5-29-2013

Plug-and-Play Priors for Model Based Reconstruction

Singanallur V. Venkatakrishnan
Electrical and Computer Engineering, Purdue University, West Lafayette Campus, svenkata@purdue.edu

Charles A. Bouman
Purdue University, bouman@purdue.edu

Brendt Wohlberg
Los Alamos National Lab, Los Alamos, NM, brendt@lanl.gov

Venkatakrishnan, Singanallur V.; Bouman, Charles A.; and Wohlberg, Brendt, "Plug-and-Play Priors for Model Based Reconstruction" (2013). *ECE Technical Reports.* Paper 448.
http://docs.lib.purdue.edu/ecetr/448

Plug-and-Play Priors for Model Based Reconstruction

Singanallur V. Venkatakrishnan

Charles A. Bouman

Brendt Wohlberg

Purdue University

School of Electrical and Computer Engineering

465 Northwestern Avenue

West Lafayette, IN 47907-1285

Plug-and-Play Priors for Model Based Reconstruction

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg

Abstract

Model-based reconstruction is a powerful framework for solving a variety of inverse problems in imaging. The method works by combining a forward model of the imaging system with a prior model of the image itself, and the reconstruction is then computed by minimizing a functional consisting of the sum of two terms corresponding to the forward and prior models.

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. A wide range of methods including non-local means, dictionary-based methods, 3D block matching, TV minimization and kernel-based filtering have proven that it is possible to recover high fidelity images even after a great deal of noise has been added. 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 report, we propose 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. 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.

S. V. Venkatakrishnan and C. A. Bouman are with the School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Ave., West Lafayette, IN 47907-2035, USA. Tel: 765-494-6553, Fax: 765-494-3358, E-mail: {svenkata,bouman}@purdue.edu.

B. Wohlberg is with the Theoretical Division, Los Alamos National Lab, Los Alamos, NM, USA. E-mail: brendt@lanl.gov.

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. For example, a typical approach is to compute the maximum *a posteriori* (MAP) estimate as the minimum of the sum of the log likelihood forward model and the log probability of the prior distribution.

In recent years, there have been enormous advances in the solution of a particular inverse problem generally referred to as image denoising [1], [2]. The problem of image denoising is to recover an image that has been corrupted by noise, the most commonly considered noise model being additive white Gaussian noise. 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. Some denoising algorithms are based on an explicit regularized inversion formulation using, for example, a MAP estimate; but in other cases, the denoising algorithms are directly formulated as ad hoc non-linear estimates of the noiseless image. In fact, a number of very novel and effective approaches have recently emerged for image denoising. Examples of new methods include a wide variety of patch based non-local means approaches [3], generalizations of bi-lateral filtering approaches [1], patch-based dictionary learning methods such as K-SVD [4], block-matching with transform-based denoising such as BM3D [5], and a variety of total-variation [2], and Markov random field (MRF) based approaches [6]. These new methods have demonstrated that it is possible to vastly improve on what was previously believed to be possible.

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 [7] to microscopes [8]. Research in this field has demonstrated that model-based inverse methods can greatly improve the quality of reconstructed images [9]. 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 imaging problems such as tomography [10], [11] and MRI [12]. Furthermore, while BM3D may not naturally lend itself to formulation as a prior, Danielyan et al. [13], [14] have adapted the BM3D [5] denoising for the inverse problem of image deblurring. However, this approach is not directly applicable to a general inverse problem. So, while some significant 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 report, 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. For denoising algorithms based on well-behaved optimization criteria (e.g. - closed, proper and convex functions [15]), it can be easily shown that the Plug-and-Play framework is convergent to the MAP estimate of the reconstruction. In more general cases, we empirically demonstrate that the method converges robustly to a good solution. We show with some simple examples how Plug-and-Play priors can be used to mix and match a wide variety of existing denoising models with the type of complex forward model that is typically used in applications such as tomographic reconstruction. Consequently, this new approach can greatly expand the range of possible models used in model-based inversion.

