Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Ruoqiao Zhang

Entitled

ADVANCED STATISTICAL MODELING FOR MODEL-BASED ITERATIVE RECONSTRUCTION FOR SINGLE-ENERGY AND DUAL-ENERGY X-RAY CT

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

CHARLES A. BOUMAN
Chair

KEN D. SAUER

MARY L. COMER

THOMAS M. TALAVAGE

JEAN-BAPTISTE THIBAULT

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyrightmaterial.

Approved by Major Professor(s): CHARLES A. BOUMAN

Approved by: Venkataramanan Balakrishnan 11/13/2015

Head of the Departmental Graduate Program Date

ADVANCED STATISTICAL MODELING FOR MODEL-BASED ITERATIVE RECONSTRUCTION FOR SINGLE-ENERGY AND DUAL-ENERGY X-RAY CT

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Ruoqiao Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

For my parents, Mr. Shu Zhang and Ms. Xiaoyun Yu,
and my grandmother, Ms. Xiaoshan Qiu.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Prof. Charles A. Bouman, whose constructive suggestions have shaped me into a better engineer and researcher. His enthusiasm and persistence toward research have always been inspiring. In addition, his appreciable sense of humor has made my time in the lab enjoyable. Furthermore, I would like to thank Prof. Bouman's wife, Ms. Cristina L. Bouman, for the uplifting conversations along with the delicious food.

I would like to thank Prof. Ken D. Sauer from University of Notre Dame and Dr. Jean-Baptiste Thibault from GE Healthcare, for being excellent research collaborators and committee members. Not only have their knowledgeability and intelligence contributed to my research immensely, but their exceptional vision and great personality have also had great influence in my life.

I would like to thank my committee members, Prof. Mary L. Comer, for sharing her perspectives on my research and future career path, and Prof. Thomas M. Talavage, for his time and input.

I thank my collaborator, Dr. Debashish Pal from GE Healthcare, for his unstinting help and guidance that improved the quality of my research significantly. I also thank my colleagues from University of Notre Dame, Jiao Wang and Zhiqian Chang, for the time we spent together in weekly teleconferences and at GE Healthcare as research interns. In particular, I would like to thank my academic siblings, Dr. Zhou Yu and Dr. Guangzhi Cao. Throughout these years, they have offered me valuable advice and immeasurable support on both research and life. Moreover, I thank my collaborators at GE, including Dr. Jiang Hsieh, Dr. Somesh Srivastava, Dr. Jiahua Fan, Mr. Yasuhiro Imai, Dr. Xue Rui, and Dr. Lin Fu, for their generous suggestion and assistance.

I thank all my lab mates in the Integrated Imaging Lab at Purdue, especially Eri Haneda, Jordan Kisner, William Wong, Haitao Xue, Singanallur Venkatakrishnan, Dong Hye Ye, Aditya Mohan, Dilshan Godaliyadda, Zeeshan Nadir, Suhas Sreehari, and Ben Foster, for all the pleasurable conversations that may or may not be relevant to research. In particular, I would like to thank my office mates, Yandong Guo and Pengchong Jin, for all insightful discussions. Their delightful personality has also made my days in the office so memorable.

I would like to thank my friends from Prof. Jan P. Allebach's group, including Ming Chen, Cheng Lv, Weibao Wang, Jianyu Wang, Yucheng Liu, Lu Zhang, Ni Yan, Chuohao Tang, Huizhi Lu, and Jia Zhang, for organizing various activities that make my life at Purdue entertaining . Thanks to Lingxing Zheng, Xia Chen, Wei Zheng, Na Wu, Xinyue Chang, Rong Fu, and Jason Chen, for being my friends ever since I came to Purdue. I also thank my friends, Qiong Nian, Kuo Guo, Mao Mao, Yuanyuan Wang and Haining Chen, for being my badminton and poker mates.

Finally, I would like to express the deepest appreciation to my family, Mr. Shu Zhang, Ms. Xiaoyun Yu, and Ms. Xiaoshan Qiu, for their unwavering love, support, and encouragement.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES viii

ABSTRACT xv

1 Introduction 1

1.1 MBIR for dual-energy X-ray CT 3

1.2 Advanced prior model for MBIR 3

2 MODEL-BASED ITERATIVE RECONSTRUCTION FOR DUAL-ENERGY X-RAY CT USING A JOINT QUADRATIC LIKELIHOOD MODEL . . 5

2.1 Introduction 5

2.2 MAP Reconstruction Framework 9

2.2.1 Measurement Preprocessing 10

2.2.2 Forward Model 13

2.2.3 Quadratic Joint Likelihood Model 15

2.2.4 Likelihood Model For Fast kVp Switching Modality 17

2.2.5 Prior Model 20

2.2.6 Constrained Optimization 23

2.3 Optimization Algorithm 25

2.4 Experimental Results 30

2.5 Conclusion 41

3 GAUSSIAN MIXTURE MARKOV RANDOM FIELD FOR MODEL-BASED ITERATIVE RECONSTRUCTION 48

3.1 Introduction 48

3.2 Gaussian mixture Markov random field 51

3.3 MAP estimation with GM-MRF prior 54

Page

3.3.1 Surrogate prior 55
3.3.2 Optimization 57
3.4 Covariance Control for GM-MRF 59
3.5 Experiments and Results 64
3.5.1 Training 64
3.5.2 Results 67
3.6 Conclusion 76
REFERENCES 81
A Single voxel update by using the KKT condition 88
B Proof of Lemma: surrogate functions for logs of exponential mixtures . . 92
VITA 93

LIST OF TABLES

Table Page

2.1 Comparison of FBP, independent DE-MBIR and JDE-MBIR for measurement of noise and in-plane resolution for the images in Fig. 2.9. The resolution measured by using the cyclic bars method is made in the 70 keV monochromatic images at the three lowest spatial frequencies. . . . 36

2.2 Theoretical densities of iodine and water for the inserted rods in the GSI contrast phantom as shown in Fig. 2.16. 40

3.1 Parameter selection for GM-MRF model. 62

3.2 Partition of the training data. Each image patch was classified into one of the six groups based on its mean and standard deviation. The number of GMM components for each group was empirically chosen. 65

3.3 Mixture weights of GMMs trained from different groups. The mixture weights are determined by the proportions of corresponding groups in the whole training data and will be used when combining different GMMs to form a single model. 65

LIST OF FIGURES

Figure Page

2.1 This figure illustrates a model for fast kVp switching technique. A single X-ray tube alternates the voltage between low- and high-kVp from view to view. Thus each view contains only one measurement, either low- or high-energy. 8

2.2 Plots of the true log-likelihood function, the independent approximation, and the proposed joint approximation, with contours plotted underneath. 19

2.3 Figure plots contours of the true log-likelihood function and different approximations. Independent and joint models are compared at three attenuation levels. Within each subfigure, we plot the contours of the true log-likelihood (blue), approximation without interpolation error (red), and with 5% interpolation error (green). Each plot covers two standard deviations of water and iodine projections. 21

2.4 Pseudocode of JDE-MBIR for DECT that uses fast kVp switching. First, we interpolate the missing sample for each projection. Second, we perform material decomposition and also compute the statistical weighting matrix. Finally, we use the decomposed sinograms and weighting matrices to reconstruct the images iteratively. The subroutine ImageRecon is described in Fig. 2.8. 22

2.5 This figure illustrates the feasible values of a voxel, $m_j = [m_{j,\mathrm{W}}, m_{j,\mathrm{I}}]$, where $m_{j,\mathrm{W}}$ is the water-equivalent density and $m_{j,\mathrm{I}}$ is the iodine-equivalent density. The yellow region shows the constrained solution set, which is formed by the intersection of only two half planes, one defined by $n_{\max}$ and the other by $n_{\min}$. The green arrows show the attenuation vectors at intermediate energies. 23

2.6 This figure illustrates the desired substitute function. The substitute function $q(\Delta;\Delta^*)$ upper bounds the true function $\rho(\Delta)$ except for the fixed point $\Delta = \Delta^*$, where two functions are equal. Thus, the true function must decrease when the substitute function is minimized. 27

Figure Page

2.7 Pseudocode for solving the quadratic minimization problem in (2.48) with the KKT condition. We first test the KKT condition on the unconstrained solution. If it fails, we solve the minimization problem on either boundary of the feasible set, and then we test the corresponding solution with the KKT condition. Once the KKT condition is met, the particular solution becomes the updated value. The origin will be the only feasible solution if no qualified solution is found in the previous cases. The derivation for the solution is provided in the appendix. 29

2.8 Pseudocode for reconstructing the image by using generic ICD algorithm. We initialize m with the raw FBP images and also initialize the error sinogram. Within each iteration, for each selected voxel, we first compute the column of the forward projection matrix. Second, we compute first two derivatives of the log-likelihood function. Third, we compute the coefficients for the surrogate prior. Fourth, we compute the first two derivatives of the quadratic cost function. Fifth, we solve the optimization problem with the KKT condition to obtain the voxel update. Finally, we update the error sinogram and the image. We define δ_j as an $N \times 1$ vector that is 1 for element j and 0 otherwise. The subroutine KKTSolve is described in Fig. 2.7. 30

2.9 Comparison of FBP, independent DE-MBIR (ind. DE-MBIR) and JDE-MBIR reconstructions from a GEPP scan. From top to bottom: water density image, iodine density, and 70 keV monochromatic image. From left to right: FBP, independent DE-MBIR, JDE-MBIR. Display window for water: [650 1250] mg/cm^3; for iodine: [-8.0 7.0] mg/cm^3; for mono images: [-300 300] HU. The white box on the 70 keV FBP image (first at the third row) indicates the region where the noise standard deviation is evaluated. 33

2.10 Resolution bars in the 70 keV monochromatic images from a GEPP scan reconstructed with: (a) FBP; (b) independent DE-MBIR; (c) JDE-MBIR. Display window: [-500 -300] HU. Each image zooms in to the resolution bars of the monochromatic images shown in Fig. 2.9 with a different display window. 34

2.11 Profile plot across the resolution bars on the GEPP 70 keV monochromatic images for FBP, independent DE-MBIR, and JDE-MBIR. Left: profile line on the image; right: attenuations along the profile line with FBP (blue), independent DE-MBIR (green), and JDE-MBIR (red). 35

Figure Page

2.12 Comparison of FBP, independent DE-MBIR (ind. DE-MBIR), and JDE-MBIR monochromatic images of the GEPP at different energies. From top to bottom: photon energy at 50 keV and 130 keV. From left to right: FBP, independent DE-MBIR, JDE-MBIR. Display window for mono images: [-300 300] HU. These monochromatic images are synthesized using the reconstructed material densities in Fig. 2.9 based on (2.18). 37

2.13 Resolution and noise of the monochromatic images across various energy levels with different reconstruction methods. 38

2.14 MTF measured at the cyclic bars of three different spatial frequencies in the monochromatic images across various photon energies. The JDE-MBIR produces higher MTF values than the other two methods at all three frequencies. 38

2.15 Comparison of the convergence speed of the JDE-MBIR and the standard single-energy MBIR with the GEPP reconstruction in Fig. 2.9. The cost for the single-energy MBIR is scaled such that it reaches the same final cost as JDE-MBIR. 39

2.16 Top row shows the JDE-MBIR reconstructions of the GE GSI contrast phantom. This phantom consists of a water phantom base and several cylindrical rods, each of which contains certain concentrations of iodine and water. Display window for water: [700 1300] mg/cm^3; for iodine: [-1.25 16.25] mg/cm^3. Bottom row shows the reconstruction accuracy of FBP and JDE-MBIR for iodine and water. 40

2.17 Comparison of water images reconstructed by FBP, independent DE-MBIR and JDE-MBIR methods for an abdominal clinical scan. Display window for water images: [850 1150] mg/cm^3; for water difference images: [-100 100] mg/cm^3. 42

2.18 Comparison of iodine images reconstructed by FBP, independent DE-MBIR and JDE-MBIR methods for an abdominal clinical scan. Display window for or iodine images: [-2.25 15.25] mg/cm^3; for iodine difference images: [-4.5 4.5] mg/cm^3. 43

2.19 Comparison of 70 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU. 44

2.20 Comparison of 60 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU. 45

Figure Page

2.21 Comparison of 100 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU. 46

2.22 Comparison of 130 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU. 47

3.1 2-D illustration of the tiling method. Each blue grid represents one of nine distinct tilings with 3×3 patches on a 6×6 grid, i.e., $L = 9$, with the center pixel of each patch marked in red. Toroidal boundary condition is considered in this illustration. Note that there are exactly 9 distinct phase shifts of the tiling, each of which is determined by the center pixel of the first patch in the upper-left corner, which corresponds to a distinct pixel location in the shadowed patch. 53

3.2 Figure illustrates the lemma with a 1-D GM distribution. The quadratic function $q(x; x')$ is a surrogate function for the negative log of the GM distribution $f(x)$ at point x'. The surrogate function is a weighted sum of the quadratic exponents of the exponential functions in the GM distribution. The weights π_1, π_2, π_3 give the posterior probabilities of the point x' belonging to different GM components. 57

3.3 The covariance scaling defined in (3.32) and (3.33) with various values of p and α on a log scale. The black dotted line shows the case when no scaling is present, i.e., $p = 0$. When $0 < p \leq 1$, the average eigenvalues $\bar{\tilde{\lambda}}_k$ are "compressed" toward α^2, where eigenvalues further away from α^2 lead to greater change and a larger p results in greater compression. For a fixed value of p, increasing α increases the covariance of each GM component. 61

3.4 1-D illustration of the covariance scaling in (3.32). The 1-D energy function is given by $u(x) = -\log(\phi(x))$ with the Gaussian mixture distribution $\phi(x) = \sum_k \frac{1}{\sqrt{2\pi\lambda_k/\sigma_k^2}} \exp\left\{-\frac{1}{2\lambda_k/\sigma_k^2}(x-\mu_k)^2\right\}$, with λ_k the original variance and the scaling $\sigma_k = (\lambda_k/\alpha^2)^{p/2}$. Left column: (a) varying p with α fixed, (c) the resulting distribution, and (e) the resulting energy function; right column: (b) varying α with p fixed, (d) the resulting distribution, and (f) the resulting energy function. 63

Figure Page

3.5 Partition of the training data. (a): a 2-D axial slice from the 3-D image volume where training patches were extracted; (b)-(g): partition of the data based on the criteria in Table 3.2. Display window: (a) [-160, 240] HU, (b) [-1250, -750] HU, (c) [-1400, 200] HU, (d)-(f) [-210, 290] HU, (g) [-300, 700] HU. Notice that different groups roughly capture different materials or tissue types in the image, as group 1 for air, group 2 for lung tissue, group 3 for smooth soft tissue, group 4 for low-contrast soft-tissue edge, group 5 for high-contrast edge, and group 6 for bone. 66

3.6 The covariances of the originally trained GMM and the adjusted GMM with $p = 0.5$ and $\alpha = 33$ HU in (3.32), for a $5 \times 5 \times 3$ patch case. More precisely, the figure plots the square root of geometrically-averaged eigenvalues $\bar{\lambda}_k$ of the GMM covariances R_k, as $\bar{\lambda}_k = |R_k|^{\frac{1}{L}}$ with $L = 75$. Numbers within the figure correspond to the indices of the training groups. Within each group, the GMM components are sorted from the most probable to the least probable. The adjusted model increases the eigenvalues for group 1, 3, 4, and decreases the eigenvalues for most of group 2, 5, and 6. 68

3.7 Denoising results with different methods. The RMSE value between each reconstructed image and the ground truth is reported. Display window: [-100 200] HU. GM-MRF methods achieve lower RMSE values and better visual quality than q-GGMRF method and K-SVD method. Though having a slightly higher RMSE value, GM-MRF with original model preserves some real texture in soft tissue without creating severe artifacts, while BM3D tends to over-smooth the soft tissue and introducing some artificial, ripple-like structure. In addition, though compromising the RMSE than the original model, GM-MRF with the adjusted model produces better visual quality, especially for the soft-tissue texture. 69

3.8 GEPP reconstruction with data collected under 290 mA. From left to right, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, (c) MBIR with original GM-MRF, and (d) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. The left column shows the wire section and the right column shows the resolution bars with the display window as [-85 165] HU. 73

Figure Page

3.9 Quantitative measurements for GEPP reconstructions. Four different magnitudes of X-ray tube current were used in data acquisition to achieve different X-ray dose levels. The mean values along with noise were measured within two different ROIs in (a). The MTF values were measured at the tungsten wire. Figure demonstrates that MBIR with GM-MRF priors improve the in-plane resolution in (b) while producing comparable or even less noise than FBP and MBIR with the q-GGMRF prior in (c) and (d), without affecting the reconstruction accuracy in (e) and (f). 74

3.10 An abdominal axial slice of the normal-dose clinical reconstruction. From top to bottom, the rows represent (a) FBP (noise: 19.14 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 14.43 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU (noise: 14.02 HU). The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is [-110 190] HU. Note the reduced jagged appearance in the GM-MRF reconstruction. 77

3.11 A sagittal view of the normal-dose clinical reconstruction in bone window. From top to bottom, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Display window is [-300 900] HU. Note the honeycomb structure of the trabecular bone reconstructed by the GM-MRF prior, which is missing in other methods. 78

3.12 An abdominal axial slice of the low-dose clinical reconstruction. From top to bottom, the rows represent (a) FBP (noise: 62.15 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 25.46 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU (noise: 18.41 HU). The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is [-160 240] HU. Note the suppression of speckle noise in soft tissue and improvement of sharpness in bone provided by MBIR with the GM-MRF prior. 79

3.13 A coronal view of the normal-dose clinical reconstruction in lung window. From top to bottom, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows zoomed-in FOVs for lung (top) and spine (bottom). Display window is [-1400 400] HU. Note the lung fissure and bone structure are reconstructed much more clearly by using the GM-MRF prior as compared to the other methods. 80

ABSTRACT

Zhang, Ruoqiao PhD, Purdue University, December 2015. Advanced Statistical Modeling for Model-Based Iterative Reconstruction for Single-Energy and Dual-Energy X-Ray CT. Major Professor: Charles A. Bouman.

Model-based iterative reconstruction (MBIR) has been increasingly broadly applied as an improvement over traditional, analytical image reconstruction methods in X-ray CT, primarily due to its significant advantage in drastic dose reduction without diagnostic loss. Early success of the method in conventional CT has encouraged the extension to a wide range of applications that includes more advanced imaging modalities, such as dual-energy X-ray CT, and more challenging imaging conditions, such as low-dose and sparse-sampling scans, each requiring refined statistical models including the data model and the prior model. In this dissertation, we developed an MBIR algorithm for dual-energy CT that included a joint data-likelihood model to account for correlated data noise. Moreover, we developed a Gaussian-Mixture Markov random filed (GM-MRF) image model that can be used as a very expressive prior model in MBIR for X-ray CT reconstruction. The GM-MRF model is formed by merging individual patch-based Gaussian-mixture models and therefore leads to an expressive MRF model with easily estimated parameters. Experimental results with phantom and clinical datasets have demonstrated the improvement in image quality due to the advanced statistical modeling.

1. INTRODUCTION

As a systematic approach of image reconstruction, the model-based iterative reconstruction (MBIR) algorithm has been increasingly widely applied as an improvement over traditional, deterministic image reconstruction methods for single-energy X-ray CT [1–6]. It has been demonstrated that MBIR can substantially improve image quality by simultaneously reducing noise and improving resolution [7–11]. Importantly, this primary advantage of MBIR can potentially lead to the clinical benefit of drastically reduced X-ray dosage without diagnostic quality loss.

The power of MBIR methods is due to the synergy that results from modeling both the sensor (i.e., forward model) and the image being reconstructed (i.e., prior model). To reconstruct the unknown image x from the measurements y, MBIR algorithms typically work by computing the maximum *a posteriori* (MAP) estimate of x given y, by

$$\hat{x} \leftarrow \arg\min_{x \in \Omega} \{ -\log p(y|x) - \log p(x) \} \ . \tag{1.1}$$

In this framework, $p(y|x)$ is the conditional probability density function of y given x, which comprises the forward model of the measurement process. The density $p(x)$ is the prior model for x, which describes the characteristics of the object being reconstructed.

The achievement of MBIR in conventional CT has encouraged the extension to a wide range of applications in the field of medical imaging. Dual-energy X-ray CT (DECT), for example, is one of the advanced applications that may benefit from applying MBIR methods. Unlike the traditional single-energy CT, by using two different X-ray spectra for imaging, DECT allows reconstruction of the distinctive energy-dependent X-ray attenuation and is therefore capable to identify the materials being reconstructed. However, the traditional deterministic reconstruction approaches for

DECT require a decomposition process that significantly amplifies the noise. Moreover, accurate reconstruction of the data collected with advanced imaging modalities such as the fast kV switching can be challenging due to incomplete and mismatched measurements. Thus, it is interesting to design an MBIR algorithm for DECT that includes an accurate forward model to account for data missing and mismatching, to improve the image quality and accuracy in the reconstructions.

On the other hand, the early success of MBIR has naturally resulted in an increasing number of scans at low dose. Existing MBIR algorithms typically exploit the conventional Markov random field (MRF) model with limited number of parameters as a prior. The simple parameterization of the conventional MRF reduces the model complexity but limits its ability to capture comprehensive characteristics of images. Though very effective under normal condition, this simple structure may be inadequate for challenging imaging conditions such as low-dose imaging, where the signal-to-noise ratio is low. Moreover, due to the simple parameterization, traditional MRFs can only provide single regularization strength for the whole reconstruction. Therefore, it is difficult for MBIR with traditional MRFs to produce a reconstruction that is "optimal" for different tissues. For example, one will need to reduce the total amount of global regularization to render sharper bone structures, which consequently compromises the image quality in soft tissue with a non-negligible increase of noise. Thus, it will be desirable to have an MRF model that is expressive enough to capture important image features while the parameter estimation remains simple.

In this dissertation, we propose MBIR algorithms for the applications of image reconstruction in low-dose single-energy X-ray CT and dual-energy X-ray CT. Firstly, we will present the MBIR algorithm for DECT using a joint quadratic likelihood model that accounts for the correlated noise in the measured data. Secondly, we will present an advanced patch-based image model that can be used as an expressive prior model in MBIR algorithms, which can be extremely useful for challenging imaging conditions such as low-dose imaging.

1.1 MBIR for dual-energy X-ray CT

Recently, DECT has drawn much attention from both academia and industry. By using two different X-ray spectra, DECT can potentially reduce artifacts and improve contrast as compared to conventional single-energy X-ray CT. Moreover, one might also expect to improve resolution and reduce noise by applying MBIR to dual-energy data. However, the direct implementation of MBIR for DECT requires the use of a nonlinear forward model to account for the broad spectra, which increases not only complexity but also computation. Alternatively, one may employ simplified forward models that process the material-decomposed sinograms separately, but these approaches do not fully take into account the statistical dependencies in the sinogram entries.

In Chapter 2, we propose an algorithm for joint dual-energy MBIR (JDE-MBIR), which simplifies the forward model while still takes into account the complete noise correlation in the material-decomposed sinogram components. The JDE-MBIR approach employs a quadratic approximation for the polychromatic log-likelihood function, separate MRFs as the prior for material separation, and a simple but exact non-negativity constraint in the image domain based on the real physical property. The optimization is performed by using iterative coordinate descent (ICD) algorithm with well-known KKT conditions. We demonstrate that our method is particularly effective when the fast kVp switching technique is used in the DECT system, since in this case the model takes into account the potential inaccuracy caused by interpolated sinogram components. Note that this method was presented in a conference [12] and was later published as a journal article [13].

1.2 Advanced prior model for MBIR

Most MBIR algorithms use conventional MRFs as prior models to reconstruct the underlying image. While MRFs provide a simple and often effective way to model the spatial dependencies in images, they suffer from the fact that parameter estimation

is difficult. In practice, this means that MRFs typically have very simple structure that cannot completely capture the subtle characteristics of complex images.

In Chapter 3, we propose a Gaussian-mixture Markov random field (GM-MRF) model that can be used as a very expressive prior model for inverse problems such as denoising and reconstruction. The GM-MRF forms a global image model by merging together individual Gaussian-mixture models (GMMs) for image patches. In addition, we present a novel analytical framework for computing MAP estimates using the GM-MRF prior model through the construction of exact surrogate functions that result in a sequence of quadratic optimizations. We also introduce a simple but effective method to adjust the GM-MRF so as to control the sharpness in low- and high-contrast regions of the reconstruction separately. Note that this method was presented in a conference [14] and has been submitted to a journal for review [15].

2. MODEL-BASED ITERATIVE RECONSTRUCTION FOR DUAL-ENERGY X-RAY CT USING A JOINT QUADRATIC LIKELIHOOD MODEL

2.1 Introduction

Dual-energy CT (DECT) scanners, which acquire X-ray projections with two distinct spectra, are of great interest in applications such as medical imaging [16, 17], security inspection [18, 19], and nondestructive testing [20]. The objective of DECT reconstruction is to determine the energy-dependent attenuation at each voxel. Fortunately, for most materials, the energy-dependent attenuation is accurately approximated as a linear combination of two basis functions corresponding to photoelectric absorption and Compton scattering [21]. In practice, it is usually more convenient to reparameterize the energy-dependent attenuation as a linear combination of two basis materials or components [22] such as water and iodine. So in this case, our objective is then to accurately reconstruct cross-sections corresponding to the equivalent densities of, say, water and iodine.

Early work on dual-energy reconstruction focused on decomposing the dual-energy measurements into two independent sinograms, each of which corresponds to a basis component or material. This can be done by first applying a material-decomposition function to the two energy measurements. This material-decomposition function then produces two sinograms corresponding to the two basis materials. Many methods have been proposed over the years for experimentally determining this function. Alvarez and Macovski [21] proposed the numerical inversion of a polynomial approximation to the polychromatic measurement process. Other approaches directly approximate the material-decomposition function as a polynomial [22–27], or compute the decomposition through an iterative estimation process [19, 28–30].

Alternatively, other approaches to dual-energy reconstruction work by first reconstructing images from the low- and high-energy sinograms using filtered back projection (FBP) method, and then performing image-domain material decomposition [31–34]. However, while sometimes effective, this type of image-domain reconstruction makes substantial approximations, particularly when the X-ray spectrum for each measurement is broad. So the resulting reconstructions may be quantitatively inaccurate and suffer from artifacts. Recently, an iterative FBP method [35] has been proposed to account for the polychromatic spectra. It repeats the process that performs back projection, image-domain material decomposition, and forward projection of the decomposed results with a calibrated nonlinear model. This method can be applied to the case where one of the dual-energy measurements is missing for each ray.

Recently, statistical reconstruction based on iterative methods has been found to be very effective in single-energy CT reconstruction [3–6]; and in particular, model-based iterative reconstruction (MBIR) methods [1–3,6], have demonstrated the ability to reduce noise and improve resolution [7–11].

Several statistical iterative approaches have been proposed for DECT reconstruction. These methods can be mainly classified into two categories, the direct-inversion methods and the decomposition-based methods. The direct-inversion methods reconstruct images directly from dual-energy measurements [36–40]. In particular, Fessler *et al.* [37] formulated the likelihood function of the detector output by using a Poisson model. Huh and Fessler [39] applied a penalized weighted least square (PWLS) approach to DECT with fast kVp switching acquisition and used an approximate Gaussian noise model for the log-transformed measurements. These approaches generally include a highly nonlinear forward model in the likelihood function to model the polychromatic measurement process, so this formulation increases complexity and consequently complicates the optimization.

