Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance

JFPJ Abascal, N Ducros, F Peyrin

Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France

E-mail: `juan.abascal@creatis.insa-lyon.fr`

October 2017

Abstract. In this paper, we address the resolution of material decomposition, which is a nonlinear inverse problem encountered in spectral computed tomography (CT). The problem is usually solved in a variational framework but, due to the nonlinearity of the forward operator, the objective function may be nonconvex and standard approaches may fail. Regularized iterative schemes based on the Bregman distance have been suggested for improving global convergence properties. In this work, we analyze the convexity of the material decomposition problem and propose a regularized iterative scheme based on the Bregman distance to solve it. We evaluate our Bregman iterative algorithm and compare it with a regularized Gauss-Newton (GN) method using data simulated in a realistic thorax phantom.

First, we prove the existence of a convex set where the usual data fidelity term is convex. Interestingly, this set includes zero, making it a good initial guess for iterative minimization schemes. Using numerical simulations, we show that the data fidelity term can be nonconvex for large values of the decomposed materials. Second, the proposed Bregman iterative scheme is evaluated in different situations. It is observed to be robust to the selection of the initial guess, leading to the global minimum in all tested examples while the GN method fails to converge when the initial guess is not well chosen. Moreover, it is found to avoid the selection of the regularization parameter for little extra computation.

In conclusion, we have provided a suitable initialization strategy to solve the nonlinear material decomposition problem using convex optimization methods and evaluated a Bregman iterative scheme for this problem. The improvement in global convergence of Bregman iterative scheme combined with other interesting properties of the Bregman distance appears as a compelling strategy for nonlinear inverse problems.

Keywords: Bregman distance, convexity, spectral computed tomography

1. Introduction

We consider regularized methods for solving nonlinear inverse problems. Successful methods for linear inverse problems have been generalized for the case of nonlinear inverse problems. However, in general, objective functions involving nonlinear operators are nonconvex, so classical optimization procedures may fail. Regularized iterative procedures have been proposed to ensure that a global minimizer exists. The idea behind these methods is to replace the original regularized problem with a sequence of problems, each of which satisfies stronger convergence properties [1, 2].

A popular and efficient iterative regularization procedure for the $\ell 1$- and total variation (TV)-problems is based on the use of the Bregman distance and has been widely studied for linear measurements [3–8]. Replacing a convex regularization functional with the Bregman distance associated with this functional leads to an iterative scheme that converges to the solution of the constrained optimization problem. The use of the Bregman distance provides additional appealing properties. For instance, it has been shown that minimizing the Bregman distance associated with TV can recover the contrast lost that may appear when solving the TV problem [9]. The split Bregman method proposed for image processing and compressed sensing combines the Bregman distance with a splitting technique to efficiently deal with non differentiable convex functional, such as total variation [7,10,11]. It has been successfully applied to solve the total variation problem in image denoising, image impainting, and compressed sensing [7, 12–17]. The use of Bregman distance for the TV minimization subject to nonlinear measurements enjoys further benefits but it has not been widely explored [18,19]. In [20], Bregman iterative schemes based on classical reconstruction methods were proposed to improve global convergence for nonlinear inverse problems.

In this work, we are concerned with the nonlinear material decomposition problem that appears in spectral computed tomography (CT). Spectral CT is gaining increasing attention due to the recent development of energy discriminating photon-counting detectors that simultaneously count photons and resolve their energy [21, 22]. This technology addresses the limitations of conventional CT and provides material decomposition capabilities, which allows, for instance, characterizing lesions for breast imaging [23] and atherosclerotic plaque [24, 25].

The spectral CT material reconstruction problem can be solved in one step by recovering the material mass densities directly from the energy projections [26–29]. A two-step alternative is first to decompose the energy projections into material projections and then to perform a tomographic reconstruction from the material projections. The one-step approach can incorporate prior knowledge about the material density –the quantity of interest– while determining prior knowledge for the projected mass densities of the two-step material decomposition is less natural. However, the two-step approach has several advantages. First, it enables to split a large inversion problem into smaller problems. Second, it offers a physics-based parallelization scheme where all projection views are decomposed in parallel and all material are reconstructed in parallel. Third, it

separates linear (i.e., CT reconstruction) from nonlinear (i.e., material decomposition) mixing, allowing to reuse state-of-the art CT reconstruction approaches. In this paper, we focus on the material decomposition problem and adopt a model-based approach which models source spectrum and energy responses of the detector [30]. For alternative approaches to material reconstruction see the following works [14, 31–38]. We remark that most of these approaches adopted some linearized forms or focused on the linear tomographic problem.

Most previous studies that addressed spectral CT have ignored nonconvexity or adopted some assumptions to make the problem convex [39]. For dual-energy CT and for the negative log-likelihood functional, a hybrid first and second order variational approach was proposed to provide a more robust algorithm [40]. Recently, a regularized Gauss-Newton (GN) method with specific regularization applied to each material was applied to solve the material decomposition problem [41]. This method provided significantly superior results than the non regularized maximum-likelihood method that is commonly used in practice [30].

The goal of this work is to investigate convexity for the material decomposition inverse problem and to propose a regularized Bregman iterative scheme to solve this problem. To this goal we build from the previously proposed regularized GN method [41] and modify it in order to minimize the Bregman distance associated with the same regularization functional. We put particular emphasis to study convexity and to provide a suitable initialization strategy for convex optimization algorithms. In addition, we assess the benefits of a Bregman iterative algorithm to provide a global minimum by studying its dependence on the initial guess.

The paper is organized as follows. In section 2, we present the formulation of the forward and inverse problem. In section 3, we show that the data fidelity term is convex in a defined set and that the zero solution belongs to this region. In section 4, we describe the numerical approaches adopted to solve this problem, the regularized Bregman iterative scheme and a regularized Gauss-Newton method. In section 5, we show the numerical results. First, we analyze convexity using numerical examples. Then, we assess the Bregman iterative scheme in comparison with the Gauss-Newton approach using a numerical thorax phantom. Section 6 contains the discussion, further work and conclusion.

2. Material decomposition in spectral CT

2.1. Spectral CT

Let $s_{i,\theta} \in \mathbb{R}^P$ represent the photon counts measured in the i-th energy bin of the detector under the θ-th projection angle and $\rho_m \in \mathbb{R}^N$ represent the mass density of the m-th material in the object. Spectral CT consists in recovering the mass densities of the M materials in the object from the photon counts measured in I energy bins for Θ views. Solving the spectral CT problem in one step requires the inversion of the nonlinear

system

$$s = \mathcal{G}(\rho) \tag{1}$$

where $s = (s_{1,1}, \ldots, s_{I,\Theta})$ is the full measurement vector, $\rho = (\rho_1, \ldots, \rho_M)$ is the material vector, and $\mathcal{G}$ models both spectral mixing and tomographic acquisition. Alternatively, the two-step approach [30, 41] first decomposes all projection views by inverting

$$s_\theta = \mathcal{F}(a_\theta), \quad 1 \le \theta \le \Theta, \tag{2}$$

where $s_\theta = (s_{1,\theta}, \ldots, s_{I,\theta})$, $a_\theta = (a_{1,\theta}, \ldots, a_{M,\theta})$ represents the *projected* mass densities for the θ-th view, and $\mathcal{F}$ models spectral mixing only. This material decomposition step is followed by the tomographic reconstruction of all materials, i.e., the inversion of

$$a_m = \mathcal{X}(\rho_m), \quad 1 \le m \le M, \tag{3}$$

where $a_m = (a_{m,1}, \ldots, a_{m,\Theta})$ and $\mathcal{X}$ is the X-ray transform. While the one-step approach requires inverting the $\mathbb{R}^{MN} \to \mathbb{R}^{I\Theta P}$ nonlinear mapping defined in (1), the two-step approach split it into multiple inversions of $\mathbb{R}^{MP} \to \mathbb{R}^{IP}$ mappings followed by multiple inversions of $\mathbb{R}^N \to \mathbb{R}^P$ mappings. The two-step approach not only reduces the size of the problems, offering a natural parallelization scheme (e.g., decomposing all angles in parallel, reconstructing all materials in parallel), but also separate nonlinear (i.e., spectral) from linear (i.e, tomographic) mixing. In the following, we focus on the material decomposition problem defined in (2).

2.2. Spectral mixing

We now consider a fixed θ in (2) and drop the θ index for the sake of simplicity. Let s_i^p represents the photon counts in the p-th pixel of the detector and a_m^p the projected mass density in the same pixel. Assuming that no inter-pixel spectral mixing occurs, the system (2) is equivalent to

$$s_i^p = \mathcal{F}_i^p(a_1^p, \ldots, a_M^p), \quad 1 \le i \le I \text{ and } 1 \le p \le P \tag{4}$$

where $s_i = (s_i^1, \ldots, s_i^P)$ and $a_m = (a_m^1, \ldots, a_m^P)$. The nonlinear mapping $\mathcal{F}_i^p : \mathbb{R}^M \to \mathbb{R}$ can be modelled by [30]

$$\mathcal{F}_i^p(a_1^p, \ldots, a_M^p) := \int_{\mathcal{E}} n_0(E) d_i(E) \exp\left[-\sum_{m=1}^M a_m^p \tau_m(E) \right] dE \tag{5}$$

where $\mathcal{E} = [0, E_{\max}]$ is the energy range delivered by the X-ray tube, $n_0(E)$ is the X-ray source energy spectrum, $d_i(E)$ is the energy-dependent response function of the i-th bin of the detector, and $\tau_m(E)$ is the energy-dependent mass attenuation of the m-th constituent of the object.

2.3. Inverse problem

Within a variational framework, material decomposition is formulated as the following optimization problem

$$\min_a \ C(a) = \min_a \ L(s^\delta, a) + \tilde{\alpha} R(a) \tag{6}$$

where s^δ represents the noisy measurements, L is a data fidelity term, R is a regularization term encoding *a priori* knowledge about the solution, and $\tilde{\alpha}$ is the regularization parameter. Assuming that the measurements are Poisson distributed, the following approximation of the discretised Kullbach-Leibler divergence is considered

$$L(s^\delta, a) = \frac{1}{2} \|\mathcal{F}(a) - s^\delta\|_W^2, \tag{7}$$

where $\|x\|_W$ is the weighted norm defined by $\|x\|_W^2 = x^\top W^\top W x$. Here, the weight matrix W is chosen as

$$W = \mathsf{diag}\left(\frac{1}{\max\left(\sqrt{s^\delta}, \epsilon\right)}\right), \tag{8}$$

where ϵ is a small parameter that is added to the denominator to avoid division by zeros.

It is important to note that although the regularization functional R is convex the cost function C in (6) can be nonconvex whenever the data fidelity term L is nonconvex, which usually occurs with nonlinear measurements.

2.4. Gauss Newton method

A wide variety of approaches can be used to solve the nonlinear problem (6) [42, 43]. Among them the Gauss Newton method has fast (quadratic) convergence and has been recently shown to be very effective for material decomposition [40, 41]. The Gauss Newton update is computed as

$$a^{n+1} = a^n - \lambda^n H^{-1}(a^n) C'(a^n) \tag{9}$$

where λ^n is the step length, C' is the gradient of C, and H is an approximation for the Hessian of C, which is given by

$$H(a^n) = \mathcal{F}'(a^n)^\top W^\top W \mathcal{F}'(a^n) + \tilde{\alpha} R''(a^n) \tag{10}$$

where $\mathcal{F}'$ denotes the Jacobian of $\mathcal{F}$ and R'' the Hessian of R. The step length is usually found by a line search strategy, i.e., minimizing the cost function along the descent direction provided by $H^{-1}(a^n) C'(a^n)$.

