



Coded aperture optimization in compressive X-ray tomography: a gradient descent approach

ANGELA P. CUADROS AND GONZALO R. ARCE*

Department of Electrical and Computer Engineering, University of Delaware Newark, DE 19716, USA

*arce@udel.edu

Abstract: Coded aperture X-ray computed tomography (CT) has the potential to revolutionize X-ray tomography systems in medical imaging and air and rail transit security - both areas of global importance. It allows either a reduced set of measurements in X-ray CT without degradation in image reconstruction, or measure multiplexed X-rays to simplify the sensing geometry. Measurement reduction is of particular interest in medical imaging to reduce radiation, and airport security often imposes practical constraints leading to limited angle geometries. Coded aperture compressive X-ray CT places a coded aperture pattern in front of the X-ray source in order to obtain patterned projections onto a detector. Compressive sensing (CS) reconstruction algorithms are then used to recover the image. To date, the coded illumination patterns used in conventional CT systems have been random. This paper addresses the code optimization problem for general tomography imaging based on the point spread function (PSF) of the system, which is used as a measure of the sensing matrix quality which connects to the restricted isometry property (RIP) and coherence of the sensing matrix. The methods presented are general, simple to use, and can be easily extended to other imaging systems. Simulations are presented where the peak signal to noise ratios (PSNR) of the reconstructed images using optimized coded apertures exhibit significant gain over those attained by random coded apertures. Additionally, results using real X-ray tomography projections are presented.

© 2017 Optical Society of America

OCIS codes: (110.7440) X-ray imaging; (340.0340) X-ray optics; (340.7430) X-ray coded apertures; (170.1630) Coded aperture imaging; (110.6955) Tomographic imaging.

References and links

1. T. M. Buzug, *Computed tomography: from photon statistics to modern cone-beam CT* (Springer, 2008).
2. R. Rangayyan, A. P. Dhawan, and R. Gordon, "Algorithms for limited-view computed-tomography: an annotated bibliography and a challenge," *Appl. Opt.* **24**, 4000–4012 (1985).
3. K. Härdt, J. L. A. J. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, "Sparse tomography," *Comput. Meth. Sci. Eng. SIAM* **35**, B644–B665 (2013).
4. I. Reiser and S. Glick, *Tomosynthesis Imaging* (Taylor and Francis, 2014).
5. D. J. Brady, A. Mrozack, K. MacCabe, and P. Llull, "Compressive tomography," *Adv. Opt. Photon.* **7**, 756–813 (2015).
6. W. E. Smith, R. G. Paxman, and H. H. Barrett, "Image reconstruction from coded data: I. Reconstruction algorithms and experimental results," *J. Opt. Soc. Am. A* **2**, 491–500 (1985).
7. K. Choi and D. J. Brady, "Coded aperture computed tomography," *Proc. SPIE* **7468**, 74680B (2009).
8. Y. Kaganovsky, D. Li, A. Holmgren, H. Jeon, K. MacCabe, D. Politte, J. O'Sullivan, L. Carin, and D. J. Brady, "Compressed sampling strategies for tomography," *J. Opt. Soc. Am. A* **31**, 1369–1394 (2014).
9. A. P. Cuadros, C. Peitsch, H. Arguello, and G. R. Arce, "Coded aperture optimization for compressive X-ray tomosynthesis," *Opt. Express* **23**, 32788–32802 (2015).
10. A. P. Cuadros, K. Wang, C. Peitsch, H. Arguello, and G. R. Arce, "Coded aperture design for compressive X-ray tomosynthesis," in *Imaging and Applied Optics 2015*, OSA Technical Digest, Optical Society of America, paper CW2F.2.
11. E. Mojica, S. Pertuz and H. Arguello, "High-resolution coded-aperture design for compressive X-ray tomography using low resolution detectors," *Opt. Commun.* (posted 30 June 2017, in press).
12. A. Parada-Mayorga and G. R. Arce, "Colored coded aperture design in compressive spectral imaging via minimum coherence," *IEEE Trans. Comput. Imag.* **3**(2), 202–216 (2017).
13. E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008).
14. M. Lustig, D. L. Donoho and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.* **58**, 1182–1195 (2007).

15. G. R. Arce, A. Ramirez, H. Rueda, H. Arguello and C. Correa, "Compressive Spectral Imaging," in Wiley Encyclopedia of Electrical and Electronics Engineering, (John Wiley & Sons, Inc., 2016).
16. A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging* (Society for Industrial and Applied Mathematics, 2001).
17. J. S. Jorgensen, E. Y. Sidky and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-Ray CT," *IEEE Trans. Med Imag* **32**(2), 460–473 (2013).
18. F. Natterer, *The mathematics of computerized tomography* (Vieweg Teubner Verlag, 1986).
19. E. Y. Sidky, C. M. Kao, and X. H. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. Technol.* **14**, 119–139 (2006).
20. W. Hou and C. Zhang, "Analysis of compressed sensing based CT reconstruction with low radiation," in Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), (2014), pp. 291–296.
21. A. Parada-Mayorga, A. P. Cuadros, G. R. Arce, "Coded aperture design for compressive X-ray tomosynthesis via coherence analysis," in Proceedings of IEEE International Symposium on Biomedical Imaging (2017).
22. G. R. Arce, D. J. Brady, L. Carin, H. Arguello and D. S. Kittle, "Compressive coded aperture spectral imaging: An Introduction," in *IEEE Signal Process. Mag* **31**(1), 105–115 (2014).
23. K. Hänninen, A. Harhanen, A. Kallonen, A. Kuusankangas, E. Niemi, and S. Siltanen, "Tomographic X-ray data of a walnut," <http://arxiv.org/abs/1502.04064>.
24. H. Rauhut, "Compressive sensing and structured random matrices," *Radon Series Comp. Appl. Math* **XX**, 1–94 (2011).
25. M. A. Davenport, M. F. Duarte, Y. C. Eldar and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*, (Cambridge University, 2011).
26. M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Sig. Process.* **55**(12), 5695–5702 (2007).
27. J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process* **18**(7), 1395–1408 (2009).
28. X. Ma and G. R. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," *J. Opt. Soc. Am. A* **25**, 2960–2970 (2008).
29. X. Ma and G. R. Arce, "Computational lithography," (John Wiley & Sons, Inc, 2010).
30. W. V. Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Express* **24**, 25129–25147 (2016).
31. Y. Mejia and H. Arguello, "Filtered gradient reconstruction algorithm for compressive spectral imaging," *Optical Engineering*, **56**(4), 041306 (2016).
32. T. A. Bubba, A. Hauptmann, S. Huotari, J. Rimpelainen, and S. Siltanen, "Tomographic x-ray data of a lotus root filled with attenuating objects," <https://arxiv.org/abs/1609.07299>.

1. Introduction

X-ray tomography has become essential in medical imaging, airport security, and industrial testing. The filtered back-projection algorithm (FBP) is the algorithm of choice for image reconstruction given its speed. However, irregular geometries of scanners and missing projection data produce artifacts and noise that make the reconstructions obtained with the FBP inadequate for diagnosis [1]. Furthermore, the numerous projections needed to obtain a high-quality reconstruction in medical X-ray computed tomography (CT) lead to high levels of radiation dose; thus, strategies to reduce the number of measurements are critical. Image estimation from incomplete data has been extensively researched since the introduction of CT [2]. Sparsity-promoting and total variation regularization algorithms have been recently used to alleviate the ill-posed inverse problem, obtaining better image reconstructions [3]. Most strategies, however, rely on optimizing the hardware settings by lowering the number of angles at which projections are taken [4]. These approaches, however, fall short when the number of measurements is further reduced leading to artifacts in the reconstructions. Brady et al. provide a study on how structured illumination can overcome these limitations while potentially reducing radiation exposure, decreasing the required number of detectors, and the acquisition time [5]. Particularly, the latter is achieved when using coded apertures in conjunction with multiplexed illumination, that is, the simultaneous illumination of the object using multiple sources. The acquisition time is reduced given that projections from more than one source may be acquired in a single step. Tomographic imaging using structured illumination was first introduced by Smith et al. in the 1980's [6]. How-