Alternatively, decomposition-based statistical approaches reconstruct images from material-decomposed sinograms [41–43] with a simplified forward model. Fessler *et*

al. [41,43] applied PWLS approaches with diagonal weighting matrices for the pair of decomposed sinogram entries. These methods, which we refer to as independent dual-energy MBIR (independent DE-MBIR), model the decomposed sinogram entries as statistically independent when conditioned on image content. The independent DE-MBIR methods are computationally simpler than the direct-inversion methods, but the decoupled likelihood functions ignore the correlation in sinogram entries that are caused by the decomposition process [44, 45]. Perhaps the most closely related work is Kinahan, Alessio, and Fessler's [42] method for dual-energy PWLS reconstruction in PET/CT attenuation correction. This framework also allowed for the potential correlation of sinogram entries, but left open the specifics of how the entries should be weighted.

In this chapter, we develop a novel joint dual-energy MBIR (JDE-MBIR) method to reconstruct basis material densities from the decomposed sinograms. In Sec. 2.2.3, we introduce a key novelty of the JDE-MBIR method, which is a quadratic approximation to the joint likelihood model. This quadratic approximation weights the decomposed sinogram entries by non-diagonal matrices that explicitly model the noise correlation in the decomposition domain. The proposed method also incorporates a prior model that accounts for the separation into materials and introduces a simple but exact non-negativity constraint that accurately reflects the true physical constraint of non-negative X-ray attenuation. We use the iterative coordinate descent (ICD) algorithm to compute the solution. We note that this method and associated experimental results have been published in [12, 13].

An important novelty of JDE-MBIR is that it achieves computational efficiency by reconstructing from material-decomposed sinograms while retaining an accurate forward model and noise model in the decomposition domain. In particular, the JDE-MBIR models the interdependence in decomposed sinogram entries that result from the decomposition process. This model leads to reconstructions with less noise than those of the independent DE-MBIR methods.

The JDE-MBIR also allows for accurate modeling of DECT data collected using fast kVp switching techniques. Fig. 2.1 illustrates a model for the fast kVp switching technique, in which the system alternates between low- and high-energy measurements from view to view. In this case, each view contains either low- or high-energy measurements, whereas the material decomposition requires both to be available. Although the angular difference between the low- and high-energy measurements is small, an additional interpolation step needs to be performed for the traditional decomposition-based approaches to ensure perfect registration. The JDE-MBIR method also exploits sinogram interpolation; however, the reconstruction principally depends on only the real measurements, which makes it robust to interpolation error and capable of preserving more resolution than other decomposition-based approaches.

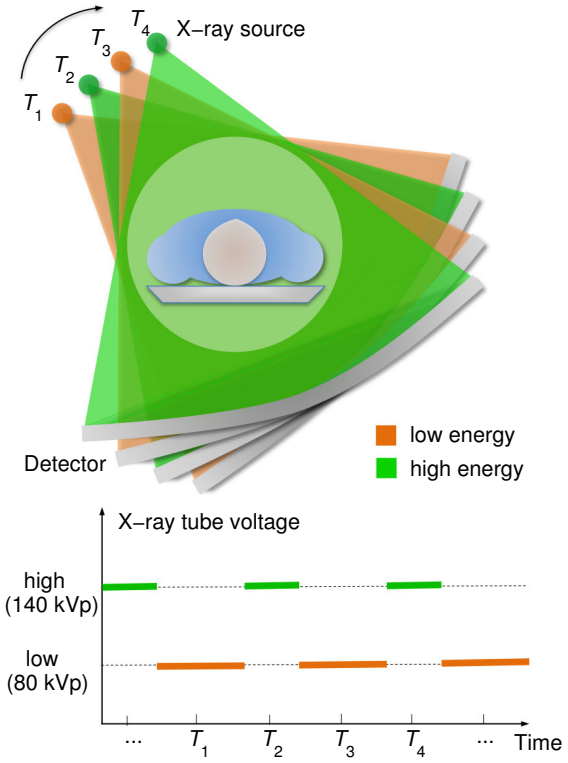


Fig. 2.1. This figure illustrates a model for fast kVp switching technique. A single X-ray tube alternates the voltage between low- and high-kVp from view to view. Thus each view contains only one measurement, either low- or high-energy.

We also propose a novel non-negativity constraint for the DECT reconstruction. Previous approaches have enforced non-negativity constraints directly on reconstructed material densities [36, 39, 46], which is not generally physically correct. We proposed a non-negativity constraint that is applied in the attenuation space. This constraint can be enforced as two simple linear constraints on the material images and accurately reflects the true physical constraints of X-ray attenuation.

In our experiments, we evaluate the performance of the JDE-MBIR by using phantom and clinical data. The experimental results show that the JDE-MBIR significantly improves resolution and reduces noise in the reconstructed material density images and the synthesized monochromatic images.

The chapter is organized as follows. Sec. 2.2 describes the formulation of the JDE-MBIR. Sec. 3.3 gives the ICD solution to the optimization problem. Sec. 2.4 presents the experimental results on phantom and clinical data to demonstrate the image quality improvement achieved by JDE-MBIR as compared to FBP and independent DE-MBIR.

2.2 MAP Reconstruction Framework

Let $y \in \Re^{M\times 2}$ be the set of dual-energy CT measurements, where each row, $y_i = [y_{i,\mathrm{l}},\ y_{i,\mathrm{h}}]$, specifies the low- and high-energy projection measurements for the i^{th} ray. We use subscript "l" for "low-energy" and subscript "h" for "high-energy". Furthermore, let $m \in \Re^{N\times 2}$ be the reconstructed density images of the scanned object for the selected material basis pair, where each row, $m_j = [m_{j,\mathrm{W}},\ m_{j,\mathrm{I}}]$, represents the water- and iodine-equivalent densities for the j^{th} voxel. We use subscript "W" for "water" and subscript "I" for "iodine". In this chapter, we choose water and iodine since they are frequently used as basis materials for separation into low and high X-ray attenuation characteristics, respectively. However, the use of other material pairs is equally valid. The integer M specifies the number of distinct ray paths during data acquisition, and N specifies the number of voxels in the reconstructed volume.

Our goal is to reconstruct the material density images, m, from the measurements, y. One typical approach is to compute the maximum *a posteriori* (MAP) estimate of m given by

$$\hat{m} = \arg\max_{m \in \Omega} \{\log P(y|m) + \log P(m)\}, \tag{2.1}$$

where $P(y|m)$ is the conditional distribution of y given m, $P(m)$ is the prior distribution of m, and Ω is the constraint set for the reconstruction.

The following sections develop the theoretical framework for the JDE-MBIR algorithm from the basic physical models. Section 2.2.1 presents a noise model for the dual-energy detector measurements based on photon statistics. Section 2.2.2 then derives the forward model for the dual-energy data using widely accepted models of polychromatic X-ray attenuation through materials. With this framework in place, Section 2.2.3 then introduces the primary innovation of our technique, which is a quadratic approximation to the log-likelihood function in the projection domain. Section 2.2.4 then shows how this innovative model can be applied to the important problem of fast kVp switching data and provides a theoretical analysis of JDE-MBIR's advantages in this application.

2.2.1 Measurement Preprocessing

In the X-ray transmission problem, we measure the photon flux after object attenuation, which is denoted by $\lambda_{i,k}$ for ray i and source spectrum k, where $k \in \{\mathrm{l}, \mathrm{h}\}$. We also measure the air scan photon flux, $\lambda_{i,k,0}$, which counts the detected photons with no object present. The air scan counts can be calibrated accurately by repeated scans and therefore are assumed noiseless. The projection measurement, y_i, is then computed as the negative log of the photon measurement normalized by the corresponding air scan photon count,

$$y_i = [y_{i,\mathrm{l}},\ y_{i,\mathrm{h}}] \triangleq \left[-\log\left(\frac{\lambda_{i,\mathrm{l}}}{\lambda_{i,\mathrm{l},0}}\right),\ -\log\left(\frac{\lambda_{i,\mathrm{h}}}{\lambda_{i,\mathrm{h},0}}\right) \right]. \tag{2.2}$$

We model $\lambda_{i,k}$ as the summation of a Poisson random variable with conditional mean $\bar{\lambda}_{i,k}$ and a Gaussian random variable with mean zero and variance σ_e^2. In fact, $\bar{\lambda}_{i,k}$ is the conditional mean of $\lambda_{i,k}$ given the image m, and the Gaussian random variable presents additive electronic noise in the detector system. From this, the conditional mean and variance of $\lambda_{i,k}$ are given by

$$\begin{aligned} \mathrm{E}[\lambda_{i,k}|m] &= \bar{\lambda}_{i,k}, & (2.3) \\ \mathrm{Var}(\lambda_{i,k}|m) &= \bar{\lambda}_{i,k} + \sigma_e^2. & (2.4) \end{aligned}$$

Then we approximate the conditional mean of $y_{i,k}$ as

$$\mathrm{E}[y_{i,k}|m] \cong -\log\left(\frac{\bar{\lambda}_{i,k}}{\lambda_{i,k,0}}\right). \tag{2.5}$$

To compute the conditional variance of $y_{i,k}$, we first perform a first-order Taylor series expansion to the expression of $y_{i,k}$ in (2.2) about $\bar{\lambda}_{i,k}$,

$$\begin{aligned} y_{i,k} &= -\log\left(\frac{\lambda_{i,k}}{\lambda_{i,k,0}}\right) \\ &\cong -\log\left(\frac{\bar{\lambda}_{i,k}}{\lambda_{i,k,0}}\right) - \frac{1}{\bar{\lambda}_{i,k}}(\lambda_{i,k} - \bar{\lambda}_{i,k}) \\ &\cong \mathrm{E}[y_{i,k}|m] - \frac{1}{\bar{\lambda}_{i,k}}\left(\lambda_{i,k} - \bar{\lambda}_{i,k}\right), & (2.6) \end{aligned}$$

which yields the approximation we will use for the conditional variance of $y_{i,k}$ [47],

$$\mathrm{Var}(y_{i,k}|m) \cong \frac{\mathrm{Var}(\lambda_{i,k}|m)}{\bar{\lambda}_{i,k}^2} = \frac{\bar{\lambda}_{i,k} + \sigma_e^2}{\bar{\lambda}_{i,k}^2} \cong \frac{\lambda_{i,k} + \sigma_e^2}{\lambda_{i,k}^2}, \tag{2.7}$$

where $\bar{\lambda}_{i,k}$ is approximated by its observation, $\lambda_{i,k}$. Thus, we will model the conditional mean and covariance of y_i as

$$\begin{aligned} \mathrm{E}[y_i|m] &= [\mathrm{E}[y_{i,\mathrm{l}}|m],\ \mathrm{E}[y_{i,\mathrm{h}}|m]]\,, & (2.8)\\ \mathrm{Cov}(y_i|m) &= \begin{bmatrix} \mathrm{Var}(y_{i,\mathrm{l}}|m) & 0 \\ 0 & \mathrm{Var}(y_{i,\mathrm{h}}|m) \end{bmatrix}. & (2.9) \end{aligned}$$

Note that the off-diagonal elements of the covariance matrix are zeros since we assume that the low- and high-energy measurements are made independently.

Assuming y_i is conditionally Gaussian with mean and covariance given by (2.8) and (2.9), the distribution of y_i is given by

$$-\log P(y_i|m) = \frac{1}{2}\left(y_i - \mathrm{E}[y_i|m]\right) W_i \left(y_i - \mathrm{E}[y_i|m]\right)^T + C, \tag{2.10}$$

where C is a normalizing constant, and W_i is the inverse covariance of y_i,

$$W_i = \begin{bmatrix} w_{i,\mathrm{l}} & 0 \\ 0 & w_{i,\mathrm{h}} \end{bmatrix} \triangleq \mathrm{Cov}^{-1}(y_i|m), \tag{2.11}$$

where

$$\begin{aligned} w_{i,\mathrm{l}} &= \frac{1}{\mathrm{Var}(y_{i,\mathrm{l}}|m)} \cong \frac{\lambda_{i,\mathrm{l}}^2}{\lambda_{i,\mathrm{l}} + \sigma_e^2}, & (2.12)\\ w_{i,\mathrm{h}} &= \frac{1}{\mathrm{Var}(y_{i,\mathrm{h}}|m)} \cong \frac{\lambda_{i,\mathrm{h}}^2}{\lambda_{i,\mathrm{h}} + \sigma_e^2}. & (2.13) \end{aligned}$$

With the assumption of measurements at distinct projections being conditionally independent, the distribution of the data given the object information is given by,

$$-\log P(y|m) = \frac{1}{2}\sum_{i=1}^{M}\left(y_i - \mathrm{E}[y_i|m]\right) W_i \left(y_i - \mathrm{E}[y_i|m]\right)^T + C. \tag{2.14}$$

However, this function is still a nonlinear function of m because the conditional expectation, $\mathrm{E}[y_i|m]$, is in general a nonlinear function of the argument m. In Section 2.2.3, we will use this result to construct a fully quadratic approximation to the log-likelihood in (2.14).

2.2.2 Forward Model

We next need to formulate a physical model for $\mathrm{E}[y_i|m]$. Given the linear attenuation coefficients, μ, the conditional mean of $\lambda_{i,k}$ is computed by integrating the photon attenuation over the source spectrum,

$$\mathrm{E}[\lambda_{i,k}|\mu] = \bar{\lambda}_{i,k} = \int_{\Re} \lambda_{i,k,0}\ S_k(\mathcal{E}) e^{-\int_{\mathrm{ray}_i} \mu(r,\mathcal{E})\mathrm{d}r} \mathrm{d}\mathcal{E}, \tag{2.15}$$

where $\mathcal{E}$ (keV) is the photon energy, $S_k(\mathcal{E})$ is the normalized photon energy distribution for source spectrum k, and $\mu(r,\mathcal{E})$ (cm^{-1}) is the linear attenuation coefficient as a function of location r and energy $\mathcal{E}$, representing X-ray photon absorption per unit distance. Since μ and m contain the same information, note that $\mathrm{E}[\cdot|\mu] = \mathrm{E}[\cdot|m]$.

If we discretize $\mu(r,\mathcal{E})$, then this leads to the expression

$$\mathrm{E}[\lambda_{i,k}|\mu] = \int_{\Re} \lambda_{i,k,0}\ S_k(\mathcal{E}) e^{-\sum_{j=1}^{N} A_{i,j}\mu_j(\mathcal{E})} \mathrm{d}\mathcal{E}, \tag{2.16}$$

where $A \in \Re^{M\times N}$ is the projection matrix, with its element, $A_{i,j}$ (cm), representing the intersection of ray i with voxel j. We use the distance driven approach [48] to compute A. Then from (2.5), the conditional mean of the projection measurement is given by

$$\mathrm{E}[y_i|\mu] = -\log\left(\int_{\Re} S(\mathcal{E}) e^{-\sum_{j=1}^{N} A_{i,j}\mu_j(\mathcal{E})} \mathrm{d}\mathcal{E}\right), \tag{2.17}$$

where $y_i = [y_{i,\mathrm{l}}, y_{i,\mathrm{h}}]$ and $S(\mathcal{E}) = [S_{\mathrm{l}}(\mathcal{E}), S_{\mathrm{h}}(\mathcal{E})]$.

Moreover, the linear attenuation coefficient can be expressed as a linear combination of the mass attenuation functions of two or more basis materials [22]. With water and iodine as the basis, the linear attenuation function can be decomposed as

$$\mu_j(\mathcal{E}) = m_{j,\mathrm{W}}\varphi_{\mathrm{W}}(\mathcal{E}) + m_{j,\mathrm{I}}\varphi_{\mathrm{I}}(\mathcal{E}), \tag{2.18}$$

where $m_{j,s}$ (mg/cm^3) is the equivalent density for basis material s at voxel j, where $s \in \{\mathrm{W}, \mathrm{I}\}$, and $\varphi_s(\mathcal{E})$ (cm^2/mg) is the known energy-dependent mass attenuation function for basis material s, which represents the photon absorption per unit distance for the particular material with 100% concentration under standard temperature and pressure. Note that the reconstructions, $m_{j,\mathrm{W}}$ and $m_{j,\mathrm{I}}$, do not depend on energy.

Then, by substituting (2.18) into (2.17), we have

$$\begin{aligned}\mathrm{E}[y_i|m] &= -\log\left(\int_{\Re} S(\mathcal{E})e^{-\sum_{j=1}^{N} A_{i,j}\left(m_{j,\mathrm{W}}\varphi_{\mathrm{W}}(\mathcal{E})+m_{j,\mathrm{I}}\varphi_{\mathrm{I}}(\mathcal{E})\right)}\mathrm{d}\mathcal{E}\right) \\ &\triangleq -\log\left(\int_{\Re} S(\mathcal{E})e^{-p_i(\varphi(\mathcal{E}))^T}\mathrm{d}\mathcal{E}\right),\end{aligned} \tag{2.19}$$

where $\varphi(\mathcal{E}) \triangleq [\varphi_{\mathrm{W}}(\mathcal{E}),\ \varphi_{\mathrm{I}}(\mathcal{E})]$, and p_i (mg/cm^2) is the material density projection defined as

$$p_i \triangleq [p_{i,\mathrm{W}},\ p_{i,\mathrm{I}}] \triangleq \left[\sum_{j=1}^{N} A_{i,j}m_{j,\mathrm{W}},\ \sum_{j=1}^{N} A_{i,j}m_{j,\mathrm{I}}\right] \triangleq [Am]_i. \tag{2.20}$$

The quantity p_i represents the line integral of material densities along ray i. We then define a vector-valued function, $h : \Re^2 \to \Re^2$, as

$$h(p_i) \triangleq -\log\left(\int_{\Re} S(\mathcal{E})e^{-p_i(\varphi(\mathcal{E}))^T}\mathrm{d}\mathcal{E}\right), \tag{2.21}$$

which models the nonlinear relationship between the material density projections and the expected photon attenuation. From this, we have

$$\mathrm{E}[y_i|m] = h([Am]_i). \tag{2.22}$$

The function h does not depend on particular ray paths generally; however, it can be a function of the ray index, i, if the source spectrum, $S(\mathcal{E})$, varies among rays. This is the case in practice with systems including bowtie filters to shape the X-ray beam to a particular scanned object.

Thus, substituting (2.22) into (2.14), we have the negative log-likelihood function,

$$-\log P(y|m) = \frac{1}{2}\sum_{i=1}^{M} \left(y_i - h([Am]_i)\right) W_i \left(y_i - h([Am]_i)\right)^T + C. \tag{2.23}$$

This is the likelihood function used in the direct-inversion methods.

While the forward model of (2.23) could be used directly for MBIR reconstruction, it is not practical for a number of reasons. First, the function h is generally not measured on real CT systems. In practice, real CT systems require a knowledge of the material decomposition function, h^{-1}, as described in the following section. This is because h^{-1} is required for implementation of standard direct reconstruction methods such as FBP; so it is accurately measured using calibration procedures. However, h is not easily computed from h^{-1} and would require a completely separate calibration procedure. Second, direct nonlinear optimization of the MAP cost function using (2.23) would be very complex and potentially very computationally expensive since it does not have a quadratic form. So our goal will next be to derive a quadratic function that accurately approximates (2.23).

2.2.3 Quadratic Joint Likelihood Model

In this section, we introduce a quadratic approximation to the negative log-likelihood function, $-\log P(y|m)$, which reduces the complexity of the reconstruction algorithm while still retaining an accurate model of the noise correlation in the decomposition domain.

We first define the inverse function, $h^{-1} : \Re^2 \rightarrow \Re^2$, as

$$h^{-1}(h(p_i)) \triangleq p_i \ . \tag{2.24}$$

In practice, the h^{-1} function is called the "material decomposition function". There are a variety of means to determine this function. One may employ a polynomial approximation to the h function and then solve numerically for p_i [21], or directly approximate the h^{-1} function as a polynomial [22–27]. The coefficients of the polynomial approximations can be determined empirically by system calibration. Possible calibration methods include a projection-domain calibration [26,49], or an image-domain approach [27]. One may also compute the decomposition through an iterative estimation process [19, 28–30]. In practice, we approximate the h^{-1} function as a high order polynomial through calibration, which will be described in detail in Sec. 2.4.

With the h^{-1} function, we can then compute the decomposed sinogram entries, $\hat{p}_i = [\hat{p}_{i,\mathrm{W}}, \hat{p}_{i,\mathrm{I}}]$, as

$$\hat{p}_i \triangleq h^{-1}(y_i), \tag{2.25}$$

with $\hat{p}_i$ an estimate of the material density projection, p_i. Performing a Taylor series expansion of $[Am]_i$ at y_i yields,

$$\begin{aligned} [Am]_i &= h^{-1}(h([Am]_i)) \\ &\cong h^{-1}(y_i) + (h([Am]_i) - y_i) \left[\nabla h^{-1}(y_i)\right] \\ &= \hat{p}_i + (h([Am]_i) - y_i) \left[\nabla h^{-1}(y_i)\right], \end{aligned} \tag{2.26}$$

where $\nabla h^{-1}(y_i)$ is the gradient of function h^{-1} at y_i. For our problem, $\nabla h^{-1}(y_i)$ is a 2×2 invertible matrix. This results in the linear approximation we will use in the model,

$$y_i - h([Am]_i) \cong (\hat{p}_i - [Am]_i) \left[\nabla h^{-1}(y_i)\right]^{-1}. \tag{2.27}$$

Thus by substituting (2.27) into (2.23), we approximate the true log-likelihood function in (2.23) by

$$-\log P(y|m) \cong \frac{1}{2} \sum_i (\hat{p}_i - [Am]_i) B_i (\hat{p}_i - [Am]_i)^T + C', \tag{2.28}$$

where the estimated material projection, $\hat{p}_i$, is given by (2.25), and the statistical weighting matrix, B_i, is given by

$$B_i \triangleq [\nabla h^{-1}(y_i)]^{-1} W_i [\nabla h^{-1}(y_i)]^{-T}. \tag{2.29}$$

Each B_i is a 2×2 symmetric matrix representing the inverse covariance of the decomposed sinogram entries, $\hat{p}_i$. Each B_i is therefore also positive semi-definite and has a zero eigenvalue if and only if the diagonal matrix, W_i, has a zero eigenvalue, which implies that $\lambda_{i,\mathrm{l}} = 0$ in (2.12) or $\lambda_{i,\mathrm{h}} = 0$ in (2.13).

The equation (2.28) gives the likelihood model we use in the proposed JDE-MBIR method. In contrast to the direct-inversion methods, our model has a simple quadratic form, so it allows for direct application of existing quadratic optimization methods for the computation of the MAP reconstruction.

It should also be noted that our weighting matrix, B_i, is in general non-diagonal for every projection. The off-diagonal elements of B_i provide significant information about the noise correlation between distinct decomposed sinogram entries.

2.2.4 Likelihood Model For Fast kVp Switching Modality

Our proposed model is particularly well suited for CT systems that use fast kVp switching to acquire dual-energy data. Fig. 2.1 graphically illustrates a model for the fast kVp switching technique, in which the system alternates between low- and high-energy measurements between adjacent views. Fast kVp switching requires high-speed detectors and X-ray sources and generators that allow for fast switching, but it offers the advantage that low- and high-energy measurements are interlaced closely in time and space so that misalignments due to motion or other effects are minimized. Notice that a fast kVp switching system has the capability of varying the duty cycle between low- and high-energy views to optimize dose. This is graphically depicted by the fact that the low-energy (orange) line may be longer than the high-energy (green) line. For typical scans, the low-energy dwell time is greater than the high-

energy dwell time, but this depends on many details of the scan parameters. While the dwell time may be different from view to view, data acquisition still results in consecutive views alternating between low- and high-energy measurements. Dwell time characteristics are taken into account in h^{-1} by calibration. The dwell time of the view generally will affect the resulting noise variance, with longer dwell times reducing noise variance and short ones increasing variance. However, this change in variance is fully accounted for by the estimates of noise variance given in Sec. 2.2.1 and more specifically equation (2.7).

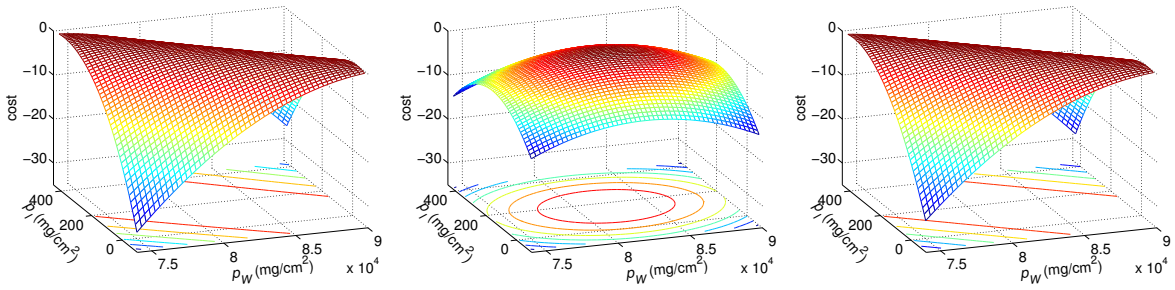
For fast kVp switching, each projection contains either low- or high-energy measurements. Therefore, for each projection, either $y_{i.\mathrm{l}}$ or $y_{i,\mathrm{h}}$ is missing. In the case of the true likelihood of equation (2.23), this missing measurement can be accommodated by setting the weighting matrix to be

$$W_i = \begin{cases} \begin{bmatrix} w_{i,\mathrm{l}} & 0 \\ 0 & 0 \end{bmatrix}, & \text{for low-energy projections;} \\ \begin{bmatrix} 0 & 0 \\ 0 & w_{i,\mathrm{h}} \end{bmatrix}, & \text{for high-energy projections.} \end{cases} \tag{2.30}$$

So in this case, the missing measurement is always weighted by zero.

However, in the case of the joint approximation in (2.28), we still must determine a value for the weighting matrix B_i from equation (2.29) and the estimated projection, $\hat{p}_i$, from equation (2.25). Unfortunately, both these values depend on the missing measurement. In order to solve this problem, we interpolate the missing value of y_i, and use this interpolated value to compute both the gradient of h^{-1} used for the weight matrix B_i, and the projection $\hat{p}_i$. While this interpolation process does introduce error, this error is relatively minor when the joint log-likelihood approximation is used, which allows using relatively straightforward interpolation techniques without significant concern for the quality of the reconstructed images. In order to see this, consider the plots of Figs 2.2 and 2.3.

Fig. 2.2 graphically illustrates the importance of using the joint log-likelihood approximation rather than the simplified independent approximation using some typical values of $[y_{i,\mathrm{l}}, y_{i,\mathrm{h}}] = [3.9, 3.8]$ at 80 kVp and 140 kVp. In the independent approximation, the off-diagonal entries of B_i are set to zero, so the errors in $\hat{p}_{i,\mathrm{W}}$ and $\hat{p}_{i,\mathrm{I}}$ are modeled as independent, and the approximated log-likelihood function has ellipsoidal level curves. This is a very poor approximation of the true log-likelihood and artificially imposes a penalty for any deviation from its unique maximum. With incorporation of the off-diagonal terms in B_i, the joint approximation is much more accurate. The joint approximation appropriately retains the under-specified nature of the maximum-likelihood projection estimate, allowing it to move along its level lines without change in cost.



(a) true log-likelihood (b) independent approximation (c) joint approximation

Fig. 2.2. Plots of the true log-likelihood function, the independent approximation, and the proposed joint approximation, with contours plotted underneath.

