Denoising Message Passing for X-ray Computed Tomography Reconstruction

Alessandro Perelli, Michael Lexa, Ali Can and Mike Davies, Fellow, IEEE

Abstract—X-ray Computed Tomography (CT) reconstruction from a sparse number of views is a useful way to reduce either the radiation dose or the acquisition time, for example in fixed-gantry CT systems, however this results in an ill-posed inverse problem whose solution is typically computationally demanding. Approximate Message Passing (AMP) techniques represent the state of the art for solving undersampling Compressed Sensing problems with random linear measurements but there are still not clear solutions on how AMP should be modified and how it performs with real world problems. This paper investigates the question of whether we can employ an AMP framework for real sparse view CT imaging? The proposed algorithm for approximate inference in tomographic reconstruction incorporates a number of advances from within the AMP community, resulting in the Denoising Generalised Approximate Message Passing algorithm (DCT-GAMP). Specifically, this exploits the use of state of the art image denoisers to regularise the reconstruction. While in order to reduce the probability of divergence the (Radon) system and Poisson non-linear noise model are treated separately, exploiting the existence of efficient preconditioners for the former and the generalised noise modelling in GAMP for the latter. Experiments with simulated and real CT baggage scans confirm that the performance of the proposed algorithms are comparable with, and can even outperform traditional statistical CT optimisation solvers.

Keywords—X-ray Computed Tomography, Compressed Sensing, Approximate Message Passing, Image denoising, Preconditioning, Iterative algorithms

I. INTRODUCTION

X-ray Computed Tomography is one of the most widely used imaging techniques for medical diagnosis, image-guided radiotherapy, material characterization and security applications. Reducing X-ray radiation exposure is an important concern in particular for diagnostic CT where patients are subjected to repeated scans. Furthermore, CT scanners employing Dual Energy (DE) systems tend to either reduce the acquisition data per energy or increase the dose and acquisition time. To lower the X-ray dose, two different strategies can be implemented: reducing the X-ray flux toward each detector element, i.e. the milliampere per seconds (low-mAs) per projection, or decrease the number of projections (sparse-views) per rotation. Similarly fixed gantry systems, e.g. [1], designed to accelerate scan time tend to further restrict the set of projections that can be acquired.

CT image reconstruction from sparse views and low dose, achieved by conventional filtered back projection (FBP) algorithms, is generally affected by noticeable streaking artifacts, due to insufficient sampling, and is not of acceptable quality for diagnostic purposes [2]. There is therefore a need in CT imaging applications for high quality image reconstruction algorithms that can accommodate sparse views and low dose. Many approaches have been proposed to solve this problem [3]. In particular, state of the art statistical image reconstruction typically aims to minimize a cost function defined as a sum of a data fidelity term that takes into account the measurement’s statistical model and the geometry of the acquisition system, and a regularization term that imposes a prior model on the solution. Generally, the cost function for X-ray CT is either the negative log-likelihood function [4] or a penalized weighted least-squares (PWLS) cost function with a weighted quadratic approximation of the Poisson measurement noise model [5], [6]. Although several types of iterative algorithms have been designed to solve the statistical X-ray CT problem which can provide images with enhanced resolution and reduced artifacts compared to the FBP [7], in general current methods require many iterations to converge yielding a high computation time, and are often not suitable for clinical/industrial CT uses [8].

A large number of iterative algorithms have been utilized for statistical CT reconstruction, among these are coordinate descent [9], preconditioned conjugate gradient [10] and ordered subsets [11]. Recently researchers have developed new algorithms with faster convergence by using splitting techniques [12], alternating direction method of multipliers based algorithm [13] or combining Nesterov momentum techniques with ordered subsets to accelerate gradient descent methods [14]. In general, any first-order iterative method requires at each iteration the computation of at least one forward and back projection operator, together with a proximal mapping to account for the regularization term. These represent the main contributions to the overall computational time. In order to accelerate the reconstruction, it is therefore necessary to either design faster CT operators or develop iterative algorithms that can converge in fewer iterations.

In this work, we investigate the use of an emerging reconstruction method from Compressed Sensing (CS), called Approximate Message Passing (AMP) [15], for sparse view CT reconstruction. AMP based inference refers to a family of iterative algorithms first proposed in [15] for Compressed Sensing problems with an i.i.d. random Gaussian system matrix and a sparse signal model. AMP is a form of approximate Bayesian inference based on the notion of message passing or loopy belief propagation and is also strongly connected to the family of Expectation Propagation and Expectation Consistent

...
approximation algorithms [16]. In essence, message passing algorithms work by iteratively updating marginal probabilities on the unknown variables until a locally consistent posterior probability model is obtained. The compelling aspect of the AMP family of algorithms is that they are designed to work in the large system limit (for random systems) which enables the central limit theorem to be invoked. This in turn simplifies the messages to be Gaussian distributions, requiring the algorithm to only pass means and variances. The result is a very efficient algorithm that is remarkably similar to the more traditional iterative shrinkage algorithm but with an additional "Onsager correction term" [15]. It also has many similarities to the Alternating Direction Method of Multipliers (ADMM) algorithm [17].

Today, AMP based algorithms provide the state-of-the-art performance in CS reconstruction both in terms of computation and reconstruction performance, e.g. [18]–[20]. For Gaussian measurements the algorithm’s convergence can be accurately quantified through its state evolution (SE) equations [15] and exhibits exponential convergence – in practice converging in very few iterations. AMP can also incorporate non-Gaussian noise models through Rangan’s Generalized Approximate Message Passing (GAMP) [21] and can approximate the Minimum Mean Square Error (MMSE) estimator by using a correctly matched prior or by exploiting learning structures such as Expectation Maximization [18] or SURE [19]. It has even been shown to be capable of incorporating sophisticated black box denoising algorithms in place of a signal prior model, resulting in the Denoising AMP (D-AMP) framework [20]. However, a key criticism directed at AMP, and its generalizations, is that they are specialist algorithms for i.i.d. and related measurement matrices and hence it is unclear to what extent they can be successfully applied to real world sensing problems. There has been some work exploring the convergence properties of AMP and its generalizations to other matrix classes [22], Vector AMP [23], S-AMP [24], and linking the algorithm with more classical optimization strategies such as ADMM [17]. A key problem is that when the measurement matrix is poorly conditioned and/or contains a significant mean offset the algorithm tends to diverge. One strategy for tackling this is commonly used in loopy belief propagation is to incorporate damping to help stabilize the algorithm. [25], [26]. However, damping comes at the cost of significantly reducing the algorithm’s convergence speed. It is also not clear what the value of the Onsager term is for general (deterministic) measurement matrices and whether the SE equations still provide a good prediction of the algorithm’s performance. In summary, we do not know whether AMP based techniques can provide a competitive reconstruction framework to state-of-the-art methods for general real world imaging problems. The aim of this paper is to explore these issues for the specific case of sparse view CT imaging.

A. Main Contributions

Our approach to develop an AMP based algorithm for CT reconstruction builds on a number of the recent developments in the field and, in particular, it makes use of the following key points: i) the design of a good preconditioner for the system based on the forward measurement model; ii) the inclusion of a non-linear Poisson noise model through the GAMP formulation; and iii) the incorporation of a broader class of signal prior than sparsity based models, through the D-AMP framework to enable the exploitation of state-of-the-art image denoising functions. We also demonstrate empirically the value of the Onsager term in the resulting algorithm and the accuracy of the generalized state evolution equations [21] even in this non-random setting.