ever, it was not further studied until the introduction of compressive sensing (CS) algorithms, which provided a strong mathematical framework to solve the resulting ill-posed problem. For instance, K. Choi et al. introduced coded aperture X-ray tomosynthesis, which outperforms sparse regularization due to the direct acquisition of compressive measurements [7]. Similarly, in [8], Kaganovsky et al. introduced subsampling strategies for transmission tomography in medical CT scanner geometries. These systems require coded apertures placed either, in front of the X-ray source to create structured X-ray bundles that interrogate the object, or at the detector to block the X-ray projections in some fashion. Interestingly, it was shown in [8] that randomly subsampling the detectors in each view angle presents the reconstructions with the lowest error. However, such strategy for coded aperture design relied on random coded apertures, which is suboptimal, as it does not exploit the structure of the sensing matrix of the system. Cuadros et al. proposed an optimization approach to design the structure of the coded apertures in a multiple source compressive X-ray tomosynthesis imaging system based on a uniform energy criteria on the voxels and detector elements, so that the object was uniformly sensed and the elements of the detector plane uniformly sensed the information [9, 10]. This approach for code optimization, however, cannot be applied directly to compressive X-ray CT in circular geometries since the characteristics of the underlying sensing matrix are radically different. Furthermore, the mathematical analysis of uniform sensing with respect to the restricted isometry property (RIP) and the coherence of the sensing matrix was not established. Besides the coded aperture optimization work for tomosynthesis in [9, 10], there has been only one attempt to design coded apertures for conventional X-ray CT systems. This approach introduced by Mojica et al. in [11] aims to optimize the coded apertures for a multiple shot CT system, where the coded apertures are modified multiple times per each view angle. Their formulation is based on reducing the mutual coherence of a high-resolution CT system matrix whose dimensions are constrained to be squared and invertible, which in most practical CT systems is not realistic [1]. Furthermore, their analysis aims at reducing the coherence of the CT system matrix without taking into account the representation basis of the object of interest, thus the coherence being minimized is not that of the sensing matrix used in the image reconstruction.

The contribution of this paper is the introduction of a rigorous coded aperture optimization framework for compressive X-ray CT. The proposed coded aperture design aims to minimize the coherence of the sensing matrix, as this parameter is often used as a metric to describe the effectiveness of compressed sensing projections [12, 13]. Furthermore, the coherence of the sensing matrix accounts for the sparse basis representation of the object. The design relies on the steepest descent algorithm in order to achieve incoherence for successful CT reconstruction from incomplete projection data. The cost function is based on the point spread function (PSF) of the system, introduced by Lustig et al. [14] as a measure of the sensing matrix quality based on the RIP and mutual coherence of the sensing matrix. Additionally, two regularization functions are introduced to obtain binary coded apertures and control the transmittance of the system while sensing all the voxels of the object uniformly. Given the set of coded measurements, CS reconstruction algorithms are used to recover the image of interest. The methods presented in this paper for compressive X-ray CT can be extended to other scanning geometries such as, projection radiography, cone-beam CT, and helical CT. Additionally, by adapting the cost function to the corresponding system matrix, the method can be extended to other imaging systems based on the X-ray transform. Such as, coded aperture snapshot spectral imaging (CASSI) [15], compressive X-ray tomosynthesis and spectral computed tomography (SCT).

This paper is organized as follows. In Section 2, the mathematical model of the coded aperture compressive X-ray CT system is proposed. Section 3 introduces the gradient descent optimization algorithm proposed for the design of the coded apertures as well as the regularization constraints proper of the system. In Section 4, a simulation study for fan beam compressive X-ray CT is presented to evaluate the performance of the optimized codes obtained using the proposed

approach and compare them to the best random subsampling strategy introduced in [8]. The reconstructed images obtained from optimized coded aperture projections present higher peak-signal-to-noise ratio (PSNR) than the reconstructions obtained using random coded apertures. Additionally, reconstructions using real X-ray projection data are shown for both optimized and random coded apertures. The normalized absolute error of these reconstructions with respect to the reconstruction obtained when using the complete set of projections (no coded apertures) reveals that the proposed optimization framework leads to fewer artifacts in the reconstructed image. Finally, Section 5 presents the conclusions and future work.

2. Forward projection model

The X-ray transmission imaging model is given by the Beer-Lambert law for monochromatic X-rays $I = I_0 \cdot e^{-\int_0^\infty \mu(\ell) d\ell}$, where I_0 is the intensity of a particular X-ray originated from the X-ray source passing through the object, I is the measured intensity in the detector, and $\mu(\ell)$ is the linear attenuation coefficient at location ℓ [16]. If such X-ray source is located at position $\vec{s}$ and illuminates an object in direction $\vec{\omega}$, the data function for the imaging model is given by: $y(\vec{s}, \vec{\omega}) = -\ln(I/I_0)$. Thus, the Beer-Lambert law can be rewritten as $y(\vec{s}, \vec{\omega}) = \int_0^\infty f(\vec{s} + \ell\vec{\omega}) d\ell$, where f corresponds to the two-dimensional linear attenuation coefficient map. This continuous-to-continuous imaging model is known as the X-ray transform of the object. For a fan-beam X-ray CT system, the source location $\vec{s}$ varies continuously in a circular trajectory around the object, and the ray direction $\vec{\omega}$ varies at each $\vec{s}$ in the plane of the source trajectory [17]. In real systems, only a discrete number of measurements are available. Furthermore, in order to use iterative image reconstruction algorithms to solve the inverse problem, the imaging model needs to be discretized. The discrete-to-discrete formulation of the problem leads to a finite linear system of equations of the form $\mathbf{y} = \mathbf{H}\mathbf{f}$, where $\mathbf{f} \in \mathbb{R}^{N^2}$ is a vectorized form of the attenuation coefficients of the $N \times N$ object to be reconstructed and each of the elements in $\mathbf{H} \in \mathbb{R}^{MP \times N^2}$ represents the intersection length of each ray with every pixel inside the image, with M the number of detector elements in the linear array for each of the P view angles. Specifically, the row index of the matrix $\mathbf{H}$ corresponds to a line integral and the column index of the matrix $\mathbf{H}$ corresponds to an image pixel [18]. The number of X-ray projections needed to reconstruct an image is determined by the Shannon-Nyquist sampling theorem [17, 18]. However, iterative reconstruction techniques may be used in conjunction with compressive sensing to reconstruct the image from fewer samples [3, 17, 19]. This approach uses ℓ_1 regularization to recover the object after subsampling the number of view angles or detectors. Although reducing the number of projection angles has been studied due to its straightforward implementation, reducing the number of detectors produces higher quality reconstructions for equivalent amounts of radiation as shown in [8, 20]. Coded aperture compressive X-ray CT was proposed by Kaganovsky et al. in [8] as an implementation of detector subsampling. It places coded apertures in front of the fan-beam source to modulate its energy producing a particular coded projection onto the detector strip. The coded apertures have the same number of elements as the detector array and the coded aperture pitch is fixed to obtain one to one correspondence with the detector elements. Different coded aperture patterns are used for each view angle position s_p with $p = 1, \dots, P$ as shown in Fig. 1(a), which depicts the coded aperture compressive X-ray CT system for a fan beam architecture.

Each row of the CT system matrix, $\mathbf{H}$, indexes a particular detector element at a given view angle, from the many detector arrays at all view angles in the system. Thus, as shown in Fig. 1(a) when a coded aperture is placed in front of the source, the coded aperture elements select the rows of the matrix $\mathbf{H}$ associated with the detector elements that correspond to the unblocking coded aperture elements. Mathematically, the coded aperture matrix $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{MP}]$ is defined as a binary matrix, where each D -long column vector $\mathbf{c}_i$ with $i = 1, \dots, MP$, is associated with a particular detector element. If the coded aperture element associated with

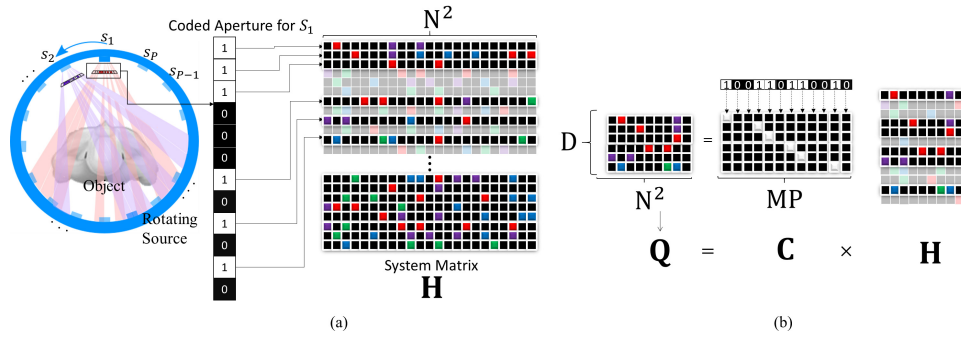


Fig. 1. (a) Coded aperture compressive fan beam X-ray CT system. Different coded apertures are placed in front of the rotating source for each location s_p . Each row in the system matrix $\mathbf{H}$ describes the sensing for a particular detector element. The unblocked coded aperture elements select the rows of the matrix $\mathbf{H}$ associated with the detector elements that correspond to the unblocked coded aperture elements. (b) Sensing matrix $\mathbf{Q} \in \mathbb{R}^{D \times N^2}$, of a system with $N^2 = 9$, $P = 4$, $D = 6$ and $M = 3$. $\mathbf{Q}$ is formed by the rows selected by the unblocked elements of the coded apertures. Each column of the coded aperture matrix $\mathbf{C}$ is associated with a detector element. If the coded aperture element associated with the i^{th} detector element is 0 then $\mathbf{c}_i = \mathbf{0}$. Otherwise, the vector $\mathbf{c}_i$ is a natural basis in $\mathbb{R}^D$.

