

# Nonlinear Compressed Sensing for Multi-Emitter X-Ray Imaging

Maria Klodt and Raphael Hauser<sup>1</sup>

Mathematical Institute, University of Oxford, Oxford OX2 6GG, U.K.  
`hauser@maths.ox.ac.uk`

**Abstract.** Compressed sensing is a powerful mathematical modelling tool to recover sparse signals from undersampled measurements in many applications, including medical imaging. A large body of work investigates the case with linear measurements, while compressed sensing with nonlinear measurements has been considered more recently. We continue this line of investigation by considering a novel type of nonlinearity with special structure that occurs in data acquired by multi-emitter X-ray tomosynthesis systems with spatio-temporal overlap. In [16] we proposed a nonlinear optimization model to deconvolve the overlapping measurements. In this paper we propose a model that exploits the structure of the nonlinearity and a nonlinear tomosynthesis algorithm that has a practical running time of solving only two linear subproblems at the equivalent resolution. We underpin and justify the algorithm by deriving RIP bounds for the linear subproblems and conclude with numerical experiments that validate the approach.

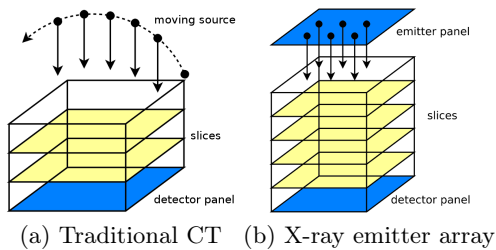
**Keywords:** nonlinear compressed sensing, restricted isometry property, sparse reconstruction, image reconstruction, medical imaging, X-ray, tomosynthesis, computed tomography

## 1 Introduction

Reconstructing a 3D image from a set of 2D projections is a typical inverse problem with applications in magnetic resonance imaging (MRI) [18,6] and X-ray computed tomography (CT) [7]. The goal of tomosynthesis is to recover the material densities in a three-dimensional domain, usually represented as a set of 2D slices. Traditionally, X-ray measurements are taken from a single moving source, yielding non-overlapping measurements (see Fig. 1 (a)), which can be formulated as linear constraints [7]. A novel type of measurements arises from recent research on X-ray emitter arrays with multiple simultaneously emitting sources [13] (see Fig. 1 (b)): Spatio-temporally overlapping X-rays, where rays from more than one emitter reach the same detector at the same time (see Fig. 2 (b)), result in nonlinear measurements [16]. When using linear image reconstruction methods, this type of nonlinear measurements has to be avoided. The scanner therefore has to be designed so that spatio-temporal overlap cannot occur, and this necessitates small collimation angles and/or stand-off distances,

thus imposing severe restrictions on the scanner geometry. Restrictions on the stand-off distance may be difficult to satisfy in hand-adjusted or hand-held devices for which the relative position of the emitter and detector panels cannot be controlled with high precision. Image reconstruction methodologies that can deal with spatio-temporal overlap as well as an appropriate understanding of how the reconstruction quality is affected as a function of the degree of overlap are therefore essential to improve the robustness and flexibility of the geometric design of these new X-ray tomosynthesis systems.

Since measurements are undersampled, we base our image reconstruction methodology on sparsity-regularized optimization models, or compressed sensing problems. Most of the literature studies linear compressed sensing problems, but more recently a growing body of work on nonlinear compressed sensing problems startet to appear: In 1-bit compressed sensing, the signal is constrained to lie on a hyper-sphere. [12,20] consider nonlinearity arising from noise and a generalization to quasi-linear compressed sensing, [17,22] consider compressed sensing with quadratic constraints. [4] considers more general models of nonlinearity and derives restricted isometry inequalities for nonlinear measurements as a function of a distance to a linearization. In our setting a different type of nonlinearity is produced by measurements from overlapping X-rays, where the nonlinearity arises from geometry. In [16] a method based on directly solving the nonlinear constraints in a least-squares sense is presented. In this paper we continue this investigation by exploiting the concavity of the nonlinear measurements, which makes it possible to achieve linearization via a corrective factor rather than using Taylor approximation, as proposed in other papers. We then propose an algorithm based on alternative optimization of the linear subproblems and updates of the corrective factors. The linearization renders the optimization stable and allows the use of efficient linear solvers on the linear subproblems. The practicability of the method is established in experiments both with simulated and real-world data.



**Fig. 1.** Tomosynthesis scanners: (a) traditional CT with moving source, (b) multi-emitter panel based system without moving parts.

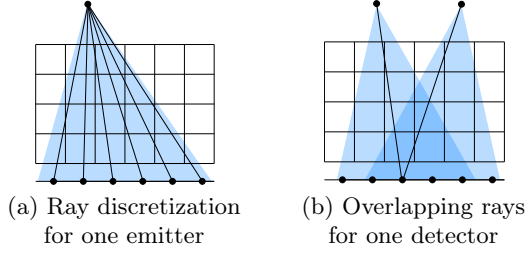
We consider X-rays emitted from conoidal sources arranged on a multi-emitter panel. A three-dimensional region of interest in which the object to reconstruct

