Statistical models and regularization strategies in statistical image reconstruction of low-dose X-ray CT: a survey

Hao Zhang¹², Jing Wang³, Jianhua Ma⁴, Hongbing Lu⁵, and Zhengrong Liang¹²

¹ Department of Radiology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA
² Department of Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794, USA
³ Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
⁴ School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong 510515, China
⁵ Department of Biomedical Engineering, Fourth Military Medical University, Xi’an, Shaanxi 710032, China

Email: zhanghao@mil.sunysb.edu

Abstract

Statistical image reconstruction (SIR) methods have shown potential to substantially improve the image quality of low-dose X-ray computed tomography (CT) as compared to the conventional filtered back-projection (FBP) method. According to the maximum a posteriori (MAP) estimation, the SIR methods can be typically formulated by an objective function consisting of two terms: (1) data-fidelity (or equivalently, data-fitting or data-mismatch) term modeling the statistics of projection measurements, and (2) regularization (or equivalently, prior or penalty) term reflecting prior knowledge or expectation on the characteristics of the image to be reconstructed. Existing SIR methods for low-dose CT can be divided into two groups: (1) those that use calibrated transmitted photon counts (before log-transform) with penalized maximum likelihood (pML) criterion, and (2) those that use calibrated line-integrals (after log-transform) with penalized weighted least-squares (PWLS) criterion. Accurate statistical modeling of the projection measurements is a prerequisite for SIR, while the regularization term in the objective function also plays a critical role for successful image reconstruction. This paper reviews several statistical models on CT projection measurements and various regularization strategies incorporating prior knowledge or expected properties of the image to be reconstructed, which together formulate the objective function of the SIR methods for low-dose X-ray CT.

Keywords: X-ray CT; low-dose; statistical image reconstruction; statistical model; regularization; maximum a posteriori; optimization
Contents

List of acronyms
Symbols and definitions
1. Introduction
2. X-ray emission and projection data acquisitions
   2.1. X-ray photon emission and beam filtration
   2.2. Energy spectrum of incident X-ray photons
   2.3. Intensity distribution of incident X-ray photons
   2.4. X-ray photons interaction inside the patient
   2.5. Signal model for energy-integrating detector
   2.6. Degrading factors and signal preprocessing
3. Statistical image reconstruction (SIR) framework
   3.1. Statistical modeling on projection measurements
      3.1.1. Calibrated transmitted photon counts
      3.1.2. Calibrated line integrals
   3.2. Discrete forward models
   3.3. Maximum likelihood (ML)/weighted least-squares (WLS) criterions
      3.3.1. Poisson model for calibrated transmitted photon counts
      3.3.2. Poisson +Gaussian model for calibrated transmitted photon counts
      3.3.3. Gaussian model for calibrated line integrals
   3.4. Maximum a posteriori (MAP) estimation
4. Regularization strategies
   4.1. Spatially independent priors
   4.2. Markov random field (MRF) model-based priors
      4.2.1. Pair-wise Gibbs priors
      4.2.2. Adjusting weighting coefficients
      4.2.3. Median-based priors
   4.3. Compressed sensing (CS)-based regularizations
      4.3.1. L1 norm regularization
      4.3.2. L-p norm (0<p<1) based regularization
      4.3.3. L0 norm based regularization
   4.4. Nonlinear neighborhood filters (NNFs)-based regularizations
4.5. Patch-based roughness regularizations

4.6. Dictionary learning (DL)-based regularizations

4.7. Previous scan-induced regularizations

4.6.1. Via the PICCS method

4.6.2. Via the NLM method

4.6.3. Via global dictionary learning

4.6.4. Via predicted MRF coefficients

5. Discussions and conclusions
**List of acronyms**

| Acronym | Description                                      |
|---------|--------------------------------------------------|
| 2D      | two-dimensional                                  |
| 3D      | three-dimensional                                |
| AD      | adaptive dictionary                              |
| ADMM    | alternating direction method of multipliers      |
| ALARA   | *as low as reasonably achievable*                |
| ART     | algebraic reconstruction technique               |
| AwTV    | adaptive-weighted total variation                |
| CG      | conjugate gradient                               |
| CS      | compressed sensing                               |
| CT      | X-ray computed tomography                        |
| DL      | dictionary learning                              |
| DR-PICCS| dose reduction-prior image constrained compressed sensing |
| ED      | exponential dispersion                           |
| EM      | expectation-maximization                         |
| EPTV    | edge-preserving total variation                  |
| FBP     | filtered back-projection                         |
| FOV     | field of view                                    |
| GD      | global dictionary                                |
| GMRF    | Gaussian Markov random field                     |
| GPU     | graphical processing unit                        |
| ICD     | iterative coordinate descent                     |
| keV     | kiloelectron volt                                |
| kV      | kilovoltage                                      |
| MAP     | maximum *a posteriori*                           |
| mAs     | milliampere-second                              |
| ML      | maximum likelihood                               |
| MP      | median prior                                     |
| MRF     | Markov random field                              |
| MRP     | median root prior                                |
| Abbreviation | Full Form |
|-------------|-----------|
| NCPICCS     | non-convex prior image constrained compressed sensing |
| NLM         | nonlocal means |
| NNFs        | nonlinear neighborhood filters |
| OS          | ordered subsets |
| OSL         | one-step-late |
| PDF         | probability density function |
| PET         | positron emission tomography |
| PICCS       | prior image constrained compressed sensing |
| pML         | penalized maximum likelihood |
| preHQ       | previous high-quality image |
| PWLS        | penalized weighted least-squares |
| q-GGMRF     | q-generalized Gaussian Markov random field |
| SAGE        | space-alternating generalized expectation-maximization |
| SD          | steepest descent |
| SIR         | statistical image reconstruction |
| SPECT       | single photon emission computed tomography |
| SPS         | separable parabolic surrogates |
| SW          | search-window |
| TF          | tight frame |
| TRI-PICCS   | temporal resolution improvement-prior image constrained compressed sensing |
| TV          | total variation |
| WLS         | weighted least-squares |
Symbols and definitions