As far as the authors are aware, this is the first work aimed at designing a denoising message passing based algorithm for CT reconstruction. A key challenge in applying GAMP to CT is the fact that the CT measurement operator for parallel or fan beam geometry has the form of a Radon type transform and is very ill-conditioned. This would require a significant level of damping to stabilize it and would be extremely slow [22]. The solution that we follow here is to replace the ill-conditioned operator with a much better conditioned one through preconditioning, exploiting the filtered back projection property of the system model [27]. The same procedure can be applied for different CT geometries like 2D fan-beam and 3D helical.

Another key challenge for CT reconstruction is how to accurately represent the Poisson noise model in the system. This can be approximated as a weighted $L_2$ error criterion [11], but then the preconditioner needs to account for both the system operator and the weighting matrix. While such preconditioners have been proposed, e.g. [14], they do not exploit the geometry of the measurement system and the subsequent system remains poorly conditioned, resulting in only modest improvements in convergence. In contrast, we will see that in the GAMP framework [21] the system operator and the noise process are naturally decoupled. This allows us a fully exploit a geometric preconditioner [27].

The final ingredient of our algorithm, which we call DCT-GAMP, is the incorporation of a Non local Means (NLM) denoiser [28] to implicitly define a signal model through a state of the art denoiser, rather than simply a sparse factorizable prior distribution [20]. We will see that the flexibility of using such a denoiser within GAMP yields to a better reconstruction of the image structure compared to more popular regularization, such as Total Variation (TV) minimization.

B. Relation to Existing Work

The main issue of stabilizing AMP algorithms for non i.i.d. measurement matrices has already received attention in the literature. As previously discussed, damping is a popular solution [22], [25], [26] and, for example, has been applied successfully to hyperspectral imaging reconstruction [29]. Schemes have also been proposed for modifying the algorithm when the matrix contains a significant non-zero offset [26]. These approaches are fundamentally different from the one we present here where both issues are solved through our choice of a geometric preconditioner. Other aspects of our algorithm, such as the exploitation of general denoisers [20], and the use of generalized noise models [21] have already
appeared in the literature. Here we combine these to define a state of the art algorithm for sparse view CT reconstruction.

During the manuscript preparation, the authors also became aware of a new class of AMP algorithms called Vector AMP (VAMP) [23] (and the similar orthogonal AMP in [30]) that directly tackle the ill-conditioning problem in AMP by exploiting the singular value decomposition (SVD) of the measurement matrix. Such algorithms exhibit impressive performance and have provable reconstruction guarantees for the class of right-orthogonally invariant random matrices characterized by a scalar SE equation. The main intuition for such algorithms is that using the SVD of the measurement matrix, \( \Phi = U S V^T \), the right-orthogonal random component, \( V^T \) can be decoupled from the poorly conditioned component, \( U \) which is dealt with via a linear MMSE estimator component within the VAMP iteration [23]. While this significantly increases the class of matrices for which AMP techniques can be applied it still requires the calculation of the SVD. For large imaging problems, such as 2D or 3D CT imaging such a calculation is not practical as the operators themselves are computed on the fly and not stored in matrix form. In contrast, the approach we propose here similarly removes the ill-conditioning, but by right-multiplying by an easy to compute preconditioner, thus making it more attractive to large scale CT imaging applications. Another difference from VAMP is therefore that our preconditioner modifies the signal space and thus the signal model is defined in the preconditioned space rather than the original image space.

Finally, it is useful to draw a link with the existing literature on model based iterative reconstruction (MBIR) for CT imaging. Current state-of-the-art MBIR solutions for CT are based on minimizing a regularized negative log-likelihood (NLL) cost function or its approximation using penalized weighted least squares (PWLS), [4], [11], [31], which can be interpreted as a Bayesian maximum a posteriori (MAP) estimator. This MAP framework can also be modified to incorporate denoising functions [32]. In contrast, our proposed algorithm takes the MMSE estimator perspective on AMP which is consistent with the SE equation predictions. Furthermore, as MAP estimation reduces to an optimization problem, the conditioning effects of the noise and system models are intertwined such that typical preconditioning has only a limited benefit. Using a preconditioned GAMP framework allows us to decouple these two effects.

### C. Notation

Matrices and column vectors are written respectively in capital and normal boldface type, i.e. \( A \) and \( a \) to distinguish from scalars and continuous variables written in normal weight. \( ( \cdot ^T ) \) refers to the transpose of a matrix and \( I \) refers to a vector of ones. Non-random quantities and random realizations are not distinguished typographically while random variable are written with capital letters. The conditional probability density function of \( y \) given \( x \) is denoted alternatively by \( p_{Y|X}(y|x) \) or \( p(y|x) \). A Gaussian random variable \( x \) with vector mean \( a \) and isotropic variance \( b \) is denoted by \( x \sim N(a, b) \). \( (a, b) = b^T a \) refers to the vectors inner product.

### D. Structure of the Paper

The remainder of this paper is structured as follows: Section II briefly describes the physical model of transmission X-ray CT from the continuous to discrete domain, introduces the Poisson non linear noise model and the approximations that lead to the classical PWLS statistical CT reconstruction problem. The section concludes with a discussion on the effects of the system and noise models on the conditioning of the problem. Section III reviews the original AMP algorithm for CS reconstruction, while Section IV-B presents the proposed DCT-GAMP algorithm highlighting the innovations which consist in utilizing the preconditioning for the Radon operator and incorporating the non linear CT Poisson noise model. Furthermore, we show empirical results for the SE of DCT-GAMP. Finally, in Sections V and VI comprehensive results of DCT-GAMP on a numerical phantom and experimental acquisitions of cargo luggage are shown together with a comparison of its performance with state-of-the-art algorithm for model-based CT reconstruction.

### II. X-RAY COMPUTED TOMOGRAPHY MODEL

#### A. Continuous-to-discrete model

X-ray CT produces images of attenuation coefficients of the object or patient being scanned. A typical geometry of a CT scanner involves an incoherent source of X-ray radiation and a detector array recording the intensity of the radiation exiting the object along a number of paths. If the intensity of the source of radiation, \( I_0 \), passing through the object is known, then Beer’s law provides the expected intensity after transmission, \( I_i \) of the \( i \)-th ray as:

\[
I_i = I_0 e^{- \int_{L_i} \mu(x) \, dl} + \epsilon_i \tag{1}
\]

where \( \int_{L_i} \, dl \) is the line integral along \( L_i \) which is the path of the \( i \)-th ray from the source to the detector. \( \mu(i) \) is the spatial distribution of attenuation and \( \epsilon_i \) models the scatter and other background noise in the \( i \)-th measurement. Equation (1) assumes a monoenergetic X-ray source which does not usually hold in practice. However, a common effective strategy for dealing with this consists of applying a polychromatic-to-monochromatic source correction pre-processing step [33], and in the rest of the paper we will therefore assume that we have a monoenergetic source or that it has already been appropriately corrected.

To obtain a discrete model, we should approximate the continuous attenuation function, \( \mu(\vec{r}) \in L_2(\mathbb{R}^2) \), here defined over the 2D domain, using a finite basis expansion:

\[
\mu(\vec{r}) \approx \sum_{j=1}^{N} \mu_j b_j(\vec{r}) \tag{2}
\]

where \( \mu = [\mu_1, \ldots, \mu_N]^T \) is the vector of attenuation coefficients and \( b_j(\vec{r}) \) define the \( N \) basis functions associated with a discrete sampling on a \( \sqrt{N} \times \sqrt{N} \) Cartesian grid.
According to the parameterization in Eq. (2), the line integral becomes a summation:

$$\int_{L_i} \mu(\vec{v})d\vec{v} \approx \sum_{j=1}^{N} \mu_j \int_{L_i} b_j(\vec{v})d\vec{v} = \sum_{j=1}^{N} a_{ij} \mu_j. \quad (3)$$

Repeating this over all lines defines the full view linear tomographic system matrix $A = [a_{ij}]$, where we assume that a sufficient density of lines has been taken such that the operator, $A$, is one-to-one and hence invertible on its range, e.g. [34].