is located is discretized to a Cartesian grid of  $n$  voxels. A schematic 2D view is given in Fig. 2 (a). A series of temporally separated measurements may be taken, actuating a different pattern of several emitters to emit photons during each acquisition interval. As long as spatio-temporal overlap does not occur, this process leads to the same measurements as if each emitter were actuated sequentially in isolation, but with the measurements taken in parallel. Each measurement  $j$  then corresponds to a pair of an emitter and a detector, the emitter emitting  $I_{E_j}$  photons in total in the direction of the detector, and the detector detecting the arrival of  $I_{D_j}$  photons emitted from the detector, the remaining photons having been absorbed by the material of the object that is to be imaged. We discretize the paths of photons between the emitter and detector, by collapsing all these paths to a single ray  $j$  subtended by the centers of the emitter and the detector involved in measurement  $j$ . Denoting the absorption coefficient of the material occupying voxel  $i$  by  $x_i \geq 0$ , and writing  $\xi_{ji}$  for the length of the intersection between ray  $j$  and voxel  $i$ , the *Beer-Lambert law* implies  $I_{D_j} = I_{E_j} \exp(\sum_{i=1}^n -\xi_{ji}x_i)$  for all  $j$ , where we assumed no ray scattering, no noise, and no spatio-temporal overlap between measurements. Taking logarithms the measurement equation can be reformulated as a linear system of equations  $\log I_{D_j} - \log I_{E_j} = \sum_{i=1}^n -\xi_{ji}x_i$ ,  $j = 1, \dots, m$ , which can be written as a linear system  $Ax = b$  of size  $m \times n$ , where  $a_{ji} = -\xi_{ji}$ ,  $b \in \mathbb{R}^m$  with  $b_j = \log(I_{D_j}/I_{E_j})$  and where the voxel densities were stacked into a vector  $x \in \mathbb{R}^n$  with  $n$  the number of voxels in the chosen voxelization. Note that  $A$  is very sparse, because each ray intersects only a small number of voxels.

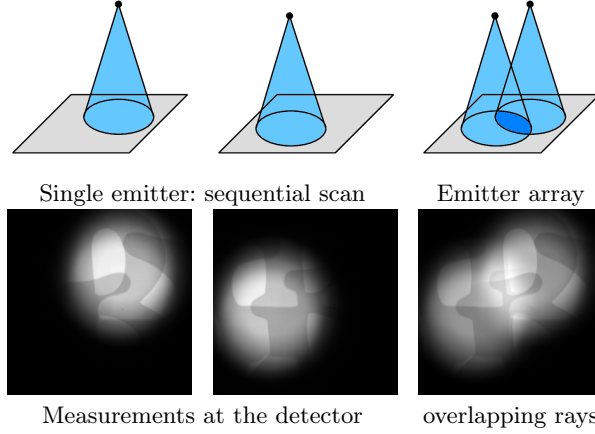
Since measurements are typically severely undersampled, prior information needs to be used to recover the original signal. Most objects to be imaged consist of material that has either density in only a small proportion of its spatial extent or has sharp interfaces between regions of high and low densities. It is known from the work of [5,9] and others that the compressed sensing problem  $\min_{x \geq 0} \{\|x\|_1 : Ax = b\}$  and related models is able to reconstruct the sparsest solution  $x$  to an underdetermined linear system  $Ax = b$  if matrix  $A$  satisfies the sufficient condition of a restricted isometry property (RIP), defined by the existence of  $\delta_s \in (0, 1)$  for each sparsity level  $s$  such that  $(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$  holds for all  $s$ -sparse vectors  $x$ . Note that the RI constants depend on the scaling of matrix  $A$ , while the reconstructability does not. In practice, one may thus optimize the RIC over a scaling factor.

## 2 Nonlinear Measurements with Overlap

If multiple X-ray sources emit rays simultaneously, it may happen that two or more emitter cones can partially overlap. Figure 3 shows a schematic overview as well as example measurements at the detector for a sequential scan, and the same measurements with overlap. In the case of simultaneous emission, rays can reach the same detector at the same time (see Fig. 2 (b)). When several rays  $k \in K(j)$  reach the same detector  $j$  simultaneously, the attenuation of measurement  $j$  equals  $c_j := I_{D_j} / \sum_{k \in K(j)} I_{E_k}$ , where the emitter intensity of the



**Fig. 2.** Scanner set-up with emitters above the reconstruction domain, and detectors below.



**Fig. 3.** Sequential versus parallel measurements, the latter inducing overlap of emitter cones.

$k$ -th source is denoted by  $I_{E_k}$  and  $I_{D_j}$  is the total radiation measured at detector  $j$ . The *Beer-Lambert law* once again describes the relation between measurements and densities, but the constraints arising from overlapping measurements are no longer log-linear,

$$\psi_j(x) := \sum_{k \in K(j)} \lambda_{jk} \exp \left( \sum_{i=1}^n -\xi_{ki} x_i \right) = c_j \quad (1)$$

where each of the summands corresponds to a different ray  $k \in K(j)$ . For simplicity of notation, we denote the fraction of intensity that emitter  $k$  contributes to measurement  $j$  by  $\lambda_{jk} := I_{E_k} / \sum_{l \in K(j)} I_{E_l}$ . The coefficients  $\xi_{ki}$  correspond to the length of intersection between voxel  $i$  and the  $k$ -th ray of measurement  $j$ . We represent them with vectors  $r_{jk} \in \mathbb{R}^n$ ,  $r_{jk} := (-\xi_{k1}, \dots, -\xi_{kn})^\top$ , for all  $j = 1, \dots, m$  and  $k \in K(j)$ . The projection of densities along one ray is then given

by  $r_{jk}^\top x = \sum_{i=1}^n -\xi_{ki}x_i$ . The vectors  $r_{jk}$  are sparse because each ray traverses only a small number of voxels. We remark that constraints similar to (1) are used in polychromatic reconstruction [14], where the nonlinearity arises from the polychromatic nature of the X-ray source. In our setting they arise from the geometry of rays, which yields different  $\xi_{ki}$  and  $x_i$  in each summand, in contrast to the setting of polychromatic reconstruction.