the i^{th} detector element is 0 (blocking element), the corresponding vector $\mathbf{c}_i$ is a zero vector. Otherwise $\mathbf{c}_i$ is a natural basis in $\mathbb{R}^D$. Measurements are not multiplexed in the detector array; thus, the vectors composing the matrix $\mathbf{C}$ are disjoint. The resulting system matrix $\mathbf{Q}$ is formed by the rows of the matrix $\mathbf{H}$ that correspond to the detectors selected by the non-zero entries of matrix $\mathbf{C}$, that is $\mathbf{Q} = \mathbf{CH} \in \mathbb{R}^{D \times N^2}$, where D is the number of unblocked X-rays. The construction of the matrix $\mathbf{C}$ from a set of coded apertures is shown in Fig. 1(b). Additionally, the sparse structure of the sensing matrix $\mathbf{Q}$ and the system matrix $\mathbf{H}$ for a real CT system using random coded apertures, with transmittance 12.5 %, is depicted in Fig. 2(a). The discretized set of line integrals $\mathbf{y} \in \mathbb{R}^D$, referred to as the sinogram is given by $\mathbf{y} = \mathbf{Q}\mathbf{f}$. Each row of the sinogram represents the set of measurements taken with a particular detector at a particular view angle, Fig. 2(b) shows the spatially coded sinogram obtained by sampling 50% of the data required by the Shannon-Nyquist theorem for a 64×64 Shepp-Logan phantom. The coded aperture pattern placed in front of the source at the position s_p will be reflected in the p^{th} row of the sinogram. Figure 2(b) depicts a zoom of a particular row of the coded sinogram. The black elements denote the blocking elements of the coded aperture for that particular view angle, while the other detectors reflect the information of the corresponding spatial position of the un-coded sinogram depicted in Fig. 2(c). The coded aperture optimization problem reduces to finding the best pattern of blocked elements in the sinogram so that the sensing matrix has low coherence.

In order to reconstruct the object $\mathbf{f}$ from the coded sinogram, the ill-conditioned problem can be solved using CS algorithms. In general, the function $\mathbf{f}$ can be recovered provided the function $\mathbf{f}$ is sufficiently sparse in some basis Ψ , and such basis is incoherent with the sensing matrix [13]. Let $\mathbf{f}$ be represented by: $\mathbf{f} = \Psi\mathbf{z} \in \mathbb{R}^{N^2}$, where $\mathbf{z}$ is the sparse coefficient representation of the object, and Ψ is the basis representation. The sensing function can be rewritten as: $\mathbf{y} = \mathbf{Q}\Psi\mathbf{z} = \mathbf{A}\mathbf{z}$, where $\mathbf{A} = \mathbf{Q}\Psi \in \mathbb{R}^{D \times N^2}$ is the sensing matrix. The sparse representation of the object $\mathbf{f}$ can be obtained by solving the following nonlinear optimization problem [13]

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1, \quad (1)$$

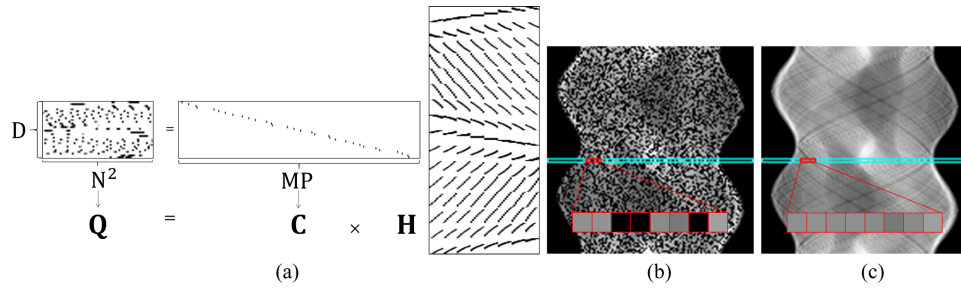


Fig. 2. (a) Sparse structures of the system matrix $\mathbf{Q}$, the coded aperture matrix $\mathbf{C}$, and the complete CT system matrix $\mathbf{H}$ for a fan beam coded aperture compressive X-ray CT system with $P = 16$ view angles, $M = 16$ detectors per view angle and a $N \times N = 8 \times 8$ image. The coded apertures have 12.5% transmittance, that is $D = 32$. Non-zero entries are depicted as black pixels for visualization purposes. (b) Under-sampled sinogram of 64×64 Shepp-Logan Phantom with $P = 128$, $M = 128$ detectors per view angle and 50% under-sampling. A zoom of a particular row of the coded sinogram is depicted. The black elements denote the blocking elements of the coded aperture for that particular view angle, while the other detectors reflect the information of the corresponding spatial position of the complete uncoded sinogram depicted in (c).

where λ is a regularization constant and $\|\cdot\|_1$ and $\|\cdot\|_2$ correspond to the ℓ_1 and ℓ_2 norms, respectively.

3. Coded aperture optimization

A possible approach to arranging the blocking and unblocking elements of the coded apertures is to distribute them randomly as proposed in initial models of coded aperture compressive X-ray CT. Kaganovsky et al. performed an extensive study regarding random subsampling of both the view angles and the detectors in [8]. Four subsampling strategies were considered: (i) Uniform view (UV) subsampling, in which view angles were selected uniformly and the complete set of detectors was used; (ii) Random view (RV) subsampling, in which the view angles were selected randomly and the complete set of detectors was used; (iii) Uniform detector (UD) subsampling, in which all the view angles were used while the set of detectors in each view angle was uniformly sampled; and (iv) Random detector (RD) subsampling, in which all the view angles were used while the detectors were randomly selected for each view. RD subsampling was shown to yield the best performance. The best results in [8] pointed to the random detector (RD) strategy, where the subsampling is chosen at random. The sensing matrices of compressive X-ray CT, however, are sparse and highly structured as depicted in Fig. 2(a). As such, random coded apertures are not optimal [9, 10, 21]. Even though coded aperture design based on the RIP and the coherence of the sensing matrix has been proposed in [12, 22] for other imaging systems such as the CASSI and the colored CASSI; these results do not extend directly to CT given that the sensing matrices of the systems are different. While the system matrix in the CASSI system has well-structured diagonal elements [22], CT matrices do not have a regular structure as shown in Fig. 2(a). To illustrate the advantages of optimized coded apertures over the UD and UV sampling strategies proposed in [8], Fig. 3 depicts the reconstruction of a 128×128 Walnut phantom Fig. 3(a), made available by Hämäläinen et al. in [23], using: (b) Optimized coded apertures obtained using the proposed algorithm, (c) Uniform subsampling of line integrals (UD strategy [8]), and (d) Uniform subsampling of view angles (UV strategy [8]). In all scenarios, the number of line integrals is fixed to $D = 16429$ for a fair comparison. Note the PSNR of the reconstruction obtained when using optimized coded apertures Fig. 3(b) is more than 20 decibels (dBs) higher than the PSNR

obtained using a uniformly distributed set of line integrals or a uniformly distributed set of view angles, Figs. 3(c) and 3(d) respectively. As it can be seen in the highlighted region of the original image and the reconstructions in Fig. 3, the reconstructions obtained by uniformly subsampling the measurements present numerous artifacts compared to using optimized coded apertures. This example illustrates that structured illumination leads to a better performance when reconstructing the object from highly under-sampled data.

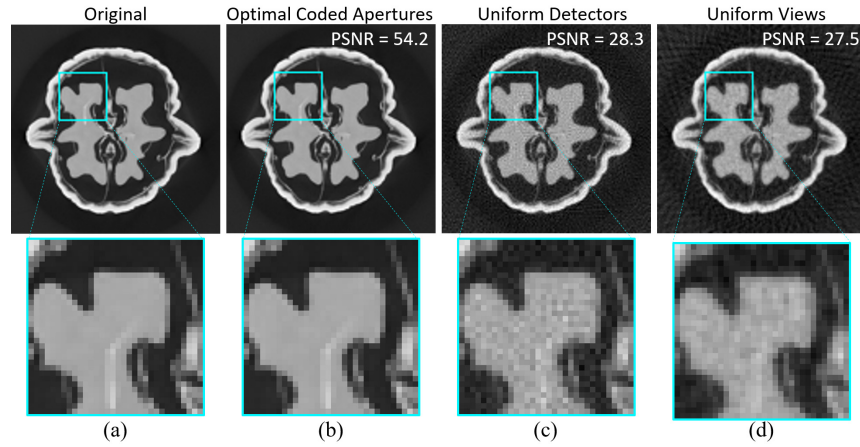


Fig. 3. (a) 128×128 Walnut phantom [23] and reconstructions for different subsampling scenarios using: (b) Optimized coded apertures obtained using the proposed algorithm, and (c) UD, and (d) UV strategies [8]. Measurements were simulated for $N = 128$, $M = P = 256$ and $D = 16429$. A zoom of the highlighted region is depicted to emphasize the error in the reconstructions. Figure (b) was reconstructed using the gradient projection for sparse reconstruction (GPSR) algorithm, Figs. (c) and (d) were reconstructed using a TV regularization algorithm.