Our proposed Plug-and-Play framework is based on a direct application of the alternating directions method of multipliers (ADMM) [15] that has recently become popular for the solution of a variety of MAP estimation/regularized inverse problems [16]–[21]. 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, this framework can be used to solve a MAP reconstruction problem even when the explicit cost function corresponding to the prior model is not known. Moreover, we also demonstrate empirically that the method can be used with denoising algorithms such as BM3D that are not explicitly formulated as optimization problems.

In order to demonstrate the approach, we apply a wide range of denoising algorithms as Plug-and-Play priors for a simple tomographic reconstruction problem using the well-known Shepp-Logan phantom.

The results indicate that methods such as K-SVD [4], BM3D [5], PLOW [22], q-GGMRF [23], TV [2], and discrete reconstruction (DR) [24] can all be easily applied as priors through direct use of software implementations of the denoising algorithms. The implementation of the Plug-and-Play priors is very simple and robust, and the observed convergence speed of the resulting iterative algorithm is comparable to the convergence speed of a tightly integrated prior model, as is traditionally done with model based reconstruction.

The organization of the rest of this report is as follows. In Section II we introduce the cost function corresponding to the MAP estimation for tomographic reconstruction. In Section III we briefly discuss the variable splitting and ADMM algorithm for solving the MAP estimation problem. In Section IV we apply the algorithm on a phantom data set with different denoising algorithms (priors) and in Section V we draw our conclusions.

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

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

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:

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

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

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

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 separate the forward and prior terms in the MAP cost function, we first split the variable x into two new variables x and v, and reformulate equation (3) as the following constrained optimization problem [16], [17].

$$(\hat{x}, \hat{v}) \quad \leftarrow \quad \operatorname*{argmin}_{x,v} \{l(y; x) + \beta s(v)\}$$

$$\text{subject to } x = v \,. \tag{4}$$

We then solve (4) by forming the augmented Lagrangian function and using the ADMM technique [15]. The augmented Lagrangian for this problem is given by

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

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.

$$\hat{x} \quad \leftarrow \quad \operatorname*{argmin}_{x} L_\lambda(x, \hat{v}, u)$$

$$\hat{v} \quad \leftarrow \quad \operatorname*{argmin}_{v} L_\lambda(\hat{x}, v, u)$$

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

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

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

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

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

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

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

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

[25] 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.

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

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

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

The overall algorithm is summarized in Fig. 1. 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 [15] 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 [16], [17], [21], [26]. 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 report 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 [15], 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.

```
function [x̂] ← RECONSTRUCT(y, β, λ)

    % Inputs: Measurements y, Regularization β, Augmented Lagrangian parameter λ

    Initialize x̂

    v̂ ← x̂, u ← 0                              // Initialize v̂ and u

    σ²ₙ ← β/λ                                  // Variance for denoising algorithm

    while Stopping criteria are not met do

        x̃ ← v̂ − u

        x̂ ← 𝔽(y, x̃; λ)                        // Only dependent on forward model

        ṽ ← x̂ + u

        v̂ ← ℍ(ṽ; σ²ₙ)                          // Denoising operator only dependent on prior

        u ← u + (x̂ − v̂)                        // Update the scaled dual variable u

    end while

end function
```

Fig. 1. Pseudocode for Plug-and-Play priors framework. In each iteration an alternating minimization is done. The first minimization depends only on the likelihood function while the second minimization only requires the application of a denoising algorithm. Thus introducing a new prior only requires introducing a new denoising software module.

IV. EXPERIMENTAL RESULTS

In this report 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. This type of forward model is widely used in electron tomography [8], [27] where, due to mechanical constraints, the sample can only be tilted in a limited range. We compare reconstructions using the Plug-and-Play priors framework by experimenting with six different denoising techniques/priors - K-SVD [4], BM3D [5], PLOW [22], Total Variation (TV) [28], q-GGMRF [23] and discrete reconstruction (DR) [24]. The regularization parameter β is adjusted for achieving the minimum root mean square error (RMSE) between the reconstruction and phantom. 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

TABLE I

<small>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.</small>

Algorithm	RMSE	β
K-SVD [4]	2.32	4.32
BM3D [5]	2.51	1.39
PLOW [22]	2.70	1.50
TV [28]	3.42	0.47
q-GGMRF [23]	4.46	0.28
Discrete Recon [24]	1.32	1.00

iterative coordinate descent (ICD) [29] with random order updates [7]. 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. The value of λ is set to $1/75$ for all experiments except the DR prior, in which case it is set to $1/20$. 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. Further details of the parameters used for different denoising algorithms are given in Appendix A.