Measurements are typically severely undersampled, requiring the incorporation of prior information to recover the original signal. Assuming the signal  $x$  is sparse, minimization with  $\ell_1$  sparsity prior yields the reconstruction problem

$$\min_{x \geq 0} \|x\|_1 \quad \text{s.t.} \quad \psi_j(x) = c_j, \quad \forall j = 1, \dots, m. \quad (2)$$

As discussed above, the  $\ell_1$  prior is designed to produce sparse solutions to constraints  $\psi_j(x) = c_j$  ( $j = 1, \dots, m$ ). Although this is justified in some applications of the X-ray scanner, such as trauma, a higher image quality is generally achieved by using a regularization term that produces sharp interfaces between regions of high and low absorption density in the object that is to be imaged. Such a term may be formulated by using the  $\ell_1$ -norm based total variation (TV) norm, or by representing the absorption density by a D4-wavelet decomposition and imposing sparsity in the wavelet domain, thus replacing the voxelization by a wavelet discretization. We remark however that both cases can be treated in the same mathematical framework as the  $\ell_1$  prior discussed below, since in the case of the TV norm the term  $\|x\|_1$  is replaced by  $\|\Phi x\|_1$  where  $\Phi$  is a matrix of size  $3n \times n$  that encodes a finite-difference scheme to compute gradients, and in the case of the wavelet decomposition the coefficients  $\xi_{ji}$  of the design matrix  $A$  are replaced by the integral of wavelet  $i$  along ray  $j$ . For the sake of notational simplicity we shall henceforth only discuss the  $\ell_1$  regularization term, the other cases requiring minimal changes to algorithms and codes.

We will now present two different models and algorithmic approaches to solving the image reconstruction problem from overlapping rays.

## 2.1 A Forward-Backward Splitting Approach

The Lagrangian formulation of (2) allows for noise in the data term. In [16] we proposed to impose a quadratic penalty on the nonlinear constraints (1),

$$\min_{x \geq 0} \left\{ F_1(x) := \mu \|x\|_1 + \frac{1}{2} \sum_{j=1}^m (\psi_j(x) - c_j)^2 \right\}. \quad (3)$$

The parameter  $\mu > 0$  depends on the amount of noise one is willing to tolerate in the measurements. This model interprets the error in counting the rescaled number of photons arriving at detector  $j$  as Gaussian noise, which is justified by the fact that the associated photon count is Poissonian with parameter on the order  $O(10^3)$ .  $F_1(x)$  is a composite merit function with (3) of the form

$$\min_{x \geq 0} \{ \mu f(x) + g(x) \} \quad (4)$$

with convex non-differentiable  $f(x) := \|x\|_1$  and a partially convex, differentiable noise term  $g(x) := \frac{1}{2} \sum_{j=1}^m (\psi_j(x) - c_j)^2$ . In [16] we proved that  $g$  is convex in the region  $0 \leq x \leq \hat{x}$ , where  $\hat{x}$  is a minimizer, and that  $\nabla g$  is Lipschitz continuous with Lipschitz constant  $L = 6mp^2h^2$ , where  $h$  is the voxel size (the spatial coordinates being scaled by  $1/h$  in the algorithm below).

Problem (4) can be solved via forward-backward splitting updates

$$x^{t+1} = \text{prox}_{\mu f}(x^t - \sigma \nabla g(x^t)). \quad (5)$$

The proximal operator is defined as  $\text{prox}_{\mu f}(x) = \arg \min_{u \in \mathbb{R}^n} \{f(u) + \frac{1}{2\mu} \|u - x\|_2^2\}$ . [10] showed that in the case  $f(x) = \|x\|_1$  the prox operator is given by soft thresholding,

$$\text{prox}_{\mu f}(x) = \begin{cases} x - \mu, & x > \mu \\ 0, & |x| \leq \mu \\ x + \mu, & x < -\mu \end{cases} \quad (6)$$

[8] showed that if  $f$  is lower semicontinuous convex and  $g$  is convex differentiable with Lipschitz continuous gradient, then (5) converges to a minimum of (4) at a rate  $\mathcal{O}(1/t)$ . [21] showed that an additional overrelaxation step yields slightly faster convergence.

For an optimization of (3) with only partially convex  $g$ , the initialization of  $x$  has to be smaller than a minimizer  $\hat{x}$ , for  $g$  is only guaranteed to be convex in the region  $0 \leq x \leq \hat{x}$ . Knowing that the densities  $x$  cannot be negative, we choose the lower bound  $x^0 = 0$  and constrain step sizes  $\sigma$  so as to ensure that  $x \leq \hat{x}$  for all  $t$ . We enforce this by use of a backtracking line-search algorithm [1]. Although in general,  $\hat{x}$  is unknown, the measurements  $c_j = \psi_j(\hat{x}) + \epsilon_j$  are known, hence we constrain the line-search by the necessary (though not sufficient) condition  $\psi_j(x) \geq c_j$  for all  $j$ . This leads to the following algorithm for the minimization of (3), as further discussed in [16].

**Input:**

- $c \in \mathbb{R}^m$  : vector of measurements
- $r_{jk} \in \mathbb{R}^n$  : sparse vectors of intersection lengths of rays and voxels
- $\theta \in (0, 1)$  : line search control parameter
- $L = 6mp^2h^2$  : Lipschitz constant for  $\nabla g$

**Output:**

- $x \in \mathbb{R}^n$  : approximate solution to (3).

**Initialize**  $x^0 = 0$ .

**Iterate** for  $t = 0, 1, 2, \dots$ :

1. *Compute search direction:*

$$\nabla g(x^t) = \sum_{j=1}^m \left( \sum_{k \in K(j)} \lambda_{jk} e^{r_{jk}^\top x^t} - c_j \right) \left( \sum_{k \in K(j)} \lambda_{jk} r_{jk} e^{r_{jk}^\top x^t} \right)$$

2. *Backtracking line search:*

$$\begin{aligned}
\sigma &= 1/L \\
x^{new} &= \text{prox}_{\mu f}(x^t - \sigma \nabla g(x^t)) \\
\text{while } \exists j \in \{1, \dots, m\} \text{ with } \psi_j(x^{new}) < b_j : \\
\quad \sigma &\leftarrow \theta \times \sigma \\
\quad x^{new} &= \text{prox}_{\mu f}(x^t - \sigma \nabla g(x^t))
\end{aligned}$$

3. *Update  $x$ :*  $x^{t+1} = x^{new}$ .