3. Analysis of the convexity of the problem

In this section, we study the convexity of the material decomposition problem. In particular, we show that the weighted least squares data fidelity term is convex on a

convex set that includes $a = 0$. This provides an initialization point that guarantees that unconstrained optimization algorithms converge to the global minimizer.

Convexity of the problem is analyzed deriving the following properties for the Hessian of the data fidelity functional. For simplicity, the analysis is conducted considering one pixel only (i.e. for $P = 1$) and the index p has been dropped. Since the data fidelity term L is separable with respect to pixels, the results bellow can be generalized to any pixel number (i.e., $P > 1$).

Definition 1. *Let $g_i(a, E)$ be the number of incident photons seen at energy E in the i-th energy bin of the detector. It is given by*

$$g_i(a, E) = n_0(E)d_i(E) \exp\left[-\sum_{m=1}^{M} a_m \tau_m(E)\right] \tag{11}$$

Proposition 1. *The Hessian of L is given by*

$$H(a) = -\sum_i w_i^2(s_i^\delta - \mathcal{F}_i(a))T^\top G_i(a)T + \mathcal{F}'^\top W^\top W \mathcal{F}' \tag{12}$$

where $T = (\tau_{j,m}) \in \mathbb{R}^{J \times M}$ is the mass attenuation matrix defined by $\tau_{j,m} = \tau_m(j\Delta E)$, $1 \le j \le J$ and $G_i(a) = \mathsf{diag}\left(g_i(a, E_1), \ldots, g_i(a, E_J)\right) \Delta E$.

Proof. The functions $\mathcal{F}_i \colon \mathbb{R}^M \to \mathbb{R}$, $1 \le i \le I$ are twice continuously differentiable and the Hessian of L takes the following matrix form [44]

$$H(a) = -\sum_i w_i^2(s_i^\delta - \mathcal{F}_i(a))H_i(a) + \mathcal{F}'^\top W^\top W \mathcal{F}' \tag{13}$$

where H_i denotes the Hessian of $\mathcal{F}_i$. It can be shown that $H_i(a) = T^\top G_i(a)T$ (see section Appendix A), which completes the proof. □

Proposition 2. *The data fidelity functional L is convex on the (possibly nonconvex) set $\mathcal{A}$ defined by*

$$\mathcal{A} = \{a \in \mathbb{R}^M \,|\, \mathcal{F}_i(a) \ge s_i^\delta, \forall i\} \tag{14}$$

Proof. While $\mathcal{F}'^\top W^\top W \mathcal{F}'$ is clearly positive semi-definite, some or all the terms $w_i^2(s_i^\delta - \mathcal{F}_i(a))T^\top G_i(a)T$ can contribute against convexity. Since $G_i(a)$ is a diagonal matrix with positive elements, $T^\top G_i(a)T$ is positive definite. If $s_i^\delta - \mathcal{F}_i(a)$ is negative for all i, then $H(a)$ is the sum of positive definite matrices, hence it is positive definite. □

Proposition 3. *(Exact Data) Let a^{true} be the solution and $s^\delta = s^{\mathrm{true}} = \mathcal{F}(a^{\mathrm{true}})$. The data fidelity functional L is convex on the convex set $\mathcal{B}$ defined by*

$$\mathcal{B} = \{a \in \mathbb{R}^M \,|\, a_m \le a_m^{\mathrm{true}}, \forall m\} \tag{15}$$

Proof. Noting that all $\mathcal{F}_i \colon a_m \mapsto \mathcal{F}_i(a_m)$ are monotonically decreasing, choosing $a_m \le a_m^{\mathrm{true}}$ for all m guaranties that $\mathcal{F}_i(a) \ge \mathcal{F}_i(a^{\mathrm{true}})$. □

Remark 4. *(Exact Data) The particular point $a = 0$ lies within the convex region, i.e., $0 \in \mathcal{B}$.*

Proof. Physically, solutions are positive. □

4. Resolution of the problem

4.1. Iterative regularization scheme based on the Bregman distance

The Bregman distance generalizes the concept of distance for convex regularization functionals R [45]. The generalized Bregman distance associated with the functional R for two points a and $\tilde{a}$ is defined as

$$D_\xi^R(a, \tilde{a}) = R(a) - R(\tilde{a}) - \langle \xi, a - \tilde{a} \rangle, \tag{16}$$

where ξ is a subgradient of R at $\tilde{a}$. Even $D_\xi^R(a, \tilde{a}) \geq 0$, it is not a distance in a strict sense as in general it is not symmetric and it does not satisfy the triangle inequality.

The use of Bregman distance became popular after its application to construct iterative regularization schemes for total variation regularization [3]. Since then, it has been widely applied to image restoration, deconvolution and compressed sensing problems with linear measurement operators and with convex regularization functional [4, 5, 7, 9, 13, 46]. A generalization of the Bregman distance for nonconvex regularization functionals has been also suggested [47]. However, its application to solve problems with nonlinear measurements is less known.

In [20], authors proposed to replace the problem (6) by a sequence of problems

$$a^{k+1} = \arg\min_a \frac{1}{2} \|\mathcal{F}(a) - s^\delta\|_W^2 + \alpha D_{\xi^k}^R(a, a^k), \tag{17a}$$

$$\xi^{k+1} = \xi^k - \alpha^{-1}(W\mathcal{F}'(a^{k+1}))^\top(W\mathcal{F}(a^{k+1}) - Ws^\delta), \tag{17b}$$

where $\alpha > 0$ is an hyperparameter, which is *not* a regularization parameter and that can be chosen as a large constant. The iterative scheme (17) turns to be a natural extension of the Bregman iterative method for linear measurements [3]. Each of the subproblems (17a) is an optimization problem that has the same form as (6) but where the regularization functional R is replaced by the Bregman distance between the unknown, a, and its previous estimate, a^k. Equation (17b) provides un update of the subgradient derived from the first order optimality condition of (17a), which enforces the data fidelity term iteratively [7].

The Bregman iterative scheme (17) constitutes then a sequence of subproblems where each of them has the same complexity as the original problem (6), but there is a substantial gain, as each subproblem is locally convex for sufficiently large values of α [20]. This provides a scheme to find a global minimizer for nonconvex functional associated with nonlinear operator equations.

An additional modification to the iterative scheme (17) was proposed in [20] to ensure local convexity for total variation regularization. A zero-order Tikhonov term $\frac{\kappa}{2}\|a\|_2^2$ is added to the original total variation regularization functional R to ensure local convexity around the current estimate a^k. The value of κ must be sufficiently small in order not to modify the properties of the regularization functional R. Substituting the

Bregman distance (16) into (17a), the iterative scheme is rewritten as

$$a^{k+1} = \min_a \ \Psi^k(a) = \min_a \ \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|_W^2 + \alpha R(a) - \alpha <\xi^k, a> + \alpha\frac{\kappa}{2}\|a\|_2^2, \quad (18a)$$

$$\xi^{k+1} = \xi^k - \alpha^{-1}(W\mathcal{F}'(a^{k+1}))^\top(W\mathcal{F}(a^{k+1}) - Ws^\delta). \quad (18b)$$

A key point in the construction of these iterative schemes is to select efficient methods for solving the subproblem (18a) [20]. We propose to solve (18a) using the Gauss-Newton method (9)–(10), which results in Algorithm 1. We refer to our approach as GN-Bregman (GNB). MATLAB code for material decomposition using GN and GNB is available from [48].

Algorithm 1 Regularized Gauss Newton Bregman (GNB) Algorithm

Set $\alpha, \kappa, tol, \psi, a^0$
$k \leftarrow 1$
while $\frac{1}{2}\|\mathcal{F}(a^k) - s^\delta\|_W^2 > tol$ **do** $\triangleright$ Outer iteration
    $a^k \leftarrow a^{k-1}$
    for $n = 1$ to N^k **do** $\triangleright$ Inner iteration
        $H^k \leftarrow (W\mathcal{F}'(a^k))^T W\mathcal{F}'(a^k) + \alpha R''(a^k) + \alpha\kappa I$
        $g^k \leftarrow (W\mathcal{F}'(a^k))^T W(\mathcal{F}(a^k) - s^\delta) + \alpha R'(a^k) - \alpha\xi^k + \alpha\kappa a^k$
        $p \leftarrow -(H^k)^{-1}g^k$
        $\lambda \leftarrow \arg\min_\lambda \Psi(a^k + \lambda p)$
        $a^k \leftarrow a^k + \lambda p$
    $\xi^k \leftarrow \xi^k - \alpha^{-1}(W\mathcal{F}'(a^k))^T(W\mathcal{F}(a^k) - Ws^\delta)$ $\triangleright$ Subgradient update
    $k \leftarrow k + 1$

4.2. Algorithm stopping criterion

Note that GNB (algorithm 1) numerically involves two iterative procedures: the (outer) Bregman iteration indexed by k and the (inner) Gauss Newton iteration that we denote with the index n. We define N^k the total number of GN iterations required to minimize Ψ^k and $N^{tot} = \sum_k N^k$ the total number of GN iterations of the GNB algorithm. The stopping criterion for the minimization of Ψ^k is based on the relative cost function decrease, i.e, the algorithm is stopped at N^k iterations when the relative cost function is smaller than ψ. The stopping criterion for the Bregman iterative scheme is the discrepancy principle, i.e., the algorithm is stopped when $\frac{1}{2}\|\mathcal{F}(a^k) - s^\delta\|_W^2 \leq tol$ [7, 20].

4.3. Selection of the initial guess

Proposition 3 and remark 4 provide a suitable initialization point for convex optimization algorithms. In particular for GNB, setting $a^0 = 0$ and $\xi^0 = 0 \in \partial R(a^0)$ leads to $D_R^{\xi^0}(a, a^k) = R(a)$ and $\Psi^0(a) = L(a, s) + \alpha R(a)$, so the first inner problem in GNB (algorithm 1) is equivalent to the original unconstrained problem (6) but with a large

regularization parameter. As a^0 belongs to the convex region, the first iterate a^1 is an excessively regularized solution that lies in the convex region and is closer to a^{true}. In [20] it is shown that when solving subproblem (17a) exactly, iterates a^k satisfy monotonicity of the residuals and of the error, so a^k for $k > 1$ remain in the convex region and get monotonically closer to a^{true}.

The Bregman iteration enjoys additional properties that improve global convexity for nonlinear inverse problems. In section 5.4, we show using numerical simulations that GNB converges to an optimal solution even if the initial guess is on the nonconvex region.

4.4. Regularization functional

In this work, we consider the separable regularizer proposed in [41] and defined by

$$R(a) = \sum_m R_m(a_m), \tag{19}$$

where R_m is the regularization functional for the m-th material. This choice allows to incorporate *a-priori* information for each material independently. Here, we consider Tikhonov regularization to promote smooth materials, i.e.,

$$R_m(a_m) = \|D^\ell a_m\|_2^2 \tag{20}$$

where D^ℓ represents the ℓ-th spatial derivative operator, and a smooth approximation of total variation to promote piecewise constant materials, i.e.,

$$R_m(a_m) = \sqrt{(Da_m)^2 + \epsilon^2} - \epsilon. \tag{21}$$

where ϵ is a small hyperparameter. Although only differentiable regularizers are used, Bregman iterative algorithms are also valid for the space of functions of bounded variations and other convex sparse regularization functional that uses ℓ_1-norm to impose sparsity in a given orthonormal basis, such as wavelet transform or framelets.