Considering the sparse view scenario, the sub-sampled CT operator can now be represented as the application of a row sub-selection operator, $\hat{S}$, to $A$, such that the linear part of the measurement system can be described in matrix form by

$$\Phi = \hat{S}A \in \mathbb{R}^{M \times N} \quad (4)$$

with an effective undersampling ratio given by $M/N$.

In the case of normal exposure, the transmitted photon flux, $I_i$, follows a Poisson distribution. Using the discrete parameterization, Eqs. (2) and (3), we obtain the following discrete generalized linear model:

$$Y_i \sim \text{Poisson}\{I_0 e^{-z_i} + \epsilon_i\}, \quad i = 1, \ldots, M \quad (5)$$

where $z_i$ represents the discrete (linear) projection of the $i$th ray such that, $z = \Phi \mu$.

### B. Sparse view CT reconstruction

The sparse view CT reconstruction problem aims to estimate the attenuation coefficients, $\mu$, from the measurements $y = [y_1, \ldots, y_M]^T$ subject to Eq. (5) and any additional regularization. The negative log-likelihood (NLL) function for (5) given $y$ has the form [4]:

$$-L(\mu) = \sum_{i=1}^{M} \left\{ y_i \log \left[ I_0 e^{-[\Phi \mu]_i} + \epsilon_i \right] - I_0 e^{-[\Phi \mu]_i} - \epsilon_i \right\}. \quad (6)$$

In the case of high/normal exposure a common practice is to use a quadratic approximation of Eq. (6) which leads to a Penalized Weighted Least Squares (PWLS) approximation [4] based on taking the logarithm of the data, $l_i = \log \left( \frac{I_0}{y_i - \epsilon_i} \right)$. This is equivalent to observing $z$ corrupted with a data-dependent Gaussian noise, $\epsilon$, with inverse covariance $W = \text{diag}\left[ \frac{(y_i - \epsilon_i)^2}{y_i} \right]$:

$$1 = z + \epsilon = \Phi \mu + \epsilon \quad (7)$$

The NLL can then be approximated as:

$$-L(\mu) \approx \text{const.} + (\Phi \mu - 1)^T W (\Phi \mu - 1). \quad (8)$$

For low dosage the logarithm cannot be utilized since the argument may not be non-negative.

### C. Conditioning in sparse view CT

It is instructive to consider the issues in minimizing (8). Most popular reconstruction algorithms solve a regularized form of (8) to further incorporate prior information of the image to be reconstructed:

$$\min_{\mu \in \mathbb{R}^M} \frac{1}{2} ||y - \Phi \mu||_W^2 + \lambda P(\mu) \quad (9)$$

with $P$ usually a convex and possibly non-smooth regularization function. Assuming (9) is convex, many first order methods can be applied to solve the optimization problem, including iterative shrinkage (IST) and its accelerated variants (FISTA, M-FISTA). However the convergence rate of such methods is highly dependent on the conditioning of the problem which in turn is a function of the Lipschitz constant of the data fit term $L = \sigma_{\text{max}}(\Phi^T W \Phi)$ where $\sigma_{\text{max}}$ is the maximum eigenvalue. A large value of $L$ requires the use of a small step-size to ensure stability and results in slow convergence.

If the weighting matrix $W \propto I$, we are faced with the challenge of finding a preconditioner for the system matrix $\Phi = \hat{S}A$ and fortunately there exist good preconditioners for this scenario based on the geometry of the tomographic problem. For example, this has been used in [27] where solutions for the direct inversion of $A$ through a filter back projection operator are exploited. Indeed, both $W$ and $\Phi^T W$ are separately easy to precondition. However, together, as in the PWLS framework, it is much more challenging. One approach that has been proposed [14] is to construct a diagonal preconditioner, $D$, that majorizes the matrix, $\Phi^T W \Phi$:

$$D = \text{diag}\left( \Phi^T W \Phi \right) > \Phi^T W \Phi. \quad (10)$$

This solution exploits the non-negativity property of the measurement matrix $\Phi$. Unfortunately, this type of preconditioner does not take into account the geometric structure in the system and therefore typically only provides modest speed improvements.

We will see that the GAMP framework enables us to avoid such problems by decoupling the measurement and noise components of the system. We are therefore able to exploit a preconditioner designed specifically for $A$ which we detail next.

### D. Preconditioning of the Radon operator

The aim is to replace the poorly conditioned operator, $A$, with a new operator, $\tilde{A}$, that has a small condition number, i.e. it is a nearly tight frame, by mapping to a preconditioned image space. For 3D CT with parallel projections or fan-beam with appropriate resampling, our proposed solution is to use a cone filter applied in the image domain that amplifies high spatial frequencies, as has previously been used to accelerate reconstruction convergence for PWLS [35].

In order to construct a discrete preconditioner, while staying geometrically faithful to the continuous setting we follow the work of Averbuch et al. [34] and use the discrete pseudopolar Fourier transform (PPFT) form of the discrete Radon transform (also sometimes called a linogram [36]).
The 2D PPFT, which we denote in matrix form as \( P \), is a
4 times overcomplete radially sampled Fourier transform. In
terms of the PPFT the discrete Radon transform can be written as [34]:
\[
A = F_1^{-1} P
\]  
(11)
where the matrix operator, \( F_1 \) applies the 1D discrete Fourier
transform (DFT) separately to each of the radial lines. This
formulation has the advantage of being guaranteed to be one-
to-one. Unfortunately, \( P \) is poorly conditioned and hence so is \( A \). In order to rectify this issue [34] proposed working with a modified transform:
\[
\tilde{A} = F_1^{-1} CP
\]  
(12)
where \( C \) is a diagonal matrix that normalizes the PPFT
components by the sampling rate relative to the Cartesian
samples and is defined by
\[
C_{i,i} = \begin{cases} 
\frac{\sqrt{k(i)}}{2N}, & k(i) \neq 0 \\
\frac{1}{8N}, & k(i) = 0.
\end{cases}
\]
(13)
Here \( k(i) \) is the pseudopolar radius associated with the \( i \)-th component of the PPFT in vectorized form.
Replacing \( A \) with \( \tilde{A} \) is equivalent to working in a new
preconditioned signal space, \( x = V \mu \) via the linear transform
\[
V = P^{-1} CP.
\]  
(14)
i.e. via a high pass filtering of the PPFT spectrum of the image.
As \( P \) is a redundant operator, the inverse in (14) is to be interpreted as the inverse defined on the range of \( P \). Both the
PPFT and its inverse have fast \( O(N \log \sqrt{N}) \) implementations.
Although other Fourier based preconditioners could have been
chosen, the PPFT based preconditioner has the advantage that
the operator is assured to be one-to-one and empirically
the singular value spread of \( \tilde{A} \) is typically less than 10%.
For sparse view CT, the row sub-sampling operator \( S \in \mathbb{R}^{M \times N} \) is applied, such that the overall linear measurement
system can be expressed by
\[
\tilde{\Phi} = SA \in \mathbb{R}^{M \times N}.
\]  
(15)
An important consequence of applying such preconditioning is
that the image prior to be used in the GAMP reconstruction
framework needs to be defined on \( x \) in the preconditioned
space. It will also be necessary to apply a final post-processing
step to map the estimated vector, \( x \), back into the image
domain \( \mu \).