## 2.2 The Lagging Multiplier Approach

Our second approach is based on applying the logarithm to the constraints from overlapping rays (1), in analogy to what we did in the case without overlap. If there is little overlap, it is reasonable to expect that taking the logarithm transforms the problem into an optimization problem that is only slightly nonlinear. For non-zero measurements, the  $j$ -th constraint equation (1) thus transforms to

$$\tilde{b}_j := \log(c_j) = \log(\psi_j(x)) = \log \left( \sum_{k \in K(j)} \lambda_{jk} \exp \left( \sum_{i=1}^n -\xi_{ki} x_i \right) \right).$$

A regularized optimization problem with noisy measurements is now given by

$$\min_{x \geq 0} \left\{ F_2(x) := \mu \|x\|_1 + \frac{1}{2} \sum_{j=1}^m (\log(\psi_j(x)) - \tilde{c}_j)^2 \right\}. \quad (7)$$

The data term of this model is nonlinear, but the concavity of the logarithm implies that for any convex weights  $\lambda_{jk} \in [0, 1]$  with  $\sum_{k \in K(j)} \lambda_{jk} = 1$  the following inequality holds,

$$\sum_{k \in K(j)} \sum_{i=1}^n -\lambda_{jk} \xi_{ki} x_i \leq \log \left( \sum_{k \in K(j)} \lambda_{jk} \exp \left( \sum_{i=1}^n -\xi_{ki} x_i \right) \right) = \log(\psi_j(x)) \leq 0.$$

Therefore, for each measurement  $j$  there exists a corrective multiplier

$$\tau_j(x) := \frac{\log \left( \sum_{k \in K(j)} \lambda_{jk} \exp \left( \sum_{i=1}^n -\xi_{ki} x_i \right) \right)}{\sum_{k \in K(j)} \sum_{i=1}^n -\lambda_{jk} \xi_{ki} x_i} \in [0, 1] \quad (8)$$

such that  $\tau_j \times \sum_{k \in K(j)} \sum_{i=1}^n -\lambda_{jk} \xi_{ki} x_i = \log(\psi_j(x))$ , for  $(j = 1, \dots, m)$ . The data term in (7) can thus be replaced by  $\|\tau(x) \tilde{A}x - \tilde{b}\|^2$ , where  $\tau = \text{diag}(\tau_j)$  is

the diagonal matrix with factors  $\tau_j$  on the diagonal, the  $j$ -th row of matrix  $\tilde{A}$  is obtained as  $\tilde{a}_j = \sum_{k \in K(j)} \lambda_{jk} r_{jk}^\top$ . Problem (7) can therefore be reformulated as

$$\min_{x \geq 0} \left\{ F_2(x) = \mu \|x\|_1 + \frac{1}{2} \|\tau(x) \tilde{A}x - \tilde{b}\|_2^2 \right\}. \quad (9)$$

Note that in contrast to (3), the data term of (9) is to be interpreted as error in the absorption due to the discretization imposed by the voxelization and to scattering. It is natural to model such noise as Gaussian.

For measurements that are only slightly nonlinear,  $\tau_j(x)$  is close to 1. Instead of iteratively solving the problem (9) with its nonlinear data term, we may therefore approximate  $\tau(x^{t+1})$  by  $\tau(x^t)$ , where  $x^t$  is the solution of the previous iteration. In other words,  $\tau$  is updated with a time lag of 1 relative to the updates  $x^{t+1}$  of the densities. This leads to the following algorithm:

**Input:**

$\tilde{b} \in \mathbb{R}^m$  : vector of measurements

$\tilde{A} \in \mathbb{R}^n$  : sparse matrix of scaled intersection lengths of rays and voxels

**Output:**

$x \in \mathbb{R}^n$  : approximate solution to (9).

**Initialize**

$x^0 = \tilde{A}^\top \tilde{b}$  (used in warm start of (10))

$\tau^0 = \mathbf{I}_m$  (the  $m \times m$  identity matrix)

**Iterate** for  $t = 0, 1, 2, \dots$ :

1. *Update*  $x$ :

$$x^{t+1} = \arg \min_{x \geq 0} \mu \|x\|_1 + \frac{1}{2} \|\tau^t \tilde{A}x - \tilde{b}\|_2^2 \quad (10)$$

2. *Update*  $\tau$ :

$$\tau_j^{t+1} \leftarrow \log(\psi_j(x^t)) / \tilde{a}_j x^t, \quad \forall j \in \{1, \dots, m\} \quad (11)$$

The advantage of this method is that the update of  $x$  obtained in (10) corresponds to solving the Lasso problem and can be solved with any standard linear solver, e.g. [1]. The update of  $\tau_j$  in (11) is solved explicitly in each iteration at a constant cost. The initialization  $\tau_j = 1$  represents the case where all rays contribute equally to measurement  $j$ . In practice we observe that two outer iterations suffice, as the factors  $\tau_j$  generally converge to near machine precision after a single update. Thus, the above described algorithm has only twice the cost of an X-ray problem without spatio-temporal overlap.

### 3 Restricted Isometry Bounds for Overlapping Rays

In this section we derive an analytical argument for why (9) can be expected to yield an exact sparse reconstruction as long as the average overlap per detector



does not exceed a factor of 2. This will be backed up by experiments. It is conceptually important to investigate this phenomenon, because it justifies the use of the lagging multiplier method.

While related work on compressed sensing mainly investigates nonlinearity due to noise, overlapping rays yield a different type of nonlinearity. However, we note that the following derivation of restricted isometry properties for our image reconstruction model (9) was inspired by a case studied in [4], who uses first order Taylor approximation to approximate nonlinear measurements. Our case is similar, but the linearization is based on the corrective factors  $\tau(x)$  rather than Taylor approximation.