5. Numerical simulations and results

5.1. Numerical phantom data

We used numerical simulations to illustrate the (non) convexity of the problem and evaluate the proposed algorithms. Different datasets were simulated using realistic X-ray tube and photon-counting detector with four energy bins (see figure 1) and considering numerical phantoms made of soft tissue, bone and gadolimium (Gd). This is a relevant example in K-edge imaging for characterizing atherosclerotic plaque [24, 25]. The linear mass attenuation coefficients for the three materials are provided in figure 1c. Data for the thorax phantom are publicly available from SPRAY (Spectral X-ray image reconstruction) toolbox, a MATLAB toolbox for simulation and nonlinear image reconstruction in spectral X-ray imaging [41, 48].

Two numerical phantoms were created: a one-pixel phantom (only used to study convexity) and a realistic thorax phantom built from a segmented thoracic-abdominal CT scan [49, 50]. For the realistic phantom, a 611×167-pixel detector is simulated, leading to $P = 102,037$. Two projection images for two different projection angles were evaluated. For each view, the dataset comprised $I = 4$ projection images corresponding to the different energy bins and $M = 3$ unknowns projected mass density images were recovered. Decreasing measurement signal-to-noise ratios were considered by lowering the number of incident photons from 10^8 to 10^5. To get away from the inverse crime and further asses the robustness of our algorithm, we considered the mixed noise model $s^\delta \sim \mathcal{P}(\lambda = s) + \mathcal{N}(\mu, \sigma^2)$, where $\mathcal{P}$ is the Poisson distribution and $\mathcal{N}$ the normal distribution whose mean μ can be calibrated to be zero [51]. We used a concentration of gadolinium equal to 1 g·cm^{-3} unless otherwise stated.

We have used specific regularization for each material as described in section 4.4. We used first- and second-order Tikhonov regularization for bone and soft tissue, respectively, and smooth total variation regularization for the gadolinium image.
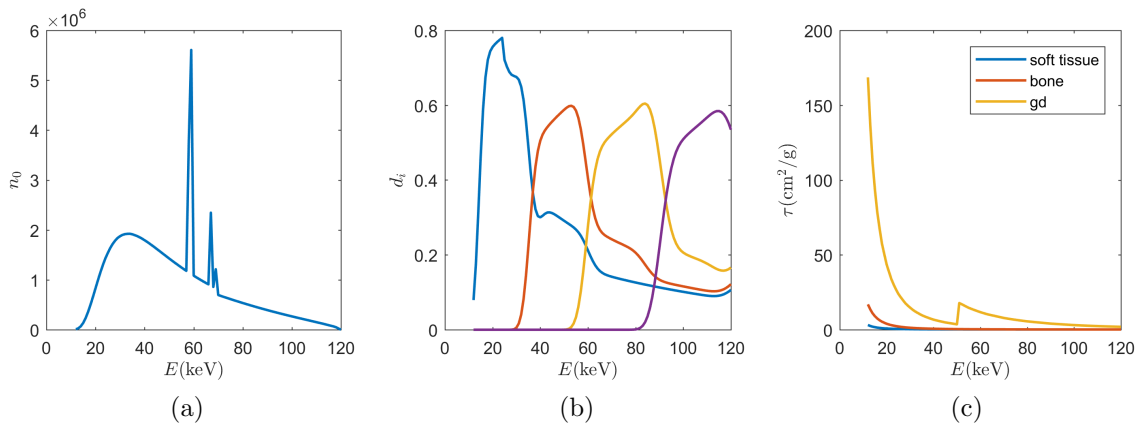
Figure 1. Forward model parameters used in the numerical simulations. (a) Source spectrum $n_0(E)$. (b) Detector response function $d_i(E)$ for each of the $I = 4$ energy bins. (c) Linear mass attenuation coefficients $\tau_m(E)$ for each of the $M = 3$ materials.

5.2. Numerical study of the data fidelity convexity

We have shown in section 3 that the problem is convex in a certain region of the solution space (proposition 3) and that the zero mass density image belongs to this region (remark 4). This is a relevant result as it can serve as a guide to initialize convex optimization algorithms. Nevertheless, there may be regions of the solution space where the data fidelity term is nonconvex. In this section, we examine convexity numerically by computing the data fidelity term (7) for two typical numerical phantoms. Figure 2 shows the data fidelity term for a one-pixel mass density image a while figure 3 displays the data fidelity term for the multiple-pixel thorax phantom. In the latter simulation, we consider only constant images for the three materials, i.e., the same value is assigned

to the entire image for each material. This has been done for simplification but it is actually relevant as constant images are commonly used as initial guesses.

These examples show that the zero mass density image belongs to the convex region and that the problem is nonconvex for large values of the three materials. It can be noticed that the data fidelity term becomes asymptotically flat for increasing values of the three materials, i.e., $\lim_{a\to\infty} L(a) = \frac{1}{2}\|s^\delta\|_W^2$.
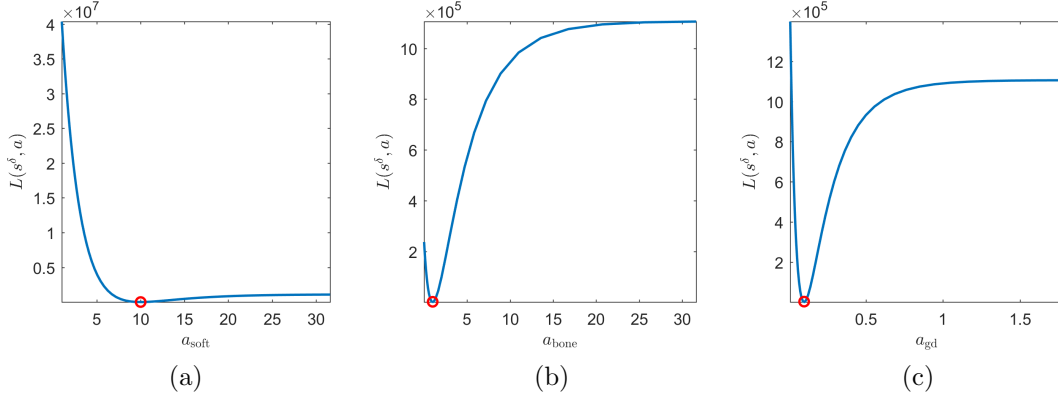

(a) (b) (c)

Figure 2. One dimensional plot of the data fidelity term for a one-pixel image $a \in \mathbb{R}^M$. The ground truth solution is set to $a_{\text{soft}}^{\text{true}} = 10$ g/cm^3, $a_{\text{bone}}^{\text{true}} = 1$ g/cm^3, and $a_{\text{gd}}^{\text{true}} = 0.1$ g/cm^3 (see red circle). Data fidelity is displayed versus the value of one material while the other two materials remain fixed and equal to the ground truth solution.

(a) (b) (c)
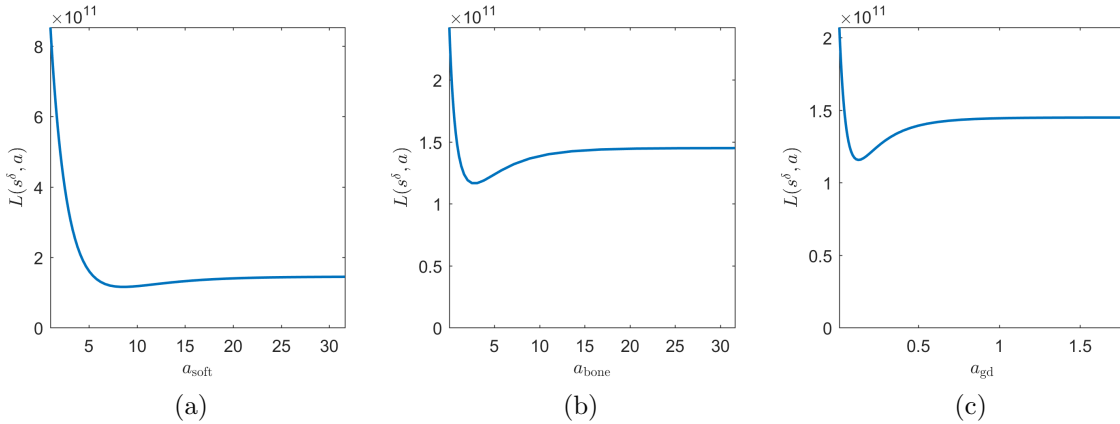
Figure 3. Data fidelity term for the realistic thorax phantom where the three materials are set to a constant image $a_{\text{soft}}^{\text{true}} = 10$ g/cm^3, $a_{\text{bone}}^{\text{true}} = 1$ g/cm^3, and $a_{\text{gd}}^{\text{true}} = 0.1$ g/cm^3. Data fidelity is displayed versus the value of one material while the other two materials remain fixed and equal to (a) $a_{\text{bone}} = 0.15$ g/cm^3 and $a_{\text{gd}} = 0.007$ g/cm^3, (b) $a_{\text{soft}} = 4$ g/cm^3 and $a_{\text{gd}} = 0.007$ g/cm^3, and (c) $a_{\text{soft}} = 4$ g/cm^3 and $a_{\text{bone}} = 0.15$ g/cm^3.

Figure 4 displays the two characteristic sets $\mathcal{A}$ and $\mathcal{B}$ that are introduced in proposition 2 and 3, respectively, for the one-pixel mass density phantom (same as figure 2). The colormap represents the number of energy bins $1 \le i \le I$ satisfying $\mathcal{F}_i(a) \ge s_i$. The solution is set to $a_{\text{soft}}^{\text{true}} = 10$ g/cm^3, $a_{\text{bone}}^{\text{true}} = 1$ g/cm^3, and $a_{\text{gd}}^{\text{true}} = 0.1$ g/cm^3 (see red circle). The data is computed using the parameters plotted in figure 1.

The set $\mathcal{A}$ is depicted in yellow (all bins satisfy $\mathcal{F}_i(a) \geq s_i$) while the set $\mathcal{B}$ is indicated by the grey dashed line.

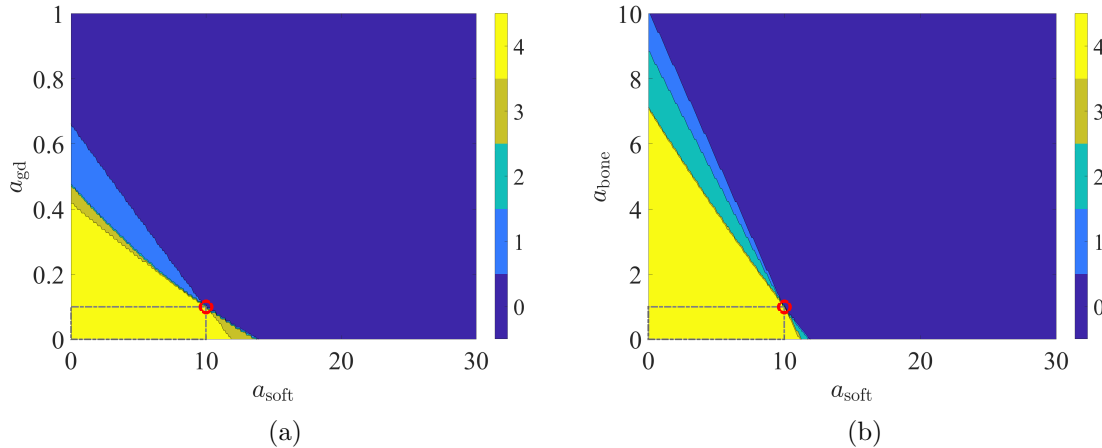(a)   (b)

Figure 4. Characteristic sets included in the solution space. The colormap represents the number of energy bins $1 \leq i \leq I$ satisfying $\mathcal{F}_i(a) \geq s_i$. The solution is set to $a_{\text{soft}}^{\text{true}} = 10$ g/cm^3, $a_{\text{bone}}^{\text{true}} = 1$ g/cm^3, and $a_{\text{gd}}^{\text{true}} = 0.1$ g/cm^3 (see red circle). The data is computed using the parameters plotted in figure 1. The set $\mathcal{A}$ is depicted in yellow (all bins satisfy $\mathcal{F}_i(a) \geq s_i$) while the set $\mathcal{B}$ is indicated by the grey dashed line. Sets are displayed versus the value of two materials while the other material remain fixed and equal to the ground truth solution.

