Plug-and-Play Priors for Model Based Reconstruction

Singanallur V. Venkatakrishnan*, Charles A. Bouman* and Brendt Wohlberg †
*School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907
†Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545

***Abstract*—Model-based reconstruction is a powerful framework for solving a variety of inverse problems in imaging. In recent years, enormous progress has been made in the problem of denoising, a special case of an inverse problem where the forward model is an identity operator. Similarly, great progress has been made in improving model-based inversion when the forward model corresponds to complex physical measurements in applications such as X-ray CT, electron-microscopy, MRI, and ultrasound, to name just a few. However, combining state-of-the-art denoising algorithms (i.e., prior models) with state-of-the-art inversion methods (i.e., forward models) has been a challenge for many reasons.**

In this paper, we propose a flexible framework that allows state-of-the-art forward models of imaging systems to be matched with state-of-the-art priors or denoising models. This framework, which we term as Plug-and-Play priors, has the advantage that it dramatically simplifies software integration, and moreover, it allows state-of-the-art denoising methods that have no known formulation as an optimization problem to be used. We demonstrate with some simple examples how Plug-and-Play priors can be used to mix and match a wide variety of existing denoising models with a tomographic forward model, thus greatly expanding the range of possible problem solutions.

I. Introduction

Model-based reconstruction is a powerful framework for solving a variety of inverse problems in imaging including denoising, deblurring, tomographic reconstruction, and MRI reconstruction. The method typically involves formulating a model for the noisy measurement system (i.e., a forward model) and a model for the image to be reconstructed (i.e., a prior model). The reconstruction is then computed by minimizing a cost function that balances a fit to these two models.

In recent years, there have been enormous advances in the solution of a particular inverse problem generally referred to as image denoising [1], [2]. Since image denoising is the simplest case of an inverse problem, the forward model being the identity operator, research in this field tends to provide a fertile environment for the creation of new prior models. In fact, a number of very novel and effective approaches have recently emerged for image denoising [1]–[7]. These new methods have demonstrated that it is possible to vastly improve on what was previously believed to be possible.

S. V. Venkatakrishnan and C. A. Bouman were supported by an AFOSR/MURI grant #FA9550-12-1-0458, by UES Inc. under the Broad Spectrum Engineered Materials contract, and by the Electronic Imaging component of the ICMD program of the Materials and Manufacturing Directorate of the Air Force Research Laboratory, Andrew Rosenberger, program manager. B.Wohlberg was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

In parallel with these efforts, researchers have been pioneering ways to create forward models for a wide array of imaging and sensing systems from medical scanners [8] to microscopes [9]. Research in this field has demonstrated that model-based inverse methods can greatly improve the quality of reconstructed images [10]. However, since this research primarily deals with the challenges of accurately modeling large and complex forward models and solving the associated optimization problems, there has been much less emphasis on the incorporation of state-of-the-art prior models. Therefore, research in model-based inversion has tended to lag behind from the perspective of advanced prior modeling; and moreover, has not fully benefited from the recent progress in denoising methods.

In fact, recent progress has been made in incorporating advanced priors into general inverse problems. For example, patch based dictionary priors have been used in inverse problems such as tomography [11] and MRI [12]. Furthermore, while the BM3D [5] denoising may not naturally lend itself to formulation as a prior, Danielyan et al. [13] have adapted the BM3D for the inverse problem of image deblurring. However, this approach is not directly applicable to a general inverse problem. Furthermore, the approach in [13] is not directly applicable to denoising algorithms/priors formulated using a nonparametric point estimation framework such as [1] and [7]. So, while some advances have been made in the integration of advanced prior and forward models, they tend to be highly customized to the problem and currently no simple turn-key approach exists to match denoising algorithms as priors for general inverse problems.

In this paper, we propose a flexible framework for using denoising algorithms as priors for model-based inversion. This framework, which we term Plug-and-Play priors, has the advantage that it simplifies software integration, and moreover, it allows state-of-the-art denoising methods that are not explicitly formulated as optimization problems to be used. Our proposed Plug-and-Play framework is based on a direct application of the alternating directions method of multipliers (ADMM) [14] that has recently become popular for the solution of a variety of MAP estimation/regularized inverse problems [15]–[18]. Our application of ADMM works by first splitting the state variable so as to decouple the prior and forward model terms of MAP estimation problem. The application of the ADMM technique to the resulting constrained minimization problem then results in two decoupled optimizations, one for the forward model and one for the prior model. We note that this allows for a completely decoupled software implementation with one module corresponding to a denoising algorithm only dependent on the prior, and a second module corresponding to a model-based inversion with l_2 regularization only dependent on the forward model. Importantly, we demonstrate empirically that this framework can be used to solve a reconstruction problem even when the explicit function corresponding to the prior model is not known.