To keep the argument simple, we assume that  $|K(j)| = p$  for all detectors  $p$ , and that all emission intensities are the same, so that  $\lambda_{jk} = 1/p$  for all  $j = 1, \dots, m$  and  $k \in K(j)$ .  $\tilde{A} \in \mathbb{R}^{m \times n}$  is then obtained from  $A \in \mathbb{R}^{mp \times n}$ , the design matrix of the corresponding temporally separated measurements, by averaging the rows with indices  $k \in K(j)$ , as writing the  $p$  indices  $k \in K(j)$  as  $j1, \dots, jp$  yields

$$A = \begin{pmatrix} -\xi_{111} & \dots & -\xi_{n11} \\ \vdots & & \vdots \\ -\xi_{11p} & \dots & -\xi_{n1p} \\ -\xi_{121} & \dots & -\xi_{n21} \\ \vdots & & \vdots \\ -\xi_{1mp} & \dots & -\xi_{nmp} \end{pmatrix}, \quad \tilde{A} = \frac{1}{p} \begin{pmatrix} \sum_{k=1}^p -\xi_{11k} & \dots & \sum_{k=1}^p -\xi_{n1k} \\ \sum_{k=1}^p -\xi_{12k} & \dots & \sum_{k=1}^p -\xi_{n2k} \\ \vdots & & \vdots \\ \sum_{k=1}^p -\xi_{1mk} & \dots & \sum_{k=1}^p -\xi_{nmk} \end{pmatrix}, \quad (12)$$

In the following we will derive inequalities that allow us to extend RIP bounds from  $A$  to  $\tau\tilde{A}$ .

**Lemma 1.** *For all  $x \in \mathbb{R}^n$  with  $x_i \geq 0$ , the norm of  $p\tilde{A}x$  is bounded by*

$$\|Ax\|_2^2 \leq \|p\tilde{A}x\|_2^2 \leq p\|Ax\|_2^2. \quad (13)$$

*Proof.* Let  $\alpha_{jk} \leq 0$  be defined as the negative distance weighted sum of densities along a ray between detector  $j \in \{1, \dots, m\}$  and emitter  $k \in \{1, \dots, p\}$ :

$$\alpha_{jk} := r_{jk}^\top x = \sum_{i=1}^n -\xi_{ijk} x_i. \quad (14)$$

Then the *Multinomial Theorem* yields the lower bound

$$\begin{aligned} \|p\tilde{A}x\|_2^2 &= \sum_{j=1}^m \left( \sum_{k=1}^p \alpha_{jk} \right)^2 = \sum_{j=1}^m \sum_{k=1}^p \alpha_{jk}^2 + \sum_{j=1}^m \left( \sum_{k_1=1}^p \sum_{k_2=k_1+1}^p 2\alpha_{jk_1} \alpha_{jk_2} \right) \\ &= \|Ax\|_2^2 + \sum_{j=1}^m \left( \sum_{k_1=1}^p \sum_{k_2=k_1+1}^p 2\alpha_{jk_1} \alpha_{jk_2} \right) \end{aligned} \quad (15)$$

$$\geq \|Ax\|_2^2 \quad (16)$$

(16) holds because the second term in (15) contains only positive summands. The upper bound can be computed starting again from (15):

$$\begin{aligned}
\|p\tilde{A}x\|_2^2 &= \|Ax\|_2^2 + \sum_{j=1}^m \left( \sum_{k_1=1}^p \sum_{k_2=k_1+1}^p 2\alpha_{jk_1}\alpha_{jk_2} \right) \\
&\leq \|Ax\|_2^2 + \sum_{j=1}^m \left( \sum_{k_1=1}^p \sum_{k_2=k_1+1}^p (\alpha_{jk_1}^2 + \alpha_{jk_2}^2) \right) \\
&= \|Ax\|_2^2 + \sum_{j=1}^m (p-1) \sum_{k=1}^p \alpha_{jk}^2 \leq \|Ax\|_2^2 + (p-1) \sum_{j=1}^m \sum_{k=1}^p \alpha_{jk}^2 = p\|Ax\|_2^2
\end{aligned} \tag{17}$$

where (17) follows from the Binomial equation  $(a-b)^2 \geq 0$  which implies  $2ab \leq a^2 + b^2$  for any  $a, b \in \mathbb{R}$ .

**Lemma 2.** *Assuming that for each measurement  $j$  there exists at least one ray  $k \in K(j)$  that traverses the reconstruction domain with minimal density  $x_{\min} > 0$ , then  $\tau_{\min} := \min_{j=1,\dots,m} \tau_j > 0$  and  $\tau_{\max} := \max_{j=1,\dots,m} \tau_j \leq 1$ .*

*Proof.* The assumptions imply that

$$\frac{1}{p} \sum_{k=1}^p \lambda_{jk} \exp \left( \sum_{i=1}^n -\xi_{ijk} x_i \right) < 1,$$

from which the result follows by strict concavity of the logarithm and the definition of  $\tau_j$  in (8).

**Lemma 3.** *For all  $x \in \mathbb{R}^n$  the following bounds apply,*

$$\tau_{\min}^2 \|p\tilde{A}x\|_2^2 \leq \|p\tau\tilde{A}x\|_2^2 \leq \|p\tilde{A}x\|_2^2. \tag{18}$$

*Proof.* Using the notation  $\alpha_{jk}$  introduced in (14), we have

$$\begin{aligned}
\|p\tau\tilde{A}x\|_2^2 &= \sum_{j=1}^m \tau_j^2 \left( \sum_{k=1}^p \alpha_{jk} \right)^2 \leq \tau_{\max}^2 \sum_{j=1}^m \left( \sum_{k=1}^p \alpha_{jk} \right)^2 = \tau_{\max}^2 \|p\tilde{A}x\|_2^2 \leq \|p\tilde{A}x\|_2^2 \\
\|p\tau\tilde{A}x\|_2^2 &= \sum_{j=1}^m \tau_j^2 \left( \sum_{k=1}^p \alpha_{jk} \right)^2 \geq \tau_{\min}^2 \sum_{j=1}^m \left( \sum_{k=1}^p \alpha_{jk} \right)^2 = \tau_{\min}^2 \|p\tilde{A}x\|_2^2.
\end{aligned}$$

Let us now derive restricted isometry properties for the nonlinear measurements with overlap  $\tau\tilde{A}$ , assuming that the design matrix  $A$  of spatio-temporally separated measurements satisfies a restricted isometry inequality

$$(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2 \tag{19}$$

for all  $s$ -sparse vectors  $x$ , an assumption that is justified by the work of [15].