Fig. 2 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 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. 3). Interestingly, the BM3D algorithm is formulated without the explicit use of an optimization framework, so the Plug-and-Play methodology provides a simple and robust framework to incorporate it as a prior.

Finally we compare the convergence of the ADMM technique to the direct implementation of the MAP estimate using a q-GGMRF prior model that is tightly integrated into the cost function [23]. The traditional approach of tight integration allows for more flexibility in the design of optimization algorithms, so it might be expected to have faster convergence; but this faster convergence is at the cost of a less modular and flexible design. Nonetheless, one would expect that both the Plug-and-Play and traditional formulation
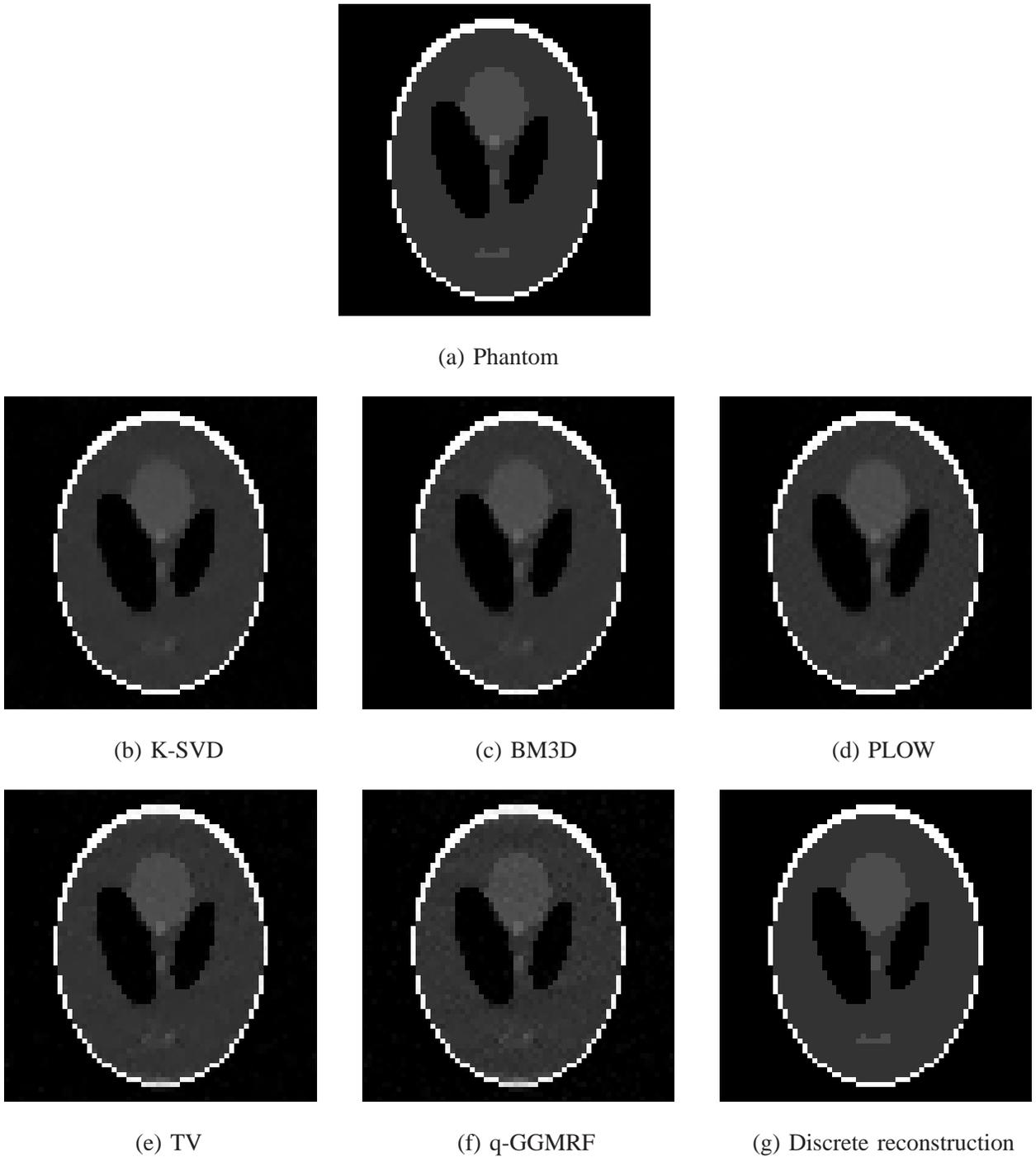
(a) Phantom