5.3. Convergence of GNB

Solution error and data fidelity versus the number of iterations are shown in figure 5 for GNB for different values of α. GNB leads to an optimal solution as long as α is sufficiently large, where the larger α the more iterations it requires [7, 13]. For these results, we used $\kappa = 10^{-6}$, initial guess $a^0 = 0$ and tolerance $tol = 1.12 \cdot 10^5$. As a comparison, we also provide the solution error for GN for the value of the regularization parameter ($\tilde{\alpha} = 0.23$) that led to the lowest solution error. For GN we set initial guess to $a^0 = 0$ and tolerance to $\psi = 10^{-4}$, which corresponded to a value of the data fidelity term of $1.12 \cdot 10^5$. GN converged in $N^{\text{tot}} = 16$ iterations for $\tilde{\alpha} = 0.23$ (see figure 5(b)). GNB solves a sequence of subproblems, so it needs a larger number of iterations. It needed $N^{\text{tot}} = 28$ and $N^{\text{tot}} = 40$ for $\alpha = 2$ and $\alpha = 10$, respectively.

Figure 6(a) shows the solution error versus the regularization parameter $\tilde{\alpha}$ for GN and versus the parameter α for GNB. While GNB provides the same solution for all α's larger than 0.5, GN depends dramatically on the choice of α. Although GN converged in only $N^{\text{tot}} = 16$ iterations (for $\tilde{\alpha} = 0.23$, see figure 5(b)), extra decompositions for different regularization parameters are required. On the contrary, GNB can be used with a large $\tilde{\alpha}$ chosen *a priori*.

Figure 6(b) shows that the number of Bregman iterations is almost constant with α and equal to $k = 15$ in the cases investigated here. Figure 6(c) shows that the total number of Gauss Newton iterations N^{tot} is linearly related to the parameter α.
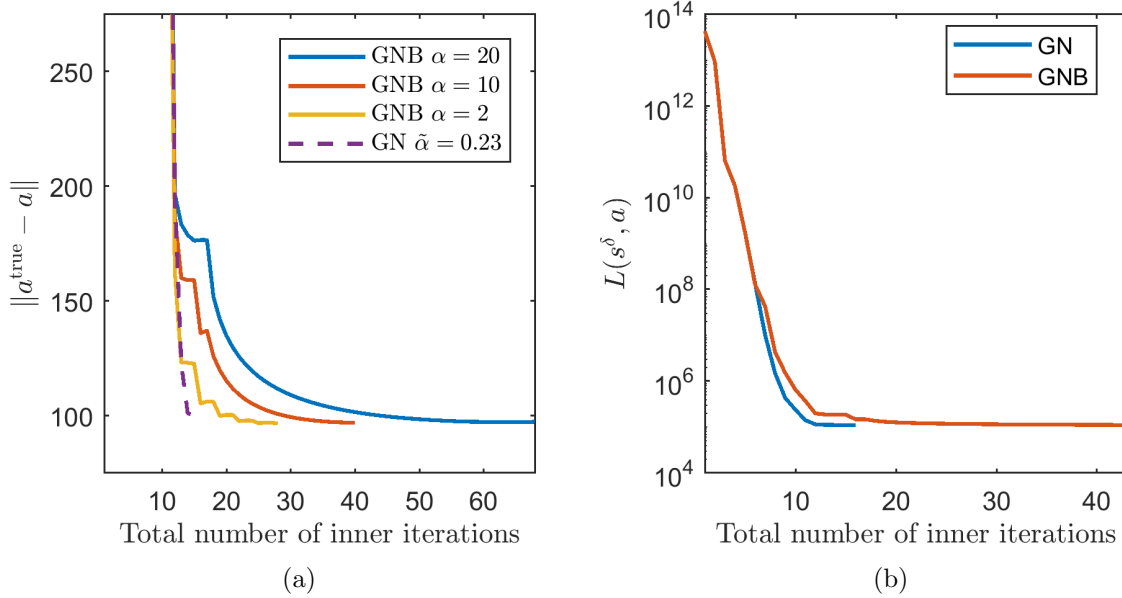
Figure 5. Convergence of the algorithms. (a) Solution error versus the number of Gauss Newton iterations for GNB (full line) and for GN (dotted line). (b) Data fidelity decrease for GNB ($\alpha = 10$) and for GN ($\tilde{\alpha} = 0.23$). We used $a^0 = 0$ as initial guess.
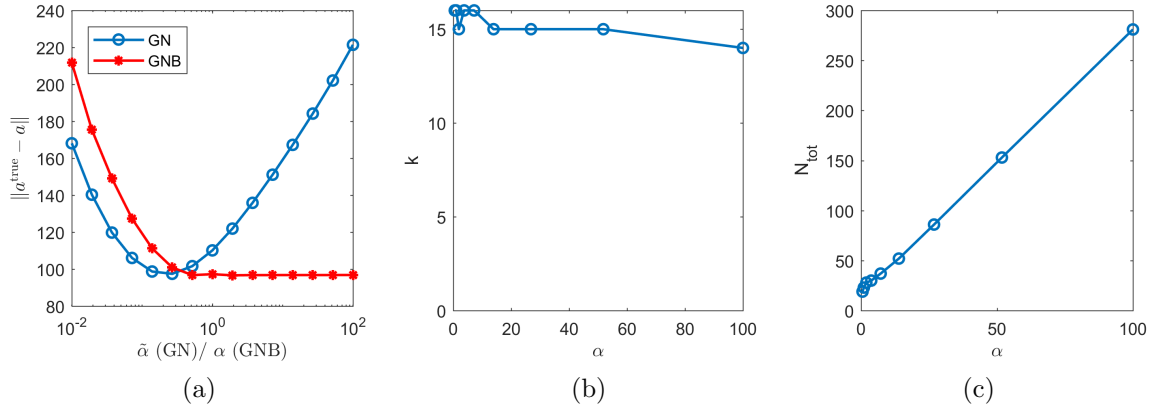

Figure 6. Influence of the hyperparameter α. (a) Solution error for GN and GNB. (b) Total number of Bregman iterations for GNB. (c) Total number of Gauss Newton iterations for GNB.

Figure 7 shows the number of Gauss Newton iterations N^k needed to minimize Ψ^k, for different values of α. An important remark is that GNB required nearly only one iteration $N^k = 1$ to solve each subproblem for $k > 1$, so it adds just few iterations with respect to GN.

Figure 8 shows decomposed images by GN and GNB. Second row in figure 8 gives decomposed images by GN for an optimal regularization parameter $\tilde{\alpha} = 0.23$. Third to fifth rows are the decomposed images by GNB for different number of iterations, setting $\alpha = 10$, $\kappa = 1 \cdot 10^{-6}$ and $a^0 = 0$. Third and fourth rows correspond to an intermediate and final solution of the first subproblem ($k = 1$). The solution of the first subproblem
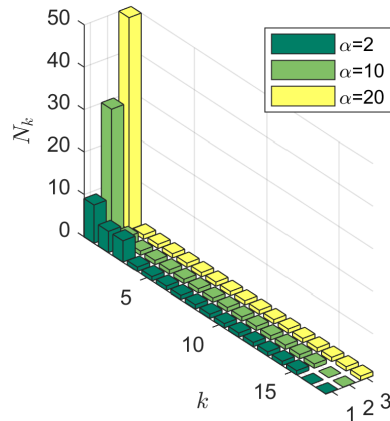
Figure 7. Number of Gauss Newton iterations, N^k, needed at each Bregman iteration, k, for different values of α.

provides a good approximation of the decomposed materials but images are smooth given the large value of α. Further Bregman iterations ($k = 25$) are needed to find the optimal solution (bottom row in figure 8).

We evaluated two different strategies for the initialization of the inner problem in GNB (algorithm 1) (figure 9). Using the solution found in the previous subproblem as initial guess in the next subproblem, i.e., setting $a^{k,n=0} = a^{k-1,N_{k-1}}$, led to an optimal solution in $N^{\mathrm{tot}} = 120$ iterations for $\alpha = 100$. Initializing each Bregman iteration to a zero image, i.e., setting $a^{k,n=0} = 0$, led to larger solution error and diverged quickly after few updates of the Bregman distance.