II. MAP COST FUNCTION FOR SOLVING INVERSE PROBLEMS

Let y be a $M \times 1$ measurement vector from which we desire to estimate the unknown x, a $N \times 1$ vector. Let $p(y|x)$ be the conditional probability density function (pdf) of the measurements y given x, and $p(x)$ be the pdf of the unknown, then the MAP estimate of x is given by

$$\begin{aligned} \hat{x} &\leftarrow \underset{x}{\operatorname{argmin}} \{-\log p(y|x) - \log p(x)\} \\ \hat{x} &\leftarrow \underset{x}{\operatorname{argmin}} \{l(y;x) + s(x)\} \end{aligned} \quad (1)$$

where $l(y;x) = -\log p(y|x)$ and $-\log p(x) = s(x) +$ terms independent of x. In the special case of $l(y;x) = \frac{1}{2\sigma_n^2}\|y-x\|_2^2 + \frac{M}{2}\log\left(2\pi\sigma_n^2\right)$ the MAP estimate corresponds to denoising designed to remove additive white Gaussian noise of variance σ_n^2. For this special case, we define $H(y;\sigma_n^2)$ to be the operator that denoises the signal y when it has been corrupted by additive Gaussian noise of variance σ_n^2. This operator is then given by the solution to the following MAP optimization problem:

$$\mathbb{H}(y;\sigma_n^2) = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2}\|y-x\|_2^2 + s(x) \right\}. \quad (2)$$

Sometimes it is useful to have an additional regularization parameter to control the relative effect of the prior model on the reconstruction. To allow for this additional control, we can rewrite the estimation problem as

$$\hat{x} \leftarrow \underset{x}{\operatorname{argmin}} \{l(y;x) + \beta s(x)\}, \quad (3)$$

where β can be used to modulate the amount of regularization applied to the inversion. Notice that when $\beta = 1$ the problem is exactly the MAP estimation problem (1).

III. VARIABLE SPLITTING AND ADMM

In order to design an algorithm for (3) that decouples the forward and prior terms, we first split the variable x into two new variables x and v, and reformulate equation (3) as the following constrained optimization problem [15].

$$\begin{aligned} (\hat{x},\hat{v}) &\leftarrow \underset{x,v}{\operatorname{argmin}} \{l(y;x) + \beta s(v)\} \\ &\text{subject to } x = v\,. \end{aligned} \quad (4)$$

We then solve (4) by forming the augmented Lagrangian function and using the ADMM technique [14]. The augmented Lagrangian for this problem is given by

$$\begin{aligned} L_\lambda(x,v,u) &= l(y;x) + \beta s(v) + \frac{\lambda}{2}\|x - v + u\|_2^2 \\ &- \frac{\lambda}{2}\|u\|_2^2. \end{aligned} \quad (5)$$

where u is a scaled dual variable and λ is the penalty parameter. The ADMM algorithm consists of repeatedly performing the following steps until convergence.

$$\begin{aligned} \hat{x} &\leftarrow \underset{x}{\operatorname{argmin}} L_\lambda(x,\hat{v},u) \\ \hat{v} &\leftarrow \underset{v}{\operatorname{argmin}} L_\lambda(\hat{x},v,u) \\ u &\leftarrow u + (\hat{x} - \hat{v})\,. \end{aligned}$$

Notice that in general λ does not effect the final result but controls the rate of convergence of the ADMM algorithm.

If $\tilde{x} = \hat{v} - u$ and $\tilde{v} = \hat{x} + u$ then each iteration of the algorithm can be written as

$$\hat{x} \leftarrow \underset{x}{\operatorname{argmin}} \left\{ l(y;x) + \frac{\lambda}{2}\|x - \tilde{x}\|_2^2 \right\} \quad (6)$$

$$\hat{v} \leftarrow \underset{v}{\operatorname{argmin}} \left\{ \frac{\lambda}{2}\|\tilde{v} - v\|_2^2 + \beta s(v) \right\} \quad (7)$$

$$u \leftarrow u + (\hat{x} - \hat{v}). \quad (8)$$

The first step only depends on the choice of forward model. The second step only depends on the choice of prior and can be interpreted as a denoising operation as in equation (2).