\( \Phi(E) \)  
normalized energy spectrum of incident X-ray photons

\( l \)  
index for spectrum energy bins (\( l=1,...,L \))

\( L \)  
total number of spectrum energy bins

\( \Phi(E_l) \)  
discrete normalized energy spectrum of incident X-ray photons

\( i \)  
index for X-ray paths (\( i=1,...,I \))

\( I \)  
total number of X-ray paths (projection measurements)

\( N_{oi} \)  
number of incident photons along the \( i \)th X-ray path

\( \bar{N}_{oi} \)  
the mean of \( N_{oi} \)

\( N_i \)  
number of transmitted photons along the \( i \)th X-ray path

\( \bar{N}_i \)  
the mean of \( N_i \)

\( \bar{N}_{oi}(E_l) \)  
mean number of incident photons with energy \( E_l \) along the \( i \)th X-ray path

\( N_i(E_l) \)  
number of transmitted photons with energy \( E_l \) along the \( i \)th X-ray path

\( \mu(E_i, \bar{r}) \)  
attenuation coefficient at position \( \bar{r} \) at reference energy \( E_i \)

\( \eta \)  
conversion factor from X-ray photon energy to signal strength in detector

\( m_{e,i} \)  
mean of electronic noise for the \( i \)th X-ray path

\( \sigma_{e,i}^2 \)  
variance of electronic noise for the \( i \)th X-ray path

\( \sigma_e^2 \)  
variance of electronic noise assuming it is the same for all X-ray paths

\( y_i \)  
line integral measurement along the \( i \)th X-ray path

\( \bar{y}_i \)  
the mean of \( y_i \)

\( \sigma_{y_i}^2 \)  
the variance of \( y_i \)

\( \mathbf{N} \)  
the vector of transmitted photon counts for all X-ray paths, \( \mathbf{N} = (N_1, ..., N_I)^T \)

\( \mathbf{y} \)  
the vector of line integral measurements for all X-ray paths, \( \mathbf{y} = (y_1, ..., y_I)^T \)

\( \bar{y} \)  
the mean of \( \mathbf{y} \)

\( j \)  
index for image voxels (\( j=1,...,J \))

\( J \)  
total number of image voxels
\( \mu_j \) attenuation coefficient at the \( j \)th voxel

\( \bm{\mu} \) the vector of attenuation coefficients of to-be-reconstructed image, \( \bm{\mu} = (\mu_1, \ldots, \mu_J)^T \)

\( a_{ij} \) contribution of the \( j \)th voxel to the \( i \)th X-ray path

\( \mathbf{A} \) system or projection matrix

\( Z_0, Z_1, Z_2 \) normalizing constant

\( \beta \) scalar control parameter

\( W_j \) MRF window of the \( j \)th pixel

\( m \) index for image pixels within the MRF window of the \( j \)th pixel

\( w_{jm} \) weighting coefficients between pixel \( j \) and pixel \( m \)

\( d_{jm} \) Euclidean distance between pixel \( j \) and pixel \( m \)

\( s, t \) horizontal/vertical index of the two-dimensional coordinate

\( p, q \) non-negative constant

\( \phi \) positive potential function

\( \Delta \) intensity difference between two pixels

\( \delta \) edge threshold in the image

\( \tau \) small positive constant

\( \xi \) error tolerance parameter for the constraints

\( \psi, \psi_1, \psi_2 \) sparsifying transform

\( \lambda, \lambda_1, \lambda_2 \) constant ranging 0–1

\( \mathbf{W} \) wavelet transform

\( \mathbf{S} \) shearlet transform

\( \tilde{\bm{\mu}} \) the vector of noisy or reference image

\( SW_j \) search-window of the \( j \)th image pixel

\( k \) index for image pixels within the search-window of the \( j \)th pixel

\( d_{jk} \) Euclidean distance between pixel \( j \) and pixel \( k \)

\( w_{jk} \) weighting coefficients between pixel \( j \) and pixel \( k \)

\( \sigma_d, \sigma_\mu \) spatial and intensity range of weighing in the bilateral filter or related regularization

\( h \) filtering parameter of the NLM filter or related regularization

\( c \) standard deviation of the Gaussian kernel
$\mathbf{P}(\hat{\mu}_j), \mathbf{P}(\hat{\mu}_k)$ the vector of a patch centered at pixel $j$ (or pixel $k$) of the image $\hat{\mu}$

$\mathbf{P}(\mu_j), \mathbf{P}(\mu_k)$ the vector of a patch centered at pixel $j$ (or pixel $k$) of the image $\mu$

$\mathbf{D}$ dictionary

$B$ the number of atoms in the dictionary

$d_b$ the $b$th atom of the dictionary $\mathbf{D}$

$G$ the number of pixels in a patch

$V$ the number of patches in the training image

$v$ index for patches in the training image ($v=1,...,V$)

$\alpha_v$ sparse representation of the $v$th patch under the trained dictionary

$\mathbf{F}_v$ operator to extract the $v$th patch from the to-be-reconstructed image

$\gamma_v$ Lagrange multiplier

$\mathbf{T}$ registration transformation operator

$\mu_{\text{preHQ}}$ the vector of attenuation coefficients of the previous high quality image

$\mu_{\text{preHQ registered}}$ the vector of attenuation coefficients of the previous high quality image registered to the to-be-reconstructed image

$\Omega_j$ MRF window of the $j$th pixel (with larger size)
1. Introduction

X-ray computed tomography (CT) has been widely exploited for various clinical applications such as diagnosis and image-guided interventions. In 2013, it was estimated that 76 million CT scans were performed across the hospitals and clinics in the United States. Recent discoveries regarding the potential harmful effects of X-ray radiation including genetic and cancerous diseases (de González and Darby 2004; Brenner and Hall 2007; de González et al 2009; Smith- Bindman et al 2009) have raised growing concerns to patients and the medical physics community. Consequently, low-dose CT with satisfactory image quality for specific clinical tasks is highly desirable. Many techniques and strategies have been proposed for radiation dose reduction of CT examinations (Hsieh 2009; McCollough et al 2009; Yu et al 2009) to achieve the as low as reasonably achievable (ALARA) principle. While improving the hardware of CT system can improve the dose efficiency, software approaches such as statistical image reconstruction (SIR) methods provide an alternative and more cost-effective means for further dose reduction while retaining satisfactory image quality.