Figure 10 shows the influence of the number of inner iterations in GNB. Previous results correspond to full convergence of the inner problem. Taking only one iteration for the inner problem ($N^1 = 1$) resulted in unstable subgradient update and did not converge. However, setting $N^k = 1$ for $k > 1$ (after the first subproblem is solved to high accuracy) almost halved the number of iterations with respect to $N^k = 2$. We remark that convergence of the inner iteration required $N^1 = 15$ for $k = 1$ and at most $N^k = 3$ for $k > 1$.
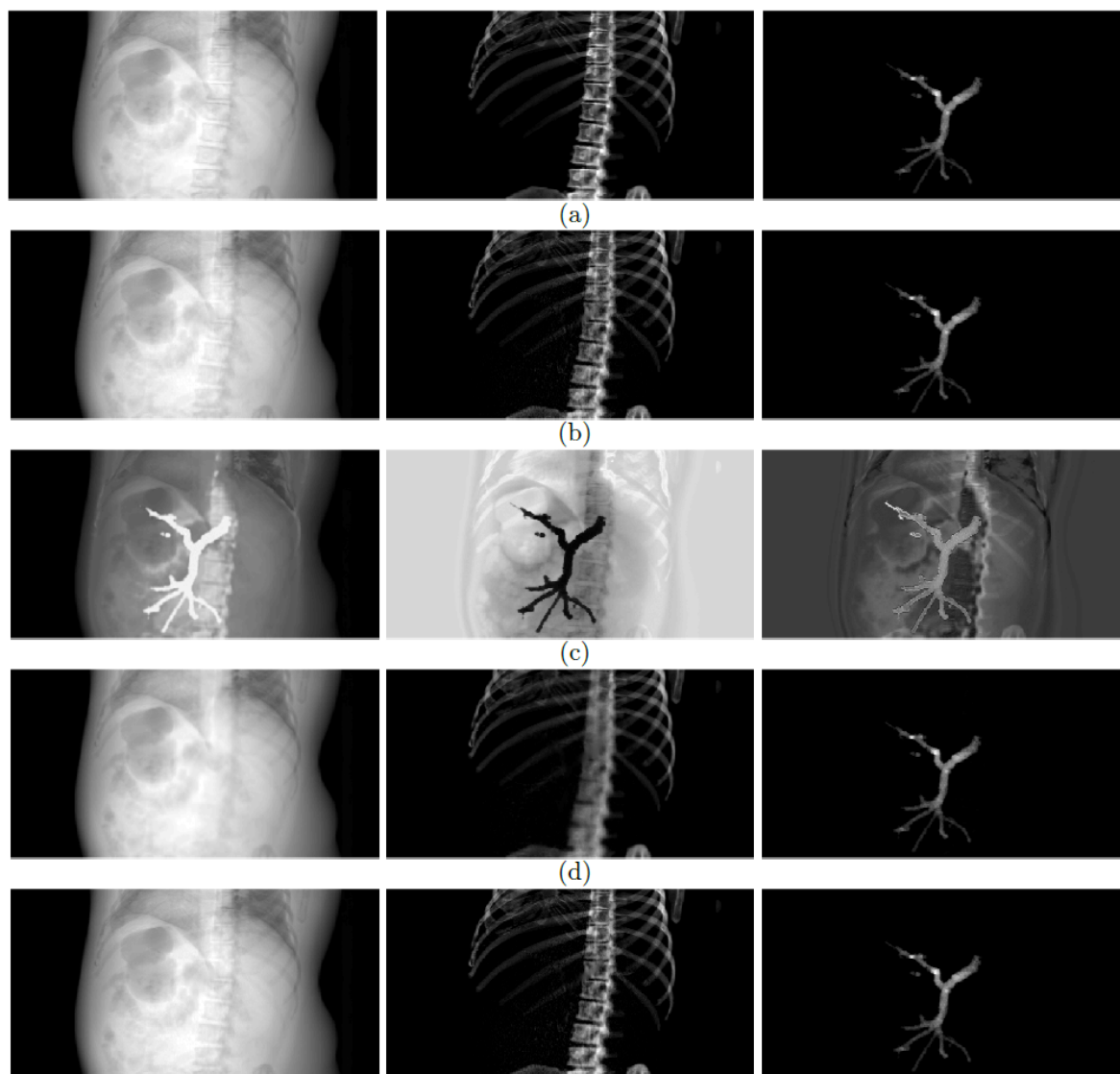
Figure 8. (a) Reference image for the thorax phantom comprising three materials: soft tissue, bone and portal vein marked with gadolinium. A particular projection view at angle of 60° is considered. (b) Material images decomposed by GN for $\tilde{\alpha} = 0.23$ and $a^0 = 0$. (c)-(e) Material images decomposed by GNB at different iterations: (c) $k = 1$, $n = 5$, $N^{\text{tot}} = 5$, (d) $k = 1$, $n = 15$, $N^{\text{tot}} = 15$ and (e) $k = 25$, $n = 1$, $N^{\text{tot}} = 44$, which is the optimal decomposition with GNB. k indicates the (outer) Bregman iteration, n the (inner) Gauss Newton iterations, and N^{tot} is the total number of Gauss-Newton iterations. We set $\alpha = 10$, $\kappa = 1 \cdot 10^{-6}$ and $a^0 = 0$.
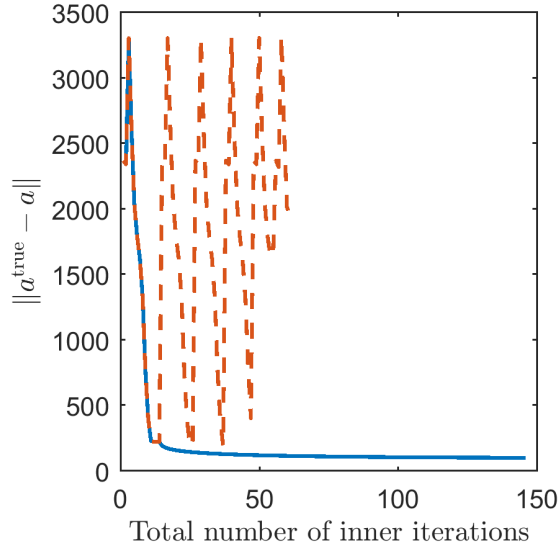
Figure 9. Influence of the initialization of the Bregman iteration. The solution error is plotted for two types of initialization for the minimization of Ψ^k: using the solution found at the previous Bregman iteration (full line) and using a zero image for all Bregman iteration (dotted line). We used $\alpha = 100$, $\kappa = 1 \cdot 10^{-6}$ and $a^0 = 0$.
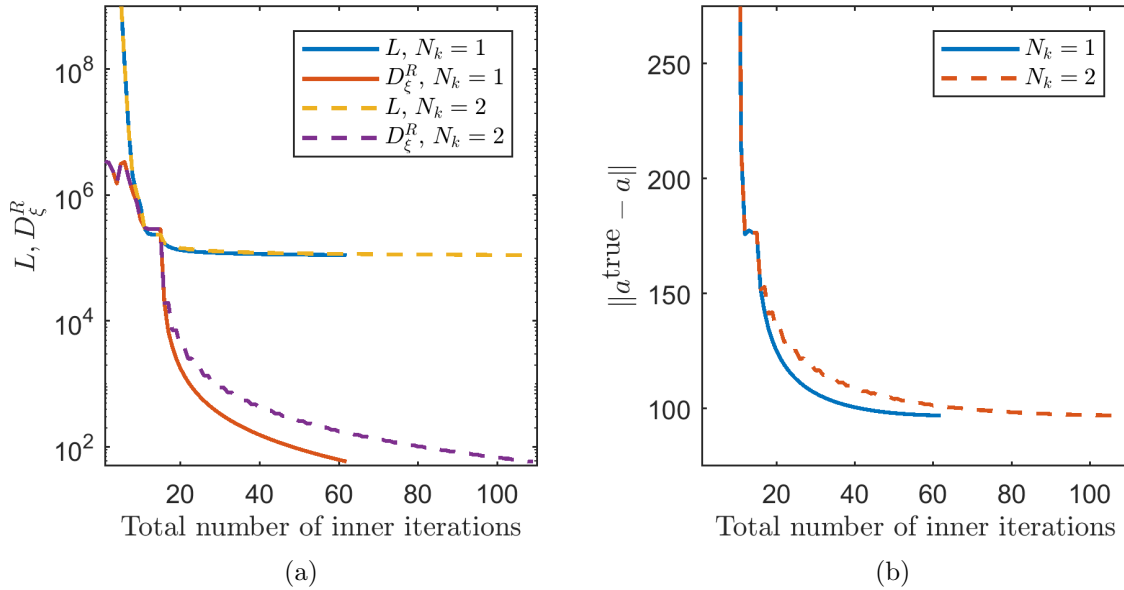

Figure 10. Influence of the number of inner iterations N^k in GNB. a) Data fidelity, L, and Bregman distance, D_ξ^R, versus the total number of inner iterations in the algorithm considering all subproblems, for $N^k = 1$ and $N^k = 2$ for $k > 1$ (the first subproblem is solved at full convergence with $N^1 = 15$). b) Solution error norm for the same parameters used in (a).
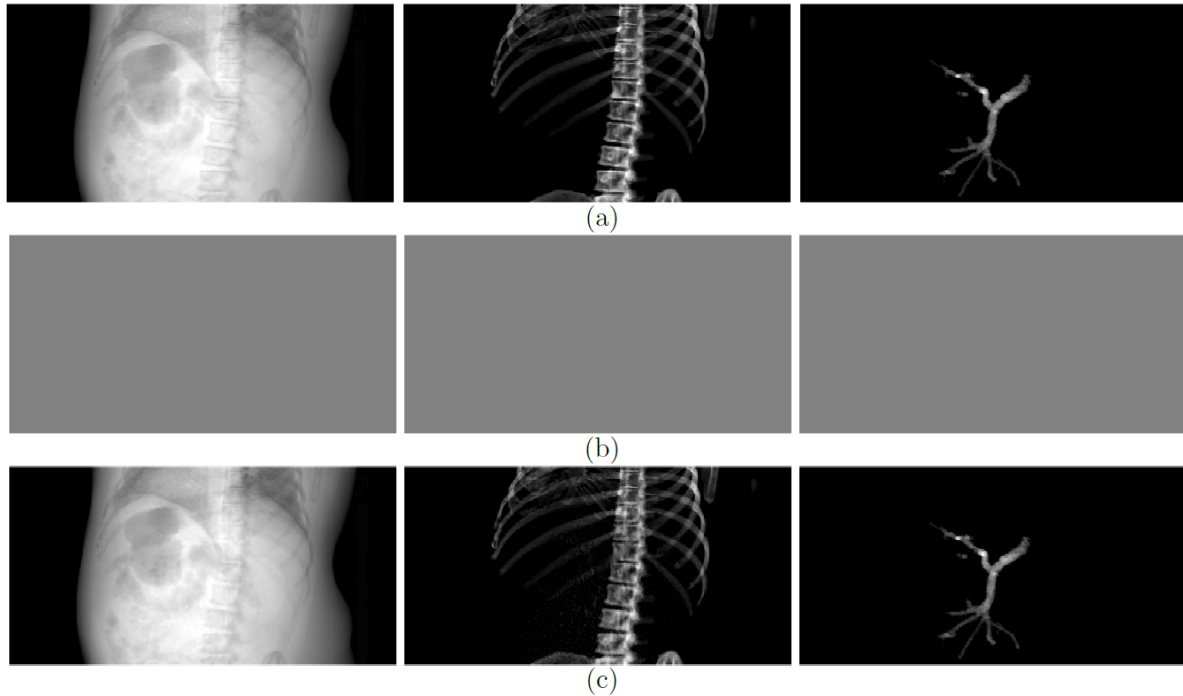
Figure 11. Influence of the initial guess. (a) Reference image and decomposed images by using (b) GN with $\tilde{\alpha} = 0.23$ and (c) GNB with $\alpha = 10$ and $\kappa = 10^{-6}$. The initial guess was $a^0 = 10$ for both methods.

5.4. Influence of the initial guess

In previous sections the initial guess was set to a zero constant image for all materials, which belongs to the convex part of the data fidelity term. To assess now the effect of the Bregman distance on global convergence, we set the initial guess to large values ($a^0 = (10, 10, 10)$) in order to make sure it belong to the nonconvex region. GNB converged to the optimal solution independently of the initial guess. On the contrary, GN did not converge, stopping at the first iteration as a descent direction could not be found (figure 11).