The reconstruction of sparse or compressible signals from a very limited number of measurements relies on properties of the sensing matrix such as the RIP. The restricted isometry constant δ_s of the system matrix with normalized columns, $\tilde{\mathbf{A}} \in \mathbb{R}^{D \times N^2}$, is defined as the smallest δ_s such that

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leq \|\tilde{\mathbf{A}}\mathbf{x}\|_2^2 \leq (1 + \delta_s) \|\mathbf{x}\|_2^2, \quad (2)$$

for all s -sparse $\mathbf{x} \in \mathbb{C}^{N^2}$. The matrix $\tilde{\mathbf{A}}$ is said to satisfy the RIP if δ_s is small for reasonably large s . As shown in [24], it holds that

$$\delta_s = \max_{S \subset [N^2], |S| \leq s} \|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{I}\|_2 \quad (3)$$

where $[N^2] := 1, 2, \dots, N^2$ and $|S|$ is the cardinality of S . $\tilde{\mathbf{A}}_S$ are the column sub-matrices of $\tilde{\mathbf{A}}$ consisting of the columns indexed by S and $\tilde{\mathbf{A}}_S^*$ corresponds to the conjugate transpose of $\tilde{\mathbf{A}}_S$. The RIP, however, is in general difficult to calculate [25]. Alternatively, instead of calculating the isometry property constant directly, Hou et al. showed in [20] that δ_s is bounded by the Frobenius norm of the difference between the PSF of the system, $\tilde{\mathbf{A}}^* \tilde{\mathbf{A}}$, and the identity matrix, that is

$$\delta_s \leq \|\tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - \mathbf{I}\|_F. \quad (4)$$

Thus, to increase the reconstruction quality, the quantity $\Gamma = \|\tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - \mathbf{I}\|_F$ needs to be minimized. In addition to the RIP, the mutual coherence also provides a measure of the quality of the

system matrix. If the system matrix is designed to have minimal coherence, then it is possible to guarantee uniqueness of the solution [26]. The mutual coherence of the matrix $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_{N^2}]$ is defined as $\mu(\mathbf{A}) = \max_{i \neq j, 1 \leq i, j \leq N^2} \left\{ \frac{\mathbf{a}_i^T \mathbf{a}_j}{\|\mathbf{a}_i\|_{\ell_2} \|\mathbf{a}_j\|_{\ell_2}} \right\}$. The Gram matrix $\mathbf{G} = \tilde{\mathbf{A}}^* \tilde{\mathbf{A}}$, where $\tilde{\mathbf{A}}$ is the equivalent sensing matrix with all its columns normalized, is an alternative way of looking at the mutual coherence. Moreover, as proposed by Elad in [26] and Duarte-Carvajalino et al. in [27], coherence minimization can be performed by making any subset of columns in $\mathbf{A}$ as orthogonal as possible, that is, making $\mathbf{G}$ as close as possible to the identity matrix. This is equivalent to minimizing $\|\tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - \mathbf{I}\|_F$, which is the same result obtained in Eq. (4). Thus, the coded aperture optimization is formulated as the search of $\mathbf{C}$, over the $D \times MP$ binary space such that

$$\hat{\mathbf{C}} = \underset{\mathbf{C}}{\operatorname{argmin}} \|\tilde{\mathbf{A}}^* \tilde{\mathbf{A}} - \mathbf{I}\|_F = \underset{\mathbf{C}}{\operatorname{argmin}} \Gamma. \quad (5)$$

In order to further expand (5), recall $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{MP}]$, where $\mathbf{c}_i$ are D -long column vectors and $\mathbf{W} = \mathbf{H}\Psi = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_{MP}^T]^T$, where each $\mathbf{w}_i$ is of dimensions $1 \times N^2$. Then the sensing matrix $\mathbf{A}$ can be represented as $\mathbf{A} = \mathbf{C}\mathbf{W} = \sum_{i=1}^{MP} \mathbf{c}_i \mathbf{w}_i$, and the element of $\mathbf{A}$ in the (m, n) position can be written as $[\mathbf{A}]_{m,n} = \sum_{i=1}^{MP} \mathbf{c}_i(m) \mathbf{w}_i(n)$. Therefore, the inner product between the m^{th} and n^{th} columns after normalization of $\mathbf{A}$ is given by:

$$[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n} = \frac{\sum_{i,j}^{MP} \langle \mathbf{c}_i, \mathbf{c}_j \rangle \mathbf{w}_i(m) \mathbf{w}_j(n)}{\left(\sum_{i,j}^{MP} \langle \mathbf{c}_i, \mathbf{c}_j \rangle \mathbf{w}_i(m) \mathbf{w}_j(m) \right)^{1/2} \left(\sum_{i,j}^{MP} \langle \mathbf{c}_i, \mathbf{c}_j \rangle \mathbf{w}_i(n) \mathbf{w}_j(n) \right)^{1/2}}. \quad (6)$$

As described in Section 2, matrix $\mathbf{C}$ is composed by a combination of zero vectors and a set of disjoint vectors that are natural bases in $\mathbb{R}^D$. As such, the inner product $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = 0 \ \forall i \neq j$ and $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = 1 \ \forall i = j$ given that $\mathbf{c}_i$ is not a zero vector otherwise $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = 0$. Therefore, (6) can be simplified as:

$$[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n} = \frac{\sum_{i=1}^{MP} b_i \mathbf{w}_i(m) \mathbf{w}_i(n)}{\left(\sum_{i=1}^{MP} b_i \mathbf{w}_i(m) \mathbf{w}_i(m) \right)^{1/2} \left(\sum_{i=1}^{MP} b_i \mathbf{w}_i(n) \mathbf{w}_i(n) \right)^{1/2}}, \quad (7)$$

where $b_i = \langle \mathbf{c}_i, \mathbf{c}_i \rangle$. Additionally, the parameter Γ is simplified as:

$$\Gamma = \sum_{m=1}^{N^2} \sum_{n=1}^{N^2} \left| \frac{\sum_{i=1}^{MP} b_i \mathbf{w}_i(m) \mathbf{w}_i(n)}{\left(\sum_{i=1}^{MP} b_i \mathbf{w}_i(m) \mathbf{w}_i(m) \right)^{1/2} \left(\sum_{i=1}^{MP} b_i \mathbf{w}_i(n) \mathbf{w}_i(n) \right)^{1/2}} - \mathbf{I}_{m,n} \right|^2. \quad (8)$$

Substituting (8) in (5), the reparametrized cost function can be written in terms of the vector $\mathbf{b} = [b_1, \dots, b_{MP}]^T$ as:

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \Gamma = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m=1}^{N^2} \sum_{n=1}^{N^2} |[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n} - \mathbf{I}_{m,n}|^2 \quad (9)$$

Furthermore, as shown in the Appendix A, the coded aperture optimization cost function in (9) can be simplified as follows

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m < n} \left(\frac{d(m, n)}{(d(m, m))^{1/2} (d(n, n))^{1/2}} \right)^2 = \underset{\mathbf{b}}{\operatorname{argmin}} F(\mathbf{b}), \quad (10)$$

where $d(m, n) = \sum_{i=1}^{MP} b_i R_i^{m,n}$ and where $R_i^{m,n} = \mathbf{w}_i(m) \mathbf{w}_i(n)$. Note the optimization problem in (10) is a combinatorial optimization since b_i can only take values 0 or 1 which prevents the

use of gradient information to explore the search space. To overcome this problem, the parameter values are relaxed to lie in the range $[0, 1]$. This is achieved by imposing the inequality constraint $0 < b_i < 1$ on the optimization problem given in (10). The resulting bound-constrained optimization problem can be further reduced to an unconstrained optimization problem using the following parametric transformation $b_i = \frac{\cos \theta_i + 1}{2}$, where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{MP}]^T$ is the unconstrained parameter vector. Such transformation was used by Ma and Arce to reduce the inverse lithography problem to an unconstrained optimization in [28, 29]. The re-parameterized cost function can be formulated in terms of the parameter vector $\boldsymbol{\theta}$ as follows

$$F(\boldsymbol{\theta}) = \sum_{m < n} \left(\frac{\sum_{i=1}^{MP} \frac{\cos \theta_i + 1}{2} R_i^{m,n}}{\left(\sum_{i=1}^{MP} \frac{\cos \theta_i + 1}{2} R_i^{m,m} \right)^{1/2} \left(\sum_{i=1}^{MP} \frac{\cos \theta_i + 1}{2} R_i^{n,n} \right)^{1/2}} \right)^2. \quad (11)$$