In the past decade, two classes of strategies have been widely explored for radiation dose reduction: (1) lower the X-ray tube current and exposure time (i.e., milliampere-second (mAs)) or the X-ray tube voltage (i.e., kilovoltage (kV)) settings to reduce the X-ray flux towards each detector bin; and (2) lower the number of projection views per rotation during projection data acquisition. The former strategy would inevitably increase the projection data noise, and the resulting image by the conventional analytical filtered back-projection (FBP) method (equipped on most state-of-the-art commercial CT scanners) may be severely degraded due to the excessive noise. The latter strategy would produce undersampled projection data, and the resulting image by the FBP method usually suffers from view-aliasing artifacts due to insufficient angular sampling. Sometimes, these two strategies are even combined, leading to both noisy and undersampled projection data, and the corresponding image reconstructed by the FBP method can be further degraded.

In order to improve the CT image quality from the abovementioned low-dose acquisitions, the SIR methods were proposed (Rockmore and Macovski 1977; Lange and Carson 1984; Sauer and Bouman 1993; Bouman and Sauer 1996) and have become an endeavor for almost all major vendors of clinical CT systems (Beister et al 2012). Actually, their origins can be traced back to the early time of CT development in the decade of 1970s (Herman 1980). In parallel to the search for analytical inversion to the Radon transform for analytical CT image reconstruction, an alternative effort was devoted to discretize the Radon transform as a system of linear equations and then invert
the system of linear equations for algebraic CT image reconstruction. A typical example of the alternative effort is an iterative approach to the solution of the linear equations, rather than directly inverting the system matrix, by consideration of the unique nature of re-projection and back-projection operations in tomographic imaging. This iterative approach was thereafter named the algebraic reconstruction technique (ART) (Gordon et al. 1970; Herman 1980), and some variations were explored later, e.g., simultaneous ART (Andersen and Kak 1984; Jiang and Wang 2003). The ART was employed for image reconstruction of the original EMI CT scanners (Hounsfield 1968) in clinic until the analytical inversion of the Radon transform was established, named FBP (Herman 1980; Kak and Slaney 1988). For low-dose CT imaging, where data statistics is an essential factor to be considered in the image reconstruction (similar to the count-limited imaging modalities of single photon emission computed tomography (SPECT) and positron emission tomography (PET)), the SIR methods are desired and iterative strategies are needed (Rockmore and Macovski 1977; Shepp and Vardi 1982; Lange and Carson 1984; Geman and McClure 1985; Sauer and Bouman 1993; Bouman and Sauer 1996; Fessler 2000). Essentially, the SIR methods search for the image or solution that makes the projection measurements the most probable. Instead of treating all the measurements equally, a statistical model provides different degrees of credibility/reliability among measurements according to the signal-to-noise ratio (Thibault et al. 2007). Figure 1 illustrates typical image reconstruction methods for X-ray CT.

**Figure 1.** List of image reconstruction methods for X-ray CT.

Because of the explicit statistics modeling and potential dose reduction benefits, the SIR methods are likely to play a dominant role in image reconstruction development for low-dose CT in
the future. However, due to the ill-posedness of the reconstruction problem, the resulting image of those SIR that directly optimizes the maximum likelihood (ML) criterion can be very noisy and unstable. Alternatively, more recent and more sophisticated SIR methods are derived from the maximum *a posteriori* (MAP) estimation from the given the measurements, which consists of two terms in the objective function: the data-fidelity (or equivalently, data-fitting or data-mismatch) term modeling the statistics of measured data, and the regularization (or equivalently, prior or penalty) term incorporating prior knowledge or expected properties of the image to be reconstructed. The statistical modeling of the projection measurements is a prerequisite for building the data-fidelity term, and the regularization term also has a strong influence on the quality of reconstructed images.

In clinical CT systems, the raw signals from detectors are always preprocessed by CT vendors for various degrading factors such as scattered radiation, beam hardening, detector non-uniformity and so on (Hsieh 2009), while the raw signals (considered proprietary by vendors) are rarely accessible to academic researchers. Therefore, the researchers generally focus on investigating the properties of preprocessed CT signals in the past decades (Lu et al 2001; Whiting 2002; Li et al 2004; Whiting et al 2006; Wang et al 2008b; Ma et al 2012a). Commonly, the accessible projection data are calibrated transmitted photon counts (before log-transform) or calibrated line integrals (after log-transform). With the monochromatic X-ray assumption, the statistics of calibrated transmitted photon counts can be described by a Poisson distribution (Macovski 1983) or 'Poisson+Gaussian' distribution with consideration of additional electronic noise (Snyder et al 1993; Snyder et al 1995). This assumption has been well accepted in the CT field under the observations that although the polychromatic X-ray quanta may follow a Compound Poisson distribution (Whiting 2002; Elbakri and Fessler 2003; Whiting et al 2006), the difference from the Poisson distribution is small and merely on the variance (Li et al 2004). Meanwhile, the noise properties of calibrated line integrals have also been validated by experimental studies of repeated scans, and the statistical analysis showed that calibrated line integrals can be fitted approximately by a Gaussian distribution with a nonlinear signal-dependent variance (Lu et al 2001; Li et al 2004; Wang et al 2008b), regardless of the consideration of electronic noise (Ma et al 2012a). Therefore, existing SIR methods for X-ray CT either use calibrated transmitted photon counts with a simple Poisson approximation (or 'Poisson+Gaussian' approximation) or calibrated line integrals with a Gaussian approximation.