GNB presented a dependence on κ when the initial guess belonged to the nonconvex region. In this case, κ was critical as the Hessian was ill-conditioned, which occurred for a large value of the initial guess. We used $\kappa = 10^{-6}$. When the initial guess belonged to the convex region, GNB led to same results independently of the value of κ as long as it was small enough $\kappa \leq 10^{-4}$.
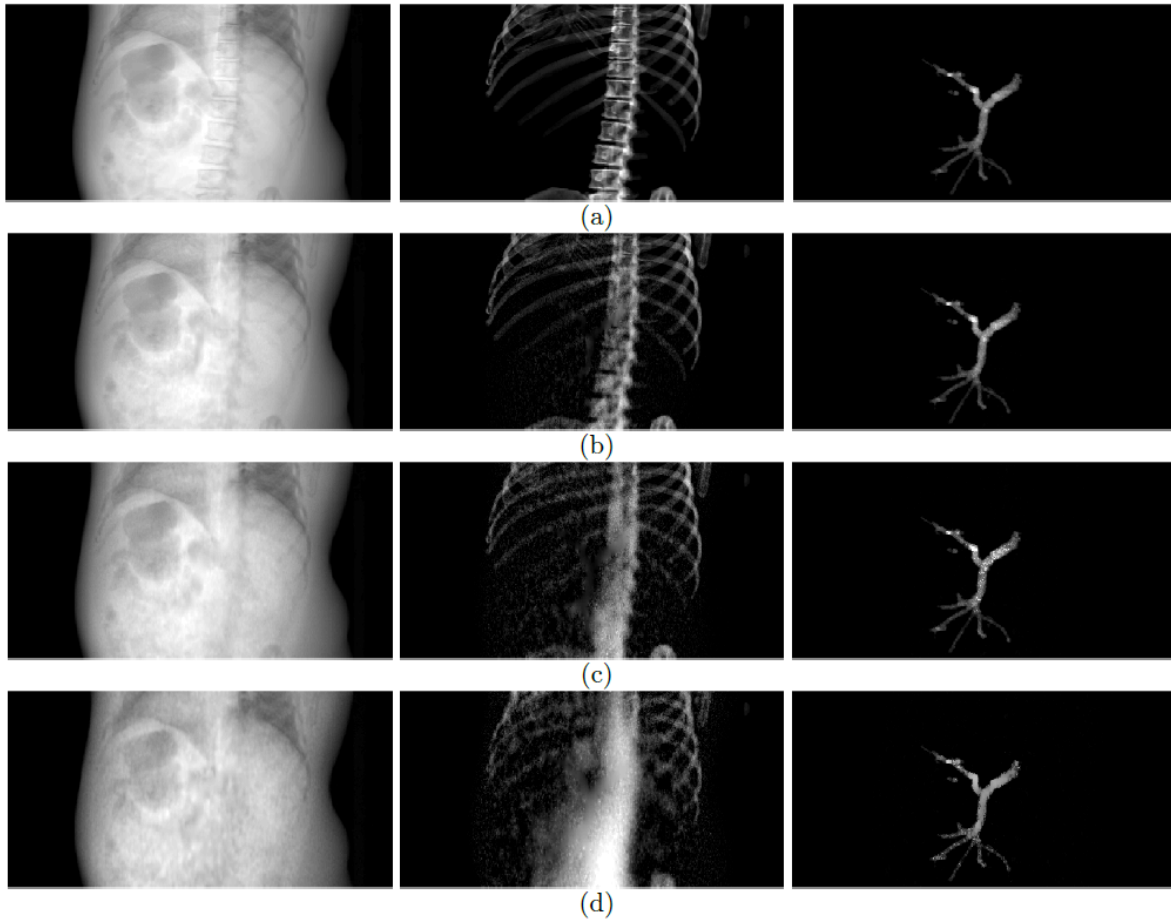
Figure 12. Material decomposed images by GNB for different number of photons n_0. Pure Poisson noise is considered, i.e., $\sigma = 0$. (a) Reference images and decomposed images for (b) $n_0 = 10^7$, (c) $n_0 = 10^6$ and (d) $n_0 = 10^5$. We used the following parameters: $\alpha = 10$, $\kappa = 10^{-6}$, and initial guess $a^0 = 0$.

5.5. Influence of noise and phantom

In spectral CT, an important problem is the low photon count in each energy bin corresponding to noisy data. To illustrate the impact of noise on GNB, we consider first pure Poisson noise. Figure 12 shows decomposed material images by GNB for different number of incident photons from $n_0 = 10^8$ to $n_0 = 10^5$. The lower the number of incident photons the more regularized the solution. The tolerance *tol* was set to $1.3 \cdot 10^5$, $1.5 \cdot 10^5$ and $1.8 \cdot 10^5$ for $n_0 = 10^7$, $n_0 = 10^6$ and $n_0 = 10^5$, respectively.

Figure 13 shows material decomposed images by GNB for mixed Poisson noise model with $n_0 = 10^7$ and different values of σ ranging from 5 to 50. GNB is able to decompose materials even with large σ but for σ larger than 20, images are noisier, which impacts the recovered concentration of gadolinium.

We also show the decompositions obtained for a different projection angle (here 180°) in figure 14.
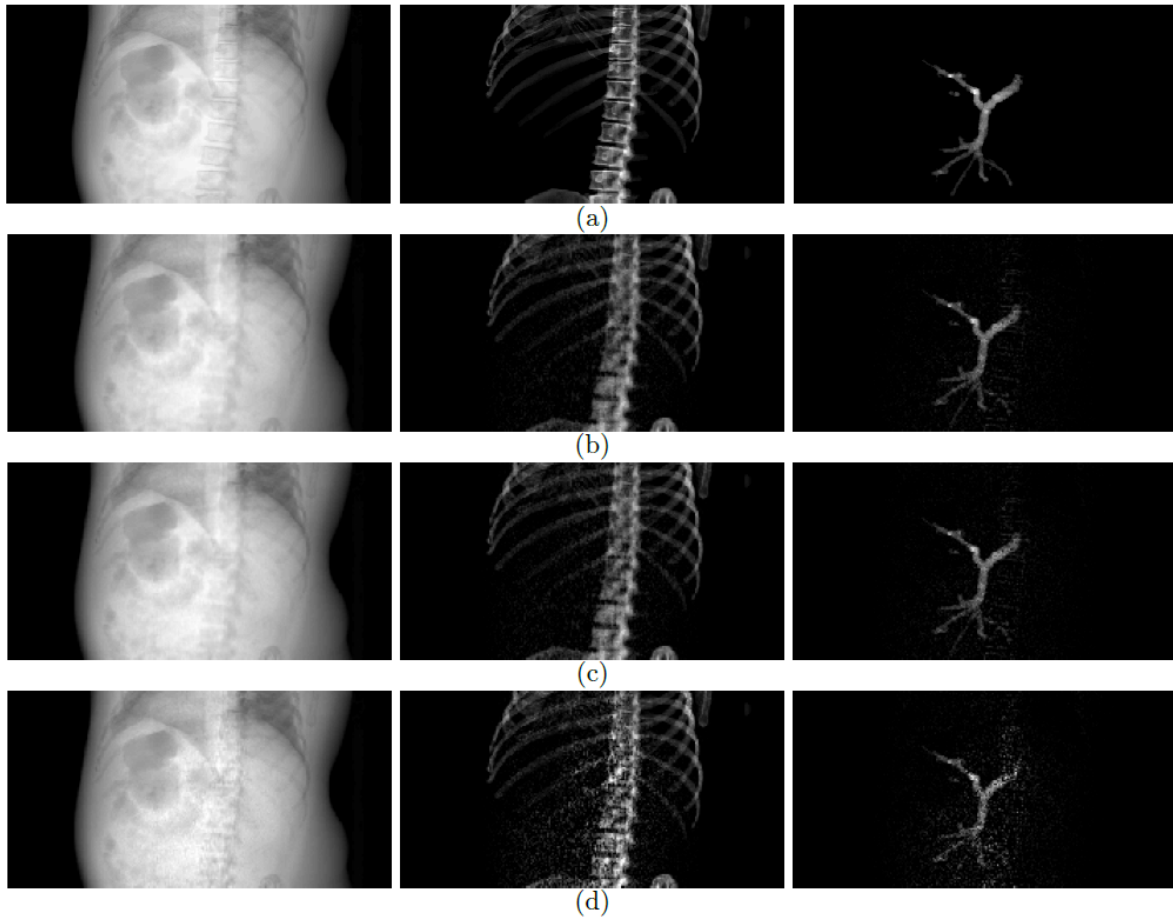
Figure 13. Material decomposed images by GNB for mixed Poisson noise model, $s^\delta \sim \mathcal{P}(\lambda = s) + \mathcal{N}(\mu, \sigma^2)$, for $n_0 = 10^7$ and concentration of gadolinium of 0.1 g·cm^{-3}. (a) Reference images and decomposed images for (b) $\sigma = 5$, (c) $\sigma = 20$, and (d) $\sigma = 50$. We used the following parameters: $\alpha = 10$, $\kappa = 10^{-6}$, $tol = 1.23 \cdot 10^5$ and initial guess $a^0 = 0$.

6. Discussion

We proposed and assessed a regularized iterative scheme based on the Bregman distance (GNB) for the solution of the material decomposition problem in spectral CT, which is a nonlinear inverse ill-posed problem and for which we consider the minimization of a regularized functional. First, we proved that the problem is convex in a region of the solution space and that the zero solution belongs to this region. Then, using numerical examples we showed that the problem is nonconvex for large values of the decomposed materials.

Using a numerical thorax phantom, we showed that zero initialization is a suitable initialization strategy for convex optimization methods such as a regularized Gauss-Newton algorithm method [41]. GN led to an optimal solution as long as the initial guess belonged to the convex region of the data fidelity term. On the contrary, GN led to a wrong solution when the initial guess is chosen far from the convex region.
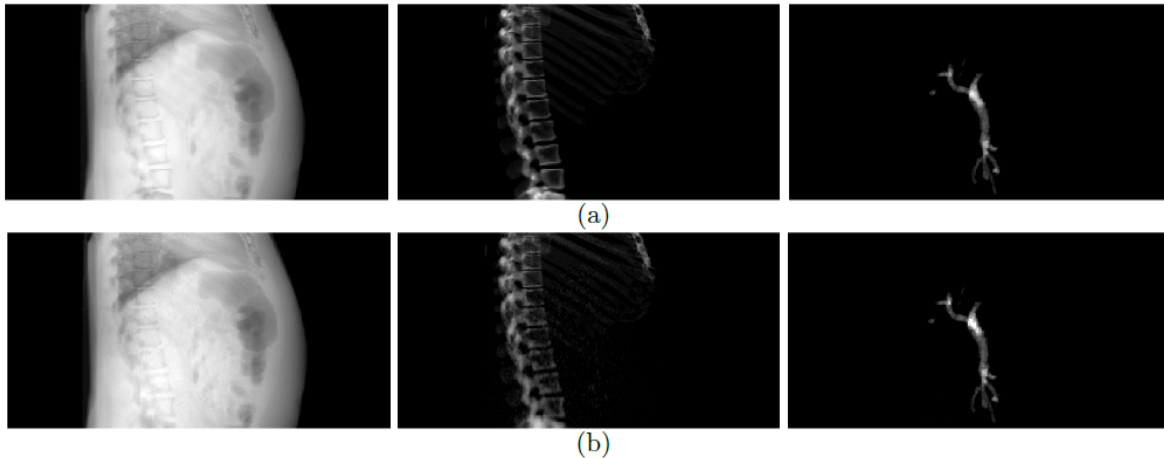
Figure 14. Material decomposed images provided for a different projection view (for an angle of $180°$). (a) The ground truth. (b) Decomposed data provided by GNB. The measurement vector is simulated for $n_0 = 10^8$ photons and the decomposition is performed using the following parameters: $\alpha = 10$, $\kappa = 10^{-6}$, and initial guess $a^0 = 0$.

We then assessed a regularized iterative scheme based on the Bregman distance. The proposed method has some advantages with respect to more standard approaches. It is robust to the choice of the initial guess, as it led to global minimum even when the initial guess was in the nonconvex region. On the contrary, Gauss-Newton method failed in this case. This is a relevant feature of the Bregman distance, which makes each subproblem locally convex around the current estimate. This makes Bregman iterative algorithms more robust but at the price of requiring more iterations than an equivalent Gauss-Newton method. Nevertheless, we found that the number of iterations required was not very large: around ten inner iterations to solve the first subproblem ($k = 1$) but generally only one to three inner iterations for $k > 1$. In addition, standard approaches such as Gauss-Newton method requires optimal selection of the regularization parameter using the L-curve or similar method [52, 53]. In this work, we selected the solution provided by GN method that led to the lowest solution error by assessing a large number of regularization parameters. As Bregman iterative methods do not require optimal selection of the regularization parameter, the extra iterations needed by these methods may not translate in much larger computational demand than standard regularization methods. In addition, we found a linear relation between the number of total iterations and the parameter α, which has been previously shown for linear measurement operators [7, 13]. Hence, α can be tuned to decrease the number of required iterations. The smaller α the lower the number of iterations but a very low value may be less robust.