Fig. 2.3 illustrates more precisely the effect of interpolation error in the joint and independent log-likelihood approximations. The figure shows contour plots of the log-likelihood function of Fig. 2.2, but also shows the effect of a 5% interpolation error in the missing sample. More precisely, without loss of generality, we fix the air scan photon flux as $[\lambda_{\mathrm{l},0}, \lambda_{\mathrm{h},0}] = [5000, 5000]$, and then simulate the low, medium, and high attenuation cases with photon measurements $[\lambda_{\mathrm{l}}, \lambda_{\mathrm{h}}]$ as $[2500, 2650]$, $[500, 550]$, and $[100, 110]$, respectively. The interpolation error has a relatively minor effect on the joint approximation, while shifting quite significantly the maximum of the inde-

pendent model. Intuitively, the joint approximation is very robust to interpolation error because the weighting matrix, B_i, has a zero eigenvalue in the direction of any interpolation error. This is due to the fact that the underlying matrix W_i of equation (2.29) has a zero in the location of the missing sample. In practice, we will see that the independent approximation results in reconstructions which lose resolution due to the interpolation process. Alternatively, the joint approximation approach preserves resolution by primarily depending on only the uninterpolated samples in the reconstruction.

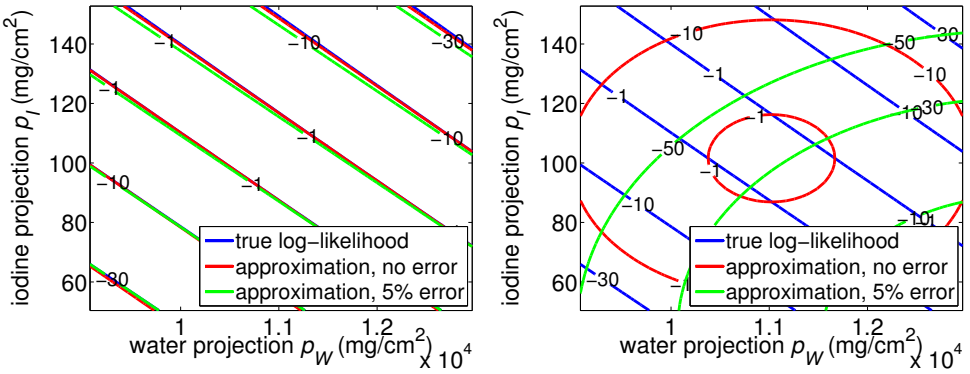
Fig. 2.4 provides a pseudocode of the joint log-likelihood approximation for fast kVp switching. First, the missing measurement is interpolated, and then the interpolated value is used to compute the material decomposition estimate, $\hat{p}_i$, and the projection weighting matrix, B_i.

2.2.5 Prior Model

We model the reconstructed density image as a Markov random field (MRF) with the following form

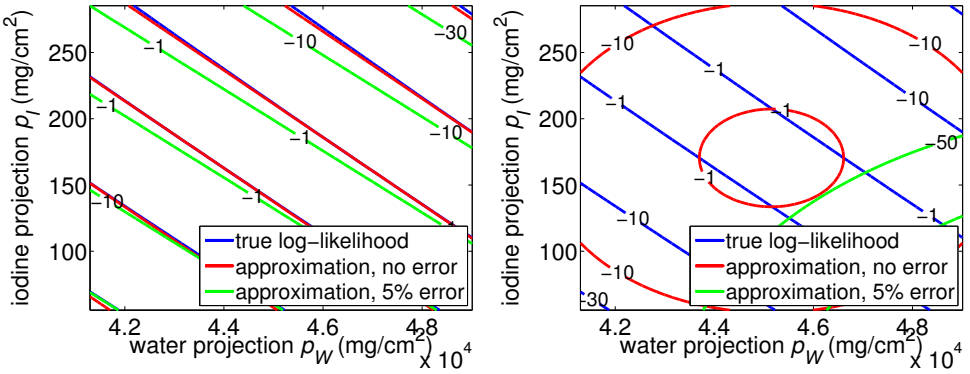
$$-\log P(m) = \sum_{s\in\{\mathrm{W,I}\}} \sum_{\{j,r\}\in\mathcal{C}} b_{jr,s}\rho_s(m_{j,s} - m_{r,s}), \tag{2.31}$$

where s is the index of material type, $\{j, r\}$ specifies a neighboring pair consisting of voxel j and voxel r, $\mathcal{C}$ represents the set of all such voxel pairs, $b_{jr,s}$ is the prior strength for voxel pair $\{j, r\}$ and material s, and $\rho(.)$ is the potential function. We choose $b_{jr,s}$ to be inversely proportional to the distance between voxel j and voxel r, and the scale of $b_{jr,s}$ can be further adjusted to balance between noise and resolution in the reconstruction. By choosing this model, we perform the regularization independently on each of the material components in the image domain.



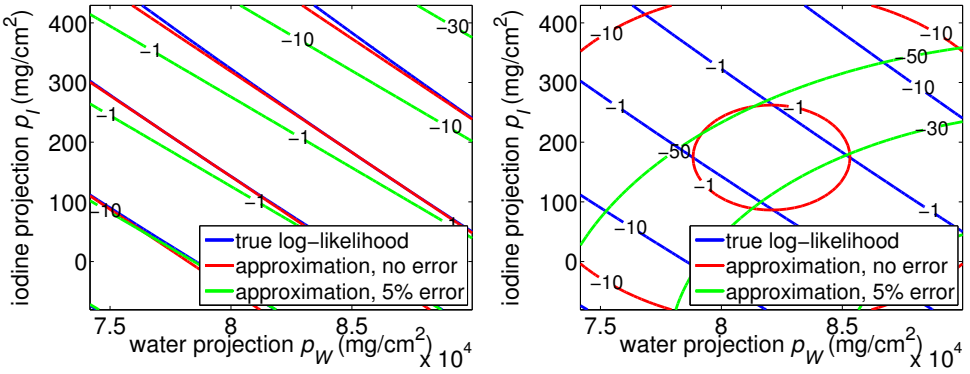
(a) joint, low attenuation

(b) independent, low attenuation



(c) joint, medium attenuation

(d) independent, medium attenuation



(e) joint, high attenuation

(f) independent, high attenuation

Fig. 2.3. Figure plots contours of the true log-likelihood function and different approximations. Independent and joint models are compared at three attenuation levels. Within each subfigure, we plot the contours of the true log-likelihood (blue), approximation without interpolation error (red), and with 5% interpolation error (green). Each plot covers two standard deviations of water and iodine projections.

JDE-MBIR($y_{\rm l}, y_{\rm h}, w_{\rm l}, w_{\rm h}, h^{-1}, \nabla h^{-1}$){
for $i = 1$ to M **do**
  if $w_{i,{\rm h}} = 0$ **then**
    $y_{i,{\rm h}} \leftarrow$ Interpolation
  else if $w_{i,{\rm l}} = 0$ **then**
    $y_{i,{\rm l}} \leftarrow$ Interpolation
  end if
  $y_i \leftarrow [y_{i,{\rm l}},\ y_{i,{\rm h}}]$
  $W_i \leftarrow diag\{w_{i,{\rm l}},\ w_{i,{\rm h}}\}$
  $\hat{p}_i \leftarrow h^{-1}(y_i)$ $\quad \{Material\ decomposition\}$
  $B_i \leftarrow [\nabla h^{-1}(y_i)]^{-1} W_i [\nabla h^{-1}(y_i)]^{-T}$
end for
$\hat{m} \leftarrow$ ImageRecon$(\hat{p}, B)$
return $\hat{m}$
}

Fig. 2.4. Pseudocode of JDE-MBIR for DECT that uses fast kVp switching. First, we interpolate the missing sample for each projection. Second, we perform material decomposition and also compute the statistical weighting matrix. Finally, we use the decomposed sinograms and weighting matrices to reconstruct the images iteratively. The subroutine ImageRecon is described in Fig. 2.8.

Our particular choice of potential function is the q-generalized Gaussian MRF (q-GGMRF) of the form

$$\rho_s(\Delta) = \frac{|\Delta|^p}{1 + |\Delta/c_s|^{p-q}}, \tag{2.32}$$

with $1 < q \leq p \leq 2$. This type of prior has shown to be effective in many tomographic reconstruction studies [3, 6, 50, 51]. With $1 < q \leq p \leq 2$, the potential function is strictly convex [3], which guarantees global convergence of the cost function and produces reconstruction as a continuous function of the data [52].

We set $p = 2.0$ and $q = 1.2$ in our application, since this particular setting has shown a desirable compromise between noise and resolution in similar clinical studies [3]. With $p = 2.0$, the potential approximates a quadratic function for small voxel differences, which preserves details in low contrast regions. The value, $q = 1.2$, approximates the behavior of a generalized Gaussian MRF [52] for large voxel

differences, which preserves edges in high contrast regions. The parameter c models the transition between low and high contrast contents. In practice, we choose $c_{\text{W}} = 10$ mg/cm^3 for water image and $c_{\text{I}} = 0.5$ mg/cm^3 for iodine image.

2.2.6 Constrained Optimization

In X-ray tomographic reconstruction problems, an important physical constraint to the reconstruction is that the linear attenuation of any material at any photon energy must be non-negative. More precisely, for all $\mathcal{E} \in [40, 140]$ keV, we know that

$$\mu_j(\mathcal{E}) = m_j \cdot \varphi(\mathcal{E}) \geq 0 \ , \tag{2.33}$$

where the photon energy range [40, 140] keV is of particular interest for medical imaging and is above the k-edge of iodine.

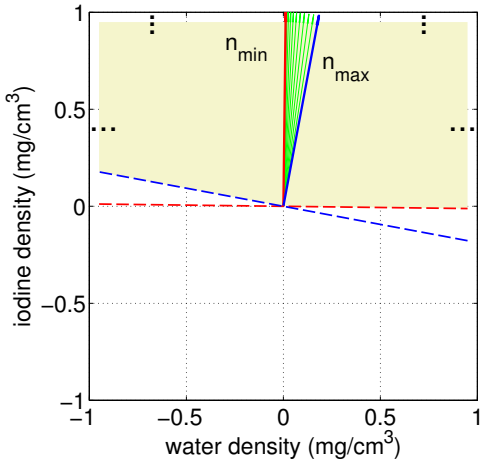


Fig. 2.5. This figure illustrates the feasible values of a voxel, $m_j = [m_{j,\text{W}}, m_{j,\text{I}}]$, where $m_{j,\text{W}}$ is the water-equivalent density and $m_{j,\text{I}}$ is the iodine-equivalent density. The yellow region shows the constrained solution set, which is formed by the intersection of only two half planes, one defined by $n_{\max}$ and the other by $n_{\min}$. The green arrows show the attenuation vectors at intermediate energies.

Let Ω' be the constraint set of a single voxel value, as

$$\Omega' = \cap_{\mathcal{E}\in[40,140]}\{m_j \in \Re^2 : m_j \cdot n(\mathcal{E}) \geq 0\}, \tag{2.34}$$

where $n(\mathcal{E}) \triangleq \frac{\varphi(\mathcal{E})}{||\varphi(\mathcal{E})||}$ is the normalized mass attenuation vector. In this way, Ω' is formed by the intersection of an infinite number of half planes. However, the form of Ω' can be dramatically simplified by observing that the direction of $n(\mathcal{E})$ moves continuously with $\mathcal{E}$. As a consequence, the constraint can be represented much more simply by the intersection of only two planes corresponding to the minimum and maximum values of $n(\mathcal{E})$, as $n_{\min} = n(40)$ and $n_{\max} = n(140)$,

$$\Omega' = \left\{m_j \in \Re^2 : m_j \cdot n_{\min} \geq 0 \text{ and } m_j \cdot n_{\max} \geq 0\right\}, \tag{2.35}$$

Fig. 2.5 illustrates the constraint set and the associated mass attenuation vectors. Then the constraint set for the entire image, denoted by Ω, is given by

$$\Omega = {\Omega'}^N, \tag{2.36}$$

where N is the number of voxels in the reconstructed volume. Clearly, Ω' is a convex set and so is Ω.

The proposed constraint allows negative values for the reconstructed densities of water and iodine. This is because the reconstructed densities are only some coefficients for the linear combination that produces the equivalent attenuation. However, in the attenuation domain, the combination of the reconstructed material densities should remain non-negative.

Combining the log-likelihood in (2.28) and the prior in (2.31) with the constraints in (2.36) yields the expression for the MAP reconstruction of equation (2.1),

$$\hat{m} = \arg\min_{m\in\Omega}\left\{\frac{1}{2}\sum_{i=1}^{M}\left(\hat{p}_i - [Am]_i\right)B_i\left(\hat{p}_i - [Am]_i\right)^T + \sum_{s\in\{\mathrm{W,I}\}}\sum_{\{j,r\}\in\mathcal{C}} b_{jr,s}\rho(m_{j,s} - m_{r,s})\right\}. \tag{2.37}$$

2.3 Optimization Algorithm

There are a wide variety of techniques that can be used to solve the optimization problem in (2.37), from which we choose the iterative coordinate descent (ICD) algorithm. The ICD algorithm has the advantages that it has rapid convergence at high spatial frequencies [1], especially when initialized with FBP to obtain a good original estimate of low frequencies. Moreover, it can easily incorporate the proposed non-negativity constraint.

The ICD algorithm sequentially updates voxels of the reconstructed image. Within each ICD iteration, every single voxel is updated with remaining voxels fixed so as to minimize the total cost function. Within each ICD update, we compute the exact solution to the constrained voxel update with the Karush-Kuhn-Tucker (KKT) condition.

More precisely, by changing only one voxel while fixing the rest of the image, we compute the voxel update, $\hat{m}_j$, from the current image, m, by

$$\hat{m}_j \leftarrow \arg\min_{u\in\Omega'} \left\{ \frac{1}{2}\sum_{i=1}^{M} \|\hat{p}_i - [Am]_i + A_{i,j}\,(m_j - u)\|_{B_i}^2 + \sum_{s\in\{\mathrm{W},\mathrm{I}\}}\sum_{r\in\partial j} b_{jr,s}\rho\,(u_s - m_{r,s}) \right\}, \tag{2.38}$$

We denote $\|x\|_B^2 = xBx^T$ for simplicity. We introduce a dummy variable $u = [u_{\mathrm{W}}, u_{\mathrm{I}}]$ to represent the voxel value being updated, to distinguish from its current value, m_j. Define the error sinogram, $e \triangleq Am - \hat{p}$. Then intuitively, the first term in equation (2.38) describes the change in the error sinogram introduced by the change in the voxel value. Equivalently, equation (2.38) can be written as

$$\hat{m}_j \leftarrow \arg\min_{u\in\Omega'} \left\{ (u - m_j)\theta_1 + \frac{1}{2}\|u - m_j\|_{\theta_2}^2 + \sum_{s\in\{\mathrm{W},\mathrm{I}\}}\sum_{r\in\partial j} b_{jr,s}\rho(u_s - m_{r,s}) + const. \right\}. \tag{2.39}$$

where θ_1 and θ_2 are the first and second derivatives of the log-likelihood function, which are given by

$$\theta_1 \triangleq \sum_{i=1}^{M} A_{i,j} B_i e_i^T, \tag{2.40}$$

$$\theta_2 \triangleq \sum_{i=1}^{M} A_{i,j}^2 B_i. \tag{2.41}$$

where $e_i = [Am]_i - \hat{p}_i$ is the i^{th} row of the error sinogram, and $A_{i,j}$ is a scalar representing the intersection of ray i with voxel j.

Solving the 2-D optimization problem in (2.39) simultaneously for both material components may be difficult, since the prior term cannot be explicitly expressed as a function of u. To address this problem, one may use a functional substitution approach [6, 53–56]. In this problem, we introduce a quadratic substitute function for the potential function. More precisely, let $\Delta = u_s - m_{r,s}$ and $\Delta^* = m_{j,s} - m_{r,s}$. Then we define the substitute function, $q(\Delta; \Delta^*)$, as

$$q(\Delta; \Delta^*) = \frac{\alpha_{jr,s}}{2}\Delta^2 + C_{jr,s}, \tag{2.42}$$

with

$$\begin{aligned} \alpha_{jr,s} &= \frac{\rho'(\Delta^*)}{\Delta^*}, \\ C_{jr,s} &= \rho(\Delta^*) - \frac{\rho'(\Delta^*)}{2}\Delta^*, \end{aligned} \tag{2.43}$$

where $C_{jr,s}$ is an offset constant and therefore can be ignored during optimization. This function, $q(\Delta; \Delta^*)$, satisfies the following two constraints for a valid substitute function [6, 53–55].

$$\begin{aligned} q(\Delta^*; \Delta^*) &= \rho(\Delta^*), \\ q(\Delta; \Delta^*) &\geq \rho(\Delta). \end{aligned}$$

Intuitively, a valid substitute function for minimization should equal the true function at the current point and upper bound the true function everywhere else. Fig. 2.6 illustrates the desired substitute function, $q(\Delta;\Delta^*)$. It is important to know that replacing the true potential function with the substitute function still guarantees monotone convergence of the cost function [6, 55].

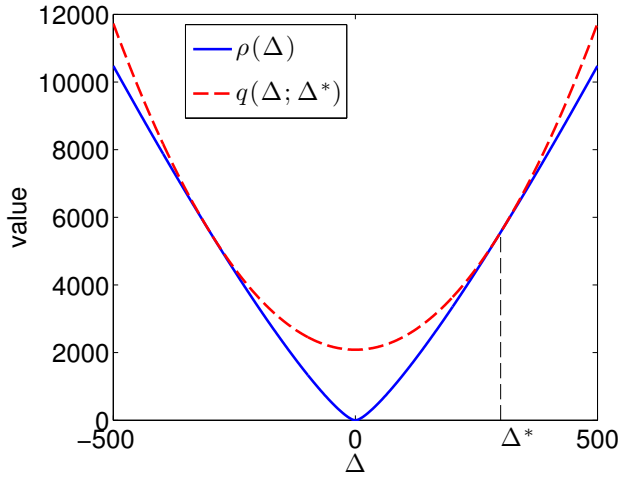


Fig. 2.6. This figure illustrates the desired substitute function. The substitute function $q(\Delta;\Delta^*)$ upper bounds the true function $\rho(\Delta)$ except for the fixed point $\Delta = \Delta^*$, where two functions are equal. Thus, the true function must decrease when the substitute function is minimized.

Replacing the potential function in (2.39) with the substitute function yields a quadratic cost function of u,

$$\hat{m}_j \leftarrow \arg\min_{u\in\Omega'} \left\{ (u-m_j)\theta_1 + \frac{1}{2}\|u-m_j\|^2_{\theta_2} + \frac{1}{2}\sum_{r\in\partial j} \|u-m_r\|^2_{\psi_r} + const. \right\}, \quad (2.44)$$

where

$$\psi_r \triangleq \begin{bmatrix} b_{jr,\mathrm{W}}\alpha_{jr,\mathrm{W}} & 0 \\ 0 & b_{jr,\mathrm{I}}\alpha_{jr,\mathrm{I}} \end{bmatrix}. \quad (2.45)$$

Furthermore, define

$$\phi_1 \triangleq \theta_1 - \theta_2 m_j^T - \sum_{r \in \partial j} \psi_r m_r^T, \tag{2.46}$$

$$\phi_2 \triangleq \theta_2 + \sum_{r \in \partial j} \psi_r. \tag{2.47}$$

By using ϕ_1 and ϕ_2, we rewrite (2.44) into a standard form,

$$\begin{aligned} \min_{u} \quad & \frac{1}{2} u \phi_2 u^T + u\phi_1 + const. \\ \text{s.t.} \quad & u \cdot n_{\min} \geq 0 \\ & u \cdot n_{\max} \geq 0 \end{aligned} \tag{2.48}$$

This is a standard quadratic minimization problem with two linear constraints. It can be solved exactly by applying the KKT condition following a standard procedure [57]. Fig. 2.7 shows the procedure for computing the solution. We first test the KKT condition on the unconstrained solution. If it fails, we solve the minimization problem on either boundary of the feasible set by rooting the derivative of the resulting 1D cost function, and then test the corresponding solution with the KKT condition. Once the KKT condition is met, the particular solution becomes our updated value. This is because the KKT condition is both necessary and sufficient in this problem, given that the cost function and the constraints are continuously differentiable and convex [58]. The origin will be the only feasible solution if no qualified solution is found in the previous cases. The derivation of the solution is provided in the appendix.

The pseudocode in Fig. 2.8 summarizes the procedure for reconstructing the image from the decomposed sinograms. We initialize the image and the error sinogram with the raw FBP images. Then within each iteration, for each selected voxel j, we first compute the j^{th} column of the forward projection matrix, which is $A_{*,j}$, by using the distance driven method [48]. Second, we compute the first two derivatives of the log-likelihood function, θ_1 and θ_2. Third, we compute the surrogate prior coefficients, ψ_r, for each of the neighboring voxels by using (2.43) and (2.45). Fourth, we compute

KKTSolve($\phi_1, \phi_2, n_{\min}, n_{\max}$){
  // unconstrained solution
  $u \leftarrow -\phi_2^{-1}\phi_1$
if $u \cdot n_{\min} < 0$ **or** $u \cdot n_{\max} < 0$ **then**
  // solve on the boundary defined by $u \cdot n_{\min} = 0$
  $u^T \leftarrow -\left(\phi_2^{-1} - \dfrac{\phi_2^{-1} n_{\min}^T n_{\min} \phi_2^{-1}}{n_{\min}\phi_2^{-1} n_{\min}^T}\right)\phi_1$
  $\lambda_1 \leftarrow \dfrac{n_{\min}^T \phi_2^{-1}\phi_1}{n_{\min}\phi_2^{-1} n_{\min}^T}$
  if $u \cdot n_{\max} < 0$ **or** $\lambda_1 \leq 0$ **then**
    // solve on the boundary defined by $u \cdot n_{\max} = 0$
    $u^T \leftarrow -\left(\phi_2^{-1} - \dfrac{\phi_2^{-1} n_{\max}^T n_{\max} \phi_2^{-1}}{n_{\max}\phi_2^{-1} n_{\max}^T}\right)\phi_1$
    $\lambda_2 \leftarrow \dfrac{n_{\max}^T \phi_2^{-1}\phi_1}{n_{\max}\phi_2^{-1} n_{\max}^T}$
    if $u \cdot n_{\min} < 0$ **or** $\lambda_2 \leq 0$ **then**
      // only feasible solution is the origin
      $u \leftarrow [0, 0]$
    end if
  end if
end if
return u
}

Fig. 2.7. Pseudocode for solving the quadratic minimization problem in (2.48) with the KKT condition. We first test the KKT condition on the unconstrained solution. If it fails, we solve the minimization problem on either boundary of the feasible set, and then we test the corresponding solution with the KKT condition. Once the KKT condition is met, the particular solution becomes the updated value. The origin will be the only feasible solution if no qualified solution is found in the previous cases. The derivation for the solution is provided in the appendix.

the first two derivatives for the quadratic cost function, ϕ_1 and ϕ_2, by using (2.46) and (2.47). Fifth, we solve the optimization problem in (2.48) by using the KKT condition to obtain the voxel update. Finally, we forward project the change in voxel to update the error sinogram and update the image as well.

```
ImageRecon(p̂, B){
  m ← raw FBP images
  A ← Compute
  e ← Am − p̂
  n_min, n_max ← Compute
repeat
  repeat
    j ← Select a voxel according to random schedule
    A_{*,j} ← Compute
    θ_1, θ_2 ← Compute by using (2.40) and (2.41)
    for each r ∈ ∂j do
      ψ_r ← Compute by using (2.43) and (2.45)
    end for
    φ_1, φ_2 ← Compute by using (2.46) and (2.47)
    m̂_j ← KKTSolve(φ_1, φ_2, n_min, n_max)
    e ← e + A_{*,j}(m̂_j − m_j)
    m ← m + δ_j(m̂_j − m_j)
  until All voxels have been visited
until Image m converges to the desired level
return m
}
```

Fig. 2.8. Pseudocode for reconstructing the image by using generic ICD algorithm. We initialize m with the raw FBP images and also initialize the error sinogram. Within each iteration, for each selected voxel, we first compute the column of the forward projection matrix. Second, we compute first two derivatives of the log-likelihood function. Third, we compute the coefficients for the surrogate prior. Fourth, we compute the first two derivatives of the quadratic cost function. Fifth, we solve the optimization problem with the KKT condition to obtain the voxel update. Finally, we update the error sinogram and the image. We define δ_j as an $N \times 1$ vector that is 1 for element j and 0 otherwise. The subroutine KKTSolve is described in Fig. 2.7.

2.4 Experimental Results

We have applied the proposed JDE-MBIR algorithm to real 3D DECT reconstruction problems. Raw data were acquired on a Discovery CT750 HD scanner (GE Healthcare, WI, USA) in a dual-energy fast switching acquisition mode, with the X-ray tube voltage alternating between 80 kVp and 140 kVp from view to view.

This spectral CT imaging technique is also referred to as Gemstone Spectral Imaging (GSI). Each scan contains approximately 2500 views per rotation, with each kVp having the same number of views, which is approximately 1250. Each scan was made with a large bowtie present. Each of the reconstructed images has a thickness of 0.625 mm, with 512×512 pixels. We reconstruct with water and iodine sinograms after material decomposition, with each material having the same number of views per rotation, which is approximately 2500. The reconstructed images represent the cross-sections corresponding to water- and iodine-equivalent densities in units of mg/cm^3. The "monochromatic" image, which specifies the cross-section corresponding to the attenuation given the photon energy, can then be generated by a linear combination of the reconstructed density images as in equation (2.18). Note that we do not generate monochromatic sinograms for reconstruction.

