–8, 13, 32).

[HR12] B. E. Hansen and J. S. Racine, "Jackknife model averaging," *J. Econometrics*, vol. 167, no. 1, pp. 38–46, 2012 (cit. on p. 7).

[HS95] J. H. Hubbell and S. M. Seltzer, "Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements $Z = 1$ to 92 and 48 additional substances of dosimetric interest," National Inst. Standards Technol., Ionizing Radiation Div., Gaithersburg, MD, Tech. Rep. NISTIR 5632, 1995 (cit. on pp. 97, 109).

[Hsi09] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, 2nd ed. Bellingham, WA: SPIE, 2009 (cit. on pp. 74, 77).

[HTWM10] Z. Harmany, D. Thompson, R. Willett, and R. F. Marcia, "Gradient projection for linearly constrained convex optimization in sparse signal recovery," in *IEEE Int. Conf. Image Process.*, Hong Kong, China, Sep. 2010, pp. 3361–3364 (cit. on pp. 7, 36).

[Hud10] W. Huda, *Review of Radiologic Physics*, 3rd ed. Baltimore, MD: Lippincott Williams & Wilkins, 2010 (cit. on p. 109).

[HYZ08] E. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008 (cit. on p. 52).

[JBS15] P. Jin, C. A. Bouman, and K. D. Sauer, "A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 200–216, 2015 (cit. on p. 4).

[KKF08] M. Krumm, S. Kasperl, and M. Franz, "Reducing non-linear artifacts of multi-material objects in industrial 3D computed tomography," *NDT & E Int.*, vol. 41, no. 4, pp. 242–251, 2008 (cit. on p. 74).

[KKL+07] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, 2007 (cit. on pp. 52, 65).

[KP02] S. Krantz and H. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Boston, MA: Birkhäuser, 2002 (cit. on p. 116).

[KR15] M. Kabanava and H. Rauhut, "Cosparsity in compressed sensing," in *Compressed Sensing and Its Applications*, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybíral, Eds. New York: Springer, 2015, pp. 315–339 (cit. on p. 55).

[KS88] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988 (cit. on pp. 73, 77, 98).

[LFP16] J. Liang, J. Fadili, and G. Peyré, "Convergence rates with inexact non-expansive operators," *Math. Program., Ser. A*, vol. 159, no. 1, pp. 403–434, 2016 (cit. on p. 9).

[LRG+14] Y. Lin, J. C. Ramirez-Giraldo, D. J. Gauthier, K. Stierstorfer, and E. Samei, "An angle-dependent estimation of CT x-ray spectrum from rotational transmission measurements," *Med. Phys.*, vol. 41, no. 6, p. 062 104, 2014 (cit. on p. 74).

[LV89] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 186–193, 1989 (cit. on p. 101).

[LWW07] G. M. Lasio, B. R. Whiting, and J. F. Williamson, "Statistical reconstruction for X-ray computed tomography using energy-integrating detectors," *Phys. Med. Biol.*, vol. 52, no. 8, p. 2247, 2007 (cit. on p. 75).

[LZZ+15] J. Liu, X. Zhang, X. Zhang, H. Zhao, Y. Gao, D. Thomas, D. A. Low, and H. Gao, "5D respiratory motion model based image reconstruction algorithm for 4D cone-beam computed tomography," *Inverse Probl.*, vol. 31, no. 11, 115007, 2015 (cit. on p. 95).

[MN89] P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. New York: Chapman & Hall, 1989 (cit. on pp. 14, 84).

[NDF+13] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Phys. Med. Biol.*, vol. 58, no. 12, R63–R96, 2013 (cit. on pp. 74, 77).

[Nes13] Y. Nesterov, "Gradient methods for minimizing composite functions," *Math. Program., Ser. B*, vol. 140, no. 1, pp. 125–161, 2013 (cit. on p. 8).

[Nes83] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Sov. Math. Dokl.*, vol. 27, no. 2, pp. 372–376, 1983 (cit. on pp. 8, 15, 28, 90).

[OC15] B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, vol. 15, no. 3, pp. 715–732, 2015 (cit. on pp. 20, 91, 93).

[OF97] J. M. Ollinger and J. A. Fessler, "Positron-emission tomography," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43–55, 1997 (cit. on pp. 13, 30, 32, 70).

[PB13] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013 (cit. on p. 76).

[PCP11] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, 2011 (cit. on p. 9).

[PL15] J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*, 2nd ed. Upper Saddle River, NJ: Pearson, 2015 (cit. on pp. 6, 13, 14).

[RFP13] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1199–1226, 2013 (cit. on pp. 9, 36).

[RPP+09] R. Redus, J. Pantazis, T. Pantazis, A. Huber, and B. Cross, "Characterization of CdTe detectors for quantitative X-ray spectroscopy," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2524–2532, 2009 (cit. on p. 74).

[Sal16] S. Salzo. (May 2016). The variable metric forward-backward splitting algorithm under mild differentiability assumptions. arXiv: `1605.00952 [math.OC]` (cit. on pp. 9, 10).

[Sch07] L. L. Schumaker, *Spline Functions: Basic Theory*, 3rd ed. New York: Cambridge Univ. Press, 2007 (cit. on p. 80).

[SRB11] M. Schmidt, N. L. Roux, and F. R. Bach, "Convergence rates of inexact proximal-gradient methods for convex optimization," in *Adv. Neural Inf. Process Syst. 24*, 2011, pp. 1458–1466 (cit. on pp. 8, 26).

[TA77] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977 (cit. on p. 52).

[Thi89] R. A. Thisted, *Elements of Statistical Computing*. New York: Chapman & Hall, 1989 (cit. on p. 90).

[Tse00] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 431–446, 2000 (cit. on p. 28).

[Vog02] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002 (cit. on p. 52).

[VSBV13] S. Villa, S. Salzo, L. Baldassarre, and A. Verri, "Accelerated and inexact forward-backward algorithms," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1607–1633, 2013 (cit. on pp. 8, 10, 12, 22, 23, 25).

[Vũ13] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013 (cit. on p. 9).

[VVD+11] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts, "Iterative correction of beam hardening artifacts in CT," *Med. Phys.*, vol. 38, S36–S49, 2011 (cit. on pp. 4, 74).

[WNF09] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009 (cit. on pp. 52, 65).

[WYD08] G. Wang, H. Yu, and B. De Man, "An outlook on x-ray CT research and development," *Med. Phys.*, vol. 35, no. 3, pp. 1051–1064, 2008 (cit. on p. 73).

[WYYZ08] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008 (cit. on p. 55).

[XT14] J. Xu and B. M. W. Tsui, "Quantifying the importance of the statistical assumption in statistical X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 61–73, 2014 (cit. on p. 75).

[XY13] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013 (cit. on pp. 4, 90, 95, 96, 116, 117).

[YW82] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: part 1—theory," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 81–94, 1982 (cit. on pp. 7, 18).

[ZBBR15] L. Zanni, A. Benfenati, M. Bertero, and V. Ruggiero, "Numerical methods for parameter estimation in Poisson data inversion," *J. Math. Imaging Vis.*, vol. 52, no. 3, pp. 397–413, 2015 (cit. on pp. 14, 84).

[ZTB+14] R. Zhang, J.-B. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 117–134, Jan. 2014 (cit. on p. 109).