**Theorem 1 (Restricted isometry for  $\tau\tilde{A}$ ).** *Let  $A$  and  $\tilde{A}$  be as defined in (12) and such that  $A$  satisfies the RIP (19). Then for all  $s$ -sparse non-negative vectors  $x \in \mathbb{R}_+^n$ , the rescaled matrix  $p^{1/2}\tau\tilde{A}$  satisfies the RIP*

$$(1 - \alpha_s)\|x\|_2^2 \leq \|p^{1/2}\tau\tilde{A}x\|_2^2 \leq (1 + \alpha_s)\|x\|_2^2$$

with  $\alpha_s = 1 - \frac{\tau_{\min}^2}{p}(1 - \delta_s)$ .

*Proof.* The lower bound follows from Lem. 1 and 3 as follows,

$$\begin{aligned} (1 - \alpha_s)\|x\|_2^2 &= \frac{\tau_{\min}^2}{p}(1 - \delta_s)\|x\|_2^2 \stackrel{(19)}{\leq} \tau_{\min}^2 \|p^{-1/2}Ax\|_2^2 \stackrel{\text{Lem. 1}}{\leq} \tau_{\min}^2 \|p^{1/2}\tilde{A}x\|_2^2 \\ &\stackrel{\text{Lem. 3}}{\leq} \|p^{1/2}\tau\tilde{A}x\|_2^2. \end{aligned}$$

Furthermore, using  $\tau_{\min}^2 \leq 1 \leq p$ , we have  $\delta_s \leq \alpha_s$ , so that the upper bound is found as follows,

$$\|p^{1/2}\tau\tilde{A}x\|_2^2 \stackrel{\text{Lem. 3}}{\leq} \|p^{1/2}\tilde{A}x\|_2^2 \stackrel{\text{Lem. 1}}{\leq} \|Ax\|_2^2 \stackrel{(19)}{\leq} (1 + \delta_s)\|x\|_2^2 \leq (1 + \alpha_s)\|x\|_2^2,$$

## 4 Numerical Experiments

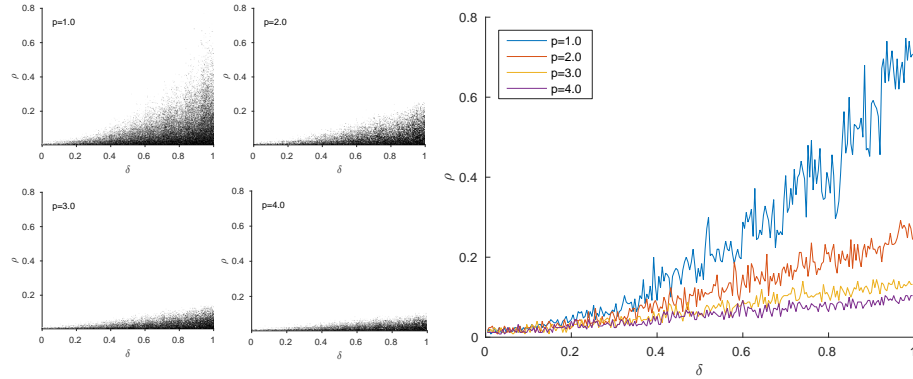
It follows from the results of [5,9] that sparse recovery occurs when  $\alpha_{2s} < \sqrt{2} - 1$ . Under the realistic assumptions that  $\tau_{\min} \approx 1$  and  $\delta_{2s} \approx 0$ , Thm. 1 then implies sparse recovery in the linear subproblems of the lagging multiplier algorithm of Sect. 2.2 generally occurs when the average overlap  $p$  per detector satisfies

$$p < \frac{\tau_{\min}^2(1 - \delta_{2s})}{2 - \sqrt{2}} \approx 1.7.$$

Our experiments indicate that this estimate is pessimistically conservative, as a little higher degree of overlap can still recover the original signal without significant deterioration of image quality.

### 4.1 Empirical Phase Transitions

Phase transitions have been introduced in [11] as a tool to analyse the recoverability of a  $k$ -sparse signal  $x \in \mathbb{R}^n$  from measurements obtained consistent with a matrix  $A \in \mathbb{R}^{m \times n}$ . They present reconstruction errors from simulated signals depending on the sampling rate  $\delta := m/n$  and the relative signal sparsity  $\rho := k/m$ . Several empirical results have shown that a relatively sharp transition can be observed between combinations of  $\delta$  and  $\rho$  where the signal can be recovered and where it cannot be recovered (with high probability) [3,19]. Phase transition analysis for random Gaussian matrices has been widely investigated, e.g. [2], while more recent work has shown empirical phase transition phenomena for X-ray measurements [15].



**Fig. 4.** Phase transitions for measurements with overlap of  $p=1,2,3$  and 4, depending on the sampling rate  $\delta$  and signal sparsity  $\rho$ .

We tested the recoverability of random signals  $x$  in a cuboidal domain which is traversed by partly overlapping X-rays. Figure 4 shows phase transitions for minimization of (2) with different degrees of overlap, defined as the number of rays simultaneously illuminating one detector. The same degree of overlap was used at all detectors. An overlap of 1 is thus equivalent to the corresponding reconstruction from sequential measurements. The four plots in the left part of the figure show phase transitions for overlap of  $p=1,2,3$  and 4, respectively. The grey values for each combination of  $\delta$  and  $\rho$  are obtained from the following experiment: We fixed  $m=100$  and computed a reconstruction of a random signal with sparsity  $k = \rho m$  and resolution of  $n = m/\delta$  unknowns. Measurements were simulated by integrating the signal along  $m$  random rays through the reconstruction domain, and the simulated photon counts of overlapping rays were summed. The number of measurements  $m/p$  of each experiment was thus determined by the number of rays  $m$  and the degree  $p$  of overlap per detector. The grey value at each data point encodes the reconstruction error as the distance of the reconstruction to the original signal (while dark encodes low and bright encodes high errors). The right part of the figure shows the observed transition between recoverability and non-recoverability for given  $\delta$  and  $\rho$ , obtained from thresholding the four individual plots from the left part of the figure. The phase transitions show that increasing sparsity and sampling rates yield increasing recoverability, while increasing overlap decreases the recoverability.