The function h^{-1} in (2.25) is approximated by using a high order polynomial with the following form,

$$[\hat{p}_{i,\mathrm{W}},\ \hat{p}_{i,\mathrm{I}}] = h^{-1}(y_{i,\mathrm{l}},\ y_{i,\mathrm{h}}) = \left[\sum_{m=0}^{L}\sum_{n=0}^{L} c_{m,n,\mathrm{W}} y_{i,\mathrm{l}}^m y_{i,\mathrm{h}}^n,\quad \sum_{m=0}^{L}\sum_{n=0}^{L} c_{m,n,\mathrm{I}} y_{i,\mathrm{l}}^m y_{i,\mathrm{h}}^n\right], \quad (2.49)$$

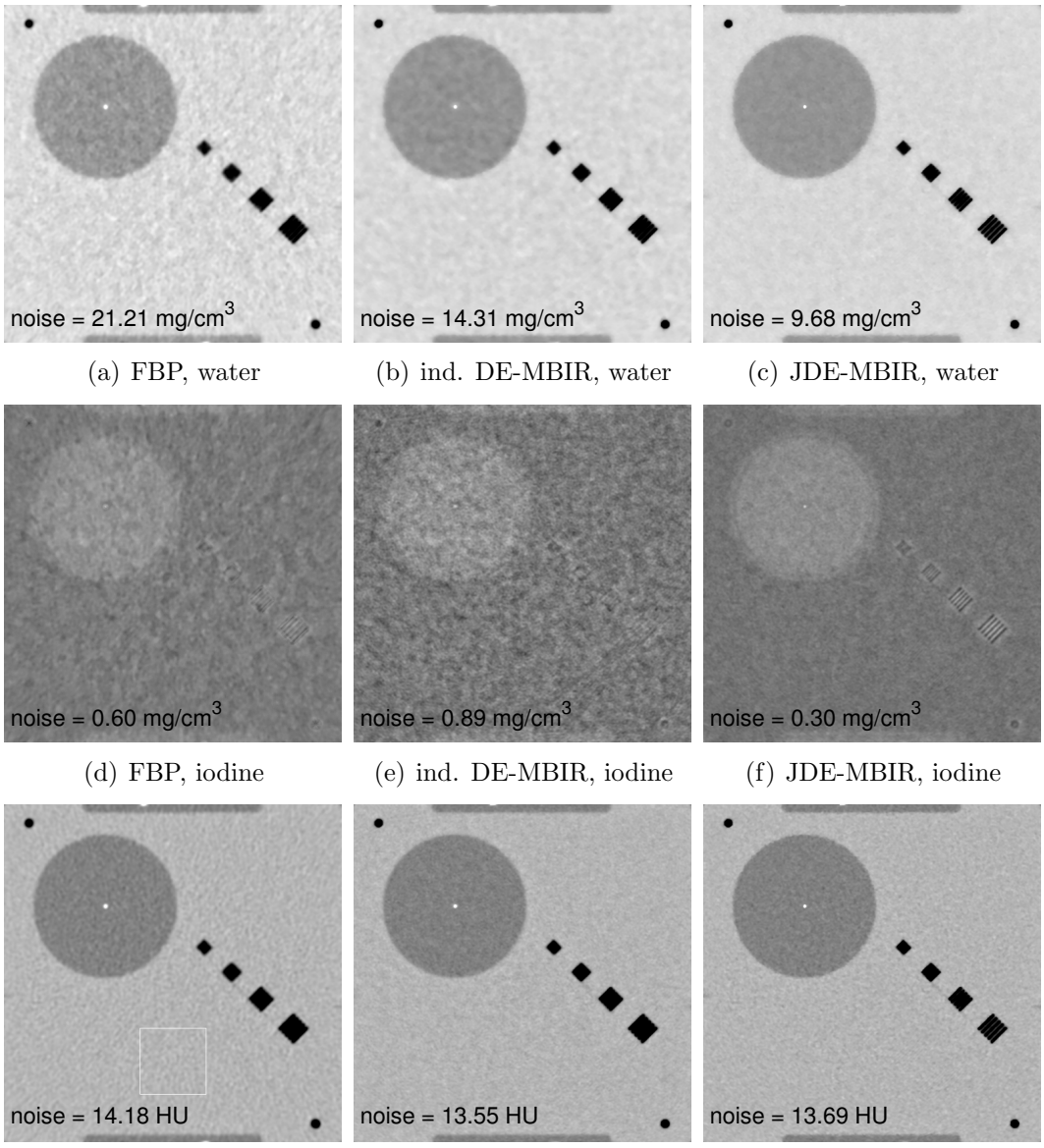
with $L = 10$. The specific coefficients for the polynomial, denoted by $\{c_{m,n,\mathrm{W}}, c_{m,n,\mathrm{I}}\}$, $\forall m \in \{0, 1, \cdots, L\}$, $\forall n \in \{0, 1, \cdots, L\}$, are computed in a calibration procedure for each device as described in [59]. As described in Sec. 2.2.2, these coefficients depend on many specific details of the device's physics including the X-ray spectra and detector sensitivity. The coefficients of the material decomposition are estimated in two stages. First, a polynomial is estimated to correct for beam hardening on a water phantom, and then the full set of coefficients are estimated for complete material decomposition.

We will compare the proposed JDE-MBIR method with two other decomposition-based methods, one using FBP reconstruction and the other using independent DE-MBIR. The FBP method consists of first obtaining two material sinograms from the material decomposition and then performing FBP on each sinogram with a stan-

dard reconstruction filter kernel. Then the resulting material density images are processed by a correlation-based noise reduction method [44, 60]. The independent DE-MBIR was implemented in the same way as described in Sec. 2.2.4. That is, the off-diagonal terms of the weighting matrix, B_i, were set to 0. All of the above methods work with the same decomposed sinograms. In practice, we implement the interleaved non-homogeneous ICD algorithm [6] for both independent DE-MBIR and JDE-MBIR. This method focuses computation where updates are mostly needed, which consequently accelerates the convergence. Both the independent DE-MBIR and JDE-MBIR are implemented on a standard 2.53 GHz clock rate 8 core Intel processor workstation with the Linux operating system. For both methods, we run 10 iterations to obtain the fully converged results.

In order to compare fairly among different reconstruction methods, for each experiment we match the noise level in 70 keV monochromatic images. That is, the difference of the noise standard deviation measured within a fixed ROI is less than 1 HU among different methods. We adjust the prior strength, $b_{jk,s}$ in (2.31), to match the noise level.

We first evaluate the performance of different methods using the phantom. For quantitative assessment, we use a 20 cm diameter GE Performance Phantom (GEPP) scanned in 64×0.625 mm helical mode at pitch 0.938:1 in 540 mAs in 300 mm field of view (FOV). The GEPP contains a Plexiglas insert with resolution bars and a 50 μm diameter tungsten wire placed in water. We measure the standard deviation within a fixed ROI in a homogeneous region of Plexiglas for noise assessment, and also measure the modulation transfer function (MTF) by using the wire for in-plane resolution assessment [61]. The width of the MTF is proportional to the spatial resolution. In this chapter, 10% MTF is chosen for comparison, since it generally reflects the visual resolution of the image. In addition to the above wire method, we also use the cyclic bar patterns to measure the spatial resolution. Following the method described in [62], we compute the MTF gain as the image modulation divided by the object modulation.



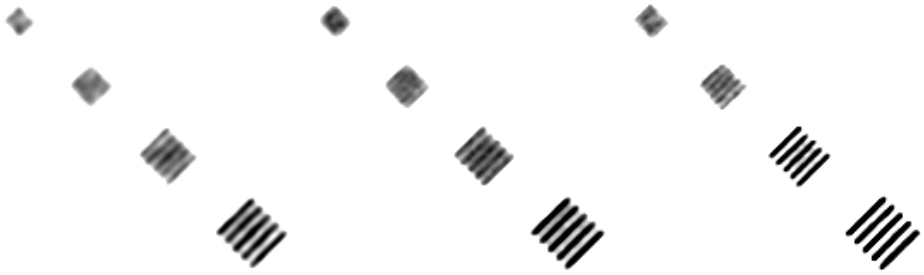
(a) FBP, water (b) ind. DE-MBIR, water (c) JDE-MBIR, water

(d) FBP, iodine (e) ind. DE-MBIR, iodine (f) JDE-MBIR, iodine

(g) FBP, 70 keV (h) ind. DE-MBIR, 70 keV (i) JDE-MBIR, 70 keV

Fig. 2.9. Comparison of FBP, independent DE-MBIR (ind. DE-MBIR) and JDE-MBIR reconstructions from a GEPP scan. From top to bottom: water density image, iodine density, and 70 keV monochromatic image. From left to right: FBP, independent DE-MBIR, JDE-MBIR. Display window for water: [650 1250] mg/cm^3; for iodine: [-8.0 7.0] mg/cm^3; for mono images: [-300 300] HU. The white box on the 70 keV FBP image (first at the third row) indicates the region where the noise standard deviation is evaluated.

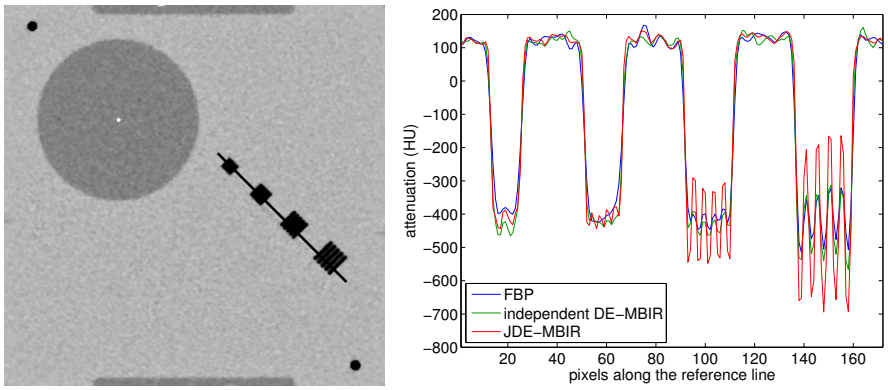
The JDE-MBIR method improves the spatial resolution and simultaneously reduces noise in the phantom study. This is illustrated by the GEPP reconstructions shown in Fig. 2.9. As shown in the figure, JDE-MBIR provides greater noise suppression than FBP and independent DE-MBIR in both water and iodine images, which leads to visually smoother homogeneous regions. Meanwhile, JDE-MBIR improves the resolution by producing a less blurred wire spot and spatially more distinguishable bars. One can also observe the resolution improvement in the monochromatic images in Fig. 2.10, which are synthesized using the reconstructed material densities in Fig. 2.9 following (2.18). The increased visual separation of the bars is illustrated by the profile plots in Fig. 2.11.



(a) FBP (b) indepedent DE-MBIR (c) JDE-MBIR

Fig. 2.10. Resolution bars in the 70 keV monochromatic images from a GEPP scan reconstructed with: (a) FBP; (b) independent DE-MBIR; (c) JDE-MBIR. Display window: [-500 -300] HU. Each image zooms in to the resolution bars of the monochromatic images shown in Fig. 2.9 with a different display window.

The visual improvement on the GEPP reconstruction is further verified by quantitative measurements in Table 2.1. With comparable noise level in 70 keV monochromatic images, JDE-MBIR significantly reduces noise as compared to FBP and independent DE-MBIR, especially for water images. JDE-MBIR also improves the in-plane resolution substantially as compared to FBP and independent DE-MBIR, according to these two different resolution metrics.



(a) profile line on the image (b) attenuations along the profile line

Fig. 2.11. Profile plot across the resolution bars on the GEPP 70 keV monochromatic images for FBP, independent DE-MBIR, and JDE-MBIR. Left: profile line on the image; right: attenuations along the profile line with FBP (blue), independent DE-MBIR (green), and JDE-MBIR (red).

Fig. 2.13 and 2.14 show the resolution and noise of the monochromatic images across various photon energies, where we match the noise level at 70 keV for different methods for comparison. Fig. 2.12 presents the monochromatic images at two distinct energies as an example. As shown in Fig. 2.13, JDE-MBIR significantly raises the resolution as compared to FBP and independent DE-MBIR. Fig. 2.14 also shows this resolution improvement by investigating the bar patterns at three different spatial frequencies. Each plot is computed using the method described in [62]. Note that for each frequency, the JDE-MBIR produces the largest contrast (i.e., closest to an ideal value of 100%) across all energies. This is consistent with the visual quality of the resolution bars in Fig. 2.10 and 2.12.

In addition, Fig. 2.13 shows that the JDE-MBIR monochromatic image contains less noise than the FBP image for all energies of diagnostic interest. It also has a more tractable noise characteristic than the monochromatic image with the independent DE-MBIR. More precisely, although the independent DE-MBIR monochromatic image appears slightly less noisy than the JDE-MBIR monochromatic image for some energy levels, the noise rises rapidly for the independent model as energy decreases.

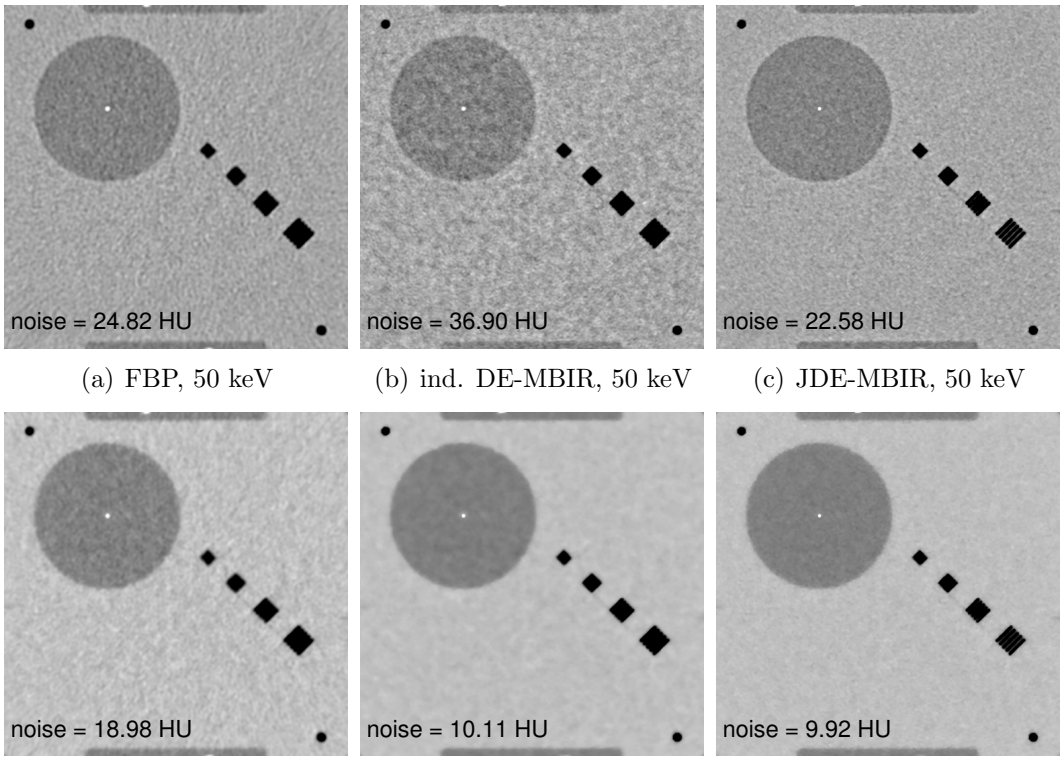
Table 2.1.
Comparison of FBP, independent DE-MBIR and JDE-MBIR for measurement of noise and in-plane resolution for the images in Fig. 2.9. The resolution measured by using the cyclic bars method is made in the 70 keV monochromatic images at the three lowest spatial frequencies.

Noise Measurement (Standard Deviation)			
	water (mg/cm^3)	iodine (mg/cm^3)	70 keV mono. (HU)
FBP	21.21	0.60	14.18
Independent DE-MBIR	14.31	0.89	13.55
JDE-MBIR	9.68	0.30	13.69

Resolution Measurement (10% MTF by the wire method)			
	water (lp/cm)	iodine (lp/cm)	70 keV mono. (lp/cm)
FBP	6.15	5.81	6.60
Independent DE-MBIR	8.61	6.35	8.90
JDE-MBIR	11.80	10.59	11.70

Resolution Measurement (MTF gain by the cyclic bars method)			
	6.25 lp/cm (%)	7.69 lp/cm (%)	10 lp/cm (%)
FBP	11.55	3.70	0
Independent DE-MBIR	15.35	3.74	0.25
JDE-MBIR	40.30	19.10	3.28

This is because the iodine component dominates the photon attenuation at low energy and the independent DE-MBIR method tends to produce noisy iodine reconstructions. Also, optimizing the prior strength for independent DE-MBIR becomes difficult due to this huge variation. This result also indicates that one can further reduce noise while still earning the advantage in spatial resolution by using the JDE-MBIR.

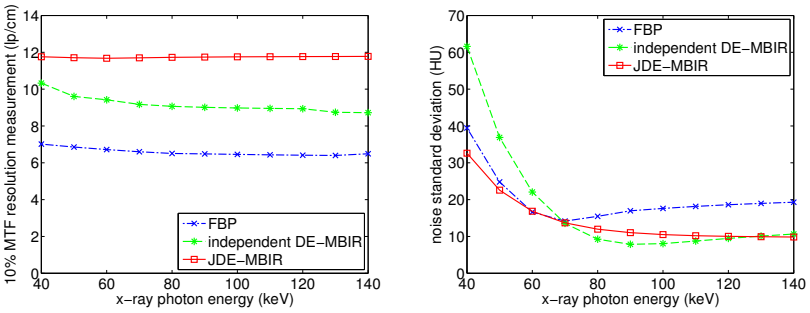


(a) FBP, 50 keV (b) ind. DE-MBIR, 50 keV (c) JDE-MBIR, 50 keV

(d) FBP, 130 keV (e) ind. DE-MBIR, 130 keV (f) JDE-MBIR, 130 keV

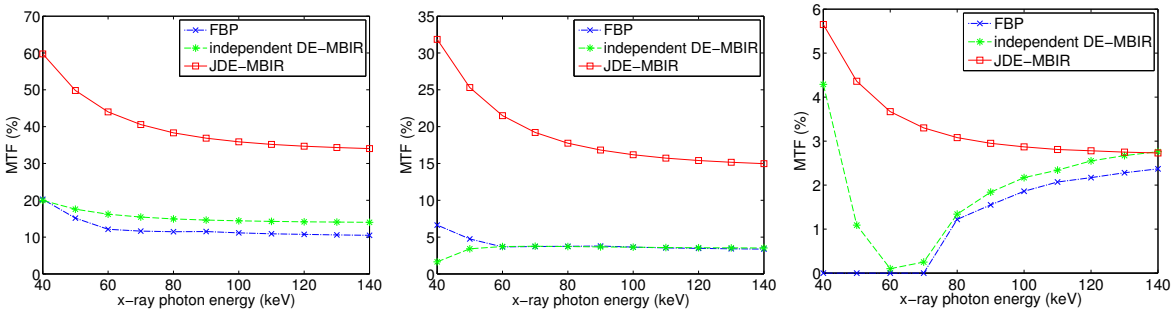
Fig. 2.12. Comparison of FBP, independent DE-MBIR (ind. DE-MBIR), and JDE-MBIR monochromatic images of the GEPP at different energies. From top to bottom: photon energy at 50 keV and 130 keV. From left to right: FBP, independent DE-MBIR, JDE-MBIR. Display window for mono images: [-300 300] HU. These monochromatic images are synthesized using the reconstructed material densities in Fig. 2.9 based on (2.18).

We also compared the convergence speed of the JDE-MBIR and the standard single-energy MBIR [6] with the GEPP reconstruction to measure the additional computational burden occasioned by the dual-energy reconstruction. For single-energy MBIR, we simply took the water sinogram and weight from the data used in the experiment of Fig. 2.9 and performed the reconstruction. In this case, the data used in JDE-MBIR and single-energy MBIR share the same scanner geometry and settings such as helical pitch, rotation speed, and local statistics for the water component. Note that the resulting single-energy MBIR reconstruction has no particularly quan-



(a) in-plane resolution (b) noise standard deviation

Fig. 2.13. Resolution and noise of the monochromatic images across various energy levels with different reconstruction methods.



(a) 6.25 lp/cm (b) 7.69 lp/cm (c) 10 lp/cm

Fig. 2.14. MTF measured at the cyclic bars of three different spatial frequencies in the monochromatic images across various photon energies. The JDE-MBIR produces higher MTF values than the other two methods at all three frequencies.

titative meaning, but it is still useful for comparing the computation time. Both algorithms were implemented on the same software platform and run on the same hardware. Fig. 2.15 shows the comparison of convergence speed between JDE-MBIR and single-energy MBIR. Since these two methods do not reach the same final cost due to different cost functions, we scale the cost of the single-energy MBIR such that it has the same final cost as JDE-MBIR, assuming full convergence has been reached in 10 iterations as usually observed in practice. As shown in the figure, both algorithms converge within 4 iterations. In this experiment, the average total computation

time per iteration for JDE-MBIR was 1.47 times the computation required for single-energy MBIR as measured across about 9 million voxels located differently in the 3D FOV. The main reason for the increase in computation stems from the fact that the sinograms for JDE-MBIR contain twice as much data as that for single-energy MBIR because of interpolation.

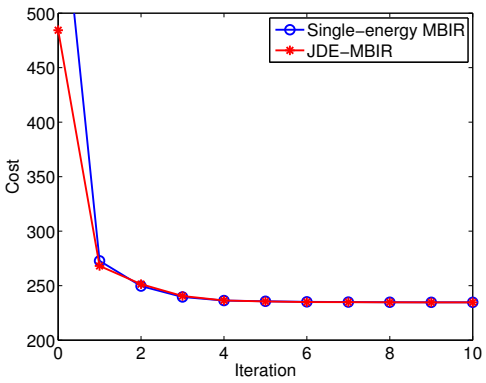


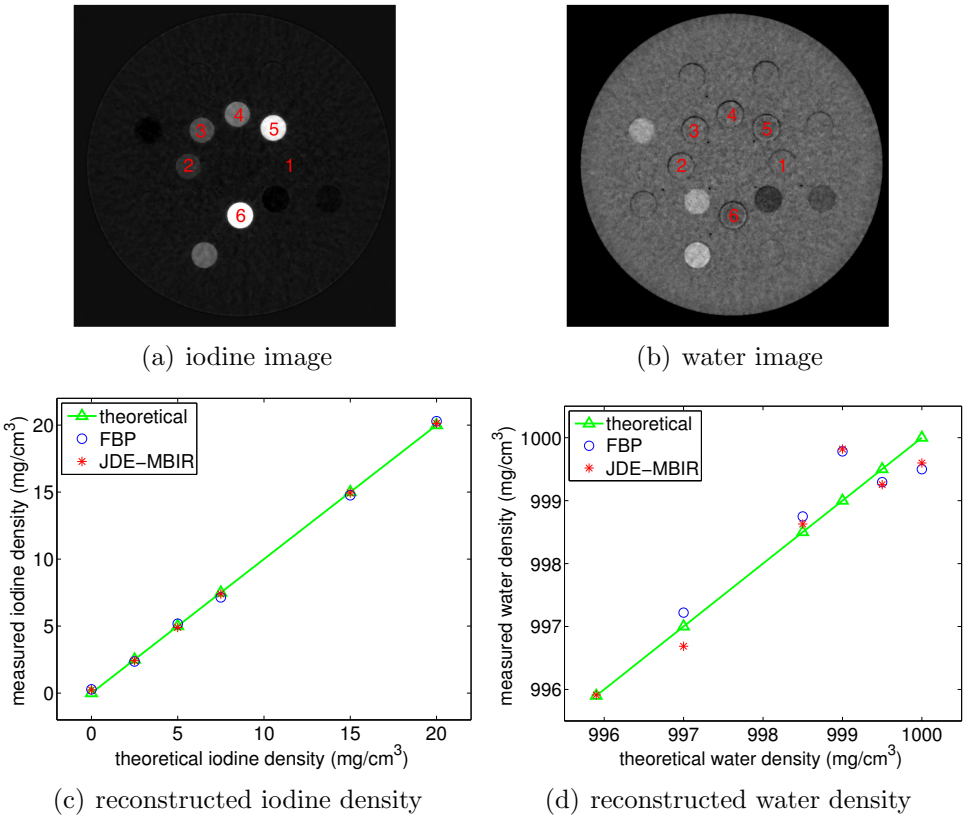
Fig. 2.15. Comparison of the convergence speed of the JDE-MBIR and the standard single-energy MBIR with the GEPP reconstruction in Fig. 2.9. The cost for the single-energy MBIR is scaled such that it reaches the same final cost as JDE-MBIR.

We also evaluated the reconstruction accuracy of JDE-MBIR by using a GE GSI contrast phantom, which was scanned in 32×0.625 mm axial mode in 384 mAs in 500 mm FOV. This phantom consists of a water phantom with several cylindrical rods inserted, each containing known concentrations of iodine and water. Fig. 2.16 shows the JDE-MBIR reconstructions of this phantom, with the theoretical iodine and water densities given in Table 2.2. Fig. 2.16 also plots the reconstructed iodine and water densities for FBP and JDE-MBIR. For each rod with known iodine concentration, we calculated the average of the reconstructed values in an ROI within the rod. As shown in the plots, FBP and JDE-MBIR produce equally accurate material densities.

We also compared FBP, independent DE-MBIR, and JDE-MBIR by using real clinical data, as shown in Figs 2.17 - 2.22. The data were collected from an abdominal scan in 64×0.625 mm helical mode at pitch 0.984:1 in 540 mAs in 500 mm

Table 2.2.
Theoretical densities of iodine and water for the inserted rods in the GSI contrast phantom as shown in Fig. 2.16.

Rod	1	2	3	4	5	6
Iodine (mg/cm^3)	0	2.5	5	7.5	15	20
Water (mg/cm^3)	1000	999.5	999	998.5	997	995.9



(a) iodine image (b) water image

(c) reconstructed iodine density (d) reconstructed water density

Fig. 2.16. Top row shows the JDE-MBIR reconstructions of the GE GSI contrast phantom. This phantom consists of a water phantom base and several cylindrical rods, each of which contains certain concentrations of iodine and water. Display window for water: [700 1300] mg/cm^3; for iodine: [-1.25 16.25] mg/cm^3. Bottom row shows the reconstruction accuracy of FBP and JDE-MBIR for iodine and water.

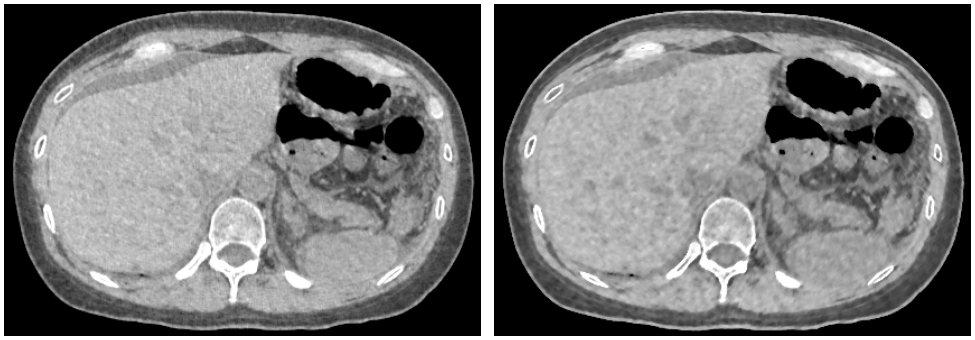
FOV. Fig. 2.17 and 2.18 show that the JDE-MBIR dramatically reduces the noise in the homogeneous regions (e.g., liver) in both water and iodine images. The bone

structures in the JDE-MBIR water image also suffer from less blooming and have sharper edges than the other two methods. Meanwhile, the JDE-MBIR improves the resolution in the iodine image as compared to the other two methods. For example, one can see details such as liver vessels more clearly in the JDE-MBIR image.

Fig. 2.19 - 2.22 present the corresponding monochromatic images at various energies. The resolution improvement can be observed in the JDE-MBIR images as compared to the other two methods, with a fixed noise level in the 70 keV monochromatic image. However, according to the resolution and noise curves shown in Fig. 2.13, one can achieve less noise while still retaining better resolution for the JDE-MBIR method as compared to the FBP method, by adjusting the prior strength. These results illustrate the potential diagnostic benefits of the JDE-MBIR method for DECT reconstruction. Note that either JDE-MBIR or independent DE-MBIR can be further improved by tuning the parameters for a particular clinical application.

2.5 Conclusion

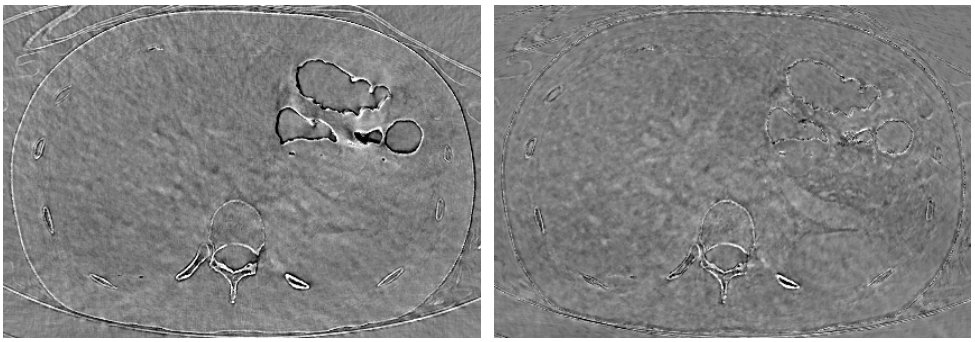
In this chapter, we have presented a JDE-MBIR approach for DECT reconstruction. The proposed method combines a joint likelihood model to account for the noise correlation in material-decomposed sinograms with MRF regularization, and features a physically realistic constraint that ensures non-negative X-ray absorptions. We also demonstrate that the JDE-MBIR method retains a more accurate model of the data likelihood than other decomposition-based statistical iterative methods when DECT uses fast kVp switching techniques. The experimental results on phantom and clinical data show that the JDE-MBIR method can reduce noise and increase resolution as compared to the FBP method and the independent DE-MBIR method. We expect that the improvement in terms of lower noise and higher resolution brought by the JDE-MBIR method may potentially reduce the CT dose required for a particular image quality. Future investigation will assess how to further improve material separation performance and investigate potential clinical benefits.