Let $d_\theta(m, n) = \sum_{i=1}^{MP} \frac{\cos \theta_i + 1}{2} R_i^{m,n}$, then (11) can be simplified to

$$F(\boldsymbol{\theta}) = \sum_{m < n} \left(\frac{d_\theta(m, n)}{(d_\theta(m, m))^{1/2} (d_\theta(n, n))^{1/2}} \right)^2. \quad (12)$$

Note that this cost function is not convex given the cosine terms in the numerator and denominator that modify the optimization variable $\boldsymbol{\theta}$. Thus, a random stochastic gradient-descent algorithm is used to solve the above problem. This requires the calculation of the first order derivatives of (12). As shown in Appendix B, the gradients $\nabla F(\boldsymbol{\theta})$ can be obtained as follows

$$\nabla F(\boldsymbol{\theta}) = \left(\sum_{m < n} \frac{-d_\theta(m, n) \mathbf{R}^{m,n}}{d_\theta(m, m) d_\theta(n, n)} + \frac{d_\theta(m, n)^2 \mathbf{R}^{m,m}}{2 d_\theta(m, m)^2 d_\theta(n, n)} + \frac{d_\theta(m, n)^2 \mathbf{R}^{n,n}}{2 d_\theta(m, m) d_\theta(n, n)^2} \right) \odot \sin \boldsymbol{\theta}, \quad (13)$$

where $\nabla F(\boldsymbol{\theta}) \in \mathbb{R}^{MP \times 1}$, $\odot$ is the element-by-element multiplication operator and $\mathbf{R}^{m,n} = [R_1^{m,n}, R_2^{m,n}, \dots, R_{MP}^{m,n}]^T$. The random stochastic gradient descent algorithm uses a sample noise term r drawn uniformly from the unit sphere. Assuming θ_i^k is the k^{th} iteration result, the update at the $k^{th} + 1$ iteration will be given by

$$\theta_i^{k+1} = \theta_i^k - \alpha [\nabla F(\theta_i^k) + r], \quad (14)$$

where α is the step size. The iterative optimization above, in general, leads to gray-valued coded aperture masks with pixel values between 0 and 1 which makes the coded aperture fabrication impractical. Hence, a post-processing step to obtain a binary coded aperture is needed. Using a global threshold parameter t_m , with $0 \leq t_m \leq 1$, the values of b_i can be quantized as $\beta_i = U\left(\sqrt{\frac{\cos(\theta_i) + 1}{2}} - t_m\right)$, where $i = 1, \dots, MP$ and $U(x)$ is the Heaviside step function of x , defined as $U(x) = 1$ for $x \geq 0$ and $U(x) = 0$ for $x < 0$. The coded aperture matrix $\mathbf{C}$ is constructed such that matrix $\mathbf{Q}$ is formed by the rows of matrix $\mathbf{H}$ corresponding to the detector elements for which $\beta_i = 1$.

3.1. Regularization

The proposed global thresholding approach to obtain binary values for the coded apertures is sub-optimal, and the transmittance of the coded apertures obtained by the optimization algorithm cannot be controlled. Thus, it is necessary to incorporate prior knowledge about the solution by constraining the solution space through the addition of appropriate regularization terms to the cost function. In general, the formulation can be described as follows:

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} F(\mathbf{b}) + \gamma_1 V_1(\mathbf{b}) + \gamma_2 V_2(\mathbf{b}), \quad (15)$$

where $F(\mathbf{b})$, defined in (10), is the data fidelity term and V_1 and V_2 are the regularization terms used to reduce the solution space and constrain the optimized results. γ_1 and γ_2 are user-defined parameters to weight the regularization.

3.1.1. Quadratic penalty

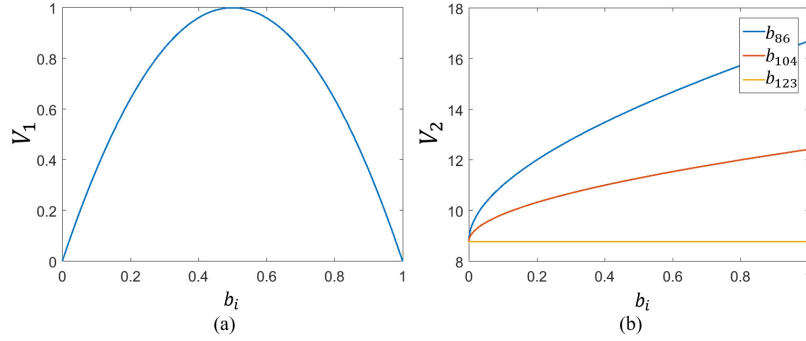


Fig. 4. (a) Quadratic penalty cost function encouraging solution in $b_i \in \{0, 1\}$ since the minimum error is obtained for binary values. (b) Transmittance and uniformity penalty cost function for $M = P = 2N = 16$ and $\mu = 2$, i.e. approximately 50 % transmittance. Note the minimum error is achieved for the desired average number of rays.

To obtain near-binary gray optimized coded apertures through the optimization process, the quadratic penalty is adopted as in [28, 29]. The quadratic penalty term is defined as $V_1(\mathbf{b}) = 4\mathbf{b}^T(\mathbf{1} - \mathbf{b})$, such that for each coded aperture element $V_1(b_i) = 1 - (2b_i - 1)^2$. The penalty is maximum for $b_i = 0.5$ and minimum for 0 and 1 as shown in Fig. 4(a), that is the penalty increases as b_i moves away from the binary region. The gradient of $V_1(\mathbf{b})$ is given by $\nabla V_1(\mathbf{b}) = (-8\mathbf{b} + 4)$, which can be used in conjunction with (13) while carrying out the steepest-descent iterations as before.

3.1.2. Transmittance penalty

Uniform sensing of the object has been used as criteria for coded aperture design in previous works [9, 10]. Particularly, uniformity in the number of rays that sense a particular voxel across the region of interest (ROI) is desirable. To this end, a second penalty term is incorporated to obtain uniformity in the ROI, Ω_C , denoted by the circle of radius N inscribed in the $N \times N$ image while using a determined number of blocking coded aperture elements, therefore obtaining a coded aperture with a fixed transmittance. The quadratic penalty term is chosen as the variance of the sum of the length of the rays that measure each pixel, that is:

$$V_2(\mathbf{b}) = \frac{1}{N_C} \sum_{k \in \Omega_C} \left[\left(\sum_{i=1}^{MP} \sqrt{b_i} \mathbf{H}_{i,k} \right) - \mu \right]^2, \quad (16)$$

where μ is the desired mean for the ray concentration, N_C is the number of pixels in the ROI Ω_C , and $\mathbf{H}_{i,k}$ is the intersection length of the i^{th} ray with every pixel k inside Ω_C . This penalty is minimum when the number of rays that sense every pixel inside Ω_C is approximately equal to μ . Figure 4(b) shows an example of the regularization function for $M = P = 16$ and 50% transmittance. The three curves show the behavior of the regularization cost function for a particular pixel in the image and 3 different coded aperture elements $i = 86, 104, 123$. Note that the behavior of this function for each coded aperture element depends on the system matrix and the values of the other coded apertures, for example for the configuration in Fig. 4(b) a change

in the coded aperture corresponding to the detector element $i = 123$ would not result in a more uniform sensing. On the other hand, values of 1 would be sub-optimal to achieve uniformity for the coded apertures corresponding to the detector elements $i = 86$, and $i = 104$. As shown in Appendix C, the gradients $\nabla V_2(\mathbf{b})$ can be obtained as follows

$$\nabla V_2(\mathbf{b}) = \frac{1}{N_C} \left[\bar{\boldsymbol{\eta}}^T \mathbf{H}_{\Omega_C} - \mu \right] \left[\mathbf{H}_{\Omega_C}^T \right] \odot \left[\bar{\boldsymbol{\eta}}^T \right]^{-1}, \quad (17)$$

where $\bar{\boldsymbol{\eta}} = \sqrt{\mathbf{b}}$ and $\mathbf{H}_{\Omega_C}$ corresponds to a matrix formed with the columns of the matrix $\mathbf{H}$ that correspond to the pixels contained in Ω_C . This gradient can be used in conjunction with (13) and $\nabla V_1(\mathbf{b})$ while carrying out the steepest-descent iterations as before. Adding the regularization terms the cost function is adjusted as

$$J(\mathbf{b}) = F(\mathbf{b}) + \gamma_1 V_1(\mathbf{b}) + \gamma_2 V_2(\mathbf{b}). \quad (18)$$