Extensive studies have also shown that the regularization term in the objective function of the SIR methods plays a critical role for successful image reconstruction (Huber 1981; Geman and
McClure 1985; Blake and Zisserman 1987; Lange et al 1987; Levitan and Herman 1987; Besag 1989; Hebert and Leahy 1989; Green 1990; Lange 1990; Stevenson and Delp 1990; Geman and Reynolds 1992; Bouman and Sauer 1993; Lalush and Tsui 1993; Charbonnier et al 1994; Alenius and Ruotsalainen 1997; Alenius et al 1998; Hsiao et al 2003; Fessler 2006; Sidky et al 2006; Xu and Tsui 2007; Chen et al 2008a; Sidky and Pan 2008; Wang et al 2008a; Chen et al 2009; Nett et al 2009; Tang et al 2009; Do et al 2010; Jia et al 2010; Yu and Wang 2010; Garduño et al 2011; Hu et al 2011; Jia et al 2011; Ouyang et al 2011; Ramirez-Giraldo et al 2011; Stayman et al 2011; Tian et al 2011; Chen et al 2012; Lauzier and Chen 2012; Liu et al 2012; Lu et al 2012; Ma et al 2012b; Vandeghinste et al 2012a; Vandeghinste et al 2012b; Wang and Qi 2012; Wang et al 2012; Xu et al 2012; Bai et al 2013; Chun and Talavage 2013; Lauzier and Chen 2013; Stayman et al 2013; Wu et al 2013; Zhou et al 2013; Zhu et al 2013; Dang et al 2014; Liu et al 2014a; Liu et al 2014b; Niu et al 2014; Zhang et al 2014a; Zhang et al 2014b; Zhang et al 2014c; Zhang et al 2014d; Zhang et al 2015a; Zhang et al 2015b). The regularization term incorporates prior knowledge or expectations of smoothness or other characteristics in the image, which can help to stabilize the solution and suppress the noise and streak artifacts. Various regularizations have been presented in the past decades based on different assumptions, models and knowledge. Although some of them were initially proposed for SIR of SPECT and PET, they can be readily employed for CT. In this survey, we group the regularizations that were, or could be, used for the SIR methods of low-dose CT into seven categories: (1) the spatially independent priors, which assume statistical independence among different voxels in the image; (2) the Markov random field (MRF) model-based priors, which assume the image is locally smooth, i.e., the neighboring voxels tend to have similar values; (3) the compressed sensing (CS)-based regularizations, which assume the medical image can have a very sparse representation after proper transform; (4) the nonlinear neighborhood filters (NNFs)-based regularizations, which assume that each voxel in the image can be represented with a weighted average of its neighboring voxels according to the similarity; (5) the patch-based roughness regularizations, which utilize patch concept to improve the traditional voxel-based roughness regularizations; (6) the dictionary learning (DL)-based regularizations, which utilize the sparse representation in terms of a redundant dictionary; and (7) the previous scan-induced regularizations, which take advantage of rich anatomical information from previous scan to improve the current low-dose scan image reconstruction. We will describe all these regularization strategies explicitly in Section 4.
The remainder of this paper is organized as follows. In Section 2, we introduce the CT projection data acquisition and preprocessing. Section 3 illustrates the statistical models of CT projection measurements and derives the SIR framework for low-dose CT accordingly. Section 4 devotes to the presentation of various regularizations which can be used for SIR of low-dose CT. After a discussion of existing issues, we draw conclusions of this survey in Section 5.

2. X-ray emission and projection data acquisitions

2.1. X-ray photons emission and beam filtration

The X-ray photons emitted from X-ray tube have a wide energy spectrum. In clinical CT systems, in order to reduce radiation dose and beam-hardening artifacts, a flat filter is usually employed to remove the low-energy photons which mostly would be absorbed by the patients and make little contribution to the detected signals (Hsieh 2009). To further optimize radiation dose utility and improve the noise homogeneity in the projection data after traversing the body, an additional bowtie filter is also commonly used to compensate for the typically oval shape of the patient in cross section (Hsieh 2009), see Figure 2. The beam filtration by the two filters modifies the energy spectrum and intensity distribution of the emitted X-ray photons across the field of view (FOV), which will be discussed in Section 2.2 and 2.3.

Figure 2. Illustration of the bowtie filter and flat filter used in clinical CT system.
2.2. Energy spectrum of incident X-ray photons

The X-ray photons after filtration and just before entering the patient are defined as the incident photons in this work. Figure 3 shows representative normalized energy spectra ($\int_{E} \Phi(E) dE = 1$) of incident X-ray photons with four different tube voltage (kV) settings. The spectra were plotted by the Spektr (Siewerdsen et al. 2004) with tungsten anode and 3mm aluminum filter. It can be observed that the maximum energy of the incident X-ray photons is determined by the tube voltage, and most of the low-energy photons (e.g., <20 keV) are removed due to the beam filtration.

![Normalized energy spectrum $\Phi(E)$](image)

**Figure 3.** Illustration of normalized energy spectra of incident X-ray photons with tube voltage at 80, 100, 120 and 140 kV.

If we discretize the polychromatic spectrum at discrete energies $E_l$ in energy bin $l$ ($l = 1, 2, ..., L$), we have:

$$\sum_{l=1}^{L} \Phi(E_l) = 1$$  \hspace{1cm} (1)

Although the normalized energy spectrum of the incident X-ray photons can be slightly different for each detector bin (Nuyts et al. 2013) due to effects such as bowtie filtration (Toth et al. 2005) and anode angulation (La Rivière and Vargas 2008), the variation is generally neglected practically (Whiting et al. 2006).