III. REVIEW OF APPROXIMATE MESSAGE PASSING
ALGORITHM

In this Section, we review the original formulation of the
AMP algorithm [15] for the CS system model \( y = \Phi \mu + e \),
where each entry of the measurement matrix \( \Phi \) is i.i.d.
random Gaussian distributed \( \mathcal{N}(0, \sigma^2) \) and the noise model
is Gaussian, i.e. \( e \sim \mathcal{N}(0, \sigma^2) \). AMP is an iterative algorithm
which proceeds according to the following equations:
\[
\begin{align*}
\mu^{t+1} &= \eta_{\delta t}^{(t)}(\mu^t + \frac{1}{d} \Phi^T r^t) \\
\sigma^t &= \eta_{\delta t}^{(t)}(\mu^t + \frac{1}{d} \Phi^T r^t) >
\end{align*}
\]  
(16)
where \( \delta = M/N \) represents the measurement rate, \( \eta_{\delta t} \)
are scalar threshold functions (applied componentwise) with
\( \eta_{\delta t}^{(t)}(\mu) = \frac{d}{\sigma^t} \eta_{\delta t}^{(t)}(\mu) \), \( < \cdot > \) is the average of a vector,
\( \mu^t \in \mathbb{R}^N \) is the current estimate of \( \mu \) and \( \frac{1}{d} \Phi^T r^t < \eta_{\delta t}^{(t)}(\mu^{t+1} + \Phi^T r^{t+1}) > \) is called, from statistical physics, the
Onsager correction term. The Onsager term has a key role since
it ensures that the input of the threshold function, \( \mu^t + \frac{1}{d} \Phi^T r^t \), is equivalent, in the large system limit, to the true \( \mu \) corrupted by
Gaussian noise with variance \( (\delta t)^2 \); therefore, \( \eta_{\delta t} \) acts as an iteration dependent threshold function that outputs an estimate of \( \mu \), given some noisy measurements
\[
\mu^t + \Phi^T r^t = \mu + \delta t \psi
\]  
(17)
where \( \psi \sim \mathcal{N}(0,1) \). Compared to previous proposed iterative
thresholding schemes, like soft-thresholding for sparse signal
reconstruction which does not include the crucial Onsager
term and the iteration dependent threshold, AMP can achieve
sparsity-undersampling trade-off matching the theoretical one
for linear programming-based reconstruction while running
dramatically faster.

The asymptotic performance of AMP can be characterized
by a SE formalism [37]:
\[
\xi[(\delta t)^2] = \lim_{N \to \infty} \frac{1}{M} \|\mu^{t+1} - \mu\|^2
\]
\[
= \mathbb{E}_\Psi[\eta_{\delta t}^{(t)}(\Phi^T(\mu + \delta t \Psi) - M)^2]
\]
\[
(\delta t)^2 = \sqrt{\frac{\sigma^2}{\delta t^2 - 1}} \xi[(\delta t)^2]
\]  
(18)
where the random variables \( \Psi \) and \( M \) have realizations re-
spectively \( \psi \) and \( \mu \) and \( M \) is drawn from the prior
probability distribution \( M \sim p_M \). Eq. (18) indicates the noise variance for
AMP iterations; the noise variance needs not necessarily be
the smallest possible, unless \( \eta_{\delta t} \) achieves the MMSE in each
iteration.

IV. DCT-GAMP: DENOISING CT WITH POISSON NOISE
BASED AMP

The proposed algorithm for CT reconstruction is built upon
the AMP framework with the following innovations: i) incor-
porate the preconditioner for the Radon operator, introduced
in Section II-D, such that the iterative algorithm is performed
in the preconditioned space together with a new operator with
a smaller condition number; ii) extend the AMP formulation
(16) for the Poisson noise model (5) by exploiting the GAMP
framework; iii) use a generic denoiser in the non linear step
to capture the data-dependent structure of complex images [20].
The benefit of employing i) and ii) relies on the property of
decoupling the measurements and noise components unlike the
solution in (10).
A. Preconditioning of the measurement operator

As described in Section II-D, the Radon operator \((11)\) can be preconditioned by using \((14)\) such that the combined operator \(\mathbf{A}\) has a condition number considerably lower than \(\mathbf{A}\). Since the preconditioner \(\mathbf{V}\) has a symmetric Toeplitz structure \([38], [39]\), it is possible to split the preconditioner and define the modified system matrix combining \((14)\) and \((4)\)

\[
\mathbf{B} = \mathbf{S} \mathbf{A} \mathbf{V}^{-\frac{1}{2}} = \Phi \mathbf{V}^{-\frac{1}{2}} \\
\mathbf{B}^T = \mathbf{V}^{-\frac{1}{2}} \mathbf{A} \mathbf{S} = \mathbf{V}^{-\frac{1}{2}} \Phi^T
\]

where \(\mathbf{V}^{-\frac{1}{2}}\) is obtained by the inverse square root of each element, since the operator \(\mathbf{V}\) multiplies the PPFT of the signal element-wise. The computational complexity of both operators, \(\mathbf{B}\) and \(\mathbf{B}^T\), is of order \(O(N \log \sqrt{N})\), since they are defined as a composition of element-wise operators with complexity \(O(N)\) and the PPFT, with complexity \(O(N \log \sqrt{N})\).

In an equivalent way, the preconditioning leads to the following change of coordinates in the signal domain within each iteration \(t\):

\[
\begin{align*}
\mu^t &= \mathbf{V}^{-\frac{1}{2}} x^t \\
x^t &= \mathbf{V}^\frac{1}{2} \mu^t
\end{align*}
\]

B. Incorporation of Poisson Noise Model in AMP

We consider the sparse views X-ray CT transmission model where the input vector \(\mathbf{z} \in \mathbb{R}^N\) is passed through the linear Radon CT operator together with the angular subsampling operator, that is modelled as

\[
\lambda_a = e^{-\mathbf{z}_a} = e^{-|\mathbf{B} \mathbf{x}_a|}, \quad a = 1, \ldots, M
\]

where the linear term is \(z = \mathbf{S} \mathbf{A} \mu = \Phi \mathbf{y}_z = \mathbf{B} \mathbf{x}\) from Eq. \((19)\) and \((20)\). Finally, each component \(\lambda_a\) randomly generates an output component \(y_a\) of the vector \(y \in \mathbb{Z}^M\). The conditional probability distribution of the i.i.d. random variable \(y\) given the linear measurement \(Z\) is an exponential-Poisson distribution \([31]\)

\[
p_{y|z}(y|z) = \prod_{a=1}^{M} \frac{1}{y_a!} e^{-(\mathbf{e}^{-\mathbf{z}_a})} e^{-y_a \mathbf{z}_a}
\]

The block diagram of the system model is shown in Fig. 1. From the reconstruction point of view, the GAMP algorithm reduces the overall MMSE estimation to a sequence of simpler MMSE estimates based on the large system limit assumption. Algorithmically, given a complete factor graph representing the linear system \(z = \mathbf{B} \mathbf{x}\), GAMP employs a MMSE estimator of \(z\), which results from a Gaussian approximation of the sum-product loopy BP on the dense graph (induced by \(\mathbf{B}\)), and it propagates these means and isotropic variances estimates backwards through \(\mathbf{B}\) to give a noisy estimate for \(x\). Then, the algorithm performs a MMSE estimate of \(x\) and propagate it forwards on the measurements again.

In this Section, we describe how to perform the MMSE estimation in the measurement domain for the nonlinear CT Poisson noise model while in Section IV-C we utilize a black box estimator which approximates the Bayesian MMSE estimator in the signal domain. For analysing the MMSE estimator related to the linear system \(z' = Bx\), the posterior probability distribution \(p(z'|p^t, y, \tau^t_p)\) of \(Z^t\) given \(Y = y\) and \(p^t = z^t - \tau^t p^{t-1}\) is given by

\[
p(z'|p^t, y, \tau^t_p) \propto e^{py_{|z'}(y|z')} e^{-\frac{\tau_p}{\tau_p}(z'-p^t)(z'-p^t)}
\]

which is the product of the likelihood distribution of the random variable \(Y\) given \(Z'\), i.e. the noise distribution \(p_{y|z'}(y|z')\), and the prior distribution for \(z'\) which is approximately Gaussian distributed with vector mean \(p^t\) and scalar variance \(\tau^t_p\), i.e. \(z' \sim N(p^t, \tau^t_p)\).