4. Simulation results

Three scenarios are considered to analyze the performance of the optimized coded apertures. Initially, a compressive X-ray CT configuration with a flat 1D detector strip composed by $M = 64$ elements, a fan-beam X-ray source, $P = 64$ view angles, and an object of interest $\mathbf{f}$ represented by a Shepp-Logan phantom of $N \times N = 32 \times 32$ pixels are used. Each of the pixels in the coded aperture corresponds to a particular detector element as detailed in Fig. 1(a). Therefore, the coded apertures placed in front of each of the sources are also composed by $M = 64$ elements. The ASTRA Tomography Toolbox (ASTRA Tomographic Reconstruction Antwerp) [30] was used to obtain the system matrix $\mathbf{H}$ as well as the projection measurements $\mathbf{y}$ of the X-ray source. The setup is adjusted to match the full sampling scenario described by Jorgensen et al. in [17]. That is, $R_0 = 40$ cm distance from the center of rotation to the X-ray source, $T_0 = 80$ cm source-to-detector-center distance, detector length 41.3 cm and $M = P = 2N$. After obtaining the system matrix using random and optimal coded apertures and the projections for the corresponding systems, the linear attenuation coefficient image is reconstructed by solving the minimization problem in (1) using a modified version of the gradient projection for sparse reconstruction (GPSR) algorithm. The modified GPSR includes a filtering step, as proposed by Mejia et al. in [31]. The linear attenuation coefficient CT image is represented on the Haar 2D Wavelet basis.

Using the algorithm described in Section 3 optimized coded apertures are obtained for different subsampling rates adjusting the regularization constants γ_1 , γ_2 and the parameter D for each case. Note 50 % subsampling is equivalent to a 50% transmittance of the coded apertures in which case $D = 0.5(MP)$, that is, the subsampling rate corresponds to the percentage of the full set of detectors used in each case. The elements of the random coded apertures used are random realizations of Bernoulli random variables, with levels of transmittance equivalent to the corresponding optimal coded aperture transmittance. The PSNR is used to compare the reconstructions obtained since it is suitable for comparing restoration results, as it does not depend strongly on the image intensity scaling. For a scenario with an image I and a reconstruction R of size $N \times N$ it is defined as $PSNR = 10 \log_{10} \left(\frac{Max_I^2}{MSE} \right)$, where Max_I is the maximum possible pixel value of the image I and the mean squared error (MSE), $MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [I(i, j) - R(i, j)]^2$. Figure 5(a) shows the PSNR of the random and optimal coded apertures for multiple subsampling rates. For all the simulations in this paper, random coded apertures correspond to the RD subsampling strategy proposed by Kaganovsky et al. in [8]. The random selection of measurements was repeated 10 times, and the mean and standard error for this ensemble were computed and are represented in the graph by the solid line and shaded strips respectively. Note that the PSNR is higher when using optimized coded

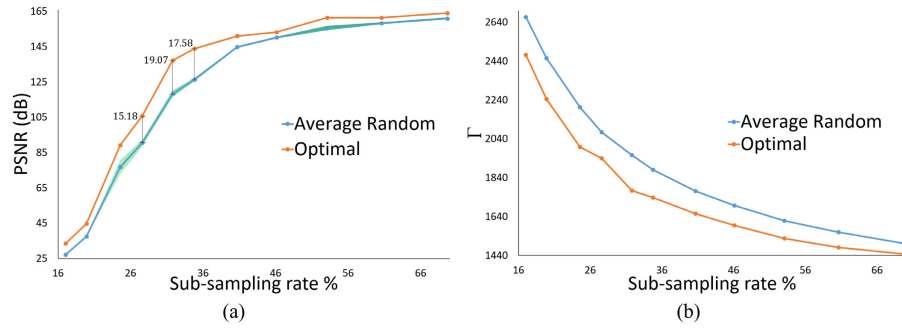


Fig. 5. (a) PSNR of the reconstructions obtained using random and optimal coded apertures with different subsampling rates. For the random coded apertures, 10 realizations were computed and the graph depicts the mean and the standard error as a solid line and shaded area respectively. (b) Γ error for optimal and random coded apertures. The subsampling rate corresponds to the percentage of the full set of detectors used in each case.

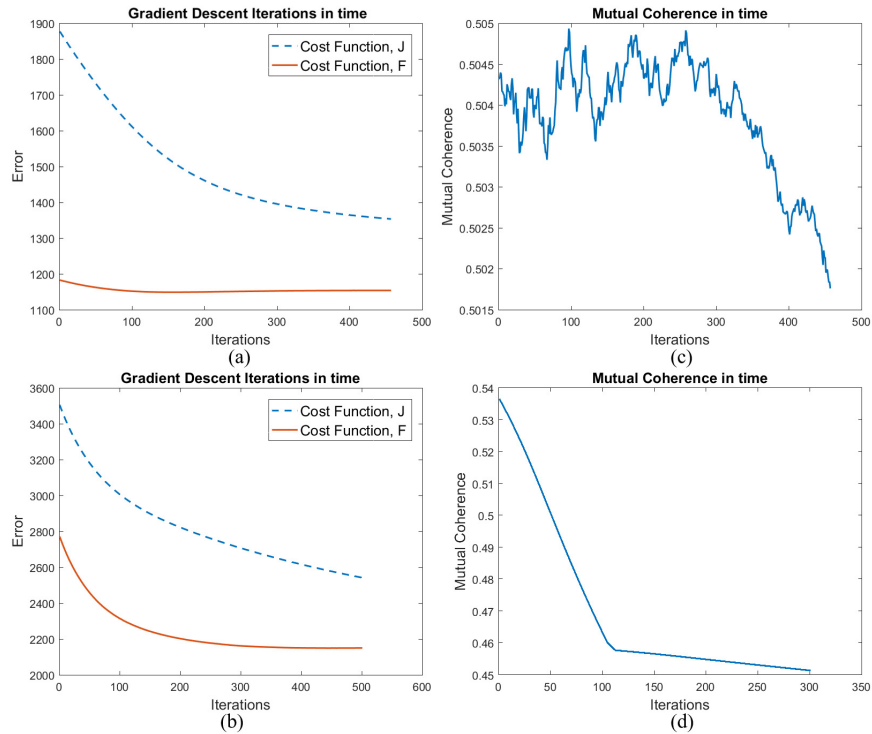


Fig. 6. Gradient descent iterations of the optimization function and regularization terms for 47.3% (a) and 20.3% (b) subsampling rates. Gradient descent iterations of the mutual coherence for 47.3% (c) and 20.3% (d) subsampling rates. The subsampling rate corresponds to the percentage of the full set of detectors used in each case. The parameters for the simulation were $M = P = 2N = 128$.

apertures for all the studied subsampling rates. Particularly, for 27.6%, 31.8%, and 34.8% subsampling rates the PSNR gain is higher than 15 dB. Furthermore, Fig. 5(b) shows that the error Γ of the system matrices obtained using optimized coded apertures is lower than the error obtained when using random coded apertures. This behavior is expected as Γ is the cost function

minimized in the proposed optimization algorithm. For all the experiments the object used was a 32×32 Shepp-Logan phantom. A second simulation scenario is considered with the same geometrical configuration for the hardware, but changing the parameters $M = P = 128$ and $N = 64$. Figures 6(a) and 6(b) show the gradient descent iterations for the error function F associated with the PSF and the complete cost function J for subsampling rates of 47.3% and 20.3% respectively. As it can be seen, the error J and the Frobenius norm of the PSF (F) are minimized over time while maintaining the coded apertures in the solution space. Furthermore, the mutual coherence of the sensing matrix is calculated in each iteration for both subsampling rates and it can be seen in Figs. 6(c) and 6(d) that the algorithm converges to a lower mutual coherence value than the initial coded aperture pattern for both 47.3% and 20.3%. For the latter subsampling rate, reconstructions using a 64×64 Walnut phantom [23] Fig. 7(a) were obtained for: (b) Optimized and (c) Random coded apertures. Figures 7(d) and 7(e) depict the normalized absolute error for the respective cases. The error images were normalized for each case by dividing all the values by the maximum error of the 2 images (reconstruction error from random and optimized coded apertures); this process was performed in order to have the same error scale for both the absolute error of the random and optimized coded aperture cases. As it