2.3. Intensity distribution of incident X-ray photons

*IRIS Laboratory, Department of Radiology, Stony Brook University*
For a given tube voltage and tube current in a given time interval (fixed kV and mAs level), the
number of incident photons along the \( i \)th X-ray path, \( N_{oi} \), is widely treated as a Poisson random
variable (Macovski 1983) with a mean \( \bar{N}_{oi} \). Since \( N_{oi} \) is a large number in most circumstances, the
noise in the incident photon counts can often be ignored (Macovski 1983), that is, \( N_{oi} \approx \bar{N}_{oi} \). \( \bar{N}_{oi} \) can
be estimated by the system calibration, e.g., by repeated air scans. In clinical CT systems, because of
the use of bowtie filter and the heel effect (Whiting et al 2006), \( \bar{N}_{oi} \) is not the same for all X-ray
paths and is depending on the detector bin position.

Figure 4 reflects the distribution of \( \bar{N}_{oi} \) of different detector bins across the FOV at a fixed mAs
level and four different kV settings for a Siemens scanner (Manduca et al 2009). Alternatively, if we
fix the X-ray tube voltage and adjust the mAs levels, we can observe similar curves of \( \bar{N}_{oi} \) as in
Figure 4.

2.4. X-ray photons interaction inside the patient

The process of X-ray photons interacting with human tissue is assumed to be a binary process,
where the photons are either attenuated (absorbed or/and scattered) or pass through without any
interaction (Macovski 1983). X-ray photons with different energies have different surviving
probabilities. Let \( \bar{N}_{oi}(E_i) \) denote the mean number of incident photons with energy \( E_i \) along the \( i \)th
X-ray path. Then, the number of transmitted photons with energy \( E_i \) along the \( i \)th X-ray path,
\( N_i(E_i) \), is a statistical independent Poisson random variable (Macovski 1983), governed by the Lambert-Beer's law:

\[
N_i(E_i) = \text{Poisson}\left\{ \tilde{N}_0(E_i) \cdot \exp\left[ -\int_{ray \cdot i} \mu(E_i, \bar{r}) d\bar{r} \right] \right\}
\]

\[
= \text{Poisson}\left\{ \tilde{N}_0 \cdot \Phi(E_i) \cdot \exp\left[ -\int_{ray \cdot i} \mu(E_i, \bar{r}) d\bar{r} \right] \right\}
\]

where \( \mu(E_i, \bar{r}) \) is the attenuation coefficient of the patient at the position \( \bar{r} \) at energy \( E_i \).

Consequently, the total number of transmitted photons along the \( i \)th X-ray path can be given as:

\[
N_i = \sum_{i=1}^{L} N_i(E_i) = \sum_{i=1}^{L} \text{Poisson}\left\{ \tilde{N}_0 \cdot \Phi(E_i) \cdot \exp\left[ -\int_{ray \cdot i} \mu(E_i, \bar{r}) d\bar{r} \right] \right\}
\]

2.5. Signal model for energy-integrating detector

In current clinical CT systems, energy-integrating detectors are commonly used. The detected signal strength is proportional to the energy that the transmitted photons carry. Let \( \eta \) be the conversion or gain factor from X-ray photon energy (keV) to detected signal. The total signal strength along the \( i \)th X-ray path can be given as (La Rivière 2005; La Rivière et al 2006):

\[
\text{Signal}_i = \sum_{i=1}^{L} \eta \cdot E_i \cdot \text{Poisson}\left\{ \tilde{N}_0 \cdot \Phi(E_i) \cdot \exp\left[ -\int_{ray \cdot i} \mu(E_i, \bar{r}) d\bar{r} \right] \right\}
\]

The energy weighted combination of Poisson random variables in Eq. (4) induces the compound Poisson statistics, which has been described by (Whiting 2002; Elbakri and Fessler 2003; Whiting et al 2006).

The detected signal is read out through detector electronics, therefore, extra uncertainty is added to the recorded signal due to the electronic noise. The electronic noise is intrinsic to the detection system and results from electronic fluctuation in the detector photodiode and other electronic components (Hsieh 2009). The electronic noise is typically modeled as additive Gaussian noise, where the mean and variance reflect the detector dark current and readout noise of electronics, respectively (La Rivière 2005; La Rivière et al 2006; Hsieh 2009; Ma et al 2012a). The mean of the electronic noise can be determined immediately before each scan by sampling the signals in unexposed detectors over some time interval, and the variance of the electronic noise can be estimated from the sample variance of a series of dark current measurements (La Rivière et al 2006; Hsieh 2009; Ma et al 2012a). Consequently, the recorded signal can be described by:
\[ \text{Signal}_i = \sum_{j=1}^{J} H_j \cdot E_j \cdot \text{Poisson} \left\{ \overline{N}_{0i} \cdot \Phi(E_j) \cdot \exp \left( -\int_{\text{ray } i} \mu(E_i, \vec{r}) d\vec{r} \right) \right\} + \text{Gaussian}(M_{e,i}, \sigma_{e,i}^2) \]  

where \( M_{e,i} \) and \( \sigma_{e,i}^2 \) denotes the mean and variance of electronic noise for the \( i \)th projection measurement, respectively.

### 2.6. Degrading factors and signal preprocessing

In reality, besides the statistical noise mentioned above, there are several other factors that degrade the recorded signals, including off-focal radiation, beam hardening, scattered radiation, detector speed and afterglow, detective quantum efficiency, detector nonlinearity, crosstalk, quantization noise, etc (Hsieh 2009). As a result, the statistics of raw CT signals can be very complicated. In clinical practice, the raw CT signals are always preprocessed/calibrated by vendors for these degrading factors, while the specific operations are proprietary in nature and unavailable in the public domain (Hsieh 2009). Therefore, the typically accessible projection data to academic researchers are calibrated transmitted photon counts (before log-transform), or calibrated line integrals (after log-transform) which are given by:

\[ y_i = \ln \left( \frac{\overline{N}_{0i}}{N_i} \right) \]  

where \( y_i \) represents the line integral measurement along the \( i \)th X-ray path, and the approximation in Eq. (6) reflects an assumption that the Lambert-Beer’s law can be applied to random values. Figure 5 illustrates the flow chart from raw CT signal acquisition, preprocessing, up to the tomographic reconstruction from calibrated transmitted photon counts or calibrated line integrals.