(b) K-SVD

(c) BM3D

(d) PLOW

(e) TV

(f) q-GGMRF

(g) Discrete reconstruction

Fig. 2. Comparison of the minimum RMSE reconstructions using different priors for the Shepp-Logan phantom projected in a limited angular range $(+/-70^{\circ})$. 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.
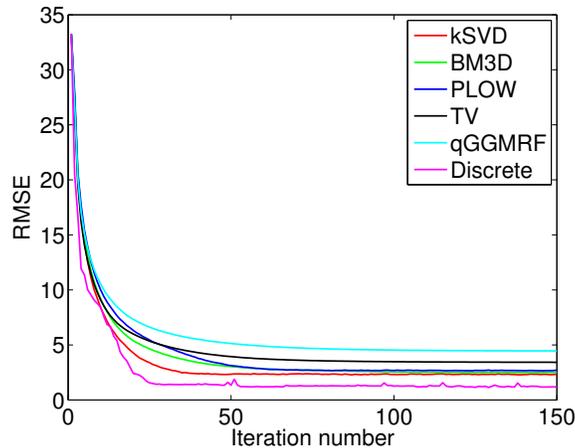
Fig. 3. Comparison of the convergence (RMSE between the reconstruction and the original phantom) as a function of iteration number for the different denoising models used. We note that the convergence for all algorithms is robust and stable. Furthermore the convergence rates across the different denoising algorithms are similar.

should yield the same result when run to convergence. Moreover, for this case, the denoising algorithm corresponds to the minimization of a convex function, and hence the ADMM technique is known to be globally convergent provided that the inner minimizations are either exact or are run for a sufficiently large number of iterations [15]. Fig. 4 shows the RMSE as a function of iteration number for the tightly integrated approach where the ICD algorithm is used to solve the tomographic inversion compared to the two step ADMM approach with a q-GGMRF denoising prior. We set the number of inner iterations in the q-GGMRF denoising step in ADMM to 1 for a fair comparison. It is possible that this will not lead to a monotonically decreasing cost function because the inner minimizations are not exact. However we still observe that the RMSE decreases with iteration. While the ADMM approach is slower than ICD for our implementation and requires greater memory to store the auxiliary variables [20], [21], it provides a more flexible framework for incorporating different priors. Furthermore several techniques have been proposed to speed up the ADMM algorithm [15], [20] which can be applied to the Plug-and-Play priors framework.

V. CONCLUSIONS

In this report, 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
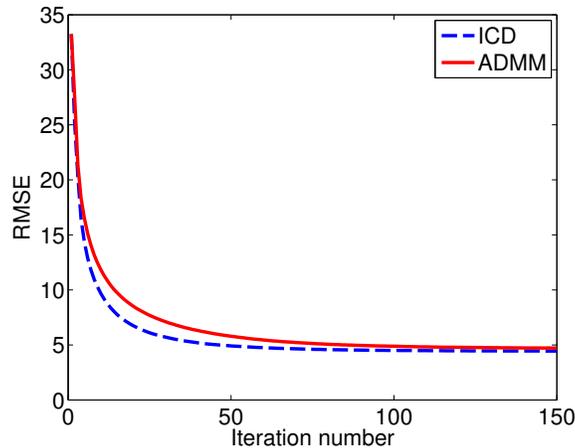
Fig. 4. Comparison of the convergence of ICD versus ADMM for the q-GGMRF prior with $p = 2, q = 1.2, c = 1/100, \sigma_x = 0.594$. The convergence is measured by using the RMSE between the reconstruction and the original phantom. The number of inner iterations in the ADMM is set to 1 for a fair comparison. We observe that ICD converges faster but ADMM's speed is comparable. However the time taken per iteration in ADMM will be higher due to the two step minimization.