The vector \(p^t\) is the estimated linear output (detailed in line (7) of the Algorithm 1) and the term \(\tau^t_p\) represents the Onsager term. Given \(p(z'|p^t, y, \tau^t_p)\), the approximate iterative BP for the MMSE problem is achieved by computing

\[
\begin{align*}
z_0^t &:= \mathbb{E}[p(z'|p^t, y, \tau^t_p)|z^t, y, \tau^t_p] \quad (24) \\
\tau^t_p(p^t, y, \tau^t_p) &:= \frac{1}{\tau^t_p} \left( z_0^t - p^t \right) \quad (25)
\end{align*}
\]

where \(z_0^t\) is the conditional expectation of \(p^t\) given \(z^t\), i.e. the MMSE estimate of \(Z^t\), and the negative derivative of \(\tau^t(p^t, y, \tau^t_p)\) with respect to \(p^t\) is given by

\[
M \tau^t_p = - \frac{\partial}{\partial p^t} \tau^t(p^t, y, \tau^t_p) = 1 - \frac{\text{Var}(z^t|p^t, y, \tau^t_p)}{\tau^t_p}
\]

To obtain \(\tau^t(p^t, y, \tau^t_p)\), we need to evaluate the expectation \(\mathbb{E}[z^t|p^t, y, \tau^t_p]\) with respect to \(p(z'|p^t, y, \tau^t_p)\), where

\[
\log p(y|z') = -(z', y) - \langle e^{-z'}, 1_M \rangle - \langle \log(y), 1_M \rangle
\]

\[
p(z'|p^t, y) = e^{-\langle z', y \rangle - \langle e^{-z'}, 1_M \rangle - \langle \log(y), 1_M \rangle - \frac{1}{\tau^t_p} ||z' - p^t||_2^2}
\]

\(z' \in \mathbb{R}_{\geq 0}^M\)
The expectation requires solving the following ratio of integrals for each element indexed with $a = 1, \ldots, M$:

$$
\mathbb{E}[z_a^t | p^t_a, y_a, \tau_p^t] = \frac{\int_{z_a \geq 0} z_a^{\log P(Y|z_a)} e^{-\frac{1}{2\tau_p^t} (z_a - p_a)^2} dz_a}{\int_{z_a \geq 0} e^{\log P(Y|z_a)} e^{-\frac{1}{2\tau_p^t} (z_a - p_a)^2} dz_a}
$$

(28)

Unfortunately no close form solution appears to exist and therefore Laplace’s method [40] is used to approximate the posterior mean $z_0^t$ and $\tau_p^t$. In Appendix A, the calculation for $z_0^t$ and $\text{Var}[z^t | p^t]$ is detailed.

It is important to highlight the difference between the Poisson NLL approximation and the procedure introduced here.

**Algorithm 1:** DCT-GAMP: Denoising Preconditioned Approximate Message Passing

1. **Initialization:** set $t = 0$, $s^0 = 0$, $\mu^0 = 0$, $\tau_p^0 = 1$

2. for $1, \ldots, T_{\text{max}}$

3. Step 1: Estimate in the measurement domain

   $z^t = \Phi(\sqrt{\tau} x^t)$

4. $\tau_p^t = \sqrt{\frac{1}{\tau_p} \| \Phi \sqrt{\tau} x^t \|^2}$

5. $p^t = \sqrt{\tau_p} \tau - 1$

6. Step 2: Poisson noise model

   $z_0^t = \mathbb{E}[z^t | p^t, y^t, \tau_p^t] = |z^t|^p \Phi V - \frac{1}{2} \| z^t - y^t \|^2$

7. $r^t = \frac{1}{\tau_p^t} [1 - \frac{\text{Var}(z^t | p^t, y^t, \tau_p^t)}{\tau_p^t}]$

8. Step 3: Estimate in the signal domain

   $\frac{1}{\tau_p^t} = \frac{1}{\tau_p} \| \Phi \sqrt{\tau} x^t \|^2$

9. $s^t = x^t + \tau_p^t \Phi^T r^t$

10. Step 4: Denoising step

    $x^{t+1} = D_{\sigma^t}(s^t)$

11. $\tau_p^{t+1} = \tau_p^t D_{\sigma^t}(s^t)$

12. $\hat{\sigma}^t = \frac{1}{\tau_p} || r^t ||^2$

13. Return $\mu^t = \sqrt{\tau} x^t$

**C. Denoising: Non-Linear Input Module**

Whilst the original GAMP algorithm was developed on a factorial (sparse) signal model, the framework has been shown to be amenable to much broader classes of estimators [20]. Since the GAMP algorithm approximates the estimate for $x$ as a noise corrupted version of the true signal with variance $(\hat{\sigma}^t)^2$ as in Eq.(17), it is meaningful to employ, instead of a prior-based non linear scalar function $r_{\sigma^t}$, a denoiser $D_{\sigma^t}$ which acts as a standard non-linear mapping

$$
D_{\sigma^t} : \mathbb{R}^N \rightarrow \mathbb{R}^N, \quad x \mapsto D_{\sigma^t}(x)
$$

(29)

that outputs an estimate of $x$, given some noisy measurements $x + \sigma \psi$ with $\psi \sim \mathcal{N}(0, 1)$. We treat $D_{\sigma^t}(-)$ as a black box MMSE estimator, i.e., we do not require knowledge of its functional form [20].

The main reason for using a generic denoiser in the non linear step is to capture the data-dependent structure of complex images, rather than a simple factorial model, obtaining a sequence of estimates eventually converging faster to the true preconditioned signal $x$; this provides the flexibility in using a variety of denoisers to reduce the noise at a voxel-level.

Given the estimated signal

$$
\hat{s}^t = x^t + \tau_p^t V - \frac{1}{2} \Phi^T r^t
$$

(30)

which is the input of the denoiser, the output vector estimates and the scalar variance are given by

$$
\begin{align*}
\hat{s}^{t+1} &= \mathbb{E}[x | s = s^t] = D_{\sigma^t}(s^t) \\
\tau_p^{t+1} &= \text{Var}(x | s = s^t) = \tau_p^t D_{\sigma^t}(s^t)
\end{align*}
$$

(31)

where the estimated noise level [15] is given by

$$
\hat{\sigma}^t = \frac{1}{\tau_p} || r^t ||^2
$$

(32)

and $D_{\sigma^t}(-)$ denotes the divergence of the denoiser.

The analytic calculation of $D_{\sigma^t}(-)$ is often not available and it is in general data-dependent, but a good approximation can be obtained through the Monte Carlo technique [41]. With this method, the calculation of the Onsager term is more efficient since it requires only one more application of the denoiser. Moreover, it follows from Eq. (31) that the denoiser $D_{\sigma^t}(-)$, introduced in Section IV-C acts on the high pass filtered image $x$, whose expression is in Eq. (20).

In Fig. 2 it is shown the block diagram for the mean calculation of the proposed DCT-GAMP algorithm; each iteration flow can be decomposed in 3 main steps: the MMSE estimation for the Poisson noise channel of the output vector $p^t$, the preconditioning, which involve a change of the signal domain, and the denoising of the signal estimate.

**D. State evolution of DCT-GAMP**

A significant characteristic of GAMP is that the MSE performance can be precisely predicted by a scalar SE analysis, with i.i.d. Gaussian random system matrices in the large system limit [42]; in particular, the GAMP SE formulation extends the SE in Eq. (18) to arbitrary noise distributions.