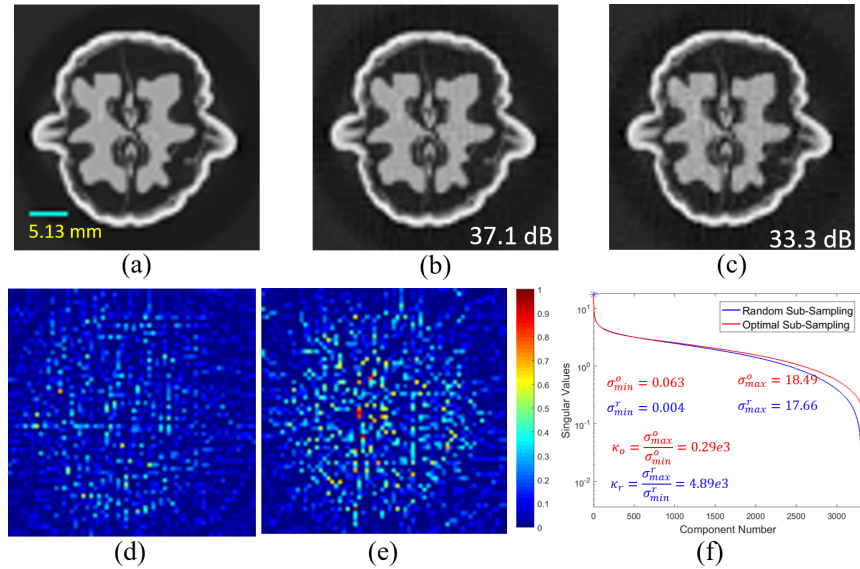


Fig. 7. (a) 64×64 Walnut phantom [23]. Reconstructions using: (b) optimized and (c) random coded apertures with subsampling rate 20.3%. Absolute error images for the (d) optimized and (e) random cases show more artifacts for random projections. (f) SVD of the CT matrix with optimized codes and random codes. The condition number of both the random κ_r and optimized κ_o cases are specified. The maximum σ_{max} and minimum σ_{min} values are specified for the random and optimal cases. Optimized projections result in a less ill-conditioned system.

is expected from the 4 dB difference between the PSNR of the reconstructions obtained using optimized and random projections, the error images have higher values and more artifacts in the latter case. The singular value decomposition (SVD) of the system matrix has been reportedly used as a tool to compare different sampling strategies [8, 9]. The SVD of the matrix $\mathbf{Q}$ for the optimized and random cases is calculated and it is depicted in Fig. 7(f). For the latter, the mean for 10 different selections is obtained. Considering there is no prior information about the object under inspection, the measurement strategy that has more singular value components lying above certain noise level is considered to outperform the others since it would capture

more orthogonal components of the object. It can be seen that the condition number (ratio of greatest singular value to the least non-zero singular value κ) of the optimal codes is better than the random coded apertures case by an order of magnitude. It can be concluded that the use of optimized codes leads to less ill-conditioned sensing matrices compared to using random codes, showing that the proposed optimization framework leads to a better conditioning of the forward operator. The maximum σ_{max} and minimum σ_{min} values are specified for the random and optimal cases and are depicted with a '*' marker in Fig. 7(f).

To illustrate the versatility of the algorithm and the motivation for using a Haar wavelet in the majority of the simulations, reconstructions using a Symlet wavelet for optimized and random coded apertures are obtained. The PSNR for the optimized case is 122.28 dB for a subsampling rate of 32.43%. While the PSNR for a subsampling rate of 32.47% using the Haar wavelet is 133.51 dB, that is an 11 dB gain. In both cases, the reconstructions obtained using optimized coded apertures outperform the reconstructions obtained using random coded apertures by approximately 8 dBs. To illustrate the difference between the reconstructions obtained from random and optimized projections in each case, Figs. 8(a) and 8(b), depict the absolute error plots for the Haar wavelet and the Symlet wavelet respectively. In both figures, the left surface plot corresponds to the normalized absolute error of the reconstruction using optimized coded apertures and the right surface plot corresponds to the normalized absolute error for the reconstructions using random coded apertures. Note in both cases, the error is higher when using random coded apertures.

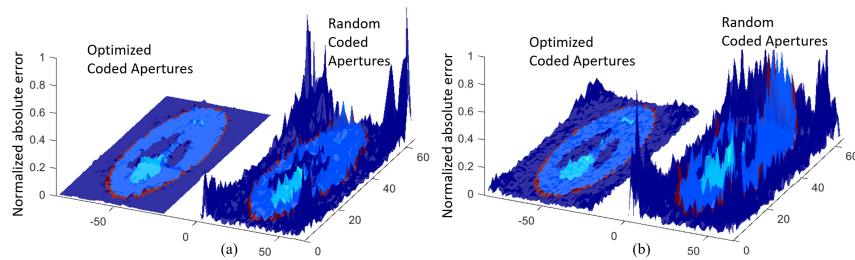


Fig. 8. Normalized absolute error plots for the reconstructions of the 64×64 Shepp-Logan phantom for subsampling rates of 32.47% using a Haar wavelet (a) and 32.4% using a Symlet wavelet (b) and $M = P = 128$.

Figure 9 depicts reconstructions for the third scenario, in which real tomographic X-ray projections of a slice of a lotus root, made available by Bubba et al. in [32], are used. The optimization of the coded apertures for this system was performed using the system matrix and the sinogram provided by the authors corresponding to $P = 120$ view angles, $M = 429$ detectors and a discretized object of 128×128 pixels ($N = 128$). Figure 9(a) depicts the reconstruction of the lotus slice from real X-ray data using Landweber iterations without using coded apertures [32], that is, $D = MP = 51480$ and Figs. 9(b) and 9(c) depict the results obtained when using coded apertures with subsampling rate 20.6%. Note that the number of projection angles does not correspond to the optimal case ($P = 2N$) which results in more artifacts for the three cases. However, as shown before, the reconstruction results obtained when using optimized coded apertures have fewer artifacts than the results obtained using random coded apertures. Note that some of the components on the left-hand side of the pictures cannot be distinguished in Fig. 9(c) but they appear in Figs. 9(a) and 9(b). Figures 9(d) and 9(e) depict the normalized absolute error of the reconstructions using random and optimized coded apertures, respectively, compared to the reconstruction using all the available data Fig. 9(a). Note the reconstruction obtained using optimal coded apertures presents a lower error for the majority of the pixels in

the image. Furthermore, the maximum error obtained for the optimized coded apertures case is 0.6 while in the random case is 1. As before, the SVD of the system matrix, as well as the condition number, were calculated for both the optimized and random coded aperture cases, the results are shown in Fig. 9(f). Note that the system matrix is less ill-conditioned when using optimized coded apertures. As mentioned in Section 3, to achieve uniform sensing, it is ideal

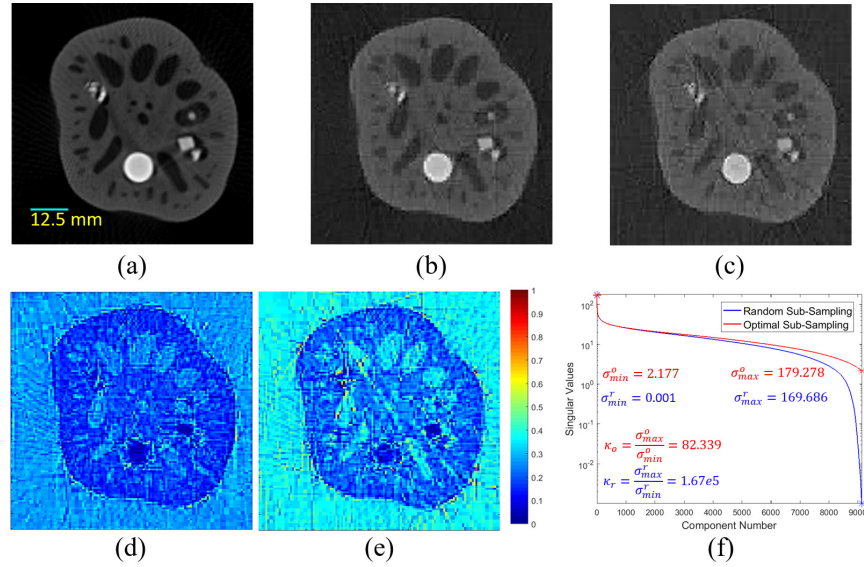


Fig. 9. Reconstructions of a slice of a lotus root from real X-ray tomographic projections, using: (a) Landweber iterations without using coded apertures [32] and (b) optimized and (c) random coded apertures with subsampling rate 20.6%. Absolute error images for the (d) optimized and (e) random cases are depicted. (f) SVD of the CT matrix with optimized codes and random codes. The condition number of both the random κ_r and optimized κ_o cases are specified. The maximum σ_{max} and minimum σ_{min} values are specified for the random and optimal cases.