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

APPENDIX A
PARAMETERS FOR DENOISING ALGORITHMS

The parameters of the algorithms are adjusted for the minimum RMSE tomographic reconstruction. In this section we specify the parameters used for the different denoising routines.

- **K-SVD** : We use the K-SVD denoising code from http://www.cs.technion.ac.il/~ronrubin/software. html. The following parameters were used for denoising:

  - Size of patch : 4×4

  - Size of dictionary : 3600 patches

  - Number of iterations: 10

  - For the β achieving minimum RMSE (see Table. I) we get $\sigma_n = \sqrt{\frac{\beta}{\lambda}} = 18$.

  The other parameters are set to the default values in the software. The K-SVD dictionary is initialized with the default settings in the software. For each subsequent outer iteration of ADMM, the K-SVD dictionary is initialized with the final dictionary from the previous iteration.

- **BM3D** : We use the BM3D code from http://www.cs.tut.fi/~foi/GCF-BM3D/index.html\#ref_software.

  - For the β achieving minimum RMSE (see Table. I) we set $\sigma_n = \sqrt{\frac{\beta}{\lambda}} = 10.192$ as input to the BM3D routine.

  The other parameters are set to the default values in the software.

- **PLOW** : The code for PLOW was downloaded from http://users.soe.ucsc.edu/~priyam/PLOW/. We use the following parameters:

  - Block size : 5

  - For the β achieving minimum RMSE (see Table. I) we set $\sigma_n = \sqrt{\frac{\beta}{\lambda}} = 10.607$.

  The other parameters are set to the default values in the software.

- **q-GGMRF** : The q-GGMRF denoising operator is given by

$$\hat{x} \leftarrow \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \|y - x\|_2^2 + \frac{1}{p\sigma_x^p} \sum_{\{i,j\} \in \chi} w_{ij} \frac{|x_i - x_j|^p}{1 + |\frac{x_i - x_j}{c}|^{p-q}} \right\} \tag{13}$$

where $p, q, c, \sigma_x, w_{ij}$ are the q-GGMRF parameters, σ_n^2 is the variance of the white noise in the data, and χ is the set of all neighboring pixels (8 point neighborhood). The weights are set to $\frac{1}{12}$ for diagonal neighbors and to $\frac{1}{6}$ for horizontal and vertical neighbors. The parameters of the denoising are set to $p = 2, q = 1.2, c = 1/100, \sigma_x = 0.29$ which was found to produce a reasonable result for images with white noise of variance σ_n^2. The random order ICD with surrogate functions [7] is used to implement the cost optimization.

  - For the β in Table. I we set $\sigma_n = \sqrt{\frac{\beta}{\lambda}} = 4.583$ as input to the denoising operator.

  - Number of ICD iterations: 10

- **Total Variation** : The code for TV denoising was downloaded from http://www.ceremade.dauphine.fr

/~peyre/matlab/image/content.html. It minimizes the cost function :

$$\hat{x} \quad \leftarrow \quad \underset{x}{\operatorname{argmin}} \left\{ \|y - x\|_2^2 + \frac{\lambda_{TV}}{2} \mathrm{TV}(x) \right\} \tag{14}$$

$$\hat{x} \quad \leftarrow \quad \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \|y - x\|_2^2 + \frac{c_{TV}}{2} \mathrm{TV}(x) \right\} \tag{15}$$

where $\lambda_{TV} = 2c_{TV}\sigma_n^2$ and $\mathrm{TV}(x)$ represents the total variation operator. We fix c_{TV} to $1/10.02$. This value was found to produce a reasonable denoising result for additive white Gaussian noise of variance σ_n^2.