![Flow chart of CT projection data preparation for image reconstruction.](image)

### 3. Statistical image reconstruction (SIR) framework

#### 3.1. Statistical modeling on projection measurements

##### 3.1.1. Calibrated transmitted photon counts

According to the analysis in Section 2, the calibrated transmitted photon counts along the \( i \)th X-ray path can be expressed as (Ma et al 2012a):

\[ N_i \sim \text{compoundPoisson}(\overline{N}_i, \Phi) + \text{Gaussian}(M_{e,i}, \sigma_{e,i}^2) \]
While the mathematical formula of the compound Poisson distribution has been explicitly derived (Whiting 2002; Whiting et al 2006), the lack of analytical probability density function (PDF) expression impedes its use in SIR method development (Lasio et al 2007; Xu and Tsui 2014). Additionally, this model has more challenges when the electronic noise is considered, which makes this exact model impractical for SIR.

Practically, this exact model can be well-approximated by a simple Poisson noise model assuming monochromatic X-ray source, and is widely used in the SIR methods:

\[ N_i \sim \text{Poisson}(\bar{N}_i) \]  

(8)

Although the simple Poisson model is acceptable in most cases, the influence of electronic noise becomes non-neglectable, and has been considered as an important factor affecting the image quality (Xu and Tsui 2009) under very low-flux acquisitions (transmitted photon counts are extremely low). In clinical CT systems, in order to reduce the effect of detector dark current, the mean of electronic noise is often calibrated to be zero in practice (Hsieh 2009; Ma et al 2012a). Also, the variance of electronic noise is assumed to be the same for all X-ray paths and thus the index can be removed. As a result, the statistics of the calibrated transmitted photon count along the \( i \)th X-ray path can be described by (Snyder et al 1993; Snyder et al 1995; La Rivièere and Billmire 2005; Ma et al 2012a):

\[ N_i \sim \text{Poisson}(\bar{N}_i) + \text{Gaussian}(0, \sigma_e^2) \]  

(9)

However, the likelihood function of the 'Poisson+Gaussian' model in Eq. (9) is still analytically intractable. To circumvent this problem, a shifted Poisson approximation can be exploited to match the first two statistical moments (Yavuz and Fessler 1998; La Rivièere and Billmire 2005). That is, the random variable \( \hat{N}_i = [N_i + \sigma_e^2]_+ \) can have the variance equal to its mean \( (\bar{N}_i + \sigma_e^2) \). Based on the relationship of variance equal to the mean, we assume that the shifted random variable \( \hat{N}_i \) follows Poisson distribution:

\[ \hat{N}_i = [N_i + \sigma_e^2]_+ \sim \text{Poisson}(\bar{N}_i + \sigma_e^2) \]  

(10)

where \( [x]_+ = x \) if \( x > 0 \) and is 0 otherwise.

3.1.2. Calibrated line integrals

The noise property of the calibrated line integrals has been investigated by analyzing experimental data of a physical phantom from repeated scans. The statistical analysis showed that
the calibrated line integrals can be fitted approximately by a Gaussian distribution with a nonlinear signal-dependent variance (Lu et al 2001; Li et al 2004; Wang et al 2008b):

$$y_i \sim \text{Gaussian}(\bar{y}_i, \sigma_{y_i}^2)$$

(11)

With the Poisson noise model of the transmitted photon counts in Eq. (8) and the Taylor's expansion, it has been shown in (Macovski 1983) that the variance of the line integral $$y_i$$ is given by:

$$\sigma_{y_i}^2 = 1/ N_i = \exp(\bar{y}_i) / N_{w}$$

(12)

Similarly, with the 'Poisson+Gaussian' noise model of the transmitted photon counts in Eq. (9), the variance of the line integral can be described by (Thibault et al 2006; Wang et al 2008c; Ma et al 2012a):

$$\sigma_{y_i}^2 = \frac{N_i + \sigma_e^2}{N_i^2} = \frac{1}{N_{o_i}} \exp(\bar{y}_i) \left( 1 + \frac{\sigma_e^2}{N_{o_i}} \exp(\bar{y}_i) \right)$$

(13)

Essentially, Eqs. (12) and (13) still have the monochromatic X-ray source assumption because of the models from which they are derived. Since the polychromatic nature of X-ray generation does not change the mean and only affect the variance of the calibrated line integrals, Eqs. (12) and (13) actually have the potential to consider the polychromatic nature by replacing $$1/ N_{o_i}$$ with a factor $$\Gamma_i$$, where $$\Gamma_i$$ is no longer exactly equal to $$1/ N_{o_i}$$ but can be measured by repeated scans and experimental data fitting (Lu et al 2001; Li et al 2004; Wang et al 2008b; Ma et al 2012a). Table 1 summarizes the statistical models for the calibrated transmitted data and the calibrated line integrals.

**Table 1.** Statistical models for CT projection measurements in SIR methods development

|                      | w/o electronic noise | w/ electronic noise |
|----------------------|----------------------|---------------------|
| **Calibrated photon counts** | $$N_i \sim \text{Poisson}(\bar{N}_i)$$ | $$N_i \sim \text{Poisson}(\bar{N}_i) + \text{Gaussian}(0, \sigma_e^2)$$ |
|                      | $$\bar{N}_i = \left[ N_i + \sigma_e^2 \right] - \text{Poisson}(\bar{N}_i + \sigma_e^2)$$ |
| **Calibrated line integrals** | $$y_i \sim \text{Gaussian}(\bar{y}_i, \sigma_{y_i}^2)$$ | $$y_i \sim \text{Gaussian}(\bar{y}_i, \sigma_{y_i}^2)$$ |
|                      | $$\sigma_{y_i}^2 = \Gamma_i \cdot \exp(\bar{y}_i)$$ | $$\sigma_{y_i}^2 = \Gamma_i \cdot \exp(\bar{y}_i) \cdot \left[ 1 + \Gamma_i \cdot \exp(\bar{y}_i) \cdot \sigma_e^2 \right]$$ |
3.2. Discrete forward models