In order to emphasize the modular structure of the ADMM update, we define the operator $\mathbb{F}(y,\tilde{x};\lambda)$ as

$$\mathbb{F}(y,\tilde{x};\lambda) = \underset{x}{\operatorname{argmin}} \left\{ l(y;x) + \frac{\lambda}{2}\|x - \tilde{x}\|_2^2 \right\}. \quad (9)$$

This function returns the MAP estimate of x given the data y, using very simple quadratic regularization to a value, $\tilde{x}$. We call $\mathbb{F}$ a simplified reconstruction operator. Notice that $\mathbb{F}$ is also the proximal mapping [19] associated with the function $\frac{1}{\lambda}l(y;x)$. Using our definition of the simplified reconstruction operator $\mathbb{F}(y,\tilde{x};\lambda)$ from (9), and our definition of the denoising operator $\mathbb{H}(y;\sigma_n^2)$ from (2), we may now reformulate the ADMM iterations as the following three steps.

$$\hat{x} \leftarrow \mathbb{F}(y,\tilde{x};\lambda) \quad (10)$$

$$\hat{v} \leftarrow \mathbb{H}(\tilde{v};\frac{\beta}{\lambda}) \quad (11)$$

$$u \leftarrow u + (\hat{x} - \hat{v}). \quad (12)$$

Importantly, using this Plug-and-Play framework, the minimization can now be written as two independent software modules - one for implementing the simplified reconstruction operator $\mathbb{F}(y,\tilde{x};\lambda)$ and the other for implementing the denoising algorithm $\mathbb{H}(\tilde{v};\sigma_n^2)$. Furthermore changing the prior model only involves changing the implementation of $\mathbb{H}(\tilde{v};\sigma_n^2)$. Thus the Plug-and-Play priors framework can be used to mix and match different denoising algorithms (priors) with the forward model of interest. Notice that the minimization corresponding to the simplified reconstruction operator and the denoising operator need not be exact. Instead, they can be replaced by the approximate operators $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{H}}$ that do not minimize the respective cost functions but instead decrease its value sufficiently. This is an important technique for speeding up the implementation of the ADMM [14] and making the algorithm useful in practical applications.

We note that the variable splitting approach discussed here has been exploited to solve a variety of inverse problems [15],

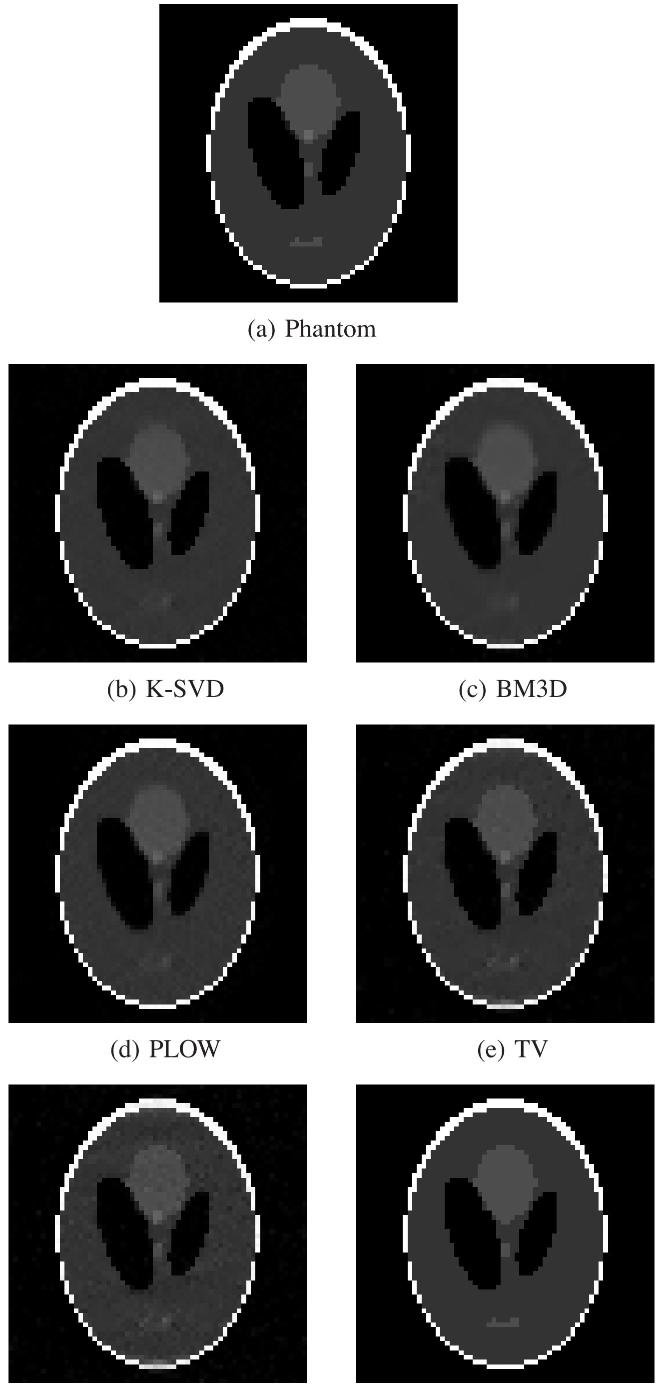
[18], [20]. However the main motivation of this research was to create better algorithms for solving the optimization problems resulting from regularized inversion. For example, this variable splitting/ADMM approach has been used to more effectively solve problems with l_1 norms, TV norms, and positivity constraints that can create difficulties in conventional gradient based optimization. In distinction to this earlier research, our primary goal is to use splitting strategies as a mechanism to create a flexible framework to easily match prior models (embodied in the form of denoising algorithms) with forward models.