Our phase transition experiments show that the sparse recoverability analysis of Sect. 3 is rather conservative: If  $A$  were to be a random matrix with i.i.d. Gaussian coefficients, one of the cases for which the relation between the RIP and the sparse recoverability is well understood via quantifiable probability bounds, then  $\tilde{A}$  would also be such a matrix and  $\tau\tilde{A}$  nearly so (if all  $\tau_j$  are close to 1), but with a sampling rate reduced by a factor  $1/p$ . If the sparse recoverability of system  $\tau\tilde{A}x = \tilde{b}$  were thus characterized by the RIP, then thresholds on the red

curve in Fig. 4 would correspond to threshold points on the blue curve at a point where the value of  $\delta$  is halved and  $\rho$  is doubled. But this is not the case, and the phase transition experiment shows that the system with overlap  $p = 2$  is able to recover signals a factor 2 less sparse than predicted under the RIP argument.

## 4.2 Comparison of Reconstruction Errors

In order to investigate how measurements with overlapping X-rays affect image reconstruction accuracies, we tested both methods discussed in this paper on simulated and real-world measurements. Simulated data allow to compute exact reconstruction errors as a distance to the ground truth, while real-world data give empirical confirmation that the methodology can be used under practical conditions that include measurement noise and scattering.

For this purpose we tested two different data sets:

1. Data set “cube” are simulated measurements from a binary test object of a cube located at the center of the reconstruction domain,
2. Data set “letters” consist of real X-ray measurements of two wooden letters stacked on top of one another. This data was kindly made available to us by Adaptix Ltd., who took the measurements in their lab in Begbroke, Oxfordshire, U.K..

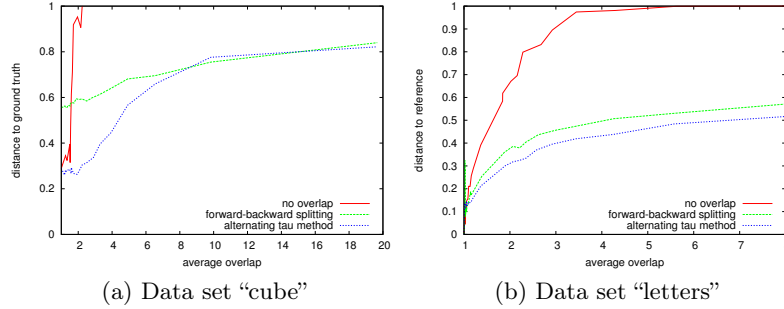
In both data sets the scanner geometry was based on a panel of emitters and a parallel panel of detectors placed above and below the reconstruction domain, with parameters given in Table 4.2. The emitters and detectors were equally spaced in the respective panels, and all emitters had the same collimation angle (the apex angle of the emitter cone). The resolution of the reconstruction domain in z-dimension was lower than the resolution in x- and y-dimension, because its direction corresponds to the main orientation of rays, resulting in reduced recoverability along the z-axis.

data set	# emitters	#detectors	# voxels	collimation angle
“cube”	5×5	15×15	20×20×20	20°
“letters”	13×14	128×128	128×128×20	~12°

**Table 1.** Parameters of scanner geometry and grid dimension for the two data sets.

Our experiments compared the following three different methods to handle overlapping X-ray measurements:

1. Discard overlapping measurements from the set of constraints and reconstruct from non-overlapping measurements only.
2. The forward-backward splitting method of Sect. 2.1.
3. The lagging multiplier method of Sect. 2.2.



**Fig. 5.** Reconstruction error for increasing degrees of overlap, comparing reconstruction from non-overlapping measurements only (red), forward-backward splitting (green) and the alternated linearization method (blue), for the letters data set. (a): Distance to ground truth for simulated data. (b): Distance to sequential reconstruction using real X-ray measurements.

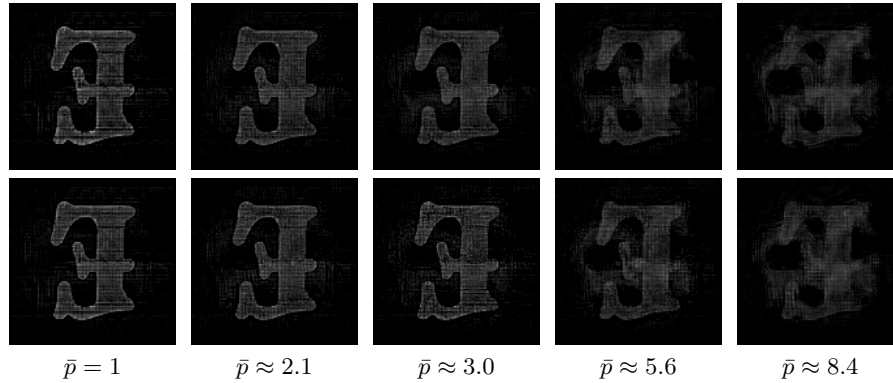
Figure 4.2 shows reconstruction errors for increasing average overlap  $\bar{p}$ , defined as the *average* number of rays that illuminate a detector simultaneously, as the number of overlapping rays was allowed to vary from detector to detector. Overlap was simulated by virtually firing an increasing number of randomly chosen emitters at the same time, and adding the respective photon counts at detector pixels with overlap. All other parameters of the experimental set-up were left unchanged. The relative distance  $d(x)$  of a reconstruction  $x$  to the reference  $\hat{x}$  was quantified as  $d(x) = \|x - \hat{x}\|_2 / \|\hat{x}\|_2$ . For the simulated data set “cube” the reference  $\hat{x}$  is given by the ground truth. Since a ground truth is not known for real data, we compare the reconstructions of the “letters” data set to the respective reconstruction from sequential measurements, that is, sequential measurements without overlap were taken using the same rays as in the case with overlap. The real world data set was acquired using nonoverlapping sequential exposures, and overlap was imputed by adding measurements of several exposures simulated as occurring simultaneously.