- Therefore for the β in Table. I we set $\sigma_n = \sqrt{\frac{\beta}{\lambda}} = 5.937$.
- Number of iterations of the optimization routine: 100

• **Discrete Reconstruction (DR) prior** : A denoising operation corresponding to the discrete reconstruction prior (one that restricts the number of classes/values taken by the pixels in the reconstruction to K) is given by

$$(\hat{\mu}, \hat{b}) \quad \leftarrow \quad \underset{\mu, b \in \{1, \cdots, K\}^M}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \sum_{i=1}^{M} (y_i - \mu(b_i))^2 + c_{\mathrm{Discrete}} \sum_{\{i,j\} \in \chi} w_{ij} \delta(b_i \neq b_j) \right\} \tag{16}$$

where y is a $M \times 1$ vector containing the noisy image, σ_n^2 is the noise variance, b is a $M \times 1$ vector of labels corresponding to each pixel, $\mu : \{1, \cdots, K\} \to \mathbb{R}$ is a function that maps each label to a discrete output level (class mean), χ consists of all pairs of neighboring pixels, w_{ij} weights the interaction between neighboring pixels , δ is an indicator function, and c_{Discrete} is a constant. The denoising operation above is a non-convex optimization problem and we will describe an algorithm to find a local minimum. To minimize the cost function in (16) we use an alternating minimization strategy. The algorithm consists of repeatedly performing the following steps for each pixel i.

- **Class label update**

$$\hat{b}_i \quad \leftarrow \quad \underset{k \in \{1, \cdots, K\}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} (y_i - \hat{\mu}(b_i))^2 + c_{\mathrm{Discrete}} \sum_{j \in \chi_i} w_{ij} \delta(b_j \neq k) \right\} \tag{17}$$

  where χ_i is the set of neighbors of pixel i.

- **Mapping function update**

  Taking derivative of the cost function in (16) with respect to each $\mu(k)$ $\forall k \in 1, \cdots, K$ and setting it to zero gives

$$\hat{\mu}(k) \quad \leftarrow \quad \frac{1}{N_k} \sum_{i=1}^{M} y_i \delta(\hat{b}_i = k) \tag{18}$$

where $N_k = \sum_{i=1}^{M} \delta(\hat{b}_i = k)$. Notice that this step simply sets the mapping for a given class to the mean value of the pixels assigned to that class.

Thus the denoising operator corresponding to the discrete reconstruction prior is given by

$$\mathbb{H}(y; \sigma_n^2) = \mu_{\hat{b}}.$$

where $\mu_{\hat{b}} = [\hat{\mu}(\hat{b}_1), \cdots, \hat{\mu}(\hat{b}_M)]^t$.

In practice we set the number of iterations to 10 as we found this to produce good results for the denoising problem. We fix c_{Discrete} to 4 as this produced a good denoising result for additive white Gaussian noise of variance σ_n^2. The weights are set to $\frac{1}{12}$ for diagonal neighbors and to $\frac{1}{6}$ for horizontal and vertical neighbors. In order to initialize the class labels b we use Otsu's method [30] on a low pass filtered version of the noisy input y.

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] 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.

[8] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. Bouman, "Bayesian Tomographic Reconstruction for High Angle Annular Dark Field (HAADF) Scanning Transmission Electron Microscopy (STEM)," in *2012 IEEE Statistical Signal Processing Workshop (SSP) (SSP'12)*, Ann Arbor, Michigan, USA, Aug. 2012.

[9] 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.

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

[11] Q. Xu, H. Yu, X. Mou, D. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 9, pp. 1682–1697, 2012.

[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, "Image deblurring by augmented Langrangian with BM3D frame prior," in *Proc. 2010 Workshop on Information Theoretic Methods in Science and Engineering, WITMSE 2010, Tampere, Finland*, Aug. 2010.

[14] ——, "BM3D frames and variational image deblurring," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1715–1728, 2012.

[15] 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.

[16] M. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681–695, 2011.

[17] ——, "Fast image recovery using variable splitting and constrained optimization," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2345–2356, 2010.

[18] 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.

[19] M. Nilchian and M. Unser, "Differential phase-contrast X-ray computed tomography: From model discretization to image reconstruction," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 2012, pp. 90–93.

[20] S. Ramani and J. Fessler, "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.

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, 1979.