Finally we note that in this paper we do not discuss theoretical convergence properties of the Plug-and-Play framework. While the ADMM is guaranteed to converge if l and s are convex, closed and proper functions and L_0 has a saddle point [14], we observe via our numerical experiments that substituting $\mathbb{H}$ with denoising algorithms that do not explicitly correspond to a convex function s or even a strict optimization problem, still produces a stable result. Thus we rely on empirical evidence from our experiments to show that our framework produces a stable result.

IV. Experimental Results

In this paper we will restrict our simulations to the case where $l(y;x) = \frac{1}{2}\|y - Ax\|_\Lambda^2$, A is a tomographic forward projector, and Λ is a diagonal weighting matrix. We will experiment with a variety of state-of-the-art denoising techniques for $\mathbb{H}$ which may or may not explicitly be formulated as prior models in a regularized optimization framework. We evaluate our method on a 64×64 Shepp-Logan phantom with values scaled between $0 - 255$. The phantom is forward projected at 141 views between $-70°$ and $+70°$ and noise is added to simulate Poisson statistics (variance is set equal to the mean). We compare reconstructions using the Plug-and-Play priors framework by experimenting with six different denoising techniques/priors - K-SVD [4], BM3D [5], PLOW [7], Total Variation (TV) [21], q-GGMRF [22] and discrete reconstruction (DR) [23]. The regularization parameter β is adjusted for achieving the minimum root mean square error (RMSE) between the reconstruction and phantom. The patch sizes for K-SVD and BM3D are set to 4×4 and for PLOW to 5×5. Instead of using the simplified reconstruction operator $\mathbb{F}$ in the ADMM loop, we use an approximate operator $\tilde{\mathbb{F}}$, which lowers the value of the cost function corresponding to $\mathbb{F}$ using N_{Iter} number of iterations of iterative coordinate descent (ICD) [24] with random order updates [8]. The algorithm is initialized with a filtered back projection reconstruction. The value of N_{Iter} is set to 1 for all algorithms except the DR prior in which case it is set to 20 for the first outer iteration. The value of λ is set to $1/20$ for all experiments. Since the DR prior is non-convex we observed that the value of λ effects the final solution. The number of levels in the case of the discrete reconstruction prior is set to 6 - the number present in the original phantom.

Fig. 1 shows the reconstructions resulting from the use of the six denoising algorithms as prior models, and Table I shows the corresponding RMSE for each prior. For this very simple Shepp-Logan image, the DR prior results in the lowest



(a) Phantom

(b) K-SVD (c) BM3D

(d) PLOW (e) TV

(f) q-GGMRF (g) Discrete reconstruction

Fig. 1. Comparison of the minimum RMSE reconstructions using different priors for the Shepp-Logan phantom projected in a limited angular range ($+/-70°$). All images are displayed in the window $[0 - 255]$. (a) Phantom (b) K-SVD (c) BM3D (d) PLOW (e) TV (f) q-GGMRF (g) Discrete reconstruction. We observe that the patch based denoising algorithms (b) - (d) work well producing qualitatively comparable reconstructions to the typically used priors like TV and q-GGMRF. Some of the features in the phantom are not reconstructed accurately due to the limited angle nature of the projection data. The discrete prior (g) produces a very accurate reconstruction for this phantom.

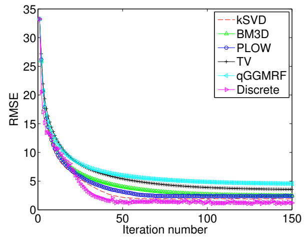


Fig. 2. Comparison of the convergence (RMSE between the reconstruction and the original phantom) as a function of iteration number for the different denoising models used. We note that the convergence for all algorithms is robust and stable. Furthermore the convergence rates across the different denoising algorithms are similar.