A key point in the construction of these iterative schemes is to select efficient methods for solving the subproblem (17a). We have proposed an efficient Gauss-Newton step and found that it required only few inner iterations for $k > 1$ (algorithm 1). This may indicate that subproblems are approximately linear when close to the solution.

Reconstructed images show that the first subproblem leads to a good estimate of the decomposed materials and then further Bregman iterations are used to improve the solution. This could be then further exploited to reduce computationally complexity. In fact, fixing the number of inner iterations to one for $k > 1$ (after the first subproblem is solved to high accuracy) converged to the optimal solution and almost halved the number of iterations with respect to taking two inner iterations.

In [20], authors proposed two iterated methods based on approximating the subproblem (17a): a Levenberg-Marquardt-type (ILM) method that linearized the forward operator, and a Landweber-type method that linearized the least square data fidelity term. There are similarities between GNB and ILM with only one inner iteration. ILM replaces (17) by the iterated scheme

$$a^{k+1} = \arg\min_a \frac{1}{2}\|\mathcal{F}(a^k) + \mathcal{F}'(a^k)(a - a^k) - s^\delta\|_W^2 + \alpha D_{\xi^k}^R(a, a^k), \tag{22a}$$

$$\xi^{k+1} = \xi^k - \alpha^{-1}\left(W\mathcal{F}'(a^k)\right)^\top W\left(\mathcal{F}(a^k) + \mathcal{F}'(a^k)(a^{k+1} - a^k) - s^\delta\right). \tag{22b}$$

Doing only one inner iteration on GNB (algorithm 1) and ILM (22) provides the same solution update a^{k+1} but subgradient updates are different. In order to compare subgradient updates given by the two methods, we rewrite ILM's update as $\xi^{k+1} \simeq \xi^k - \alpha^{-1}\left(W\mathcal{F}'(a^k)\right)^\top W(F(a^{k+1}) - s)$. While the Jacobian in ILM is evaluated on the previous estimate a^k, as the forward operator has been linearized in the data fidelity term, the Jacobian in GNB is evaluated on the current estimate a^{k+1}. This difference seems to have a significant impact when approximating the inner problem, as discussed below.

An unexpected result in the analysis was to find that fixing the number of inner iterations to one in GNB resulted in an unstable subgradient update. We observed that after the first subgradient update, the Bregman distance becomes negative and then the algorithm diverges. As the Bregman distance is given by $D_{\xi^k}^R(a, a^k) = R(a) - R(a^k) - <\xi^k, a - a^k >\ge 0$ for $\xi^k \in \partial R(a^k)$, then we assume that $\xi^k \notin \partial R(a^k)$. To further understand this issue, we computed GNB and ILM when taking one inner iteration for all k. While GNB did not converge, ILM converged to a solution close to the optimal one but results were slightly worse than those provided by GNB when solving exactly the first subproblem. Further work is needed for a rigorous comparison between GNB and ILM. From these results and the difference between GNB and ILM discussed above, it seems that given the nonlinearity, taking only few inner iterations for GNB may not be stable in general. However, if the first inner problem is solved to high accuracy, it is then feasible to take only one inner iteration. This may be due to the fact that solving the first subproblem brings the first iterate close to the optimal solution and the rest of subproblems are approximately linear.

Initialization of each Bregman iteration for GNB was critical. Using the solution found at the previous iteration led to robust convergence. On the other hand, a zero initialization strategy for each subproblem required a large number of iterations to solve each subproblem and became unstable as the number of Bregman iterations increased.

This can be explained by the fact that the problem is nonlinear and that as the number of Bregman iterations increases, the solution is further away from the zero initial guess. Nevertheless, for linear problems, zero initialization was a valid strategy for Bregman iterative methods in a wide variety of applications [13, 46].

In [20] authors included an extra ℓ_2-term (weighted by κ in (18)) to make the functional associated with total variation convex on each subproblem. Here we found a dependence on κ. When the initial guess was in the nonconvex region, this term was essential to reach to a global minimum. An optimal solution was found by using a sufficiently small value of κ ($\kappa \leq 10^{-4}$). The effect of κ can be explained by the fact that when the initial guess is in the nonconvex region the Hessian is ill-conditioned, where the nonconvex region in this problem corresponds to large values of the materials. For the case where the initial guess was in the convex region, this term did not play a role and any small value of κ (including $\kappa = 0$) led to the same solution.

In conclusion, we have studied convexity for the material decomposition problem and showed that an iterative scheme based on the Bregman distance had a positive effect to find a global minimum. In addition to this, the Bregman distance presents other benefits. For instance, it can efficiently minimize nonsmooth functional such as total variation or impose a nonnegativity constraint, which has been exploited for linear measurements [7, 12, 13]. Two variants of the proposed GNB algorithm, GNB with a nonnegativity constraint and GNB with a nondifferentiable convex functional, are included in the appendix B. Another benefit of using the Bregman distance is to correct for the lost of contrast that affects total variation [9]. This is crucial for quantitative imaging techniques such as spectral CT. In future work we will investigate these other benefits of the Bregman distance for solving the material decomposition problem in spectral CT.

Acknowledgments

Appendix A. Hessian of the forward mapping

Let $H_i(a)$ be the Hessian matrix of the operator $\mathcal{F}_i$. It is defined as

$$(H_i)_{m,m'} = \frac{\partial^2 \mathcal{F}_i}{\partial a_m \partial a_{m'}}. \tag{A.1}$$

The operator $\mathcal{F}_i$ is continuously twice differentiable and its second derivative is given by

$$\frac{\partial^2 \mathcal{F}_i}{\partial a_m \partial a_{m'}}(a) = \int_{\mathcal{E}} \tau_m(E)\tau_{m'}(E)g_i(a, E)\,\mathrm{d}E \tag{A.2}$$

where $g_i(a, E)$ is given by (11). Discretising the energy interval $\mathcal{E} = [0, E_{\max}]$ with an energy step ΔE, the integral (A.2) can be approximated by

$$\frac{\partial^2 \mathcal{F}_i}{\partial a_m \partial a_{m'}}(a) = \sum_j \tau_m(E_j)\tau_{m'}(E_j)g_i(a, E_j)\Delta E. \tag{A.3}$$

Hence, it can be noted that the Hessian takes the following matrix form

$$H_i(a) = T^\top G_i(a)T, \tag{A.4}$$

where $G_i(a) = \mathsf{diag}\left(g_i(a, E_1), \ldots, g_i(a, E_J)\right)\Delta E$ is the vector of photons incident on the detector, and $T = (\tau_{j,m}) \in \mathbb{R}^{J \times M}$ is the mass attenuation matrix defined by $\tau_{j,m} = \tau_m(E_j)$.

Appendix B. Extensions of GNB

Appendix B.1. GNB with nonnegativity constraint

Including a nonnegativity constraint in GNB can be done by combining the Bregman distance with a splitting technique, which leads to the split Bregman method [12], [13]. Proceeding as in [12], let $\mathbb{1}_{a \geq 0}$ be the indicator function of the set $\mathbb{R}^N_{\geq 0}$ and add an auxiliary variable b to allow for the splitting, the material decomposition problem with a nonnegativity constraint can be posed as follows:

$$\underset{a,b}{\arg\min}\, R(a) + \mathbb{1}_{b \geq 0} \quad \text{st.} \quad \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|^2_W \leq \sigma, b = a. \tag{B.1}$$

Defining the convex functional $U(a, b) = R(a) + \mathbb{1}_{b \geq 0}$ and the Bregman distance associated with this functional

$$D^U_{\xi^k}(a, a^k, b, b^k) = U(a, b) - U(a^k, b^k) - <\xi^k_a, a - a^k> - <\xi^k_b, b - b^k>, \tag{B.2}$$

the constrained problem (B.1) can be solved by the following iterative scheme:

$$(a^{k+1}, b^{k+1}) = \underset{a,b}{\arg\min}\, \alpha D^U_{\xi^k}(a, a^k, b, b^k) + \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|^2_W + \frac{1}{2}\|b - a\|^2$$

$$= \underset{a,b}{\arg\min}\, \alpha \left(R(a) + \mathbb{1}_{b \geq 0} - <\xi^k_a, a - a^k> - <\xi^k_b, b - b^k>\right)$$

$$+ \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|^2_W + \frac{1}{2}\|b - a\|^2 \tag{B.3a}$$

$$\xi^{k+1}_a = \xi^k_a - \alpha^{-1}(W\mathcal{F}'(a^{k+1}))^\top(W\mathcal{F}(a^{k+1}) - Ws^\delta) - \alpha^{-1}(a^{k+1} - b^{k+1}) \tag{B.3b}$$

$$\xi^{k+1}_b = \xi^k_b - \alpha^{-1}(b^{k+1} - a^{k+1}). \tag{B.3c}$$

Variables a and b in (B.3a) are de-coupled, so they can be solved in two alternating steps:

$$a^{k+1} = \arg\min_a \Psi^k(a) + \frac{1}{2}\|b^k - a\|^2 \tag{B.4a}$$

$$b^{k+1} = \arg\min_b \frac{1}{2}\|b - a^{k+1}\|^2 + \alpha\mathbb{1}_{b\geq 0} - \alpha < \xi_b^k, b - b^k > . \tag{B.4b}$$

Subproblem (B.4a) is a slight variation of subproblem (17a) where a quadratic term is added to the previous cost function $\Psi^k(a)$ and where ξ_a has an extra term. Hence, minimization of (B.4a) can be also done with GN as in algorithm 1. Subproblem (B.4b) can be solved analytically by using a thresholding operation [12], [13].

Appendix B.2. GNB for nondifferentiable functional

The minimization of convex nondifferentiable functional such as total variation can be solved efficiently using the split Bregman method, which has been widely exploited for linear measurements [7]. In specific, the split Bregman method is very efficient for solving L1-regularized problems. Let $R(a)$ be a nondifferentiable convex functional of the form $R(a) = \|\Phi(a)\|_1$, ie. total variation $R(a) = \|\nabla a\|_1$, we replace $\Phi(a)$ by d and consider the following constrained problem:

$$\arg\min_{a,d} \|d\|_1 \quad \text{st.} \quad \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|_W^2 \leq \sigma, d = \Phi(a). \tag{B.5}$$

Proceeding as above, we define $U(a,d) = \|d\|_1$, the constrained problem (B.5) can be solved by the following iterative scheme:

$$(a^{k+1}, d^{k+1}) = \arg\min_{a,d} \alpha D_{\xi^k}^U(a, a^k, d, d^k) + \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|_W^2 + \frac{1}{2}\|d - \Phi(a)\|^2$$

$$= \arg\min_{a,d} \alpha\left(\|d\|_1 - < \xi_a^k, a - a^k > - < \xi_d^k, d - d^k >\right) + \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|_W^2$$

$$+ \frac{1}{2}\|d - \Phi(a)\|^2 \tag{B.6a}$$

$$\xi_a^{k+1} = \xi_a^k - \alpha^{-1}(W\mathcal{F}'(a^{k+1}))^\top(W\mathcal{F}(a^{k+1}) - Ws^\delta) - \alpha^{-1}\Phi'(a)^T(\Phi(a^{k+1}) - d^{k+1}) \tag{B.6b}$$

$$\xi_d^{k+1} = \xi_d^k - \alpha^{-1}(d^{k+1} - \Phi(a^{k+1})). \tag{B.6c}$$

Variables a and d in (B.6a) are de-coupled, so they can be solved in two alternating steps:

$$a^{k+1} = \arg\min_a \frac{1}{2}\|\mathcal{F}(a) - s^\delta\|_W^2 + \frac{1}{2}\|d^k - \Phi(a)\|^2 - \alpha < \xi_a^k, a - a^k > \tag{B.7a}$$

$$d^{k+1} = \arg\min_d \alpha\|d\|_1 + \frac{1}{2}\|d - \Phi(a^{k+1})\|^2 - \alpha < \xi_d^k, d - d^k > . \tag{B.7b}$$

Subproblem (B.7a) is easier to solve than subproblem (17a) as it does not depend on $R(a)$ and a quadratic term is added to the previous cost function. Subproblem (B.7b) can be solved analytically by using a thresholding operation [12], [13].