(a) FBP (b) ind. DE-MBIR

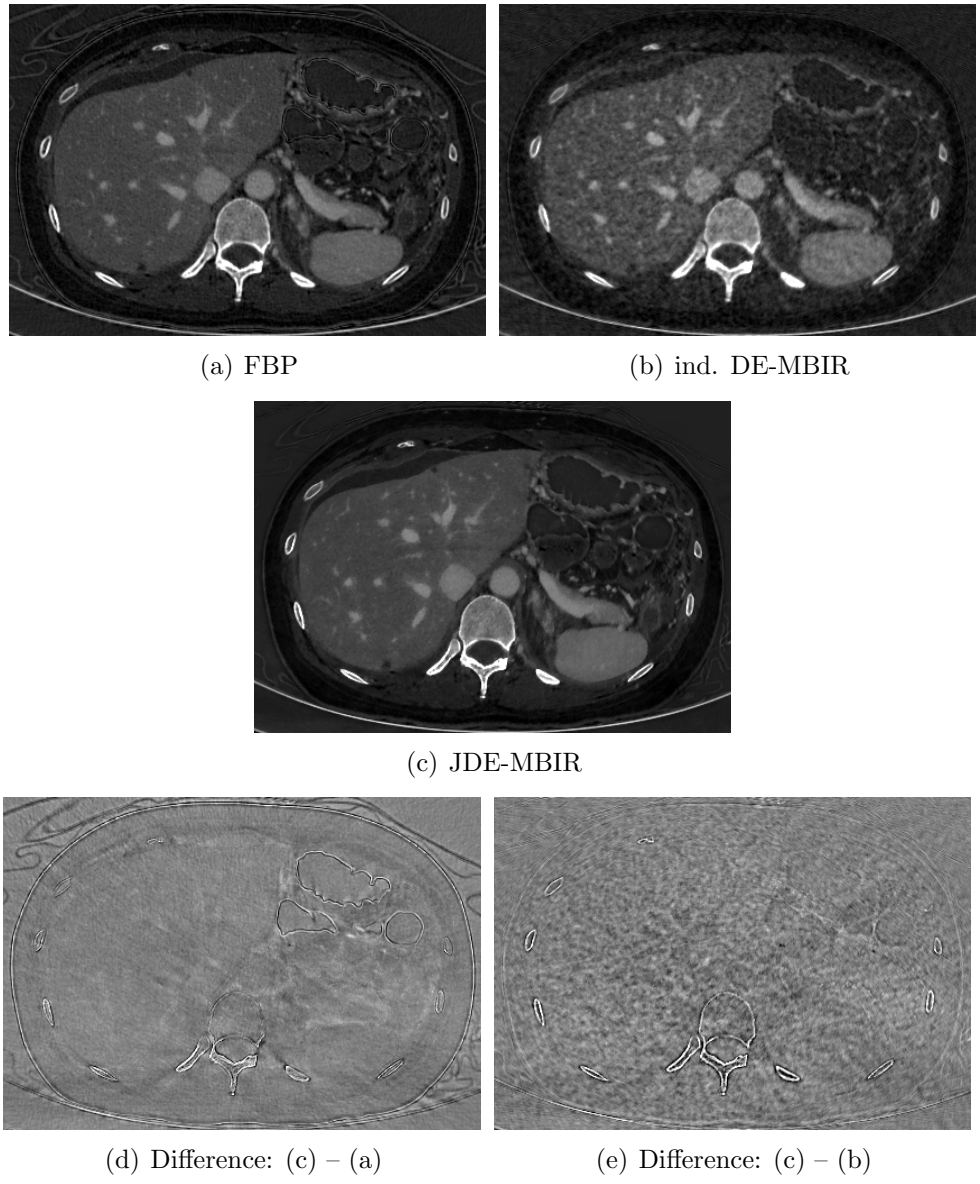


(c) JDE-MBIR



(d) Difference: (c) – (a) (e) Difference: (c) – (b)

Fig. 2.17. Comparison of water images reconstructed by FBP, independent DE-MBIR and JDE-MBIR methods for an abdominal clinical scan. Display window for water images: [850 1150] mg/cm^3; for water difference images: [-100 100] mg/cm^3.

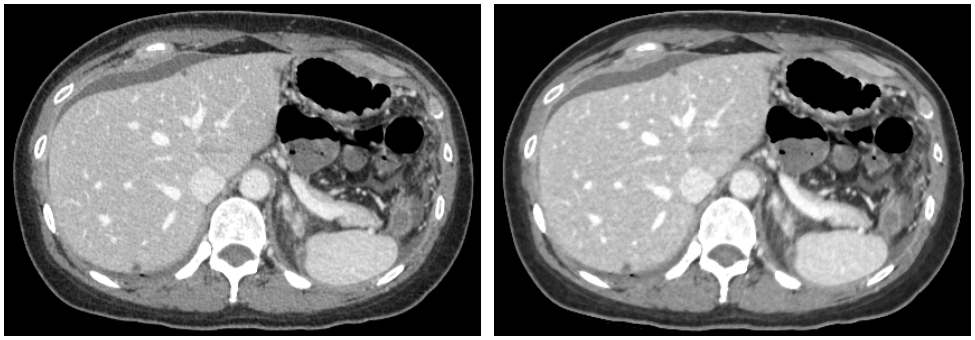


(a) FBP (b) ind. DE-MBIR

(c) JDE-MBIR

(d) Difference: (c) – (a) (e) Difference: (c) – (b)

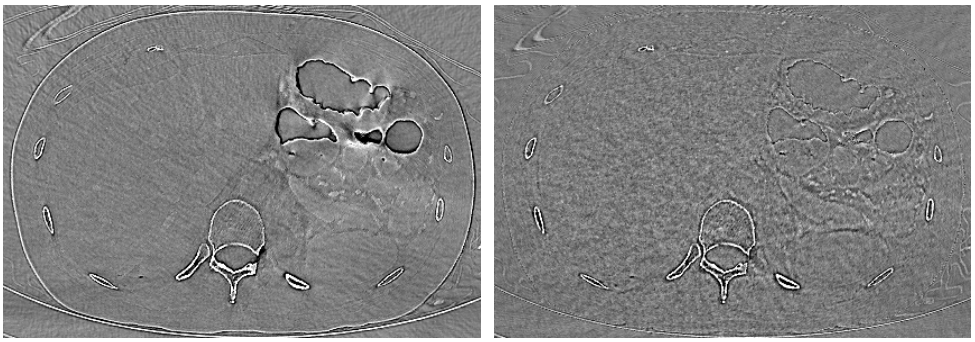
Fig. 2.18. Comparison of iodine images reconstructed by FBP, independent DE-MBIR and JDE-MBIR methods for an abdominal clinical scan. Display window for or iodine images: [-2.25 15.25] mg/cm^3; for iodine difference images: [-4.5 4.5] mg/cm^3.



(a) FBP (b) ind. DE-MBIR



(c) JDE-MBIR



(d) Difference: (c) – (a) (e) Difference: (c) – (b)

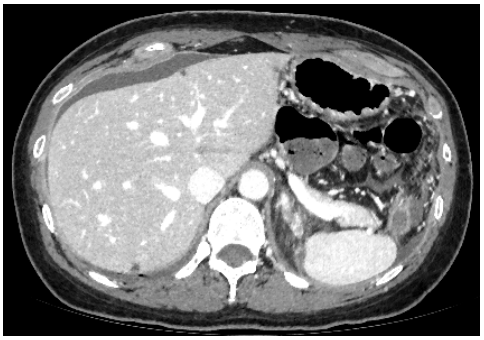
Fig. 2.19. Comparison of 70 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU.



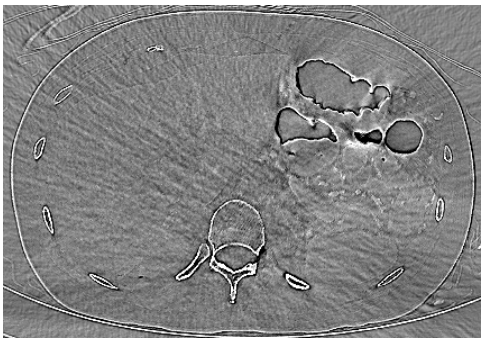
(a) FBP



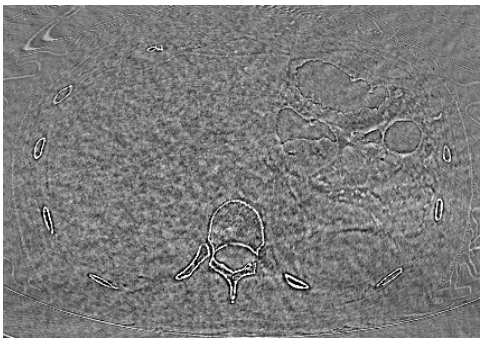
(b) ind. DE-MBIR



(c) JDE-MBIR

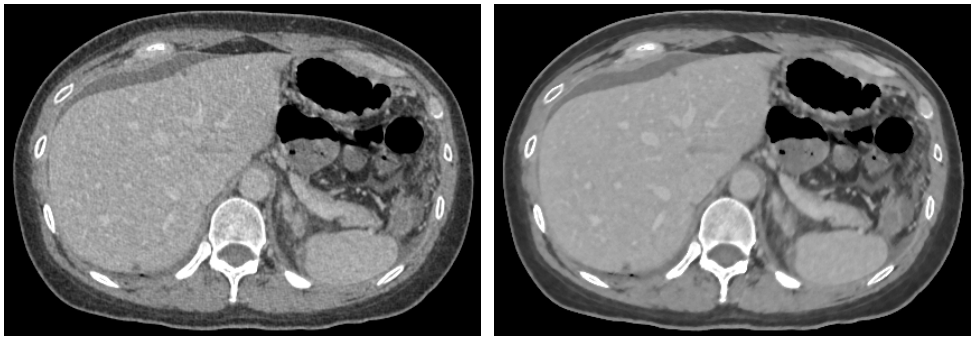


(d) Difference: (c) – (a)



(e) Difference: (c) – (b)

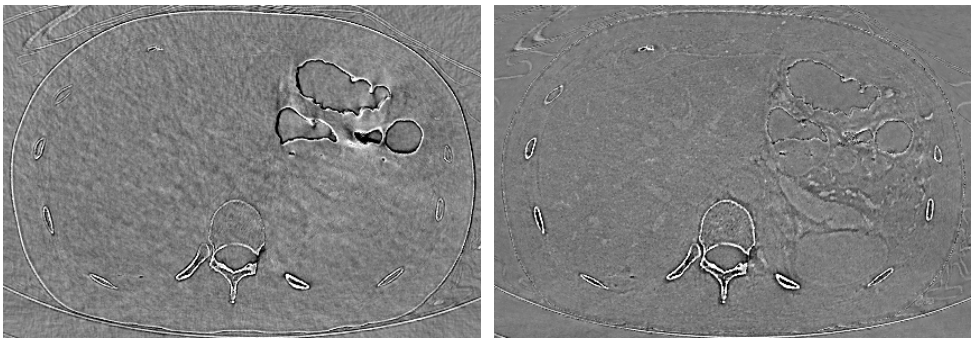
Fig. 2.20. Comparison of 60 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU.



(a) FBP (b) ind. DE-MBIR

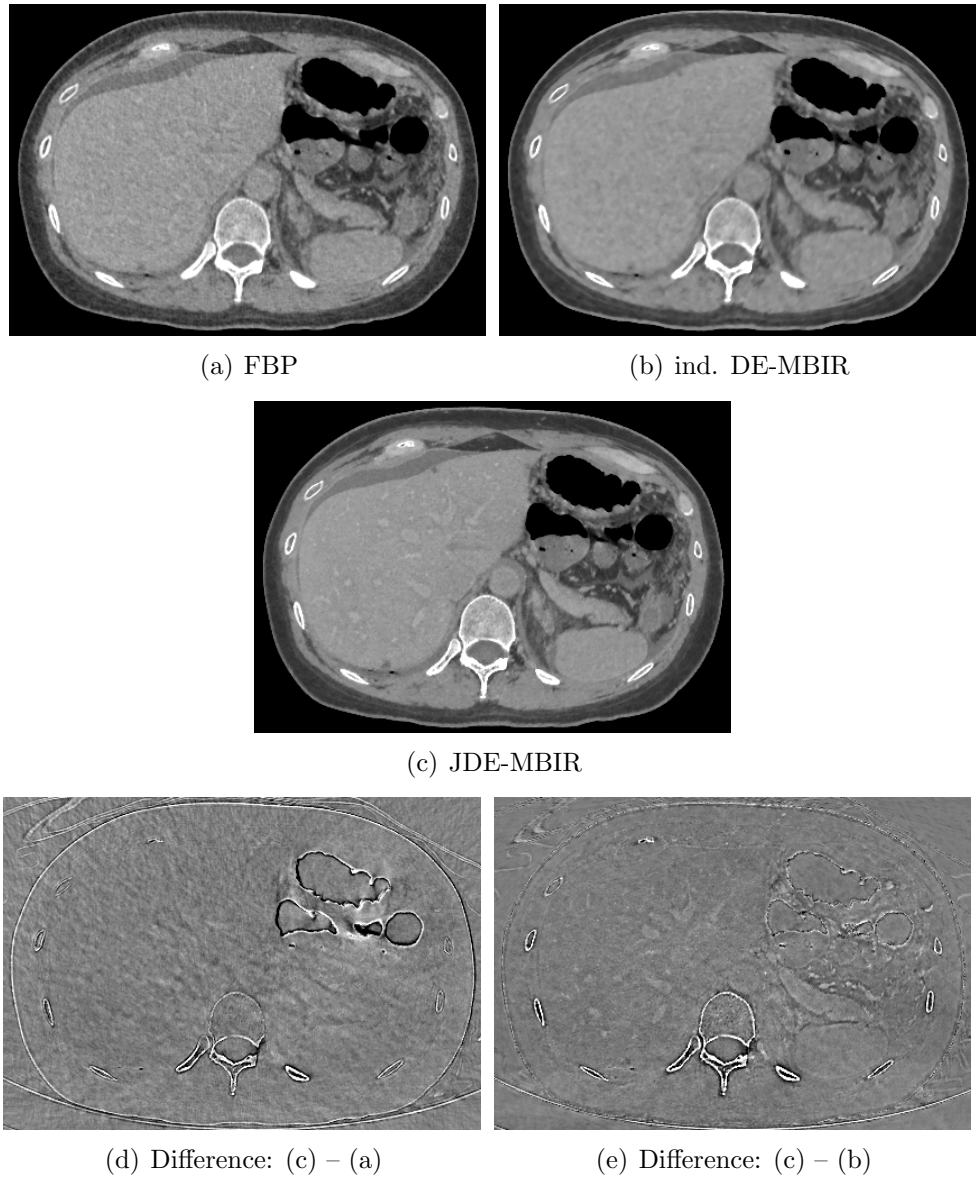


(c) JDE-MBIR



(d) Difference: (c) – (a) (e) Difference: (c) – (b)

Fig. 2.21. Comparison of 100 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU.



(a) FBP (b) ind. DE-MBIR

(c) JDE-MBIR

(d) Difference: (c) – (a) (e) Difference: (c) – (b)

Fig. 2.22. Comparison of 130 keV monochromatic images synthesized from FBP, independent DE-MBIR and JDE-MBIR reconstructions, in Figs. 2.17 and 2.18 based on (2.18). Display window for mono images: [-160 240] HU; for mono difference images: [-100 100] HU.

3. GAUSSIAN MIXTURE MARKOV RANDOM FIELD FOR MODEL-BASED ITERATIVE RECONSTRUCTION

3.1 Introduction

In recent years, model-based iterative reconstruction (MBIR) has emerged as a very powerful approach to reconstruct images from sparse or noisy data in applications ranging from medical, to scientific, to non-destructive imaging [3, 51, 63–68]. In medical applications, for example, MBIR has been demonstrated to substantially improve image quality by both reducing noise and improving resolution [7, 10, 11]. This important advantage of MBIR comes from the tight integration of the forward model and the prior model, which improves the estimation of the underlying image substantially.

While the MBIR forward model is typically based on the physics of the sensor, accurate prior modeling of real images remains a very challenging problem. Perhaps the most commonly used prior model is a very simple Markov random field (MRF) with only very local dependencies and a small number of parameters [1, 3, 69]. Alternatively, total-variation (TV) regularization approaches can also be viewed as simple MRF priors [70–73]. While these models have been very useful, their simple form does not allow for accurate or expressive modeling of real images.

More recently, methods such as K-SVD have been proposed which can be adapted as prior models in MBIR reconstruction [74–76]. Other patch-based or dictionary-based methods such as BM3D [77], non-local means [78], or bilateral filters [79] can be very effective in denoising applications, but are not directly suited for application in model-based reconstruction problems. While K-SVD can be adapted as a prior model, it does not explicitly model the multivariate distribution of the image. This can lead to drawbacks in applications. For example, the K-SVD algorithm is designed

to be invariant to scaling or average gray level of image patches. In applications such as CT reconstruction, this is a severe limitation since regions of different densities generally correspond to different tissues (e.g., bone and soft tissue) with distinctly different characteristics. In [80], Wang and Qi adapted a non-local prior model as a mechanism to capture subtler image characteristics. A variety of research also adapted the ideas of dictionary learning to the problem of prior modeling in CT reconstruction [81–83].

Another approach to prior modeling is to allow different patches of the images to have different distributions. This approach has been used by both Zoran et. al. [84] and Yu et. al. [85] to construct non-homogeneous models of images as the composition of patches, each with a distinct Gaussian distribution. The Gaussian distribution of each patch is selected from a discrete set of possible distributions (i.e., distinct mean and covariance). The reconstruction is then computed by jointly estimating both the image and a discrete class for each patch in the image. This approach can be very powerful for modeling the different spatially varying characteristics in real images. However, the approaches suffer from the need to make hard classifications of each patch. These hard classifications can lead to artifacts when patch distributions overlap, as is typically the case when a large number of classes are used.

In this chapter, we introduce the Gaussian-mixture MRF (GM-MRF) image prior along with an associated method for computing the MAP estimate using exact surrogate functions. (See [14] for an early conference version of our method). The GM-MRF model is constructed by seaming together patches that are modeled with a single Gaussian mixture (GM) distribution. The advantages of this approach are that:

- The GM-MRF prior provides a theoretically consistent and very expressive model of the multivariate distribution of the image;

- The GM-MRF parameters can be easily and accurately estimated by fitting a GM distribution to patch training data using standard methods such as the EM algorithm [86];

- MAP optimization can be efficiently computed by alternating soft classification of image patches with MAP reconstruction using quadratic regularization.

To create a consistent global image model, we seam together the GM patch models by using the geometric mean of their probability densities. This approach, similar to the product-of-experts technique [87] employed in deep-learning, produces a single consistent probability density for the entire image. Moreover, we also show that the resulting GM-MRF model is a Markov random field (MRF) as its name implies.

Of course, an accurate image model is of little value if computation of the MAP estimate is difficult. Fortunately, it can also be shown that the GM-MRF prior has an exact quadratic surrogate function for its log likelihood. This surrogate function allows for tractable minimization of the MAP function using a majorization-minimization approach [88]. The resulting MAP optimization algorithm has the form of alternating minimization. The two alternating steps are soft classification for patches followed by MAP optimization using quadratic regularization (i.e., a non-homogeneous Gaussian prior). Moreover, our approach to MAP optimization with the GM-MRF prior avoids the need for hard classification of individual patches. In practice, this means that patch-based GM-MRF models with a large numbers of overlapping mixture components can be used without adverse modeling effects. This allows for the use of very expressive models that capture fine details of image behavior.

Our results show that patch-based GM-MRF priors can be used to model complex characteristics of real images. So for example, the GM-MRF model can capture the different texture and edge characteristics of bone and soft tissue in medical images. However, our experiments indicate that when used with an accurate prior model, the MAP estimate tends to produce images that are exceedingly sharp in high-contrast bone regions and exceedingly smooth in low-contrast soft-tissue regions. Though

favorable in some situations, these MAP estimate images may not meet the specific needs of certain applications in visual quality. Therefore, to optimize the visual quality for different applications, we introduce a simple method for adjusting the GM components of the GM-MRF prior, so as to control the sharpness in low- and high-contrast regions of the reconstruction separately.

Our experimental results indicate that GM-MRF method results in improved image quality and reduced RMS error in simple denoising problems as compared to simple MRF and K-SVD priors. We also show multi-slice helical scan tomographic reconstructions from both phantom and clinical data that demonstrate that the GM-MRF prior produces visually superior images as compared to filtered back-projection (FBP) and MBIR using the traditional q-GGMRF prior [3].

The rest of the chapter is organized as follows. Sec. 3.2 describes the formulation of the GM-MRF model. Sec. 3.3 provides the usage of GM-MRF as a prior in MAP estimation with the optimization strategy. Sec. 3.4 introduces a systematic approach to adjust the GM-MRF prior in MAP estimation for better visual quality. Sec. 3.5 presents the results with a 2-D image denoising experiment and 3-D CT reconstruction experiments on phantom and clinical data.

3.2 Gaussian mixture Markov random field

Recall the typical formulation of MBIR algorithms as

$$\hat{x} \leftarrow \arg\min_{x \in \Omega} \{ -\log p(y|x) - \log p(x) \} \ , \tag{3.1}$$

where x represents the unknown image that is being reconstructed and y represents the measured data. In this framework, $p(y|x)$ is the conditional probability of y given x, which comprises the forward model of the measurement process. The density $p(x)$ is the prior model for x, which will be discussed in detail in this section.

Let $x \in \Re^N$ be an image with pixels $s \in S$, where S is the set of all pixels in x with $|S| = N$. Let $P_s \in \mathcal{Z}^{L \times N}$ be a patch operator that extracts a patch from the

image, where the patch is centered at pixel s and contains L pixels. More precisely, P_s is a rank L matrix that has a value of 1 at locations belonging to the patch and 0 otherwise. Furthermore, we assume that each patch, $P_s x$, can be modeled as having a multivariate Gaussian mixture distribution with K components,

$$g(P_s x) = \sum_{k=1}^{K} \frac{\pi_k |R_k|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}} \exp\left\{ -\frac{1}{2} \|P_s x - \mu_k\|^2_{R_k^{-1}} \right\} , \tag{3.2}$$

where parameters π_k, μ_k, R_k represent the mixture probability, mean, and covariance, respectively, of the k^{th} mixture component.

Then let $\{S_m\}, m \in \{1, \cdots, L\}$, be a partition of the set of all pixels into L sets, each of which tiles the image space. In other words, $\{P_s x\}_{s \in S_m}$ forms a set of non-overlapping patches, which contain all pixels in x. A simple 2-D example of this is when each $P_s x$ is a square $r \times r$ patch, and S_m is the set of pixels at each r^{th} row and column. Then the set of patches, $\{P_s x\}_{s \in S_m}$, tiles the plane.

Importantly, there are exactly L distinct tilings of the image space where L is the number of pixels in a patch. In order to see why this is true, consider the 2-D example in Fig. 3.1. (Note that this tiling method can be easily extended to n-D space with $n \geq 3$ by using n-D patches.) Notice that each distinct tiling of the space is determined by the position of the center pixel for the first (e.g., upper left hand) patch since the positioning of the first patch determines the phase shift of the tiling. With this in mind, there are exactly L distinct phase shifts corresponding to the L pixels in a single patch. Using this notation, we model the distribution of each tiling as the product of distributions of all its patches, as

$$p_m(x) = \prod_{s \in S_m} g(P_s x) \ . \tag{3.3}$$

In this case, $p_m(x)$ has the desired distribution for each patch. However, the discrete tiling of the space introduces artificial boundaries between patches. To remove the boundary artifacts, we use an approach similar to the product-of-experts approach

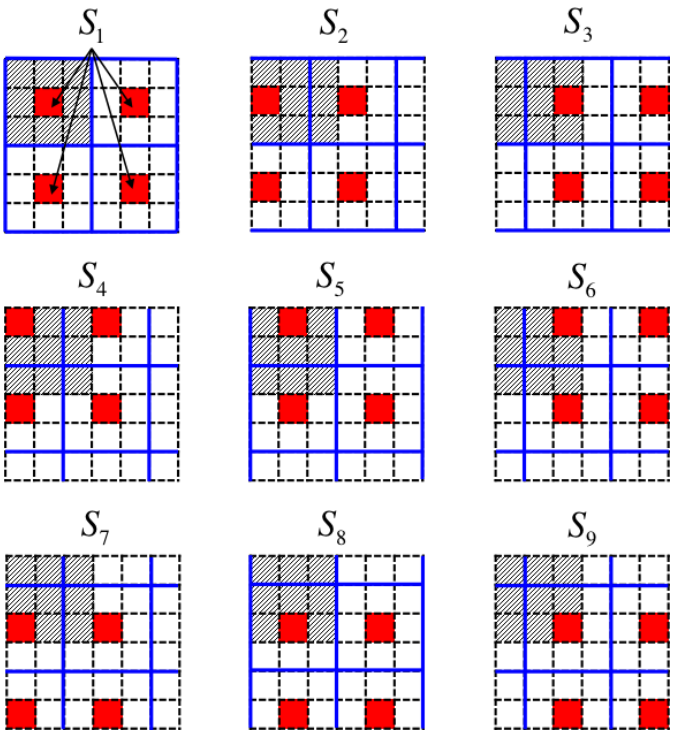


Fig. 3.1. 2-D illustration of the tiling method. Each blue grid represents one of nine distinct tilings with 3×3 patches on a 6×6 grid, i.e., $L = 9$, with the center pixel of each patch marked in red. Toroidal boundary condition is considered in this illustration. Note that there are exactly 9 distinct phase shifts of the tiling, each of which is determined by the center pixel of the first patch in the upper-left corner, which corresponds to a distinct pixel location in the shadowed patch.

in [87] and take the geometric average of the probability densities for all L tilings of the image space to obtain the resulting distribution

$$p(x) = \frac{1}{z}\left(\prod_{m=1}^{L} p_m(x)\right)^{\frac{1}{L}} = \frac{1}{z}\left(\prod_{s\in S} g(P_s x)\right)^{\frac{1}{L}}, \tag{3.4}$$

where z is a normalizing factor introduced to assure that $p(x)$ is a proper distribution after the geometric average is computed.

Let $V(P_s x) = -\log\{g(P_s x)\}$. Then we formulate a Gaussian mixture MRF (GM-MRF) model directly from (3.4) as

$$p(x) = \frac{1}{z}\exp\left\{-u(x)\right\} \ , \tag{3.5}$$

with the energy function

$$u(x) = \frac{1}{L}\sum_{s\in S} V(P_s x) \ , \tag{3.6}$$

and the potential function

$$V(P_s x) = -\log\left\{\sum_{k=1}^{K} \frac{\pi_k |R_k|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}} \exp\left\{-\frac{\|P_s x - \mu_k\|^2_{R_k^{-1}}}{2}\right\}\right\} \ . \tag{3.7}$$

Notice that $p(x)$ is a Gibbs distribution by (3.5). Therefore, by the renowned Hammersley-Clifford theorem [69], $p(x)$ is also an MRF.

3.3 MAP estimation with GM-MRF prior

For typical model-based inversion problems, the log-likelihood function may be modeled under the Gaussian assumption as

$$-\log p(y|x) = \frac{1}{2}\|y - Ax\|^2_D, \tag{3.8}$$

where $A \in \Re^{M\times N}$ is the projection matrix with M measurements and N unknowns. The weighting D is a diagonal matrix with each diagonal element inversely proportional to the variance of the corresponding measurement.