Under the assumption of monochromatic X-ray generation, the CT projection measurement can be expressed as:

\[
\overline{y}_i = \ln(\overline{N}_{oi} / \overline{N}_i) = \int_{\text{ray}_i} \mu(\bar{r})d\bar{r} \approx \sum_j a_{ij} \mu_j = [A\mu]_i
\]  

(14)

or described as a discrete linear system:

\[
\overline{y} = A\mu
\]

(15)

where \(\overline{y} \in \mathbb{R}^{I \times 1}\) denotes the vector of expected line integrals, and \(I\) is the total number of projection measurements; \(\mu \in \mathbb{R}^{J \times 1}\) represents the vector of attenuation coefficients of the object to be reconstructed, and \(J\) is the total number of image voxels; \(A \in \mathbb{R}^{I \times J}\) is the projection matrix and its element \(a_{ij}\) represents the contribution of voxel \(j\) to the \(i\)th projection ray. In CT imaging, the projector mainly accounts for the geometry of the imaging system. A variety of projectors (Herman 1980; Lo 1988; Zeng and Gullberg 1993; Zhuang et al. 1994; Matej and Lewitt 1996; De Man and Basu 2004; Ziegler et al. 2006; Long et al. 2010; Liu et al. 2013) as well as strategies to accelerate the process (Peters 1981; Siddon 1985; Han et al. 1999; Mueller et al. 1999; Riddell and Troussé 2006; Wu and Fessler 2011) have been proposed, comprehensive review of which are out of the scope of this paper. Herein, we briefly review three commonly used approaches to calculate the projection matrix: (1) voxel-driven (or pixel-driven, in 2D presentation), which connects a line from the focal spot through the center of concerned voxel and projects to the detector (Herman 1980); (2) ray-driven, which casts one ray (or several rays) from the focal spot to the center (or boundaries) of the detector bin of interest and utilize the intersected length (Herman 1980; Siddon 1985; Zeng and Gullberg 1993), intersected area (for 2D case), or intersected volume (for 3D case) as the weight; (3) distance-driven, which maps the boundaries of each voxel and detector bin onto a common axis and employ the length of overlap as the weight (De Man and Basu 2004). Figure 6 illustrates the three projection models in 2D case (De Man and Basu 2004). Although there are potential benefits of modeling the focal spot size (Beister et al. 2012), it is noted that, for the projectors in Figure 6, the focal spot is generally assumed to be infinitely small point for simplicity.
3.3. Maximum likelihood (ML)/weighted least-squares (WLS) criterions

3.3.1. Poisson model for calibrated transmitted photon counts

Let \( \mathbf{N} \in \mathbb{R}^{I \times 1} \) denote the vector of calibrated transmitted photon counts. Assuming the measurements among different bins are statistically independent, the likelihood function of the joint probability distribution, given a distribution of the attenuation coefficients, can be written as:

\[
P(\mathbf{N} | \mathbf{\mu}) = \prod_{i=1}^{N} \left( \frac{\mathbf{N}_i^{\mathbf{\mu}_i} \exp(-\mathbf{N}_i)}{\mathbf{N}_i!} \right)
\]

Due to the logarithm's monotonicity, we can take the natural logarithm, which will not change the location of the maximum. Thus, the log-likelihood function

\[
L(\mathbf{N} | \mathbf{\mu}) = \sum_{i=1}^{N} (\mathbf{N}_i \ln \mathbf{N}_i - \mathbf{N}_i - \ln \mathbf{N}_i !)
\]

(17)

Ignoring the constant terms which will not change the optimization solution, we can obtain:

\[
L(\mathbf{N} | \mathbf{\mu}) = -\sum_{i} (\mathbf{N}_0 e^{-\mathbf{\nu}_i} + \mathbf{N} \mathbf{\bar{y}}_i) = -\sum_{i} (\mathbf{N}_0 e^{-(\mathbf{\nu}_i + \mathbf{\mu}_i}) + \mathbf{N} [\mathbf{\mu}]_i )
\]

(18)

where Eq. (18) is called the ML criterion.

A second-order Taylor's expansion can be applied to \( g_i(\mathbf{\bar{y}}_i) = \mathbf{N}_0 e^{-\mathbf{\nu}_i} + \mathbf{N} \mathbf{\bar{y}}_i \) around the measured line integral \( y_i \) (Sauer and Bouman 1993; Elbakri and Fessler 2002), that is:

\[
g_i(\mathbf{\bar{y}}_i) = \mathbf{N}_0 e^{-\mathbf{\nu}_i} + \mathbf{N} \mathbf{\bar{y}}_i \approx g_i(y_i) + \frac{g_{ii}^{(y_i)}}{1!} (\mathbf{\bar{y}}_i - y_i) + \frac{g_{ii}^{(y_i)}}{2!} (\mathbf{\bar{y}}_i - y_i)^2
\]

\[
= (N_i + N_i \ln \frac{N_0}{N_i}) + 0 + \frac{N_i}{2} (\mathbf{\bar{y}}_i - y_i)^2
\]

(19)
Therefore, ignoring the constant and irrelevant terms, the log-likelihood in Eq. (18) can be approximated as:

\[
L(y | \mu) \approx -\sum_i \left\{ \frac{N_i}{2} (y_i - \bar{y}_i)^2 \right\} = -\frac{1}{2} (y - \bar{y})^T \Lambda (y - \bar{y}) = -\frac{1}{2} (y - A\mu)^T \Lambda (y - A\mu)
\]  

(20)

where \( \Lambda = \text{diag}\{N_i\} \) depends on the random variable \( N_i \). The approximate log-likelihood in Eq. (20) has computational advantages compared to Eq. (18) due to the quadratic form, but it may be biased when \( N_i \) is close to zero. Despite the potential shortcomings, the approximate log-likelihood has been used successfully in CT applications, and its negative is also widely known as the WLS criterion.