In addition, if a generic denoiser is used as in Eq. (31), it is shown empirically in [20] that the MSE can be predicted precisely by the SE. Hence, the SE equations for the proposed DCT-GAMP are based on the GAMP SE derivation [42] where the signal to estimate lies in the preconditioned domain and a denoiser is utilized as the non linear input function.

The DCT-GAMP SE equations are formulated following the GAMP SE ones which are derived according to the recursion

$$
\xi^t = \mathbb{E} [X - D_{\sigma^t}(\hat{s}^t, \tau_p^t)]^2
$$

(33)

where the denoiser $D_{\sigma^t}$ takes as input an equivalent isotropic Gaussian distributed signal with mean $\hat{s}^t$ and variance $\tau_p^t$. 
Denoising

\[ \text{Change of the signal domain} \]

\[ \text{MMSE estimator - measurement domain} \]

Fig. 2: Block diagram of the DCT-GAMP framework highlighting the 3 steps: 1) Denoising the signal estimate; 2) Preconditioning: change of the signal domain; 3) MMSE estimator for the non linear Poisson noise model.

The derivation of \( \hat{s}^t \) and \( \hat{\tau}^t \) for the DCT-GAMP is described in Appendix B. Unfortunately, the SE prediction

\[
\lim_{N \to \infty} \frac{1}{N} \| \hat{x}^t - x \|^2 = \xi^t
\]

(34)
is only valid in the random large system limit and therefore one may wonder what its relevance is in the considered CT problem. Here we argue that the empirical accuracy of the SE predictions provides an insight into the validity of DCT-GAMP approximations when applied to general linear models.

In Section VI, we present an empirical evidence for the SE of DCT-GAMP based on a real CT dataset and we show that the SE equations of DCT-GAMP provide a good prediction of the actual MSE achieved by DCT-GAMP at each iteration.

V. Simulation Results with Numerical Phantom

We discuss the numerical results for a 2-D CT reconstruction from the simulated NCAT phantom [43], shown in Fig. 3(b) of dimension \( N = 256 \times 256 \), with a fan beam geometry, depicted in Fig. 3(a); we consider 50 views in the sinogram domain, obtained from a regular angular undersampling of the full projection measurements (512 views \( \approx 2\sqrt{N} \)), resulting in, approximately, 10 times undersampling ratio.

The CT projection and back-projection are implemented using the Michigan Image Reconstruction Toolbox [44]. The simulations include the Poisson noise model with different levels of intensity: an initial intensity of \( I_0 = 10^5 \), which is referred to as normal dose in the toolbox, and \( I_0 = 10^4 \) for the low dose case. The sparse views sinograms, for the 2 levels of intensity, are shown in Figs. 3(c)-(d) where it is worth noticing the low values in case of low dose; we will show that the Gaussian approximation of the CT noise is less effective with low beam intensity.

For a quantitative comparison we have chosen the PSNR as the metric, defined as the ratio between the maximum squared value of the FBP with the full number of projections and the mean square error of the estimation.

Figures 4(a)-(b) show the FBP with ramp filter, for the normal and low photon intensities, which produces very poor reconstructions with strong streaking artifacts.

In Figures 4(c)-(d) are shown the reconstruction results for the normal dosage obtained using respectively NLM-CT-GAMP algorithm 1, whose NLM denoiser is implemented using the Matlab toolbox, with 20 iterations, and a fast gradient descent method, FISTA, for solving the PWLS objective function with a Huber regularizer on the finite difference between neighbouring pixels [11].

The image denoising algorithm NLM [28] is used as the denoiser in DCT-GAMP since it provides good reconstruction performance and keeps computation time reasonable. The algorithm for solving PWLS is run with 20 outer iterations, exploiting the ordered subsets with blocks of 41 elements and is used as the reference reconstruction algorithm to compare with our proposed method.

It is worth noting that from Figures 4(c)-(d), NLM-CT-GAMP achieves a better qualitative reconstruction compared to PWLS, whose output retains noise in the inner region probably due to the the rays intercepting the hard tissue or bones.

Furthermore, from Table I it can be seen that NLM-CT-GAMP produces a better quantitative reconstruction in terms of PSNR (dB) compared to PWLS, but it requires more computational time due to the complexity of the NLM denoiser. For computational time evaluation, the simulations are run on an Intel® Xeon 2GHz machine using 4 cores.

Finally, in Figures 5(a)-(b) the results with low dose are shown for, respectively, NLM-CT-GAMP and PWLS. It is important to highlight that, in this case, the weighted Gaussian noise approximation, is not accurate due to the presence of zero values in the sinogram related in particular to the rays intercepting the bones. Taking the logarithm of the measurement leads to errors, especially in the region surrounded by hard tissue/bones; this is also confirmed quantitatively in Table I.
Fig. 3: (a) Fan-beam geometry, (b) Original NCAT phantom, (c) Sinogram for normal dose, $I_0 = 10^5$, (d) Sinogram for low dose, $I_0 = 10^4$

![Image](image_url)

Fig. 4: (a) FBP Normal dose, (b) FBP low dose. Normal dose: (c) NLM-CT-GAMP, (d) PWLS with Huber regularizer

![Image](image_url)

Fig. 5: Low dose: (a) NLM-CT-GAMP, (b) PWLS

![Image](image_url)

VI. EXPERIMENTAL RESULTS

In this section, we investigate the reconstruction quality of DCT-GAMP on real CT data. The DCT-GAMP framework has been applied for CT reconstruction on real luggage scans obtained using Morpho CTX5500 Air Cargo dual energy system with fan beam CT geometry. This is a single-row scanner with 476 detector channels and a 80 cm field of view. For each transversal location, two slices were acquired one at 100 KVP, the other at 198 KVP; at each energy, the full acquisition of a single slice contains 720 views/projections. The reconstruction has been performed for each energy independently and here we consider only the results obtained for 100 kVp. The reconstructed image is of $256 \times 256$ array size. The results in Figure 7 show two slices from the reconstruction with DCT-GAMP using 72 views regularly undersampled out of the full set of views constituted of 720 views. In the figure it is possible to see that the scanned object contains highly resolved metal staples, bottle of fluid, wires.

The number of iterations for DCT-GAMP algorithms has been set to 15 iteration but we will see that the Mean Square Error (taking as reference the FBP with full set of views) tends to converge in under 10 iterations. As reference, in Figure 6 we show the reconstruction obtained with FBP using 72 views for one of the image slices.

### TABLE I: PSNR and time comparison

| Algorithms          | PSNR (dB) | Time  |
|---------------------|-----------|-------|
| Low photon intensity|           |       |
| FBP                 | 31.5      | 45 sec|
| PWLS - FISTA        | 54.1      | 4.9 min|
| NLM-CT-GAMP         | 61.8      | 5.7 min|
Fig. 6: Filtered Back Projection

Fig. 7: CT image reconstruction of slices 22 and 35 using: (a)-(b) NLM-CT-GAMP with Cone filter preconditioning and Poisson noise model.

Fig. 8: CT image reconstruction of slices 22 and 35 using NLM-AMP without Onsager term

A. Role of the Onsager term

Given the similarity between iterative shrinkage algorithms and the GAMP family of algorithms, it is interesting to evaluate the importance of the Onsager term $\tau^{t-1}$ in the DCT-GAMP algorithm 1 to check whether it improves the reconstruction. Without the Onsager term the GAMP algorithm behaves like a denoising iterative thresholding algorithm [20]. The reconstruction for slices 22 and 35 without the Onsager term is shown in Fig. 8 and highlights that we incur a substantial reduction in performance by its omission in both cases, as it is also quantitatively confirmed by the PSNR value in Table II. The Onsager term yields to a PSNR improvement of 9.55 dB, for this particular CT reconstruction instance.