3.3.1 Surrogate prior

By substituting (3.5) and (3.8) into (3.1), we can calculate the MAP estimate with the GM-MRF prior as

$$\hat{x} \leftarrow \arg\min_{x\in\Omega} \left\{ \frac{1}{2}\|y - Ax\|_D^2 + u(x) \right\} \ . \tag{3.9}$$

However, the function $u(x)$ is not well suited for direct optimization due to the mixture of logarithmic and exponential functions. Therefore, we will use a majorization-maximization approach, in which we replace the function $u(x)$ with a quadratic upper-bounding surrogate function.

More precisely, the objective of the majorization-maximization method is to find a surrogate function $u(x; x')$ that satisfies the following two conditions.

$$u(x'; x') = u(x') \tag{3.10}$$

$$u(x; x') \geq u(x) \tag{3.11}$$

Intuitively, these conditions state that the surrogate function upper bounds $u(x)$ and that the two functions are equal when $x = x'$. Importantly, these conditions also imply that any reduction of $u(x; x')$ also must reduce $u(x)$.

In order to construct a surrogate function for our problem, we introduce the following lemma that is proved in Appendix A. The lemma provides a surrogate function for a general class of functions formed by the log of a sum of exponential functions. Since the potential function of (3.7) has this form, we can use this lemma to construct a surrogate function for our MAP estimation problem. Fig. 3.2 illustrates the usage of the lemma for a particular case of Gaussian mixture distribution.

Lemma (surrogate functions for logs of exponential mixtures): Let $f : \Re^N \to \Re$ be a function of the form

$$f(x) = \sum_k w_k \exp\{-v_k(x)\} \ , \tag{3.12}$$

where $w_k \in \Re^+$, $\sum_k w_k > 0$, and $v_k : \Re^N \to \Re$. Furthermore $\forall (x, x') \in \Re^N \times \Re^N$ define the function

$$q(x; x') \triangleq -\log f(x') + \sum_k \tilde{\pi}_k (v_k(x) - v_k(x')) \ , \tag{3.13}$$

where $\tilde{\pi}_k = \frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}}$. Then $q(x; x')$ is a surrogate function for $-\log f(x)$, and $\forall (x, x') \in \Re^N \times \Re^N$,

$$q(x'; x') = -\log f(x') \tag{3.14}$$

$$q(x; x') \geq -\log f(x) \tag{3.15}$$

Proof: see Appendix A.

Since the function $u(x)$ specified by (3.6) and (3.7) has the same form as assumed by the lemma, we can use this lemma to find a surrogate function with the following form

$$u(x; x') = \frac{1}{2L} \sum_{s \in S} \sum_{k=1}^{K} \tilde{w}_{s,k} \|P_s x - \mu_k\|^2_{R_k^{-1}} + c(x') \ , \tag{3.16}$$

where x' is the current state of the image, $c(x')$ only depends on the current state, and the weights $\tilde{w}_{s,k}$ are given by

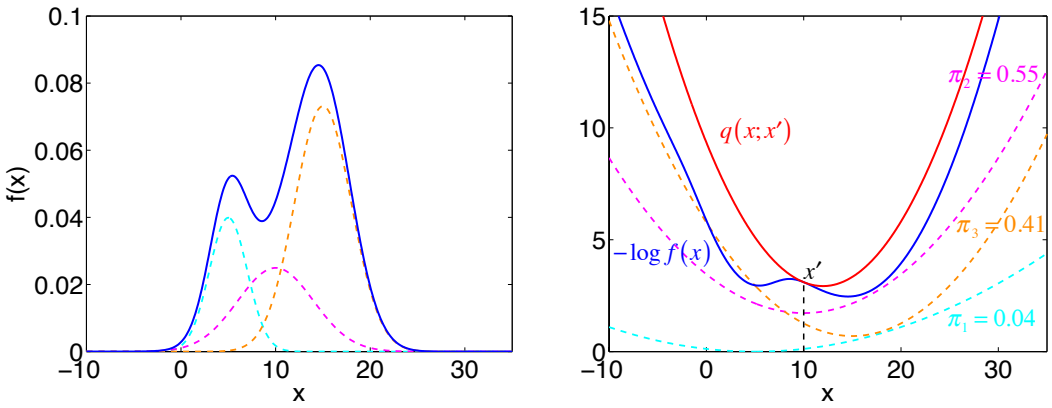
$$\tilde{w}_{s,k} = \frac{\pi_k |R_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \|P_s x' - \mu_k\|^2_{R_k^{-1}} \right\}}{\sum_{l=1}^{K} \pi_l |R_l|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \|P_s x' - \mu_l\|^2_{R_l^{-1}} \right\}} \ . \tag{3.17}$$

Note that the weights $\tilde{w}_{s,k}$ are only functions of the current image x'. Therefore, the optimization in (3.9) can be implemented as a sequence of optimizations as

$$\begin{aligned} \text{repeat}\{ \quad & \hat{x} \leftarrow \arg\min_x \left\{ \frac{1}{2} \|y - Ax\|^2_D + u(x; x') \right\} \\ & x' \leftarrow \hat{x} \quad \} \ , \end{aligned} \tag{3.18}$$

with $u(x; x')$ being a quadratic prior that adapts to the current image at each iteration.

Importantly, the weights in (3.17) represent a soft classification of the current patch into GM components. This differs from existing approaches in which each patch is classified to be from a single component of the mixture [84, 85].



(a) Gaussian mixture distribution (b) surrogate function

Fig. 3.2. Figure illustrates the lemma with a 1-D GM distribution. The quadratic function $q(x; x')$ is a surrogate function for the negative log of the GM distribution $f(x)$ at point x'. The surrogate function is a weighted sum of the quadratic exponents of the exponential functions in the GM distribution. The weights π_1, π_2, π_3 give the posterior probabilities of the point x' belonging to different GM components.

3.3.2 Optimization

Denote the cost function in (3.18) as $C(x)$,

$$\begin{aligned} C(x) &\triangleq \frac{1}{2}\|y - Ax\|_D^2 + u(x; x') \\ &= \frac{1}{2}\|y - Ax\|_D^2 + \frac{1}{2L}\sum_{s\in S}\sum_{k=1}^{K} \tilde{w}_{s,k}\|P_s x - \mu_k\|_{R_k^{-1}}^2 \ . \end{aligned} \tag{3.19}$$

Then the gradient of the cost function is given by

$$\nabla_x C(x) = A^t \Lambda (Ax - y) + \frac{1}{L} \sum_{s \in S} \sum_{k=1}^{K} \tilde{w}_{s,k} P_s^t R_k^{-1} (P_s x - \mu_k) \ . \tag{3.20}$$

Therefore, the MAP estimate of x is given by

$$\hat{x} = \left(A^t \Lambda A + \frac{1}{L} \sum_{s \in S} \sum_{k=1}^{K} \tilde{w}_{s,k} P_s^t R_k^{-1} P_s \right)^{-1} \left(A^t \Lambda y + \frac{1}{L} \sum_{s \in S} \sum_{k=1}^{K} \tilde{w}_{s,k} P_s^t R_k^{-1} \mu_k \right) \ . \tag{3.21}$$

Note that P_s is a patch operator that only involves a few pixels and therefore is very sparse. Then one may use simultaneous update methods to calculate the solution based on (3.21).

Alternatively, we use the iterative coordinate descent (ICD) algorithm [1] to solve this quadratic minimization problem in (3.18). The ICD algorithm sequentially updates each of the pixels by solving a 1-D optimization problem, as

$$\hat{x}_j \leftarrow \arg\min_{x_j} \left\{ \frac{1}{2} \| y - Ax' + A_{*j}(x'_j - x_j) \|_D^2 + u(x_j; x') \right\} \ , \tag{3.22}$$

with the surrogate prior for x_j, as

$$u(x_j; x') = \frac{1}{2L} \sum_{r \in \mathcal{S}_j} \sum_{k=1}^{K} \tilde{w}_{r,k} \| P_r x - \mu_k \|_{R_k^{-1}}^2 + c(x') \ , \tag{3.23}$$

where the weights $\tilde{w}_{r,k}$ are given by (3.17) and S_j represents a set of center pixels whose patches contain pixel j.

By rearranging the terms, we can explicitly write (3.22) as a quadratic function of x_j, as

$$\hat{x}_j \leftarrow \arg\min_{x_j} \left\{ (\theta_1 + \varphi_1) x_j + \frac{\theta_2 + \varphi_2}{2} (x_j - x'_j)^2 + c(x') \right\} \ , \tag{3.24}$$

where $c(x')$ is constant to x_j and $\theta_1, \theta_2, \varphi_1, \varphi_2$ are given by

$$\theta_1 = A_{*j}^t D(Ax' - y) \ , \tag{3.25}$$

$$\theta_2 = A_{*j}^t D A_{*j} \ , \tag{3.26}$$

$$\varphi_1 = \frac{1}{L} \sum_{r \in S_j} \sum_k \tilde{w}_{r,k} (P_r \delta_j)^t R_k^{-1} (P_r x' - \mu_k) \ , \tag{3.27}$$

$$\varphi_2 = \frac{1}{L} \sum_{r \in S_j} \sum_k \tilde{w}_{r,k} (P_r \delta_j)^t R_k^{-1} (P_r \delta_j) \ , \tag{3.28}$$

where the calculation of the projection matrix A follows the same procedure in [3]. The function $\delta_j \in \Re^{|S|}$ is a Kronecker delta function, which is a vector with a value of 1 at entry j and with 0 elsewhere. Therefore, $P_r \delta_j$ is simply an operator that extracts a particular column from a matrix corresponding to the location of the pixel j within the patch operator P_r.

Solving (3.24) by rooting the gradient, we then have

$$\hat{x}_j \leftarrow x'_j - \frac{\theta_1 + \varphi_1}{\theta_2 + \varphi_2} \ . \tag{3.29}$$

3.4 Covariance Control for GM-MRF

We will see that the GM-MRF distribution can be used to form a very accurate model of images. However, in applications such as CT reconstruction, the MAP estimate may not be visually appealing even with an accurate forward and prior model. This is because the MAP estimate tends to produce a reconstruction that is under-regularized (i.e., too sharp) in high-contrast regions and over-regularized (i.e., too smooth) in low-contrast regions. While this variation in spatial resolution may produce a lower mean squared error (MSE), in particular applications it may not be visually appealing.

In order to address this problem of spatial variation in sharpness, in this section we introduce a simple parameterization for systematically controlling the covariance of each GM component of the GM-MRF model. In the experimental results section, we

will then demonstrate that this simple parameterization can be used to effectively tune the visual quality of the MAP reconstruction. In real applications such as medical CT reconstruction, this covariance adjustment can be used to effectively fine-tune the rendering of specific tissue types, such as soft tissue, lung, and bone, which may have different desired characteristics.

We start by introducing regularization parameters, σ_x and $\{\sigma_k\}_{k=1}^K$, into the distribution given by

$$u_\sigma(x) = -\frac{1}{L\sigma_x^2}\sum_{s\in S}\log\{g_\sigma(P_s x)\}\ , \tag{3.30}$$

with the patch Gaussian mixture distribution

$$g_\sigma(P_s x) = \sum_{k=1}^{K}\frac{\pi_k|R_k/\sigma_k^2|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}}\exp\left\{-\frac{\sigma_k^2\|P_s x-\mu_k\|^2_{R_k^{-1}}}{2}\right\}\ . \tag{3.31}$$

Notice that σ_x controls the overall level of regularization and that the K values of σ_k control the regularization of each individual component of the GMM. When the value of σ_x is increased, the overall reconstruction is made less regularized (i.e., sharper) and when the value of σ_k is increased, the individual GM component is made more regularized.

Now for a typical GM-MRF model there may be many components, so this would require the choice of many values of σ_k. Therefore, we introduce a simple method to specify these K parameters using the following equation,

$$\sigma_k = \left(\bar{\lambda}_k/\alpha^2\right)^{p/2}\ , \tag{3.32}$$

where p and α are two user-selectable parameters such that $0 \le p \le 1$, $\alpha > 0$, and $\bar{\lambda}_k = |R_k|^{\frac{1}{L}}$ is the geometric average of the eigenvalues of R_k. Define $\tilde{R}_k = R_k/\sigma_k^2$ as the covariance matrix after scaling. Then its corresponding average eigenvalue is given by

$$\bar{\tilde{\lambda}}_k = \alpha^{2p}\bar{\lambda}_k^{1-p}\ . \tag{3.33}$$

Fig. 3.3 illustrates this scaling with various values of p and α.

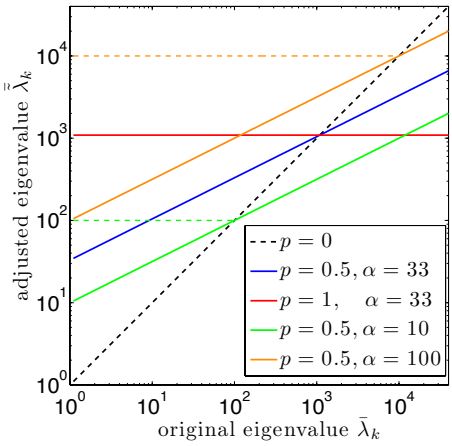


Fig. 3.3. The covariance scaling defined in (3.32) and (3.33) with various values of p and α on a log scale. The black dotted line shows the case when no scaling is present, i.e., $p = 0$. When $0 < p \leq 1$, the average eigenvalues $\bar{\bar{\lambda}}_k$ are "compressed" toward α^2, where eigenvalues further away from α^2 lead to greater change and a larger p results in greater compression. For a fixed value of p, increasing α increases the covariance of each GM component.

In this model, the parameters p and α collectively compress the dynamic range of the average eigenvalues $\bar{\lambda}_k$ of all GM covariance matrices. That is, for those GM components with large average eigenvalues of covariance, which typically correspond to high-contrast or structural regions, applying the scaling in (3.32) decreases the eigenvalues, which leads to increased regularization. Conversely, for those GM components with small average eigenvalues of covariance, which are typically associated with low-contrast or homogeneous regions, applying the scaling increases the eigenvalues and subsequently results in reduced regularization.

More specifically, p is the compression rate with a larger value resulting in greater compression of the dynamic range, and α defines a stationary point during the compression, i.e.,

$$\bar{\bar{\lambda}}_k = \bar{\lambda}_k \ , \quad \text{if} \quad \bar{\lambda}_k = \alpha^2 \ . \tag{3.34}$$

When $0 < p < 1$, the average eigenvalues $\bar{\tilde{\lambda}}_k$ are "compressed" toward α^2, with eigenvalues further away from α^2 leading to greater change. When $p = 1$, all GM components have the same average eigenvalue α^2, while they maintain the original eigenvalues when $p = 0$. Fig. 3.4(a)(c)(e) illustrate the change in the distribution and the energy function as p varies.

In addition, the parameter α controls the "smoothness" of the GM distribution. With p fixed, increasing the value of α leads to a smoother distribution of (3.31), which potentially reduces the degree of non-convexity of the energy function in (3.30). Moreover, an increased α also reduces the overall regularization. Fig. 3.4(b)(d)(f) illustrate the change in the distribution and the energy function as α varies.

Table 3.1 presents the selection of the regularization parameters and the corresponding effect. Note that the parameter α is related to the reconstruction noise and therefore has the same unit as the reconstruction. For instance, in X-ray CT reconstruction, the parameter α is in Hounsfield Unit (HU).

Table 3.1.
Parameter selection for GM-MRF model.

Parameter	Selection	Effect
σ_x	≈ 1	MAP estimate
	$>> 1$	large value $\rightarrow$ weak overall regularization
	$<< 1$	small value $\rightarrow$ heavy overall regularization
p	0	unmodified GM-MRF
	0.5	regularization strength increases for GM components with large average eigenvalues; reduces for those with small average eigenvalues
	1	same regularization strength for all GM components
α	33 (HU)	large value $\rightarrow$ smooth prior distribution $\rightarrow$ weak regularization
		small value $\rightarrow$ peaky prior distribution $\rightarrow$ heavy regularization

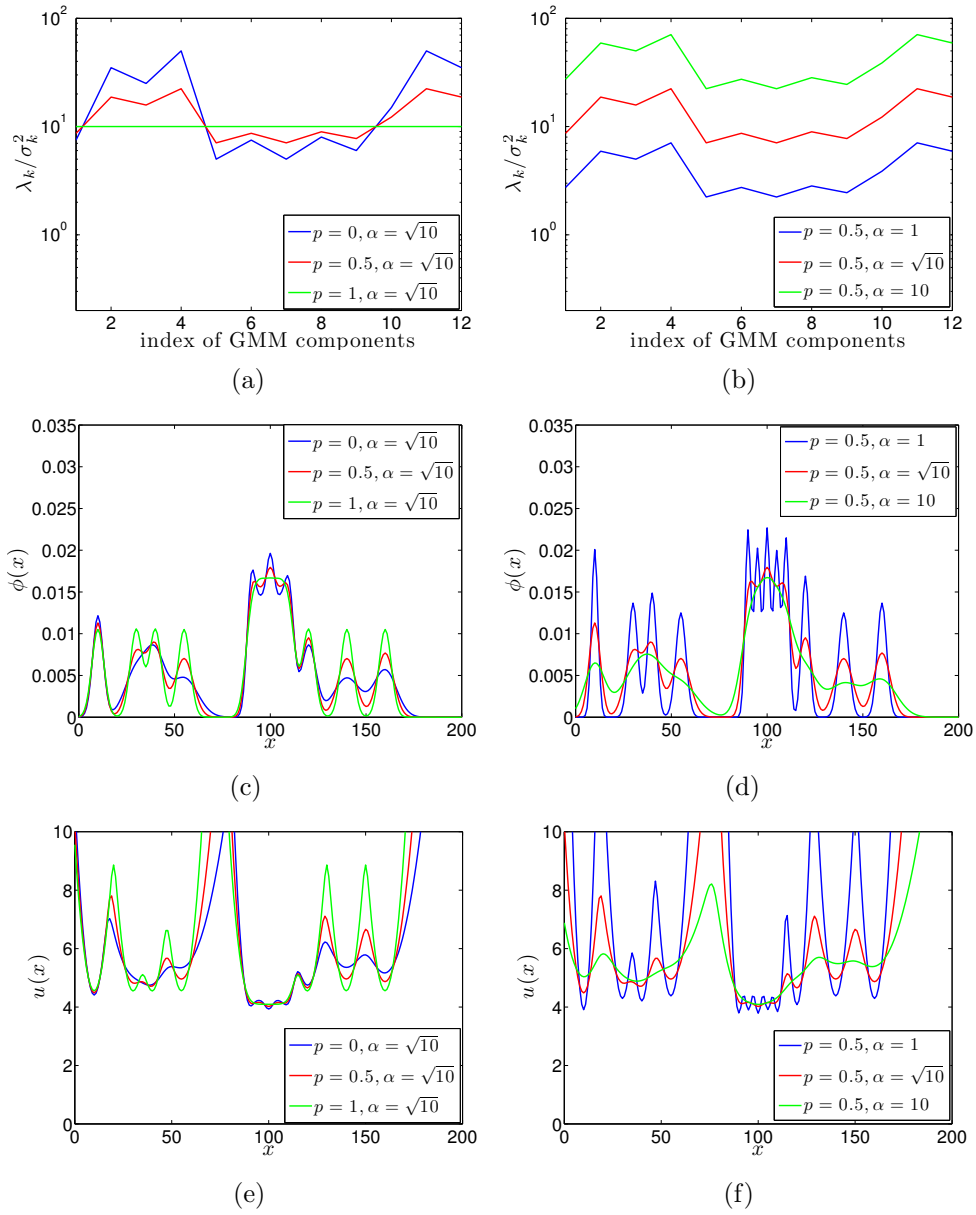


Fig. 3.4. 1-D illustration of the covariance scaling in (3.32). The 1-D energy function is given by $u(x) = -\log(\phi(x))$ with the Gaussian mixture distribution $\phi(x) = \sum_k \frac{1}{\sqrt{2\pi\lambda_k/\sigma_k^2}} \exp\left\{-\frac{1}{2\lambda_k/\sigma_k^2}(x-\mu_k)^2\right\}$, with λ_k the original variance and the scaling $\sigma_k = (\lambda_k/\alpha^2)^{p/2}$. Left column: (a) varying p with α fixed, (c) the resulting distribution, and (e) the resulting energy function; right column: (b) varying α with p fixed, (d) the resulting distribution, and (f) the resulting energy function.

3.5 Experiments and Results

In this section, we present the results of applying the GM-MRF as a prior for model-based inversion problems, including 2-D image denoising and 3-D CT reconstruction experiments.

3.5.1 Training

We trained the GMM patch distribution, $g(P_s x)$ in (3.2), on clinical CT images using the standard EM algorithm with the software in [86]. Training data consisted of 2-D or 3-D overlapping patches extracted from a 3-D image volume of a normal-dose clinical CT scan, which was collected on a GE Discovery CT750 HD scanner in 64×0.625 mm helical mode with 100 kVp, 500 mA, 0.8 s/rotation, pitch 0.984:1, and reconstructed in 360 mm field-of-view (FOV).

Instead of training one GMM using all the patches, we partitioned the patches into different groups and then trained one GMM from each of the groups. In this way, we were able to collect sufficiently many samples from underrepresented groups to obtain accurate parameter estimates, while simultaneously limiting the data size for other groups to retain training efficiency. For the i^{th} group, we trained the parameters, $\{\pi_{i,k}, \mu_{i,k}, R_{i,k}\}_k^{K_i}$, for one GMM, $g_i(P_s x)$, with K_i components. Then we merged all GMMs trained from different groups into a single GMM by weighted summation,

$$g(P_s x) = \sum_{i=1}^{I} \pi_i g_i(P_s x), \tag{3.35}$$

where the mixture weights π_i were determined by the natural proportions of corresponding groups in the whole training data.

More specifically, we partitioned the patches into six groups based on the mean and standard deviation as listed in Table 3.2, where the partition thresholds were empirically determined to roughly reflect typical tissue types in a medical CT image. Fig. 3.5 illustrates different groups on a 2-D image slice. As shown in Fig. 3.5, different

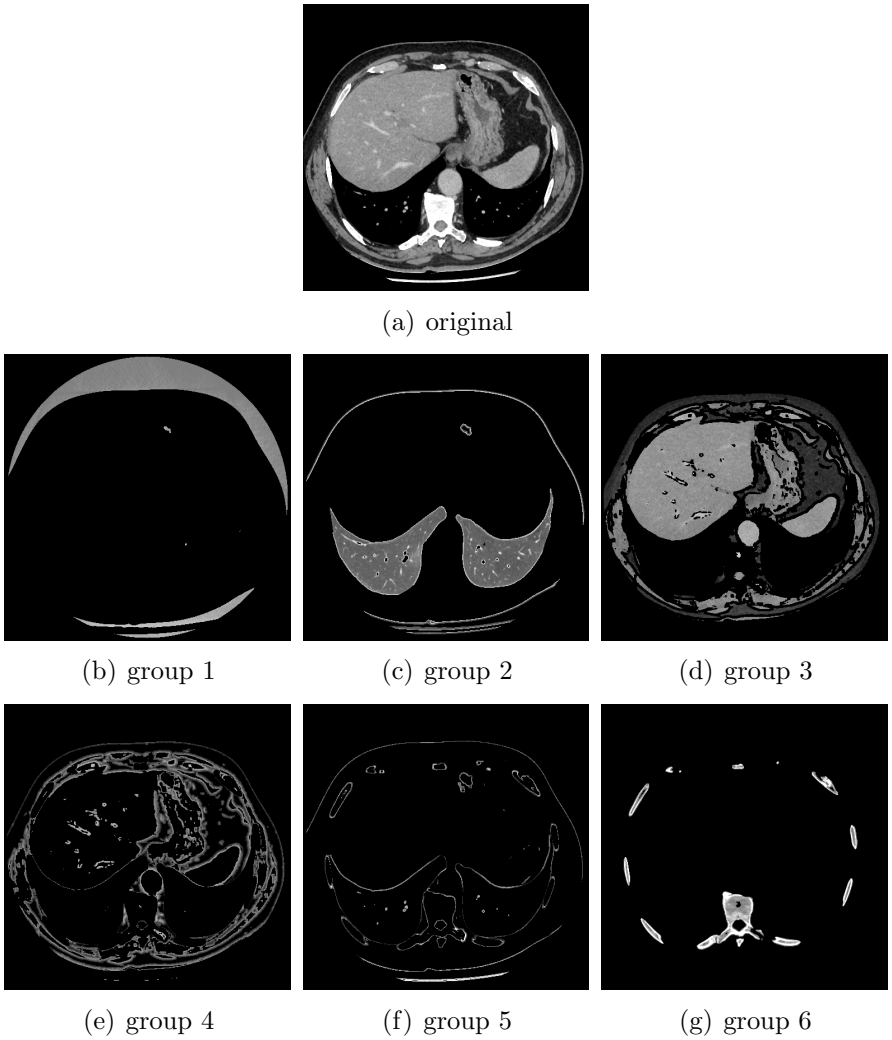
groups roughly capture different materials or tissue types in the image, as group 1 for air, group 2 for lung tissue, group 3 for smooth soft tissue, group 4 for low-contrast soft-tissue edge, group 5 for high-contrast edge, and group 6 for bone. With this partition, we were able to collect adequate patches for individual groups separately, especially for the underrepresented ones as group 4, 5, and 6. During the separate training process, we empirically fixed the number of GMM components, K_i, in the EM algorithm for each group. Table 3.3 presents the mixture weights π_i for different GMMs, which were determined by the natural proportions of corresponding groups in the whole training data.

Table 3.2.
Partition of the training data. Each image patch was classified into one of the six groups based on its mean and standard deviation. The number of GMM components for each group was empirically chosen.

Group index, i	Mean (HU)	Standard deviation (HU)	Number of patch samples	Number of GMM components, K_i
1	[-1000 -850)		5×10^3	1
2	[-850 -200)		1×10^5	15
3	[-200 200)	[0 25)	5×10^4	5
4	[-200 200)	[25 80)	1×10^5	15
5	[-200 200)	≥ 80	1×10^5	15
6	≥ 200		1×10^5	15

Table 3.3.
Mixture weights of GMMs trained from different groups. The mixture weights are determined by the proportions of corresponding groups in the whole training data and will be used when combining different GMMs to form a single model.

Group index, i	1	2	3	4	5	6
Mixture weight, π_i	0.05	0.17	0.40	0.25	0.04	0.09



(a) original

(b) group 1 (c) group 2 (d) group 3

(e) group 4 (f) group 5 (g) group 6

Fig. 3.5. Partition of the training data. (a): a 2-D axial slice from the 3-D image volume where training patches were extracted; (b)-(g): partition of the data based on the criteria in Table 3.2. Display window: (a) [-160, 240] HU, (b) [-1250, -750] HU, (c) [-1400, 200] HU, (d)-(f) [-210, 290] HU, (g) [-300, 700] HU. Notice that different groups roughly capture different materials or tissue types in the image, as group 1 for air, group 2 for lung tissue, group 3 for smooth soft tissue, group 4 for low-contrast soft-tissue edge, group 5 for high-contrast edge, and group 6 for bone.

The blue plot in Fig. 3.6 illustrates the square-rooted geometrically-averaged eigenvalues, $\bar{\lambda}_k$, of the trained GMM covariance matrices for a $5\times5\times3$ patch case. The numbers within the figure correspond to the indices of training groups in Table 3.2.