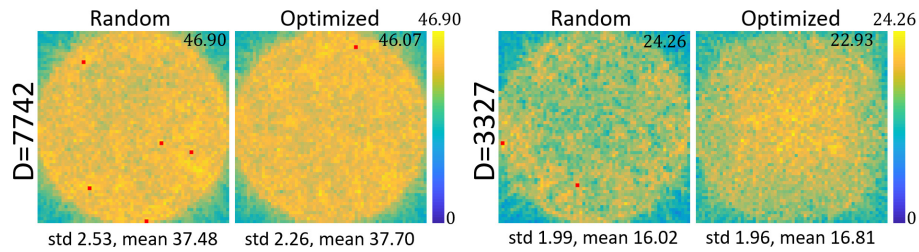


Fig. 10. Intensity graphs of the number of rays that measure certain pixel for $D = 7742$, and $D = 3327$ i.e. 47.3% and 20.3% subsampling rates respectively, with the corresponding standard deviation, mean and maximum values. Hot Spots, with radiation higher than the 95% of the maximum intensity, are located in the random case for both subsampling rates and are shown as red dots in the figure.

that the number of rays that sense a particular pixel is approximately uniform across the ROI. Figure 10 shows the intensity graphs corresponding to the number of rays that measure certain pixel for different subsampling rates using both optimized and random coded apertures. The mean and standard deviation are calculated for each case in the circle of radius $R = N$ which

denotes the ROI for fan-beam X-ray CT systems. Note that for $D = 7742$ which corresponds to $D > N^2$ and $D = 3327$ which corresponds to $D < N^2$ the intensity graphs show that the optimized case results in a more uniform sensing of the object; particularly, note that the standard deviation is lower when using optimized coded apertures. Furthermore, the maximum radiation is obtained when using random coded apertures for both subsampling rates as shown by the hot-spots in the intensity graphs. This concentration is in general not ideal. It should be noted that the regularization term V_2 could be modified to control the radiation at different points. For example, to radiate specific parts of the object in applications with multiple ROIs.

5. Conclusion

The analysis of the mutual coherence and the restricted isometry constant of the compressive X-ray CT system allows the formulation of an optimization algorithm based on the PSF of the system matrix to design the coded aperture patterns in a coded aperture compressive X-ray CT system. The performance of the proposed optimal codes has been compared to the random approach introduced in [8] using object-independent performance measures such as SVD, the mutual coherence, and the uniformity criteria. Additionally, reconstruction results for specific objects with synthetic and real data show that the coded apertures obtained through the proposed design lead to reconstructions with higher PSNR than the ones obtained using random coded apertures. This paper considers a fan-beam geometry and represents the object using Haar and Symlet wavelet bases, but the framework presented here is far more general and can be applied to other geometries and to other sparse representations of the object. Furthermore, the analysis can be extended to other imaging systems that make use of the X-ray transform, such as coded aperture snapshot spectral imaging (CASSI), compressive X-ray tomosynthesis and spectral CT, by adapting the cost function according to the corresponding system matrix. Different transmittance levels can be attained according to the constraints imposed by the system by changing a regularization constant in the proposed cost function and using the random stochastic gradient descent algorithm proposed in this work. Additionally, a regularization term can be added to the cost function to induce a more uniform sampling in multiple regions of interest. Fabrication of the coded apertures for circular tomography is under development. Calibration procedures will be used in order to mitigate the mismatching errors that might occur. Furthermore, the angular collimation produced by the implemented coded apertures can be accounted for in the system matrix $\mathbf{H}$. The optimized codes would take into account this phenomenon.

Appendix A: PSF simplification

From equation (9)

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m=1}^{N^2} \sum_{n=1}^{N^2} |[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n} - \mathbf{I}_{m,n}|^2. \quad (19)$$

Given that $\mathbf{I}$ is an identity matrix Eq. (19) can be rewritten as

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m \neq n} ([\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n} - 0)^2 + \sum_{m=n} ([\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,m} - 1)^2. \quad (20)$$

Furthermore, since the columns of $\tilde{\mathbf{A}}$ are normalized, $[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,m} = 1 \ \forall m$, and given that $[\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]$ is symmetric, Eq. (19) simplifies to

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m \neq n} ([\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n})^2 = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{m < n} ([\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}]_{m,n})^2. \quad (21)$$

Appendix B: Gradient of the cost function $F(\theta)$

From equation (12), $F(\theta) = \sum_{m < n} \left(\frac{d_\theta(m, n)}{(d_\theta(m, m))^{1/2} (d_\theta(n, n))^{1/2}} \right)^2$, deriving with respect to θ_i

$$\nabla F(\theta)_{\theta_i} = \frac{\sum_{m < n} \left[\frac{\partial d_\theta(m, n)}{\partial \theta_i} \right] 2d_\theta(m, n)d_\theta(m, m)d_\theta(n, n) - d_\theta(m, n)^2 \left[\frac{\partial d_\theta(m, m)}{\partial \theta_i} d_\theta(n, n) + \frac{\partial d_\theta(n, n)}{\partial \theta_i} d_\theta(m, m) \right]}{(d_\theta(m, m))^2 (d_\theta(n, n))^2} \quad (22)$$

Recall from Section 3, $d_\theta(m, n) = \sum_{i=1}^{MP} \frac{\cos \theta_i + 1}{2} R_i^{m, n}$, thus $\frac{\partial d_\theta(m, n)}{\partial \theta_i} = R_i^{m, n} \left(\frac{-\sin \theta_i}{2} \right)$. Let $\odot$ be the element-by-element multiplication operator, and $\underline{\mathbf{R}}^{m, n} = [R_1^{m, n}, R_2^{m, n}, \dots, R_{MP}^{m, n}]^T$. Then, substituting $\frac{\partial d_\theta(m, n)}{\partial \theta_i}$ in Eq. 22 we obtain

$$\nabla F(\theta) = \left(\sum_{m < n} \frac{-d_\theta(m, n) \underline{\mathbf{R}}^{m, n}}{d_\theta(m, m)d_\theta(n, n)} + \frac{d_\theta(m, n)^2 \underline{\mathbf{R}}^{m, m}}{2d_\theta(m, m)^2 d_\theta(n, n)} + \frac{d_\theta(m, n)^2 \underline{\mathbf{R}}^{n, n}}{2d_\theta(m, m)d_\theta(n, n)^2} \right) \odot \sin \theta. \quad (23)$$

Appendix C: Gradient of the regularization term $V_2(\mathbf{a})$

Consider the regularization cost function V_2 for $K = 2$ terms in Ω_C and $MP = 2$

$$V_2(\mathbf{b}) = \frac{1}{N_C} \sum_{k=1}^2 \left((\sqrt{b_1} \mathbf{H}_{k,1} + \sqrt{b_2} \mathbf{H}_{k,2}) - \mu \right)^2. \quad (24)$$

Deriving with respect to b_1

$$\nabla V_2(\mathbf{b})_{b_1} = \frac{1}{N_C} \left\{ 2 \left[(\sqrt{b_1} \mathbf{H}_{1,1} + \sqrt{b_2} \mathbf{H}_{1,2}) - \mu \right] \frac{1}{2\sqrt{b_1}} \mathbf{H}_{1,1} + 2 \left[(\sqrt{b_1} \mathbf{H}_{2,1} + \sqrt{b_2} \mathbf{H}_{2,2}) - \mu \right] \frac{1}{2\sqrt{b_1}} \mathbf{H}_{2,1} \right\} \quad (25)$$

Simplifying the expression in (25)

$$\nabla V_2(\mathbf{b})_{b_1} = \frac{1}{N_C} \left[\sqrt{\mathbf{b}}^T \mathbf{H}_{\Omega_C} - \mu \right] \left[\mathbf{H}_{\Omega_C}^1 \right]^T \frac{1}{\sqrt{b_1}} \quad (26)$$

Where $\mathbf{H}_{\Omega_C}^1$ corresponds to the first row of the matrix $\mathbf{H}_{\Omega_C}$, and $\mathbf{H}_{\Omega_C}$ corresponds to a matrix formed with the columns of the matrix $\mathbf{H}$ that correspond to the pixels contained in the ROI Ω_C . Generalizing the result in (26) for a vector $\mathbf{b}$ with MP entries, we obtain the following expression for the gradient of the regularization term V_2

$$\nabla V_2(\mathbf{b}) = \frac{1}{N_C} \left[\sqrt{\mathbf{b}}^T \mathbf{H}_{\Omega_C} - \mu \right] \left[\mathbf{H}_{\Omega_C} \right]^T \odot \left[\sqrt{\mathbf{b}}^T \right]^{-1} \quad (27)$$

where $\mathbf{H}_{\Omega_C}$ corresponds to a matrix formed with the columns of the matrix $\mathbf{H}$ that correspond to the pixels contained in Ω_C .

Funding

National Science Foundation (NSF) (CIF 1717578); Fulbright-Finland Foundation under the 2017 Fulbright-Nokia Distinguished Chair award.

Acknowledgments

The authors thank Jose Monsalve and Aaron Landwehr for their advice regarding the parallel implementation of the algorithm.