TABLE I
COMPARISON OF THE MINIMUM ROOT MEAN SQUARE ERROR OF THE RECONSTRUCTION WITH RESPECT TO THE ORIGINAL PHANTOM FOR VARIOUS PRIORS. WE OBSERVE THAT THE THE PATCH BASED NONLOCAL DENOISING OPERATORS GIVE A LOW RMSE RECONSTRUCTION.

Algorithm	RMSE	β
K-SVD [4]	2.13	4.32
BM3D [5]	2.46	1.39
PLOW [7]	2.35	1.50
TV [21]	3.55	0.47
q-GGMRF [22]	4.58	0.28
Discrete Recon [23]	1.20	1.00

RMSE. However, the other methods result in a comparable RMSE. Most importantly, each denoising algorithm was easily matched to the tomographic forward model and for each prior, the convergence to the fixed solution was stable and robust (see Fig. 2). Interestingly, BM3D [5] and PLOW [7] are formulated without the explicit use of an optimization framework, so the Plug-and-Play methodology provides a simple and robust framework to incorporate them as priors for model based reconstruction.

V. CONCLUSIONS

In this paper, we proposed a flexible framework that allows state-of-the-art forward models of imaging systems to be matched with state-of-the-art prior or denoising models. The framework, which is based on variable splitting and use of the ADMM algorithm, simplifies the software architecture by decoupling the forward and prior models. Furthermore the framework enables state-of-the-art denoising algorithms, even those that have no known formulation as an optimization problem, to be used as priors/regularizers for model based inversion.

VI. ACKNOWLEDGMENT

We would like to thank Ruoqiao Zhang for sharing the q-GGMRF prior code which was used to implement the q-GGMRF denoising routine.

REFERENCES

[1] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 106–128, 2013.

[2] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[3] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, 2005, pp. 60–65 vol. 2.

[4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[6] E. Haneda and C. Bouman, "Implicit priors for model-based inversion," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 3917–3920.

[7] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1635–1649, 2012.

[8] Z. Yu, J. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. on Image Processing*, vol. 20, no. 1, pp. 161 –175, Jan. 2011.

[9] S. V. Venkatakrishnan, L. F. Drummy, M. De Graef, J. P. Simmons, and C. A. Bouman, "Model based iterative reconstruction for bright field electron tomography," *Proc. SPIE 8657, Computational Imaging XI, 86570A (February 14, 2013)*, pp. 86 570A–86 570A–12, 2013.

[10] D. B. Husarik, D. Marin, E. Samei, S. Richard, B. Chen, T. A. Jaffe, M. R. Bashir, and R. C. Nelson, "Radiation dose reduction in abdominal computed tomography during the late hepatic arterial phase using a model-based iterative reconstruction algorithm: How low can we go?" *Invest Radiol.*, vol. 47, no. 8, pp. 468–474, 2012.

[11] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *IEEE International Symposium on Biomedical Imaging*, 2008, pp. 1375–1378.

[12] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 5, pp. 1028 –1041, May 2011.

[13] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1715–1728, 2012.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, Jul. 2011.

[15] M. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2345–2356, 2010.

[16] S. Ramani and J. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 3, pp. 694–706, 2011.

[17] ——, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 677–688, 2012.

[18] M. G. McGaffin, S. Ramani, and J. A. Fessler, "Reduced memory augmented Lagrangian algorithm for 3D iterative X-ray CT image reconstruction," pp. 831 327–831 327–6, 2012.

[19] R. Rockafellar, "Moreaus proximal mappings and convexity in Hamilton-Jacobi theory," in *Nonsmooth Mechanics and Analysis*, ser. Advances in Mechanics and Mathematics, P. Alart, O. Maisonneuve, and R. Rockafellar, Eds. Springer US, 2006, vol. 12, pp. 3–12.

[20] Y. Huang, M. Ng, and Y. Wen, "A fast total variation minimization method for image restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 2, pp. 774–795, 2008.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[22] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[23] T. Frese, C. A. Bouman, and K. Sauer, "Multiscale Bayesian methods for discrete tomography," in *Discrete Tomography: Foundations, Algorithms and Applications*, G. T. Herman and A. Kuba, Eds. Birkhaüser Boston, Cambridge, MA, 1999, pp. 237–261.

[24] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480 –492, Mar. 1996.