Within each group, the GM components are sorted from the most probable to the least probable based on the trained mixture probabilities, $\pi_{i,k}$. Fig. 3.6 shows that different groups present different amounts of regularization strength. Note that there is a large variation in average eigenvalues for different groups, which leads to highly varying regularization strength for different image contents. For example, group 3 has much smaller average eigenvalues than groups 2 and 6, which indicates that during the reconstruction, patches dominated by group 3, typically the smooth soft-tissue patches, will be regularized more heavily than patches dominated by group 2 and 6, typically lung and bone patches respectively, and therefore will contain less noise in the reconstructed image.

As introduced in Sec. 3.4, we will apply the simple parameterization of (3.32) to the trained GMM covariances to tune the visual quality. The red plot in Fig. 3.6 illustrates this adjusted model with $p = 0.5$ and $\alpha = 33$ HU. By adjusting the model parameters, we increase the eigenvalues of group 3 and 4, which will consequently reduce the regularization for smooth and low-contrast soft-tissue contents, while we decrease the eigenvalues of group 2, 5, and 6, which will lead to stronger regularization for lung, high-contrast edge, and bone.

3.5.2 Results

2.5.2.1 2-D image denoising

We applied the proposed GM-MRF method in a 2-D image denoising experiment. The ground-truth image in Fig. 3.7(a) was obtained from the 3-D clinical CT image volume, whose data was collected on a GE Discovery CT750 HD scanner in 64×0.625 mm helical mode with 120 kVp, 200 mA, 0.5 s/rotation, pitch 0.984:1, and reconstructed in 320 mm FOV. Then, we added Gaussian white noise to the ground truth to generate the noisy image in Fig. 3.7(b). Different denoising methods were then applied to the noisy image.

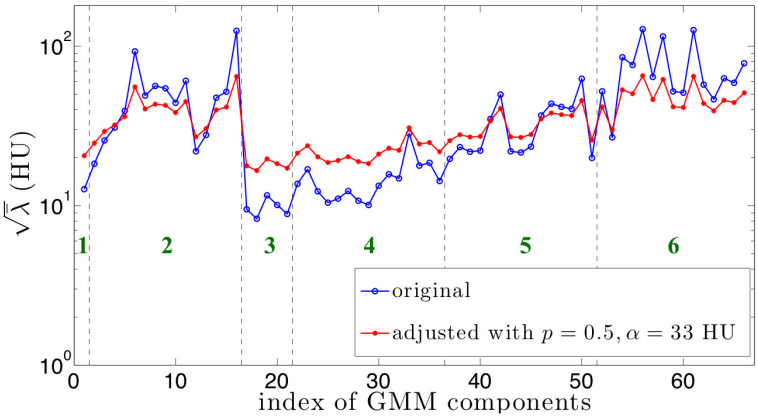
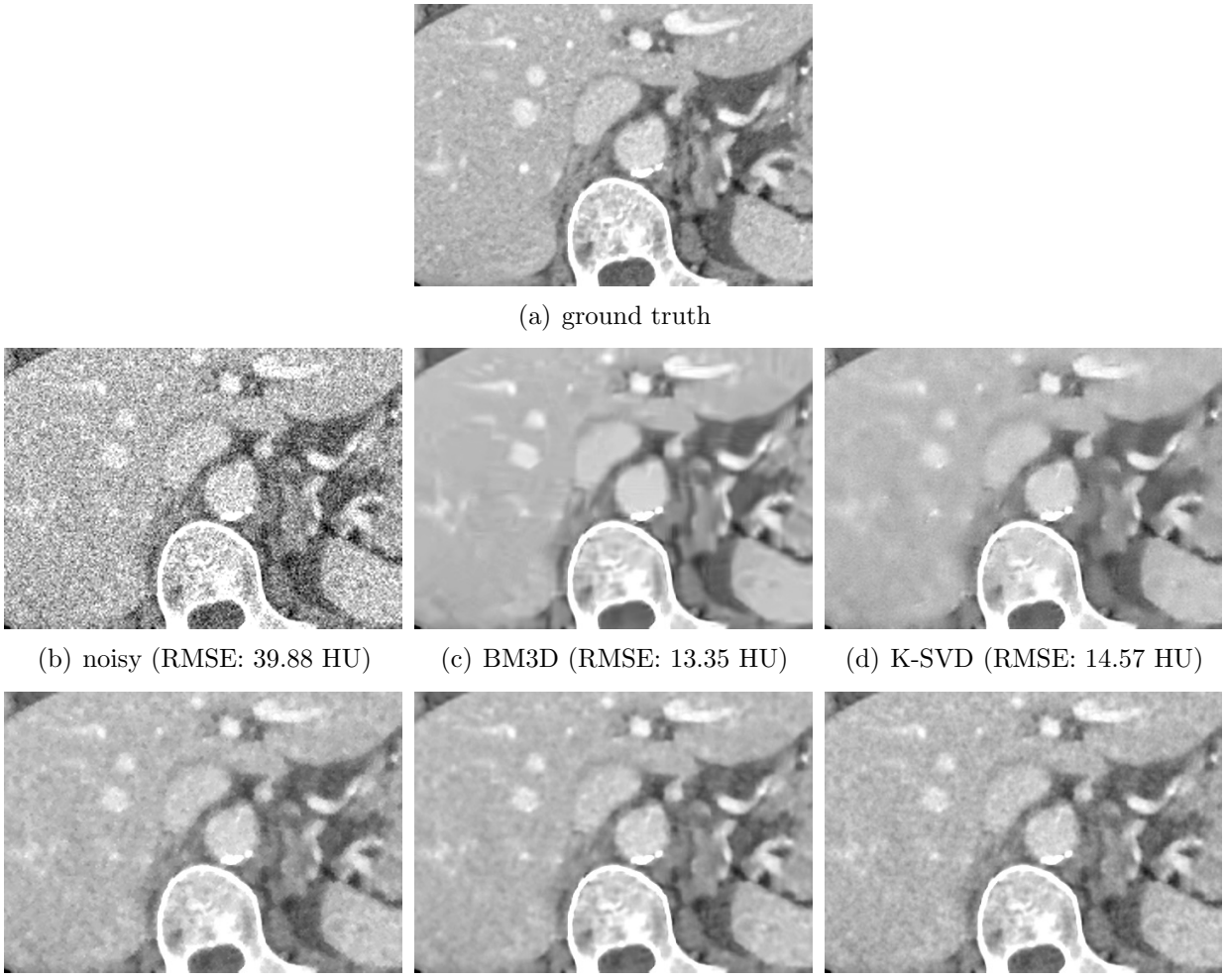


Fig. 3.6. The covariances of the originally trained GMM and the adjusted GMM with $p = 0.5$ and $\alpha = 33$ HU in (3.32), for a $5 \times 5 \times 3$ patch case. More precisely, the figure plots the square root of geometrically-averaged eigenvalues $\bar{\lambda}_k$ of the GMM covariances R_k, as $\bar{\lambda}_k = |R_k|^{\frac{1}{L}}$ with $L = 75$. Numbers within the figure correspond to the indices of the training groups. Within each group, the GMM components are sorted from the most probable to the least probable. The adjusted model increases the eigenvalues for group 1, 3, 4, and decreases the eigenvalues for most of group 2, 5, and 6.

For this experiment, we trained a 2-D GM-MRF model consisting of 66 GMM components for 5×5 image patches using the training procedure in Sec. 3.5.1. Note that the ground-truth image was not from the 3-D image volume used for training. We experimented with two different GM-MRF models, where one was the original model obtained directly from training and the other was adjusted with $p = 0.5, \alpha = 33$ HU in (3.32), so as to increase the regularization for high-contrast components and reduce the regularization for low-contrast components.

We compare the GM-MRF methods with a number of widely used methods, including the q-GGMRF method [3], K-SVD method [74], and BM3D method [77]. The q-GGMRF method is implemented with 3×3 neighborhood with parameters $p = 2$, $q = 1.2$, $c = 10$ HU. The K-SVD method is performed by using the software provided in [89] with 7×7 patch size and 512 dictionary entries. The BM3D method is performed by using the software provided in [90] with 8×8 patch. We adjust the



(a) ground truth

(b) noisy (RMSE: 39.88 HU) (c) BM3D (RMSE: 13.35 HU) (d) K-SVD (RMSE: 14.57 HU)

(e) q-GGMRF (RMSE: 15.96 HU) (f) original GM-MRF (RMSE: 13.78 HU) (g) adjusted GM-MRF (RMSE: 14.33 HU)

Fig. 3.7. Denoising results with different methods. The RMSE value between each reconstructed image and the ground truth is reported. Display window: [-100 200] HU. GM-MRF methods achieve lower RMSE values and better visual quality than q-GGMRF method and K-SVD method. Though having a slightly higher RMSE value, GM-MRF with original model preserves some real texture in soft tissue without creating severe artifacts, while BM3D tends to over-smooth the soft tissue and introducing some artificial, ripple-like structure. In addition, though compromising the RMSE than the original model, GM-MRF with the adjusted model produces better visual quality, especially for the soft-tissue texture.

regularization strength for all methods to achieve the lowest root-mean-square error (RMSE) value between the reconstructed image and the ground truth.

Fig. 3.7 presents the denoising result with different methods. It shows that the mean-square-error (MSE) achieved by GM-MRF method with original covariances is slightly higher than the BM3D method, but significantly lower than the q-GGMRF method and K-SVD method. Qualitatively, the GM-MRF method with original model produces sharper edges and less speckle noise than the q-GGMRF method, and preserves more fine structures and details than the K-SVD method. The BM3D method seems to produce more enhanced fine structures than GM-MRF due to its strong structure-preserving behavior. However, it tends to over-smooth the soft tissue region and meanwhile creating some artificial, ripple-like structures and texture. The GM-MRF method, on the contrary, is able to preserve some real texture in soft tissue without inducing severe artifacts, which can be important in some medical applications.

Interestingly, though compromising the MSE, the GM-MRF method with the adjusted model produces images with better visual quality than that with the original model. The better visual quality is achieved with improved soft-tissue texture and better rendering of high-contrast structures. This is because the MAP estimate tends to over-regularize the low-contrast regions and under-regularize the high-contrast regions in the image. Therefore, by adjusting the regularization strength in different contrast regions, we may achieve desirable visual quality but with the compromise in the MSE.

2.5.2.2 3-D CT reconstruction

We trained a 3-D GM-MRF model consisting of 66 GMM components for $5 \times 5 \times 3$ image patches, that is, a stack of 3 layers with a 5×5 patch at each 2-D layer, using the training procedure in Sec. 3.5.1. We applied the trained GM-MRF as a prior model in the MBIR algorithm. We will compare the MBIR using GM-MRF prior with two other methods: FBP using a standard kernel, and MBIR using a traditional pair-

wise MRF prior as the q-GGMRF model [3]. Note that we intentionally reduced the regularization for MBIR with the q-GGMRF prior so as to obtain higher resolution, which lead to much higher noise as well.

We first evaluate the performance of different methods using the GE Performance Phantom (GEPP) data taken with four different dose levels. The GEPP contains a plexiglas insert with cyclic water bars and a 50 μm diameter tungsten wire placed in water. We will measure the mean and standard deviation within fixed ROIs in flat regions to assess the reconstruction accuracy and noise. In addition, we will measure the modulation transfer function (MTF) using the wire to assess the in-plane resolution and contrast. We will report the 10% MTF since it reflects the visual resolution of the image, with higher value indicating finer texture, which is a desirable image quality especially for a low-dose condition. Data were collected on a GE Discovery CT750 HD scanner in 64×0.625 mm helical mode with 120 kVp, 1 s/rotation, pitch 0.516:1, with four different magnitudes of tube current as 290 mA, 145 mA, 75 mA, and 40 mA, and were later reconstructed in 135 mm FOV.

We compare the MBIR using GM-MRF prior with two widely used reconstruction methods: FBP using a standard kernel and MBIR using q-GGMRF prior with reduced regularization. We match the noise level between q-GGMRF and GM-MRF methods by adjusting the global regularization parameter, σ_x, in (3.30). More precisely, for a given dose level, we match the standard deviations in an ROI within water (ROI 1 in Fig. 3.9(a)) across a number of slices between q-GGMRF and GM-MRF methods such that the absolute difference of the two is within 1 HU.

Fig. 3.8 shows the GEPP reconstruction under normal X-ray dosage, with zoomed-in images for the tungsten wire and cyclic bars. It shows that MBIR with the traditional q-GGMRF prior produces sharper images with less noise than FBP, as indicated by smoother homogeneous regions, a smaller reconstructed wire, and more enhanced cycling bars. As a further improvement, MBIR with the GM-MRF priors produce even sharper image than MBIR with the q-GGMRF prior at a comparable noise level. The GM-MRF priors also improve the texture in smooth regions over the

q-GGMRF method by reducing the speckle noise and grainy texture. For the GM-MRF priors, the original model shows a sharper tungsten wire as compared to the adjusted model, since the adjusted model increases regularization for high-contrast edge (group 5) and bone (group 6), as shown in Fig. 3.6. However, the limited regularization for high-contrast edge and bone in the original model also leads to noisy rendering of high-attenuation objects, such as the non-circular tungsten wire and irregularly shaped small metal insertion.

The visual comparison is further verified by quantitative measurements in Fig. 3.9, which presents the measurements of reconstruction accuracy, noise, and resolution, of the GEPP reconstructions at different dose levels. It is shown that MBIR with GM-MRF priors improve the in-plane resolution (in Fig. 3.9(b)) while producing comparable or even less noise than FBP and MBIR with the q-GGMRF prior (in Fig. 3.9(c)(d)), without affecting the reconstruction accuracy (in Fig. 3.9(e)(f)).

In addition, we present the result of applying the GM-MRF priors on two clinical data sets; one is normal-dose and the other is low-dose. For the clinical data, we use the adjusted GM-MRF prior with $p = 0.5$ and $\alpha = 33$ HU for more balanced visual quality between low- and high-contrast regions. Similarly, we compare MBIR using the GM-MRF prior with FBP using a standard kernel and MBIR using the q-GGMRF prior with reduced regularization. For the normal-dose data, we adjust the global regularization in the reconstruction such that the noise measured in the aorta is matched between q-GGMRF and GM-MRF. However, for the low-dose data, it is challenging to match the noise between those two methods due to the excessive speckle noise produced by using under-regularized q-GGMRF prior. Thus, we will instead demonstrate that the GM-MRF prior achieves higher resolution with even less noise than the q-GGMRF prior.

We first present the reconstruction results with the normal-dose clinical data. The data was collected on a GE Discovery CT750 HD scanner in 64×0.625 mm helical mode with 120 kVp, 200 mA, 0.5 s/rotation, pitch 0.984:1, and reconstructed in 320 mm FOV. Figs. 3.10 - 3.11 show the reconstructed results of different methods.

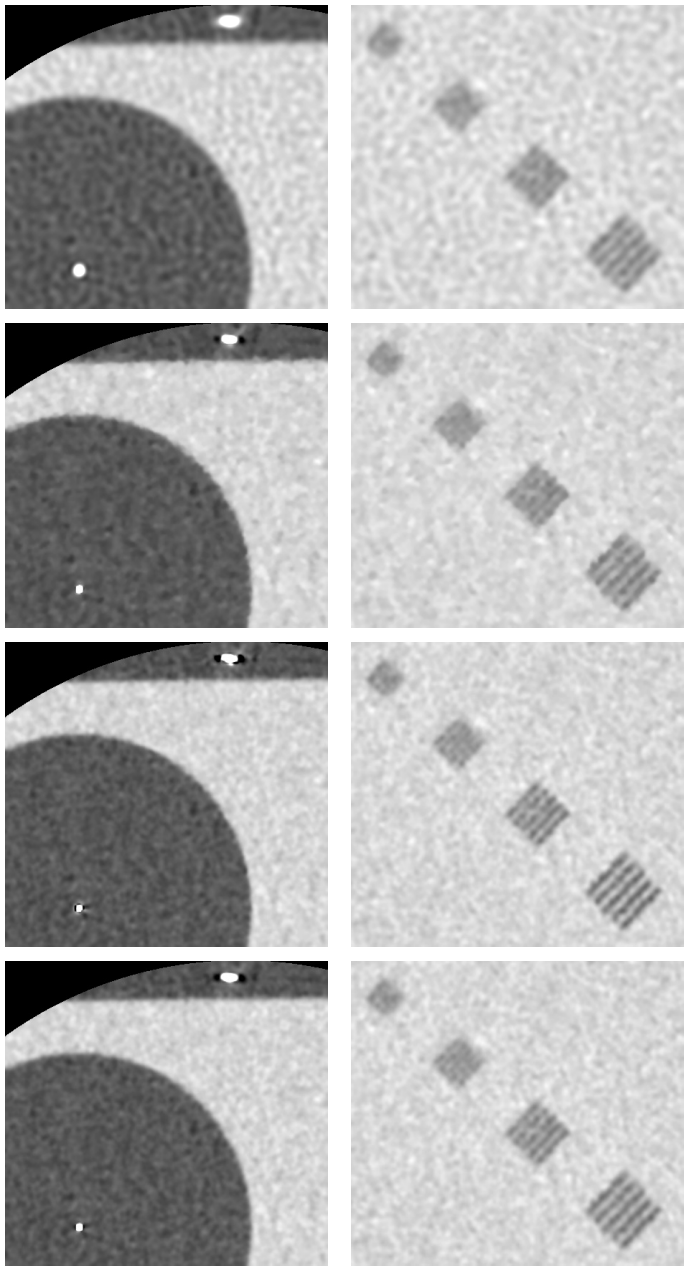
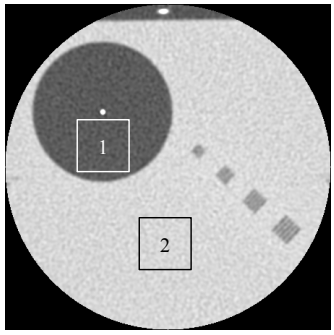
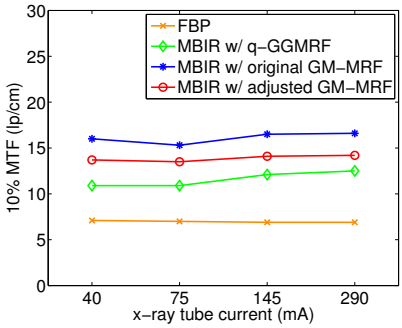


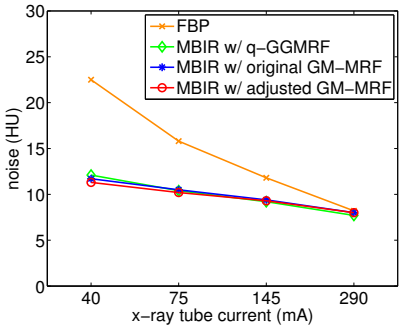
Fig. 3.8. GEPP reconstruction with data collected under 290 mA. From left to right, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, (c) MBIR with original GM-MRF, and (d) MBIR with adjusted GM-MRF with $p = 0.5, \alpha =$ 33 HU. The left column shows the wire section and the right column shows the resolution bars with the display window as [-85 165] HU.



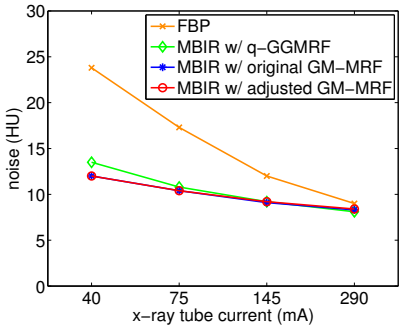
(a) measurements



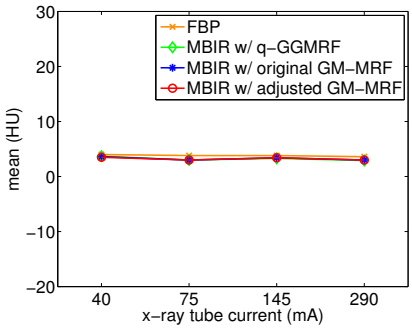
(b) 10% MTF



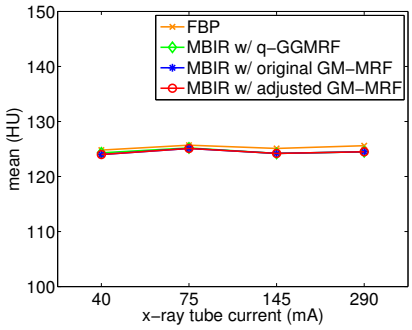
(c) noise within ROI 1



(d) noise within ROI 2



(e) mean within ROI 1



(f) mean within ROI 2

Fig. 3.9. Quantitative measurements for GEPP reconstructions. Four different magnitudes of X-ray tube current were used in data acquisition to achieve different X-ray dose levels. The mean values along with noise were measured within two different ROIs in (a). The MTF values were measured at the tungsten wire. Figure demonstrates that MBIR with GM-MRF priors improve the in-plane resolution in (b) while producing comparable or even less noise than FBP and MBIR with the q-GGMRF prior in (c) and (d), without affecting the reconstruction accuracy in (e) and (f).

As compared to FBP, MBIR with GM-MRF prior produces images with sharper bones, more lung details, as well as less noise in soft tissues. When compared to MBIR with traditional q-GGMRF prior with similar noise level, MBIR with GM-MRF prior reduces the jagged appearance in edges, as shown in the zoomed-in images in Fig. 3.10. These improvements are due to better edge definition in the patch-based model over traditional pair-wise models. Moreover, MBIR with GM-MRF prior reveals more fine structures and details in bone, such as the honeycomb structure of trabecular bone in the zoomed-in images of Fig. 3.11. This indicates that the GM-MRF model is also a very flexible prior and inherently allows different regularization strategies for different tissues in the CT images. This flexibility allows CT reconstructions with great soft-tissue quality while simultaneously preserving the resolution in regions with larger variation, such as bone and lung.

Figs. 3.12 - 3.13 present the reconstruction results with the low-dose clinical data. Experimental data was acquired from the same patient as for the training data in Sec. 3.5.1, with the same scan setting except for a lower tube current of 40 mA and a higher pitch of 1.375:1. It is shown that all the improvements revealed by experiment with normal-dose data can be observed more clearly in the low-dose situation, where the better image prior model is perhaps more valuable. Fig. 3.12 shows that the GM-MRF prior improves the texture in soft tissue without compromising the fine structures and details, as compared to the other methods. Particularly, when compared to MBIR with traditional q-GGMRF prior, MBIR with GM-MRF prior reduces the speckle noise in liver while still maintaining the normal texture and edge definition. Fig. 3.13 shows the improved resolution in lung and bone as produced by the GM-MRF prior. More specifically, the zoomed-in images show that the lung fissure reconstructed by MBIR with GM-MRF have comparable resolution as that produced by FBP, which is blurred by MBIR with q-GGMRF. The GM-MRF prior also leads to much clearer bone structure as compared to other methods. These improvements demonstrate the material-specific regularization capability of the GM-MRF prior.

3.6 Conclusion

In this chapter, we introduced a novel Gaussian-mixture Markov random field (GM-MRF) image model along with the tools to use it as a prior for model-based iterative reconstruction (MBIR). The proposed method constructs an image model by seaming together Gaussian-mixture (GM) patch models. In addition, we presented an analytical framework for computing the MAP estimate with the GM-MRF prior using an exact surrogate function. We also proposed a systematic approach to adjust the covariances of the GM components of the GM-MRF model, in order to control the sharpness in low- and high-contrast regions of the reconstruction separately. The results in image denoising and multi-slice CT reconstruction experiments demonstrate improved image quality and material-specific regularization by the GM-MRF prior.

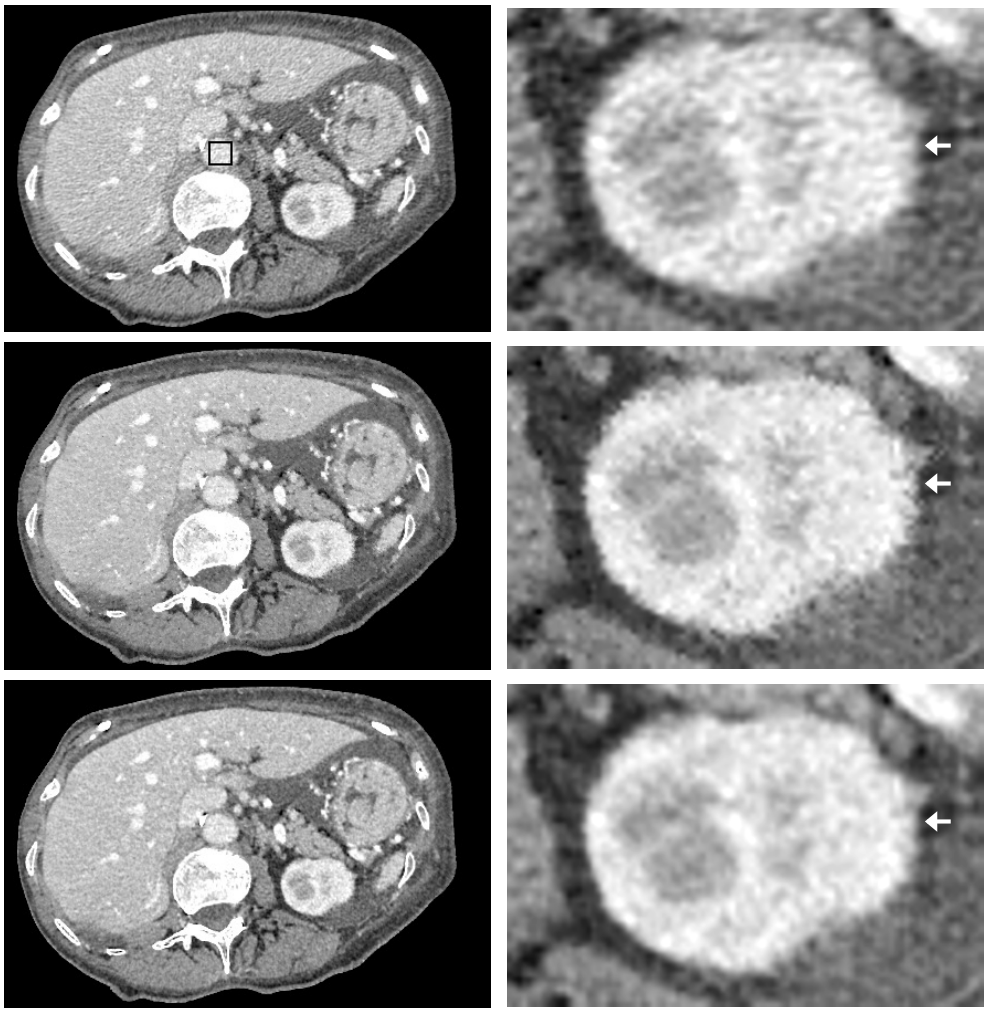


Fig. 3.10. An abdominal axial slice of the normal-dose clinical reconstruction. From top to bottom, the rows represent (a) FBP (noise: 19.14 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 14.43 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU (noise: 14.02 HU). The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is [-110 190] HU. Note the reduced jagged appearance in the GM-MRF reconstruction.

(a)

(b)

(c)



Fig. 3.11. A sagittal view of the normal-dose clinical reconstruction in bone window. From top to bottom, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Display window is [-300 900] HU. Note the honeycomb structure of the trabecular bone reconstructed by the GM-MRF prior, which is missing in other methods.