TABLE II: PSNR for of NLM-CT-GAMP with/without Onsager term

| Algorithms                      | PSNR [dB] | Time  |
|---------------------------------|-----------|-------|
| NLM-CT-GAMP                     | 69.65     | 7.4 min |
| NLM-CT-GAMP without Onsager term| 60.1      | 4.6 min |

B. Comparison against PWLS

Although the DCT-GAMP formulation allows us to exploit more complex priors, it is meaningful to present a comparison in terms of accuracy and computational cost between CT-GAMP with TV a denoiser and solving the PWLS cost function with the TV regularization term with FISTA using the ordered subsets (OS) to reduce the running time. We use FISTA with OS as a reference iterative method for PWLS; moreover, this can give an indication of the order of complexity, i.e. computational time, even if there are ongoing works in accelerating such schemes, like using ADMM.

It is worth pointing out that we are comparing 2 different methods, in particular AMP exploits a denoising function based of Total variation while PWLS incorporated the TV term as a regularizer in the cost function; this is the reason that they yield different accuracies in terms of PSNR in the Table III.

TABLE III: PSNR: of TV-PWLS, TV-CT-GAMP

| Algorithms                      | PSNR [dB] | Time  |
|---------------------------------|-----------|-------|
| TV-CT-GAMP Cone Filter          | 62.1      | 5.3 min |
| TV-FISTA PWLS                   | 60.3      | 5.1 min |

Fig. 9: Results of (a) TV-CT-GAMP, (b) TV-FISTA: FISTA to solve PWLS objective with TV regularizer.

In Figure 9, we show the reconstructed results and in Table III the computational time; for FISTA we have used 80 iterations since it was the minimum number to reach
convergence for this particular dataset. TV-CT-GAMP can be seen to have better performance and roughly comparable computational complexity. Although we have reported the computational comparison between TV-CT-AMP and TV-FISTA which shows that AMP-based algorithm tends to be comparable, this analysis cannot be considered complete since there exist recent optimization methods as mentioned at the beginning of the Section.

A coherent analysis of the trade off between accuracy and complexity of AMP compared to fast first order optimization methods is left for a future study.

C. State Evolution Analysis

An important aspect of DCT-GAMP is its internal variance estimate within the algorithm and in the SE equations. This not only provides an estimate of uncertainty with the algorithm, it also essentially allows the algorithm to adapt its “step size” on the fly [17].

It is therefore instructive to see how accurate such an estimate is. If accurate, this term should ensure a fast convergence rate of the algorithm. Given the actual MSE estimate, taking as reference the full views FBP reconstruction, we can calculate the predicted MSE at the next iteration and compare with the actual estimate.

![Fig. 10: Deterministic state evolution and MSE estimates using NLM-CT-AMP with and without Onsager term](image)

Our empirical results run on the experimental data acquired using the 2D fan beam CT geometry show that the SE prediction gives a reasonable estimate of the true MSE of NLM-CT-GAMP at each iteration as depicted in Figure 10.

In particular considering the first iterations the actual MSE of NLM-CT-GAMP is very close with the predicted one, after 10 iteration the current estimate tends to be higher, between 0.7-1.2 dB but overall we can claim that we can achieve a good prediction. Note that when the Onsager term is omitted, the state evolution is no longer a good MSE predictor and the algorithm’s performance is significantly deteriorated.

VII. Conclusions

In this work, we have presented a proof of concept for using Generalized Approximate Message Passing type of iterative algorithms for solving X-ray CT reconstruction from limited number of projections. The proposed framework relies on the design of an appropriate preconditioner for the ill-conditioned measurement matrix and a statistical model for the non linear Poisson measurement noise.

In addition, exploiting the flexibility of the GAMP framework we can decouple the action of the preconditioner from the noise model, which is not possible in the PWLS framework.

We have experimentally shown the important role of the Onsager term regarding reconstruction performance improvement and the ability of the state evolution analysis to estimate the current MSE through the iterations. Numerical results on experimental Cargo data demonstrate how the DCT-GAMP framework provides a promising alternative to optimization based iterative reconstruction algorithms for CT reconstruction. In addition DCT-GAMP allows different prior image models to be used on the signal by employing different denoisers.

Further acceleration of the DCT-GAMP may be possible utilizing the Ordered Subsets principle [11], however its implementation is not straight forward within the AMP framework and is left for future research.

ACKNOWLEDGMENT

This work was supported by U.S. Department of Homeland Security, Science and Technology Directorate, Explosives Division, BAA 13-05, Contract # HSHQDC-14-C-B0048, EP-SRC platform grant EP/J015180/1 and the ERC Advanced grant, project C-SENSE, (ERC-ADG-2015-694888). The authors thank Walter Garms, Morpho Detection Inc., for providing real CT datasets.

APPENDIX A

**LAPLACE METHOD FOR APPROXIMATING THE POSTERIOR MEAN OF THE NONLINEAR NOISE DISTRIBUTION**

In order to evaluate the expression in (28), we write the ratio of integrals in the following form (where we have not indicated the iteration $t$ for notation simplicity)

$$E(z_a|p_a, y_a, \tau_p) = \frac{\int_{\mathbb{R}} g(z_a) e^\log p(y_a|z_a) \pi(z_a) dz_a}{\int_{\mathbb{R}} e^\log p(y_a|z_a) \pi(z_a) dz_a}$$

(35)

where $\pi(z_a) = e^{-\frac{1}{2\tau_p}(z_a-p_a)^2}$ and $g(z_a) = z_a$. We set

$$L = \log \pi + \frac{1}{M} \log p(y_a|z_a)$$

(36)

$$L^* = \log z_a + L = \log g(z_a) + \log \pi(z_a) + \frac{1}{M} \log p(y_a|z_a)$$

$$= \log z_a - \frac{1}{2} (z_a-p_a)^2 + \frac{1}{M} \left[ -z_a y_a - e^{-z_a} - \log(y_a!) \right]$$

$$L^* = \log z_a - \frac{1}{2} (z_a-p_a)^2 + \frac{1}{M} \left[ -z_a y_a - e^{-z_a} + \log(y_a!) \right]$$

(37)
Therefore, the MMSE can be written as
\[
\mathbb{E}(z_a | p_a, y_a, \tau_p) = \frac{\int_{R_{z_a}} e^{M^T L' z_a} dz_a}{\int_{R_{z_a}} e^{M^T L z_a} dz_a} \tag{38}
\]

We consider the probability density function \( L(z_a) \) which we expect to have a peak at the point \( z_{a_0} \) and the Taylor-expansion of \( L(z_a) \) at \( z_{a_0} \) is
\[
L(z_a) \approx L(z_{a_0}) - \frac{1}{2} \frac{\partial^2 L(z_a)}{\partial z_a^2} (z_a - z_{a_0})^2 + \ldots \tag{39}
\]

The Laplace’s method [40] is a way to approximate \( L(z_a) \) by an unnormalized Gaussian and approximate the partition function \( Z_P = \int L(z_a) dz_a \) with the one of the Gaussian
\[
Z_Q = L(z_{a_0}) \sqrt{\frac{2 \pi}{c}} \tag{40}
\]

The Laplace approximation leads to
\[
\int e^{mL(z_a)} dz_a \approx \int e^{mL(z_{a_0}) - m(z_a - z_{a_0})^2/(2\sigma^2)} dz_a = \sqrt{2 \pi \sigma n}^{-1/2} e^{mL(z_{a_0})} \tag{41}
\]

with \( \sigma^2 = -\frac{1}{L''(z_{a_0})} \); this integral form is similar to the one in Eq. (28) for the numerator and denominator.