References

[1] Ronny Ramlau and Gerd Teschke. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Probl.*, 21(5):1571–1592, 2005.

[2] Ronny Ramlau and Gerd Teschke. A Tikhonov-based projection iteration for nonlinear Ill-posed problems with sparsity constraints. *Numer. Math.*, 104(2):177–203, 2006.

[3] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An Iterative Regularization Method for Total Variation-Based Image Restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[4] Wotao Yin, Stanley Osher, Donald Goldfarb, and Jerome Darbon. Bregman Iterative Algorithms for l1-Minimization with Applications to Compressed Sensing. *SIAM J. Imaging. Sci.*, 1(1):143–168, 2008.

[5] Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Linearized Bregman iterations for compressed sensing. *Math. Comput.*, 78(267):1515–1536, 2009.

[6] Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Split Bregman Methods and Frame Based Image Restoration. *Multiscale Modeling & Simulation*, 8(2):337–369, 2010.

[7] Tom Goldstein and Stanley Osher. The Split Bregman Method for L1-Regularized Problems. *SIAM J. Imaging. Sci.*, 2(2):323–343, 2009.

[8] Stanley Osher, Yu Mao, Bin Dong, and Wotao Yin. Fast linearized Bregman iteration for compressive sensing and sparse denoising. *Commun Math Sci.*, 8(1):93–111, 2010.

[9] Martin Burger and Stanley Osher. A Guide to the TV Zoo. *Lecture Notes in Mathematics*, 2090:1–70, 2013.

[10] Erny Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. Technical report, 2009.

[11] Chunlin Wu and Xue-Cheng Tai. Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and High Order Models. *SIAM J. Imaging. Sci.*, 3(3):300–339, 2010.

[12] S Setzer, G Steidl, and T Teuber. Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image Represent.*, 21(3):193–199, 2010.

[13] J F P J Abascal, J Chamorro-Servent, J Aguirre, S Arridge, T Correia, J Ripoll, J J Vaquero, and M Desco. Fluorescence diffuse optical tomography using the split Bregman method. *Med. Phys.*, 38(11):6275–6284, 2011.

[14] Hao Gao, Hengyong Yu, Stanley Osher, and Ge Wang. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). *Inverse Probl.*, 27(11):115012, 2011.

[15] Gerlind Plonka and Jianwei Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. *Int. J. Wavelets Multiresolution*, 09(01):79–110, 2011.

[16] J F P J Abascal, P Montesinos, E Marinetto, J Pascau, and M Desco. Comparison of total variation with a motion estimation based compressed sensing approach for self-gated cardiac cine MRI in small animal studies. *PLoS One*, 9(10), 2014.

[17] Juan F P J Abascal, Manuel Desco, and Juan Parra-Robles. Incorporation of prior knowledge of the signal behavior into the reconstruction to accelerate the acquisition of MR diffusion data. *IEEE Trans. Med. Imaging*, (in press), 2017.

[18] Hao Gao, Stanley Osher, and Hongkai Zhao. Quantitative Photoacoustic Tomography. pages 131–158. Springer, Berlin, Heidelberg, 2012.

[19] A Gholami, H R Siahkoohi, W Navidi, L Tenorio, and W Yin. Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints. *Geophys. J. Int.*, 180(2):871–882, 2010.

[20] Markus Bachmayr and Martin Burger. Iterative total variation schemes for nonlinear inverse problems. *Inverse Probl.*, 25(10):105004, 2009.

[21] Katsuyuki Taguchi and Jan S Iwanczyk. Vision 20/20: Single photon counting x-ray detectors in medical imaging. *Med. Phys.*, 40(10):100901, 2013.

[22] Bjorn J Heismann, Bernhard T Schmidt, and Thomas Flohr. *Spectral computed tomography.* SPIE, 2012.

[23] Huanjun Ding, Bo Zhao, Pavlo Baturin, Farnaz Behroozi, and Sabee Molloi. Breast tissue decomposition with spectral distortion correction: A postmortem study. *Med. Phys.*, 41(10):101901, 2014.

[24] David P Cormode, Ewald Roessl, Axel Thran, Torjus Skajaa, Ronald E Gordon, Jens-Peter Schlomka, Valentin Fuster, Edward A Fisher, Willem J M Mulder, Roland Proksa, and Zahi A Fayad. Atherosclerotic Plaque Composition: Analysis with Multicolor CT and Targeted Gold Nanoparticles1. *Radiology*, 256(3):774–782, 2010.

[25] L Boussel, P Coulon, A Thran, E Roessl, G Martens, M Sigovan, and P Douek. Photon counting spectral CT component analysis of coronary artery atherosclerotic plaque samples. *Br. J. Radiolg*, 87(1040):20130798, 2014.

[26] Y Long and J A Fessler. Multi-Material Decomposition Using Statistical Image Reconstruction for Spectral CT. *IEEE Trans. Med. Imaging*, 33(8):1614–1626, 2014.

[27] Ruoqiao Zhang, J.-B. Thibault, C A Bouman, K D Sauer, and Jiang Hsieh. Model-Based Iterative Reconstruction for Dual-Energy X-Ray CT Using a Joint Quadratic Likelihood Model. *Medical Imaging, IEEE Transactions on*, 33(1):117–134, 2014.

[28] Yunsong Zhao, Xing Zhao, and Peng Zhang. An extended algebraic reconstruction technique (E-ART) for dual spectral CT. *IEEE Trans. Med. Imaging*, 34(3):761–768, 2015.

[29] Rina Foygel Barber, Emil Y Sidky, Taly Gilat Schmidt, and Xiaochuan Pan. An algorithm for constrained one-step inversion of spectral CT data. *Phys. Med. Biol.*, 61(10):3784, 2016.

[30] J P Schlomka, E Roessl, R Dorscheid, S Dill, G Martens, T Istel, C Bãumer, C Herrmann, R Steadman, G Zeitler, A Livne, and R Proksa. Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography. *Phys. Med. Biol.*, 53(15):4031, 2008.

[31] Bo Zhao, Hao Gao, Huanjun Ding, and Sabee Molloi. Tight-frame based iterative image reconstruction for spectral breast CT. *Med. Phys.*, 40(3):031905, 2013.

[32] Darin P Clark and Cristian T Badea. Spectral diffusion: an algorithm for robust material decomposition of spectral CT data. *Phys. Med. Biol.*, 59(21):6445–6466, 2014.

[33] Mats Persson, Ben Huber, Staffan Karlsson, Xuejin Liu, Han Chen, Cheng Xu, Moa Yveborg, Hans Bornefalk, and Mats Danielsson. Energy-resolved CT imaging with a photon-counting silicon-strip detector. *Phys. Med. Biol.*, 59(22):6709, 2014.

[34] O Semerci, Ning Hao, M E Kilmer, and E L Miller. Tensor-Based Formulation and Nuclear Norm Regularization for Multienergy Computed Tomography. *IEEE Trans. Image Process.*, 23(4):1678–1693, 2014.

[35] Kyungsang Kim, Jong Chul Ye, William Worstell, Jinsong Ouyang, Yothin Rakvongthai, Georges El Fakhri, and Quanzheng Li. Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty. *IEEE Trans. Med. Imaging*, 34(3):748–760, 2015.

[36] Yan Xi, Yi Chen, Rongbiao Tang, Jianqi Sun, and Jun Zhao. United Iterative Reconstruction for Spectral Computed Tomography. *IEEE Trans. Med. Imaging*, 34(3):769–778, 2015.

[37] David S Rigie and Patrick J La Rivière. Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. *Phys. Med. Biol.*, 60(5):1741–1762, 2015.

[38] Mengheng Touch, Darin P Clark, William Barber, and Cristian T Badea. A neural network-based method for spectral distortion correction in photon counting x-ray CT. *Phys. Med. Biol.*, 61(16):6132, 2016.

[39] Jiulong Liu and Hao Gao. Material reconstruction for spectral computed tomography with detector response function. *Inverse Probl.*, 32(11):114001, 2016.

[40] Bernhard Brendel, Frank Bergner, Kevin Brown, and Thomas Koehle. Penalized likelihood decomposition for dual layer spectral CT. In *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, 2016.

[41] Nicolas Ducros, Juan Felipe Perez-Juste Abascal, Bruno Sixou, Simon Rit, and Françoise Peyrin.

Regularization of nonlinear decomposition of spectral x-ray projection images. *Med. Phys.*, 44(9):e174–e187, 2017.

[42] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.

[43] Simon Arridge and John Schotland. Optical tomography: forward and inverse problems. *Inverse Probl.*, 25(12):123010, 2009.

[44] S A Vavasis. *Nonlinear Optimization*. Princeton University Press, 1991.

[45] L M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math.*, 7(3):200–217, 1967.

[46] J F P J Abascal, M Abella, E Marinetto, J Pascau, and M Desco. A novel prior- and motion-based compressed sensing method for small-animal respiratory gated CT. *PLoS One*, 11(3), 2016.

[47] Markus Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Probl.*, 26(11):115014, 2010.

[48] Spectral X-ray image reconstruction toolbox, 2017 (accessed June 1, 2017). `https://www.creatis.insa-lyon.fr/~ducros/WebPage/spray.html/`.

[49] 3D Image Reconstruction for Comparison of Algorithm Database, 2010 (accessed May 15, 2017). `http://www.ircad.fr/research/3d-ircadb-02/`.

[50] R Kéchichian, S Valette, M Desvignes, and R Prost. Shortest-Path Constraints for 3D Multiobject Semiautomatic Segmentation Via Clustering and Graph Cut. *IEEE Trans. Image Process.*, 22(11):4224–4236, 2013.

[51] Dong Zeng, Jing Huang, Zhaoying Bian, Shanzhou Niu, Hua Zhang, Qianjin Feng, Zhengrong Liang, and Jianhua Ma. A Simple Low-Dose X-Ray CT Simulation From High-Dose Scan. *IEEE Transactions on Nuclear Science*, 62(5):2226–2233, 2015.

[52] Per Christian Hansen and Dianne Prost O'Leary. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, 1993.

[53] J F P J Abascal, S R Arridge, R H Bayford, and D S Holder. Comparison of methods for optimal choice of the regularization parameter for linear electrical impedance tomography of brain function. *Physiol. Meas.*, 29(11), 2008.