The plots show that for both sets, the forward-backward splitting and the alternated optimization with the corrective factors  $\tau$  yield significantly improved reconstruction qualities compared to discarding the overlapping measurements. The alternated  $\tau$ -method of Sect. 2.2 yields slightly better reconstruction accuracies than the forward-backward splitting.

This observation can also be confirmed visually in Fig. 6, which shows the 5th of 20 z-slices of the reconstructed densities of the data set “letters” for increasing degrees of average overlap  $\bar{p}$ .

## 5 Conclusions

This paper has shown practical methods to reconstruct images from measurements with overlapping X-rays and renders multi-emitter tomosynthesis robustly



**Fig. 6.** Reconstructed densities (slices) of wooden letters, comparing forward-backward-splitting (first row) and the alternated optimization with  $\tau$  (second row).  $\bar{p}$  is the average overlap per measurement.

applicable in applications where the multi-emitter panel of a scanning system is hand held or hand-adjusted, and where an exact stand-off distance avoiding overlap cannot always be guaranteed. Both theory and practical experiments show that a moderate degree of average overlap of up to ca 2 rays per detector has no major deteriorating effect on the quality of reconstructed images.

## Acknowledgments

We thank Adaptix Ltd for providing the X-ray measurements used in the experiments with real-world data. This work was supported by Adaptix Ltd and EPSRC EP/K503769/1, as well as by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

## References

1. A. BECK AND M. TEBoulLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Img. Sci., 2 (2009), pp. 183–202.
2. J. BLANCHARD, C. CARTIS, J. TANNER, AND A. THOMPSON, *Phase transitions for greedy sparse approximation algorithms*, Applied and Computational Harmonic Analysis, 30 (2011), pp. 188–203.
3. J. D. BLANCHARD, C. CARTIS, J. TANNER, AND A. THOMPSON, *Phase transitions for greedy sparse approximation algorithms*, Applied and Computational Harmonic Analysis, 30 (2011), pp. 188–203.
4. T. BLUMENSATH, *Compressed sensing with nonlinear observations and related nonlinear optimisation problems*, IEEE Transactions on Information Theory, 59 (2013), pp. 3466–3474.
5. E. J. CANDÉS, J. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

6. C. CHEN AND J. HUANG, *Compressive sensing MRI with wavelet tree sparsity*, in Advances in neural information processing systems, 2012, pp. 1115–1123.
7. K. CHOI, J. WANG, L. ZHU, T. SUH, S. BOYD, AND L. XING, *Compressed sensing based cone-beam computed tomography reconstruction with a first-order method*, Medical Physics, 37 (2010), pp. 5113–5125.
8. P. L. COMBETTES AND V. R. WAJS, *Signal Recovery by Proximal Forward-Backward Splitting*, Multiscale Modeling & Simulation, 4 (2005), pp. 1168–1200.
9. D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
10. D. L. DONOHO AND I. M. JOHNSTONE, *Minimax estimation via wavelet shrinkage*, Annals of Statistics, 26 (1998), pp. 879–921.
11. D. L. DONOHO AND J. TANNER, *Precise undersampling theorems*, Proceedings of the IEEE, 98 (2010).
12. M. EHLER, M. FORNASIER, AND J. SIGL, *Quasi-linear compressed sensing*, Multiscale Modeling & Simulation, 12 (2014), pp. 725–754.
13. B. GONZALES, D. SPRONK, Y. CHENG, A. W. TUCKER, M. BECKMAN, O. ZHOU, AND J. LU, *Rectangular fixed-gantry CT prototype: Combining CNT X-ray sources and accelerated compressed sensing-based reconstruction*, IEEE Access, 2 (2014), pp. 971–981.
14. R. GU AND A. DOGANDŽIĆ, *Polychromatic X-ray CT image reconstruction and mass-attenuation spectrum estimation*, arXiv preprint arXiv:1509.02193, (2015).
15. J. S. JØRGENSEN AND E. Y. SIDKY, *How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography*, Philosophical Transactions of Royal Society of London. Series A. Mathematical, Physical, and Engineering Sciences, 373 (2015), p. 20140387.
16. M. KLODT AND R. HAUSER, *3d image reconstruction from x-ray measurements with overlap*, in European Conference on Computer Vision (ECCV), 2016.
17. X. LI AND V. VORONINSKI, *Sparse signal recovery from quadratic measurements via convex programming*, CoRR, abs/1209.4785 (2012).
18. M. LUSTIG, D. L. DONOHO, J. M. SANTOS, AND J. M. PAULY, *Compressed sensing MRI*, in IEEE Signal Processing Magazine, 2007.
19. H. MONAJEMI, S. JAFARPOUR, M. GAVISH, AND D. DONOHO, *Deterministic matrices matching the compressed sensing phase transitions of gaussian random matrices*, Proceedings of the National Academy of Sciences, 110 (2013), pp. 1181–1186.
20. D. NEEDELL AND J. A. TROPP, *Cosamp: Iterative signal recovery from incomplete and inaccurate samples*, Commun. ACM, 53 (2010), pp. 93–100.
21. Y. NESTEROV, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.
22. H. OHLSSON, A. YANG, R. DONG, AND S. SASTRY, *CPRL—an extension of compressive sensing to the phase retrieval problem*, in Advances in Neural Information Processing Systems, 2012, pp. 1367–1375.