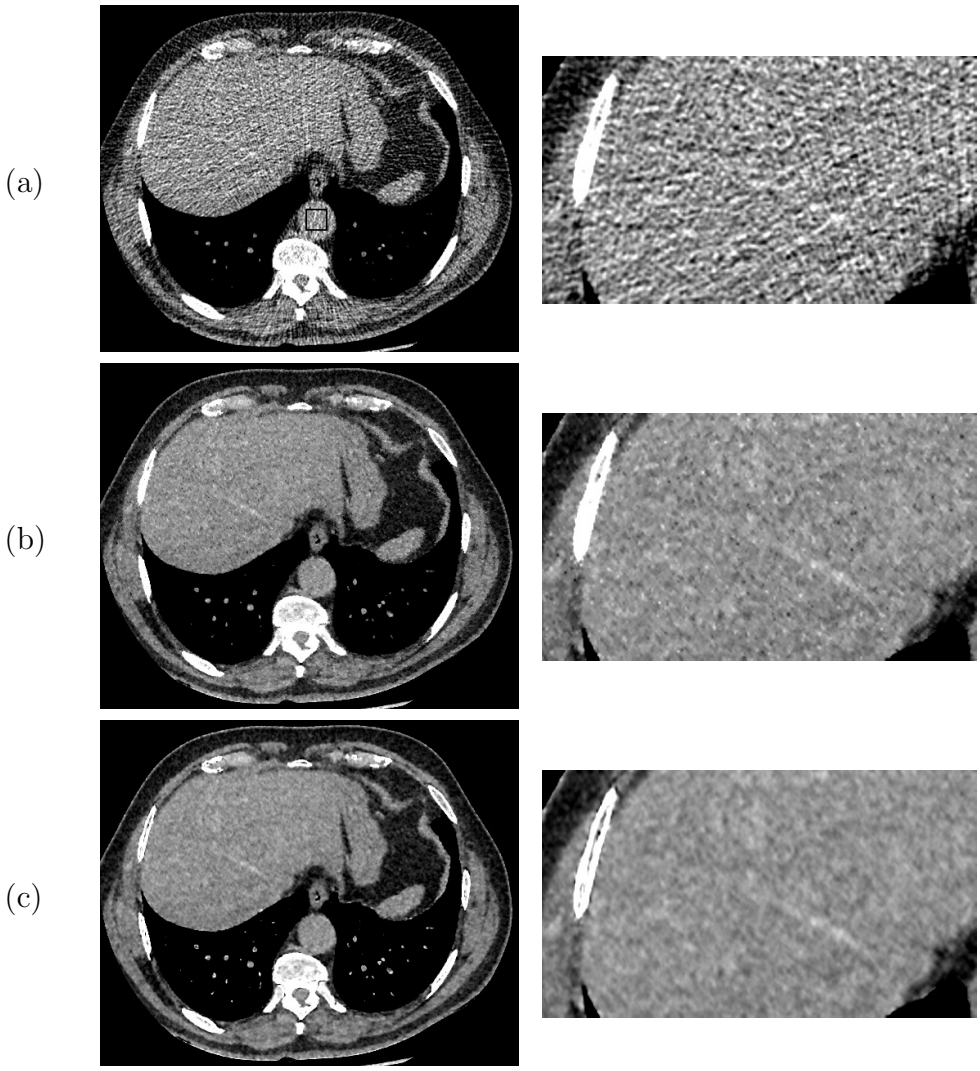


Fig. 3.12. An abdominal axial slice of the low-dose clinical reconstruction. From top to bottom, the rows represent (a) FBP (noise: 62.15 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 25.46 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU (noise: 18.41 HU). The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is [-160 240] HU. Note the suppression of speckle noise in soft tissue and improvement of sharpness in bone provided by MBIR with the GM-MRF prior.

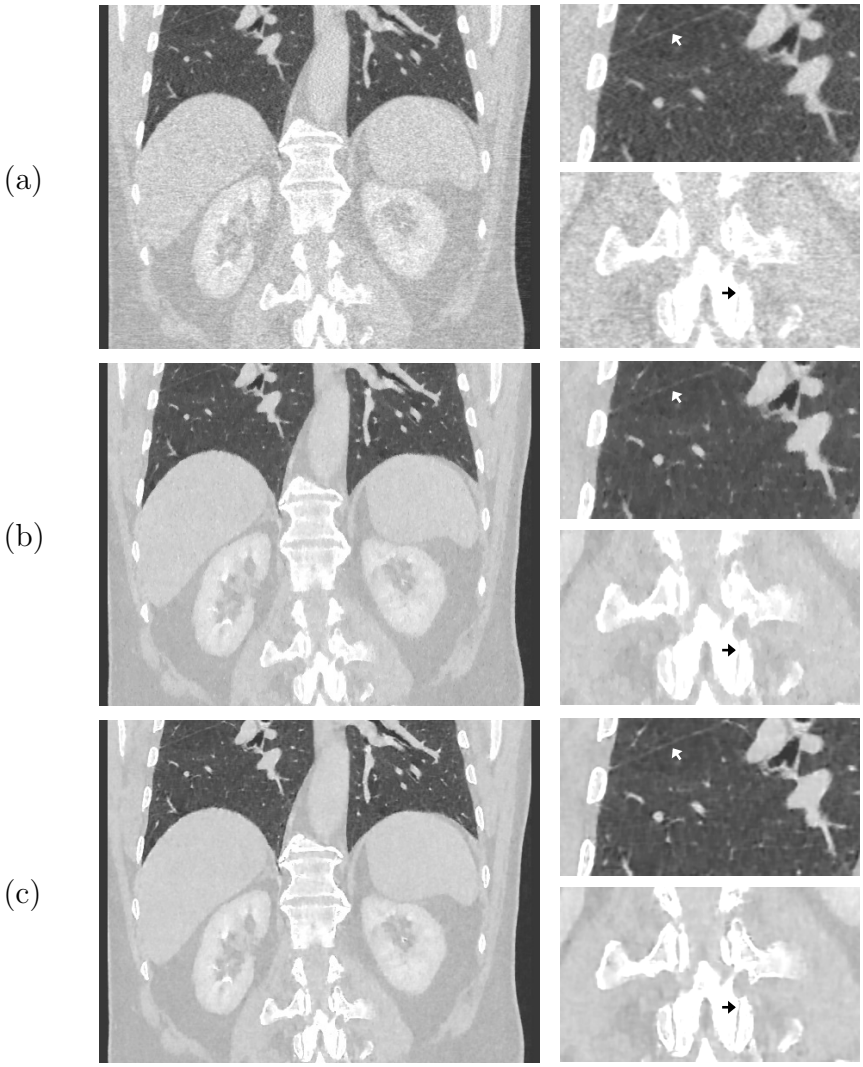


Fig. 3.13. A coronal view of the normal-dose clinical reconstruction in lung window. From top to bottom, the rows represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. The left column shows the full field-of-view (FOV) of the reconstructed images, while the right column shows zoomed-in FOVs for lung (top) and spine (bottom). Display window is [-1400 400] HU. Note the lung fissure and bone structure are reconstructed much more clearly by using the GM-MRF prior as compared to the other methods.

REFERENCES

REFERENCES

[1] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, pp. 534–548, Feb. 1993.

[2] C. A. Bouman and K. D. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Process.*, vol. 5, pp. 480–492, Mar. 1996.

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

[4] G. Wang, H. Yu, and B. De Man, "An outlook on X-ray CT research and development," *Med. Phys.*, vol. 35, no. 3, pp. 1051–1064, 2008.

[5] A. Ziegler, T. Kohler, and R. Proksa, "Noise and resolution in images reconstructed with FBP and OSC algorithms for CT," *Med. Phys.*, vol. 34, no. 2, pp. 585–598, 2007.

[6] Z. Yu, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. Image Process.*, vol. 20, pp. 161–175, Jan. 2011.

[7] R. C. Nelson, S. Feuerlein, and D. T. Boll, "New iterative reconstruction techniques for cardiovascular computed tomography: how do they work, and what are the advantages and disadvantages?" *J. Cardiovascular Computed Tomography*, vol. 5, no. 5, pp. 286–292, 2011.

[8] H. Scheffel, P. Stolzmann, C. L. Schlett, L.-C. Engel, G. P. Major, M. Károlyi, S. Do, P. Maurovich-Horvat, and U. Hoffmann, "Coronary artery plaques: cardiac CT with model-based and adaptive-statistical iterative reconstruction technique," *European J. Radiology*, vol. 81, no. 3, pp. e363–e369, 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?" *Investigative Radiology*, vol. 47, no. 8, pp. 468–474, 2012.

[10] Y. Yamada, M. Jinzaki, Y. Tanami, E. Shiomi, H. Sugiura, T. Abe, and S. Kuribayashi, "Model-based iterative reconstruction technique for ultralow-dose computed tomography of the lung: A pilot study," *Investigative Radiology*, vol. 47, no. 8, pp. 482–489, 2012.

[11] P. J. Pickhardt, M. G. Lubner, D. H. Kim, J. Tang, J. A. Ruma, A. M. del Rio, and G.-H. Chen, "Abdominal CT with model-based iterative reconstruction (MBIR): Initial results of a prospective trial comparing ultralow-dose with standard-dose imaging," *Amer. J. Roentgenology*, vol. 199, pp. 1266–1274, 2012.

[12] R. Zhang, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A model-based iterative algorithm for dual-energy X-ray CT reconstruction," in *Proc. 2nd Int. Mtg. on Image Formation in X-ray CT*, 2012, pp. 439–443.

[13] ——, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 117–134, 2014.

[14] R. Zhang, C. A. Bouman, J.-B. Thibault, and K. D. Sauer, "Gaussian mixture Markov random field for image denoising and reconstruction," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 1089–1092.

[15] R. Zhang, D. H. Ye, D. Pal, J.-B. Thibault, K. D. Sauer, and C. A. Bouman, "Gaussian mixture Markov random field for model-based iterative reconstruction," *submitted to IEEE Trans. Computational Imag.*, 2015.

[16] W. A. Kalender, W. H. Perman, J. R. Vetter, and E. Klotz, "Evaluation of a prototype dual-energy computed tomographic apparatus. I. Phantom studies," *Med. Phys.*, vol. 13, no. 3, pp. 334–339, May./Jun. 1986.

[17] T. Johnson, B. Krau, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard, C. Fink, S. Weckbach, M. Lenhard, B. Schmidt, T. Flohr, M. F. Reiser, and C. R. Becker, "Material differentiation by dual energy CT: initial experience," *European J. Radiology*, vol. 17, no. 6, pp. 1510–1517, Jun. 2007.

[18] S. Singh and M. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Process.*, vol. 83, no. 1, pp. 31–55, Jan. 2003.

[19] Z. Ying, R. Naidu, and C. R. Crawford, "Dual energy computed tomography for explosive detection," *J. X-Ray Sci. Tech.*, vol. 14, pp. 235–256, 2006.

[20] P. Engler and W. Friedman, "Review of dual-energy computed tomography techniques," *Mater. Eval.*, vol. 48, pp. 623–629, 1990.

[21] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Med. Phys.*, vol. 21, no. 5, pp. 733–744, Sept. 1976.

[22] L. A. Lehmann, R. E. Alvarez, A. Macovski, W. R. Brody, N. J. Pelc, S. J. Riederer, and A. L. Hall, "Generalized image combinations in dual KVP digital radiography," *Med. Phys.*, vol. 8, no. 5, pp. 659–667, Sept./Oct. 1981.

[23] W. R. Brody, G. Butt, A. Hall, and A. Macovski, "A method for selective tissue and bone visualization using dual energy scanned projection radiography," *Med. Phys.*, vol. 8, no. 3, pp. 353–357, May./Jun. 1981.

[24] K.-S. Chuang and H. K. Huang, "A fast deal-energy computational method using isotransmission lines and table lookup," *Med. Phys.*, vol. 14, no. 2, pp. 186–192, Mar./Apr. 1987.

[25] H. N. Cardinal and A. Fenster, "An accurate method for direct dual-energy calibration and decomposition," *Med. Phys.*, vol. 17, no. 3, pp. 327–341, May./Jun. 1990.

[26] B. Brendel, E. Roessl, J.-P. Schlomka, and R. Proksa, "Empirical, projection-based basis-component decomposition method," in *Proc. SPIE 7258, Med. Imag.: Phys. Med. Imag.*, 2009, p. 72583Y.

[27] P. Stenner, T. Berkus, and M. Kachelrieß, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Med. Phys.*, vol. 34, no. 9, pp. 3630–3641, Sept. 2007.

[28] U. Hassler, L. Garnero, and P. Rizo, "X-ray dual-energy calibration based on estimated spectral properties of the experimental system," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1699–1712, Jun. 1998.

[29] Y. Zou and M. D. Silver, "Analysis of fast kV-switching in dual energy CT using a pre-reconstruction decomposition technique," in *Proc. SPIE 6913, Med. Imag.: Phys. Med. Imag.*, 2008, p. 691313.

[30] J. Noh, J. A. Fessler, and P. E. Kinahan, "Statistical sinogram restoration in dual-energy CT for PET attenuation correction," *IEEE Trans. Med. Imag.*, vol. 28, pp. 1688–1702, Nov. 2009.

[31] R. A. Brooks, "A quantitative theory of the Hounsfield unit and its application to dual energy scanning," *J. Comput. Assist. Tomography*, vol. 1, no. 4, pp. 487–493, 1977.

[32] P. Joseph and R. Spital, "A method for correcting bone induced artifacts in computed tomography scanners," *J. Comput. Assist. Tomography*, vol. 2, no. 1, pp. 100–108, 1978.

[33] B. J. Heismann, J. Leppert, and K. Stierstorfer, "Density and atomic number measurements with spectral x-ray attenuation method," *J. Appl. Phys.*, vol. 94, no. 3, pp. 2073–2079, 2003.

[34] C. Maaß, M. Baer, and M. Kachelrieß, "Image-based dual energy CT using optimized precorrection functions: A practical new approach of material decomposition in image domain," *Med. Phys.*, vol. 36, no. 8, pp. 3818–3829, 2009.

[35] C. Maaß, E. Meyer, and M. Kachelrieß, "Exact dual energy material decomposition from inconsistent rays (MDIR)," *Med. Phys.*, vol. 38, no. 2, pp. 691–700, 2011.

[36] P. Sukovic and N. H. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy x-ray transmission tomography," *IEEE Trans. Med. Imag.*, vol. 19, pp. 1075–1081, Nov. 2000.

[37] J. A. Fessler, I. Elbakri, P. Sukovic, and N. H. Clinthorn, "Maximum-likelihood dual-energy tomographic image reconstruction," in *Proc. SPIE 4684, Med. Imag.: Image Process.*, 2002, pp. 38–49.

[38] J. A. O'Sullivan and J. Benac, "Alternating minimization algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 26, pp. 283–297, Mar. 2007.

[39] W. Huh and J. A. Fessler, "Model-based image reconstruction for dual-energy X-ray CT with fast KVP switching," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI): From Nano to Macro*, 2009, pp. 326–329.

[40] O. Semerci and E. L. Miller, "A parametric level-set approach to simultaneous object identification and background reconstruction for dual-energy computed tomography," *IEEE Trans. Image Process.*, vol. 21, pp. 2719–2734, May. 2012.

[41] P. E. Kinahan, J. A. Fessler, A. M. Alessio, and T. K. Lewellen, "Quantative attenuation correction for PET/CT using iterative reconstruction of low-dose dual-energy CT," in *Proc. IEEE Nucl. Sci. Symp.*, vol. 5, 2004, pp. 3285–3289.

[42] P. E. Kinahan, A. M. Alessio, and J. A. Fessler, "Dual energy CT attenuation correction methods for quantitative assessment of response to cancer therapy with PET/CT imaging," *Technology in Cancer Research and Treatment*, vol. 5, no. 4, pp. 319–327, Aug. 2006.

[43] W. Huh and J. A. Fessler, "Iterative image reconstruction for dual-energy x-ray CT using regularized material sinogram estimates," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI): From Nano to Macro*, 2011, pp. 1512–1515.

[44] R. Alvarez and E. Seppi, "A comparison of noise and dose in conventional and energy selective computed tomography," *IEEE Trans. Nucl. Sci.*, vol. 26, pp. 2853–2856, Apr. 1979.

[45] W. A. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual energy CT material density images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 218–224, Sept. 1988.

[46] N. H. Clinthorne, "A constrained dual-energy reconstruction method for material-selective transmission tomography," *Nucl. Ins. Met. Phys. Res. A*, vol. 353, no. 1-3, pp. 347–348, 1994.

[47] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical tomographic imaging," in *Proc. SPIE 6065, Computational Imag. IV*, 2006, p. 60650X.

[48] B. De Man and S. Basu, "Distance-driven projection and back projection in three-dimensions," *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2463–2475, Jun. 2004.

[49] A. J. Coleman and M. Sinclair, "A beam-hardening correction using dual-energy computed tomography," *Phys. Med. Biol.*, vol. 30, no. 11, pp. 1251–1256, 1985.

[50] S. J. Kisner, E. Haneda, C. A. Bouman, S. Skatter, M. Kourinny, and S. Bedford, "Limited view angle iterative CT reconstruction," in *Proc. SPIE 8296, Computational Imag. X*, 2012, p. 82960F.

[51] S. V. Venkatakrishnan, L. F. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "A model based iterative reconstruction algorithm for high angle annular dark field - scanning transmission electron microscope (HAADF-STEM) tomography," *IEEE Trans. Image Process.*, vol. 22, pp. 4532–4543, Nov. 2013.

[52] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, pp. 296–310, Jul. 1993.

[53] Z. Yu, J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "Accelerated line search for coordinate descent optimization," in *Proc. IEEE Nucl. Sci. Symp.*, vol. 5, 2006, pp. 2841–2844.

[54] J. Zheng, S. S. Saquib, K. D. Sauer, and C. A. Bouman, "Parallelizable bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Trans. Image Process.*, vol. 9, pp. 1745–1759, Oct. 2000.

[55] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, pp. 801–814, Sept. 1999.

[56] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Process.*, vol. 4, pp. 1417–1429, Oct. 1995.

[57] E. K. P. Chong and S. H. Żak, *An Introduction to Optimization Methods*, 3rd ed. New York, NY: Wiley & Sons, Inc., 2008, ch. 20.

[58] D. H. Martin, "The essence of invexity," *J. Optim. Theory Appl.*, vol. 47, no. 1, pp. 65–76, Sept. 1985.

[59] X. Wu, J. W. Leblanc, and P. Sainath, "Method for calibrating a dual-spectral computed tomography (CT) system," U.S. Patent 7 801 264, Sep. 21, 2010.

[60] X. Wu, D. A. Langan, D. Xu, T. M. Benson, J. D. Pack, A. M. Schmitz, J. E. Tkaczyk, J. Leverentz, and P. Licato, "Monochromatic CT image representation via fast switching dual kVp," in *Proc. SPIE 7258, Med. Imag.: Phys. Med. Imag.*, 2009, p. 725845.

[61] C. J. Bischof and J. C. Ehrhardt, "Modulation transfer function of the EMI CT head scanner," *Med. Phys.*, vol. 4, no. 2, pp. 163–167, 1977.

[62] R. T. Droege and R. L. Morin, "A practical method to measure the MTF of CT scanners," *Med. Phys.*, vol. 9, no. 5, pp. 758–760, 1982.

[63] D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Processing Mag.*, vol. 18, no. 6, pp. 57–75, 2001.

[64] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, p. R541, 2006.

[65] J. A. Fessler, "Model-based image reconstruction for MRI," *IEEE Signal Processing Mag.*, vol. 27, no. 4, pp. 81–89, 2010.

[66] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "Model-based iterative reconstruction for bright-field electron tomography," *IEEE Trans. Computational Imag.*, vol. 1, no. 1, pp. 1–15, 2015.

[67] P. Jin, C. A. Bouman, and K. D. Sauer, "A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT," *IEEE Trans. Computational Imag.*, vol. 1, no. 3, pp. 200–216, 2015.

[68] A. K. Mohan, S. Venkatakrishnan, J. Gibbs, E. Gulsoy, X. Xiao, M. De Graef, P. Voorhees, and C. A. Bouman, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Trans. Computational Imag.*, vol. 1, no. 2, pp. 96–111, 2015.

[69] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.

[70] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[71] K. D. Sauer and C. A. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 4, pp. 1144–1152, 1992.

[72] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.

[73] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

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

[75] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, 2011.

[76] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, 2012.

[77] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[78] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.

[79] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vision (ICCV)*, 1998, pp. 839–846.

[80] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2194–2204, 2012.

[81] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI): From Nano to Macro*, 2008, pp. 1375–1378.

[82] Y. Lu, J. Zhao, and G. Wang, "Few-view image reconstruction with dual dictionaries," *Phys. Med. Biol.*, vol. 57, no. 1, p. 173, 2012.

[83] L. Pfister and Y. Bresler, "Tomographic reconstruction with adaptive sparsifying transforms," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2014, pp. 6914–6918.

[84] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vision (ICCV)*, 2011, pp. 479–486.

[85] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, 2012.

[86] C. A. Bouman, "Cluster: An unsupervised algorithm for modeling Gaussian mixtures," Apr. 1997, available from http://engineering.purdue.edu/~bouman.

[87] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Computation*, vol. 24, no. 8, pp. 1967–2006, 2012.

[88] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[89] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Feb. 2013, available from www.cs.technion.ac.il/~ronrubin/software.

[90] K. Dabov, A. Danieyan, and A. Foi, "BM3D demo software for image/video restoration and enhancement," Jan. 2014, available from www.cs.tut.fi/~foi/GCF-BM3D/index.html#ref_software.

APPENDICES

A. SINGLE VOXEL UPDATE BY USING THE KKT CONDITION

We derive the solution to the 2D quadratic minimization problem defined in equation (2.48). The optimization problem is given by

$$\begin{aligned} \min_{u} \quad & \frac{1}{2} u \phi_2 u^T + u\phi_1 + const. \\ s.t. \quad & \\ & u \cdot n_{\min} \geq 0 \\ & u \cdot n_{\max} \geq 0 \end{aligned}$$

where $u \in \Re^2$ and

$$\begin{aligned} \phi_1 &= [\phi_1(1), \phi_1(2)]^T, \\ \phi_2 &= \begin{bmatrix} \phi_2(1,1) & \phi_2(1,2) \\ \phi_2(1,2) & \phi_2(2,2) \end{bmatrix}, \\ n_{\min} &= [n_{\min}(1), n_{\min}(2)], \\ n_{\max} &= [n_{\max}(1), n_{\max}(2)]. \end{aligned}$$

We solve this problem by using the KKT condition. The KKT condition states that a valid solution for this problem should satisfy the following conditions,

$$\begin{cases} \phi_2 u^T + \phi_1 - \lambda_1 n_{\min}^T - \lambda_2 n_{\max}^T &= 0 \\ \lambda_1 \cdot u \cdot n_{\min} &= 0 \\ \lambda_2 \cdot u \cdot n_{\max} &= 0 \\ u \cdot n_{\min} &\geq 0 \\ u \cdot n_{\max} &\geq 0 \\ \lambda_1 &\geq 0 \\ \lambda_2 &\geq 0 \end{cases} \tag{A.1}$$

where λ_1 and λ_2 are the KKT multipliers. Then we can compute the solution within four different cases, i.e., $(\lambda_1 = 0, \lambda_2 = 0)$, $(\lambda_1 > 0, \lambda_2 = 0)$, $(\lambda_1 = 0, \lambda_2 > 0)$ and $(\lambda_1 > 0, \lambda_2 > 0)$. Within each case, we compute the solution by using the equality conditions and then test the resulting solution with the inequality conditions. Since the cost function and the constraints in this problem are all continuously differentiable and convex, the KKT condition is both necessary and sufficient [58]. Thus, a solution becomes our updated value if and only if it satisfies both the equality and inequality conditions.

1. $\lambda_1 = 0, \lambda_2 = 0$. This combination gives the unconstrained solution, which can be computed from the first equation in (A.1) as

$$u = -\phi_1^T \phi_2^{-1}. \tag{A.2}$$

We need to test this solution with the following inequality conditions

$$\begin{cases} u \cdot n_{\min} &\geq 0 \\ u \cdot n_{\max} &\geq 0 \end{cases} \tag{A.3}$$

2. $\lambda_1 > 0, \lambda_2 = 0$. In this case, we solve the optimization problem on the boundary with $u \cdot n_{\min} = 0$. This combination leads to the following equations

$$\begin{cases} \phi_2 u^T + \phi_1 - \lambda_1 n_{\min}^T &= 0 \\ u \cdot n_{\min} &= 0 \end{cases} \tag{A.4}$$

The solution is given by

$$\begin{cases} u = & \left(\dfrac{-\phi_1(1) + k_1 \phi_1(2)}{\phi_2(1,1) - 2k_1\phi_2(1,2) + k_1^2 \phi_2(2,2)} \right) [1, -k_1] \\ \\ \lambda_1 = & \left(u \cdot [\phi_2(1,1), \phi_2(1,2)] + \phi_1(1) \right) / \left(n_{\min}(1) \right) \end{cases} \tag{A.5}$$

where $k_1 = n_{\min}(1)/n_{\min}(2)$. This solution needs to be tested with

$$\begin{cases} u \cdot n_{\max} &\geq 0 \\ \lambda_1 &> 0 \end{cases} \tag{A.6}$$

3. $\lambda_1 = 0, \lambda_2 > 0$. In this case, we solve the optimization problem on the boundary with $u \cdot n_{\max} = 0$. Similarly to the previous case, this combination gives the solution

$$\begin{cases} u = & \left(\dfrac{-\phi_1(1) + k_2 \phi_1(2)}{\phi_2(1,1) - 2k_2\phi_2(1,2) + k_2^2 \phi_2(2,2)} \right) [1, -k_2] \\ \\ \lambda_2 = & \left(u \cdot [\phi_2(1,1), \phi_2(1,2)] + \phi_1(1) \right) / \left(n_{\max}(1) \right) \end{cases} \tag{A.7}$$

where $k_2 = n_{\max}(1)/n_{\max}(2)$. This solution needs to be tested with

$$\begin{cases} u \cdot n_{\min} &\geq 0 \\ \lambda_2 &> 0 \end{cases} \tag{A.8}$$

4. $\lambda_1 > 0, \lambda_2 > 0$. With this combination, the only feasible solution is $u = [0, 0]$.

In practice, we test the four cases sequentially. Once all the equality and inequality conditions are met, the solution becomes the desired voxel update. This process is shown in Fig. 2.7.

B. PROOF OF LEMMA: SURROGATE FUNCTIONS FOR LOGS OF EXPONENTIAL MIXTURES

Proof:

$$\begin{aligned}
\log f(x) &= \log f(x') + \log\left\{\frac{f(x)}{f(x')}\right\} \\
&= \log f(x') + \log\left\{\sum_k \left(\frac{w_k}{f(x')}\right)\exp\{-v_k(x)\}\right\} \\
&= \log f(x') + \log\left\{\sum_k \left(\frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}}\right)\right. \\
&\left.\times \exp\left\{-v_k(x) + v_k(x')\right\}\right\} \\
&= \log f(x') + \log\left\{\sum_k \tilde{\pi}_k \exp\{-v_k(x) + v_k(x')\}\right\} \\
&\geq \log f(x') + \sum_k \tilde{\pi}_k\{-v_k(x) + v_k(x')\}
\end{aligned}$$

where

$$\tilde{\pi}_k \triangleq \frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}}.$$

The last inequality results from Jensen's inequality. Taking the negative of the final expression results in

$$-\log f(x) \leq -\log f(x') + \sum_k \tilde{\pi}_k\{v_k(x) - v_k(x')\} \triangleq q(x; x') \ ,$$

and evaluating this result at $x = x'$ results in

$$-\log f(x') = q(x'; x').$$

VITA

VITA

Ruoqiao Zhang received his Bachelor of Science degree in Engineering Physics from Tsinghua University, Beijing, China, in 2009, and is currently pursuing his Ph.D. degree in the School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA. His current research interest include computational imaging, statistical signal processing, and inverse problems with applications in computed tomography and medical image reconstruction.