Considering the denominator, we need to calculate the points where the derivative is zero in order to find \( z_{a_0} \):
\[
\frac{\partial L(z_a)}{\partial z_a} = -\frac{1}{\tau_p} (z_a - p_a) - y_a + e^{-z_a}
\]

with \( y_a \in \mathbb{Z}_+ \), \( z_a = [Bx]_a \in \mathbb{R}_\geq 0 \); then, to find \( \frac{\partial L(z_a)}{\partial z_a} = 0 \), it results
\[
-\frac{(z_a - p_a)}{\tau_p} - y_a + e^{-z_a} = 0
\]
\[
\log \left[ -\frac{(z_a - p_a)}{\tau_p} - y_a \right] = z_a
\]

Finally, the second derivative is
\[
\frac{\partial^2 L(z_a)}{\partial z_a^2} \bigg|_{z_{a_0}} = -\frac{z_{a_0}}{\tau_p} - e^{-z_{a_0}}
\]

Similar procedure for the numerator (\( \sigma^* \) and \( z_{a_0}^* \)); therefore, taking the ratio of the 2 approximations it yields to
\[
\mathbb{E}[z_a | p_a] = \frac{\sigma^*}{\sigma} e^{L^*(z_{a_0}^* - L(z_{a_0}))} \tag{42}
\]

For the variance, given the approximation 42, we can use the standard formula
\[
\text{Var}[z_a | p_a] = \mathbb{E}[z_a^2 | p_a] - \mathbb{E}[z_a | p_a]^2 \tag{43}
\]
[12] S. Ramani and J. A. Fessler, “A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction,” IEEE Transactions on Medical Imaging, vol. 31, no. 3, pp. 677–688, 2012.

[13] S. Y. Chun, Y. Dewaraja, and J. A. Fessler, “Alternating direction method of multiplier for tomography with nonlocal regularizers,” IEEE Transactions on Medical Imaging, vol. 33, no. 10, pp. 1960–1968, 2014.

[14] D. Kim, S. Ramani, and J. A. Fessler, “Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction,” IEEE Transactions on Medical Imaging, vol. 34, no. 1, pp. 167–178, 2015.

[15] D. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” Proceedings of the National Academy of Sciences, vol. 106, no. 45, pp. 18914–18919, 2009.

[16] M. Opper and O. Winther, “Expectation consistent approximate inference,” Journal of Machine Learning Research, vol. 6, no. Dec, pp. 2177–2204, 2005.

[17] S. Rangan, A. K. Fletcher, P. Schniter, and U. Mikolaj, “Inference for generalized linear models via alternating directions and Bethe free energy minimization,” in 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 1640–1644.

[18] J. P. Vila and P. Schniter, “Expectation-maximization gaussian-mixture approximate message passing,” IEEE Transactions on Signal Processing, vol. 61, no. 19, pp. 4658–4672, 2013.

[19] C. Guo and M. E. Davies, “Near optimal compressed sensing without priors: Parametric sure approximate message passing,” IEEE Trans. Signal Processing, vol. 63, no. 8, pp. 2130–2141, 2015.

[20] C. A. Metzler, A. Maleki, and R. G. Baraniuk, “From denoising to compressed sensing,” IEEE Transactions on Information Theory, vol. 62, no. 9, pp. 5117–5144, 2016.

[21] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in IEEE International Symposium on Information Theory Proceedings, 2011, pp. 2168–2172.

[22] S. Rangan, P. Schniter, and A. Fletcher, “On the convergence of approximate message passing with arbitrary matrices,” in IEEE International Symposium on Information Theory (ISIT), IEEE, 2014, pp. 236–240.

[23] S. Rangan, P. Schniter, and A. Fletcher, “Vector approximate message passing,” arXiv preprint arXiv:1610.03082, 2016.

[24] B. Cakmak, O. Winther, and B. H. Fleury, “S-AMP: Approximate message passing for general matrix ensembles,” in IEEE Information Theory Workshop (ITW), 2014, pp. 192–196.

[25] P. Schniter and S. Rangan, “Compressive phase retrieval via generalized approximate message passing,” IEEE Transactions on Signal Processing, vol. 63, no. 4, pp. 1043–1055, 2015.

[26] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, “Adaptive damping and mean removal for the generalized approximate message passing algorithm,” in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 2021–2025.

[27] M. Nilchian, C. Vonesch, P. Modregger, M. Stampanon, and M. Unser, “Fast iterative reconstruction of differential phase contrast X-ray tomograms,” Optics express, vol. 21, no. 5, pp. 5511–5528, 2013.

[28] A. Buades, B. Coll, and J-M. Morel, “A non-local algorithm for image denoising,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). IEEE, 2005, vol. 2, pp. 60–65.

[29] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, “Compressive hyperspectral imaging via approximate message passing,” IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 2, pp. 389–401, March 2016.

[30] J. Ma and L. Ping, “Orthogonal amp,” IEEE Access, vol. 5, pp. 2020–2033, 2017.

[31] J. Nyuys, B. De Man, J. A. Fessler, W. Zhijewski, and F. J. Beekman, “Modelling the physics in the iterative reconstruction for transmission computed tomography,” Physics in medicine and biology, vol. 58, no. 12, pp. R63, 2013.

[32] S. Ramani, X Wang, L. Fu, and M. Lexa, “Denoising-based accelerated statistical iterative reconstruction for x-ray ct,” in 4th International Meeting on Image Formation in X-Ray Computed Tomography, 2016, pp. 337–340.

[33] B. Whiting, P. Massoumzadeh, O. Earl, J. OSullivan, D. Snyder, and J. Williamson, “Properties of preprocessed sinogram data in X-ray computed tomography,” Medical physics, vol. 33, no. 9, pp. 3290–3303, 2006.

[34] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, and J. Walden, “Fast slant stack: A notion of radon transform for data in a cartesian grid which is rapidly compatible, algebraically exact, geometrically faithful and invertible,” SIAM Scientific Computing, 2001.

[35] N. H. Clinehorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, “Preconditioning methods for improved convergence rates in iterative reconstructions,” IEEE Transactions on Medical Imaging, vol. 12, no. 1, pp. 78–83, 1993.

[36] P. R. Edholm and G. T. Herman, “Linograms in image reconstruction from projections,” IEEE Transactions on Medical Imaging, vol. 6, no. 4, pp. 301–307, 1987.

[37] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” IEEE Transactions on Information Theory, vol. 57, no. 2, pp. 764–785, 2011.

[38] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, and Y. Shkolnisky, “A framework for discrete integral transformations I - the pseudopolar fourier transform,” SIAM Journal on Scientific Computing, vol. 30, no. 2, pp. 764–784, 2008.

[39] A. Averbuch, R. Coifman, D. L. Donoho, M. Israeli, Y. Shkolnisky, and I. Sedelnikov, “A framework for discrete integral transformations II - the 2D discrete radon transform,” SIAM Journal on Scientific Computing, vol. 30, no. 2, pp. 785–803, 2008.

[40] L. Tierney and J. B Kadane, “Accurate approximations for posterior moments and marginal densities,” Journal of the American statistical association, vol. 81, no. 393, pp. 82–86, 1986.

[41] S. Ramani, T. Blu, and M. Unser, “Monte-carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms,” IEEE Transactions on Image Processing, vol. 17, no. 9, pp. 1540–1554, 2008.

[42] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” arXiv preprint arXiv:1010.5141, 2010.

[43] W. P. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, “Realistic CT simulation using the 4D XCAT phantom,” Medical physics, vol. 35, no. 8, pp. 3800–3808, 2008.

[44] J. Fessler, “Michigan image reconstruction toolbox (MIRT),” https://web.eecs.umich.edu/~fessler/code/, 2010–2016.