3.3.2. Poisson+Gaussian model for calibrated transmitted photon counts

The likelihood function of the 'Poisson+Gaussian' model in Eq. (9) is analytically intractable. To circumvent this problem, a shifted Poisson approximation in Eq. (10) is exploited to match the first two statistical moments (Yavuz and Fessler 1998; La Rivière and Billmire 2005). With the shifted Poisson model and ignoring the constant and irrelevant terms, the corresponding log-likelihood function similar to that in Eq. (18) can be written as:

\[
L(N | \mu) = \sum_i \left\{ \left[ N_i + \sigma_{\mu}^2 \right], \ln(\bar{N}_i + \sigma_{\mu}^2) - (\bar{N}_i + \sigma_{\mu}^2) \right\} \\
= \sum_i \left\{ \left[ N_i + \sigma_{\mu}^2 \right], \ln(\bar{N}_i e^{-\gamma} + \sigma_{\gamma}^2) - (\bar{N}_i e^{-\gamma} + \sigma_{\mu}^2) \right\} \\
= \sum_i \left\{ \left[ N_i + \sigma_{\mu}^2 \right], \ln(\bar{N}_i e^{-\mu} A_{\mu}) + \sigma_{\mu}^2) - (\bar{N}_i e^{-\mu} A_{\mu} + \sigma_{\mu}^2) \right\}
\]

(21)

where Eq. (21) is the ML criterion with consideration of the electronic noise.

Similar to the Taylor's expansion in Eq. (19) and omitting the constant and irrelevant terms, the log-likelihood in Eq. (21) can be approximated as:

\[
L(y | \mu) \approx -\frac{1}{2} \sum_i \left\{ \frac{N_i^2}{N_i + \sigma_{\gamma}^2} (y_i - \bar{y}_i)^2 \right\} = -\frac{1}{2} (y - \bar{y})^T \Lambda (y - \bar{y}) = -\frac{1}{2} (y - A\mu)^T \Lambda (y - A\mu)
\]  

(22)

where \( \Lambda = \text{diag}\left\{ \frac{N_i^2}{N_i + \sigma_{\gamma}^2} \right\} \) depends on the random variable \( N_i \). Also, the negative of Eq. (22) is called as the WLS criterion in consideration of the electronic noise.

It is noted that if we neglect the influence of the electronic noise (\( \sigma_{\gamma}^2 = 0 \)), the diagonal matrix \( \Lambda \) in Eq. (22) reduces to \( \Lambda = \text{diag}\{N_i\} \), which is the same as that in Eq. (20).
3.3.3. Gaussian model for calibrated line integrals

Similarly, assuming the calibrated line integrals among different detector bins are statistically independent, the likelihood function of the joint probability distribution, given a distribution of the attenuation coefficients, can be written as:

\[ P(y | \mu) = \frac{1}{Z_0} \prod_i \exp \left( -\frac{(y_i - \bar{y}_i)^2}{2\sigma^2_i} \right) \]  

(23)

where \( Z_0 \) is a normalizing constant.

Then, ignoring the constant and irrelevant terms, the log-likelihood function can be written as:

\[ L(y | \mu) = \ln P(y | \mu) = \sum_i \left( -\frac{(y_i - \bar{y}_i)^2}{2\sigma^2_i} \right) = -\frac{1}{2} (y - \bar{y})^T \Lambda (y - \bar{y}) = -\frac{1}{2} (y - \Lambda \mu)^T \Lambda (y - \Lambda \mu) \]  

(24)

where the matrix \( \Lambda \) is diagonal and \( \Lambda = \text{diag}\{1/\sigma^2_i\} \).

As illustrated in Section 3.1.2, with the monochromatic X-ray source assumption, the variance of the calibrated line integral can be given as:

\[ \sigma^2_i = 1/\bar{N}_i, \text{ or } \sigma^2_i = \frac{\bar{N}_i + \sigma^2_e}{\bar{N}_i^2} \]  

(25)

The WLS criterion derived in Eq. (24) is consistent with that in Eq. (20) and Eq. (22) with the approximation of \( \bar{N}_i \approx N_i \). It shall be noted that this approximation holds only if the mean value \( \bar{N}_i \) is relatively large. In the cases of ultra low-dose CT imaging and presence of the electronic noise, this approximation may not hold. The gain of the matrix \( \Lambda \) of (24) over that of (20) was analyzed theoretically by (Xu and Tsui 2014). In practice, since the variance in matrix \( \Lambda \) of (24) depends on the unknown mean \( \bar{N}_i \), it is typically re-calculated during iterative image reconstruction, and the criterion in Eq. (24) is sometimes called re-weighted least-squares (Green 1984; Dollinger and Staudte 1991; Wang et al 2006).

3.4. Maximum a posteriori (MAP) estimation

Mathematically, low-dose CT image reconstruction is an ill-posed problem due to the presence of noise and other inconsistencies in the projection data. Therefore, the image estimation that directly optimizes the ML criterion can be very noisy and unstable. So researchers reformulate this problem with the MAP estimation by posing a prior term to penalize or regularize the solution. The prior term enables us to incorporate available information or expected properties of the image to be reconstructed.
Mathematically, the MAP estimator can be expressed as:

$$\mu^* = \arg\max_{\mu} P(\mu | N)$$  \hspace{1cm} (26)$$

According to the Bayesian law:

$$P(\mu | N) = \frac{P(N | \mu) P(\mu)}{P(N)}$$  \hspace{1cm} (27)$$

By taking the logarithm and omitting the irrelevant term, the MAP estimator can be simplified to:

$$\mu^* = \arg\max_{\mu} [\ln P(\mu | N)] = \arg\max_{\mu} [L(\mu | N) + \ln P(\mu)]$$  \hspace{1cm} (28)$$

By replacing the log a priori probability $\ln P(\mu)$ with a more general form, we have:

$$\mu^* = \arg\max_{\mu} [L(N | \mu) - R(\mu)] = \arg\max_{\mu} [L(N | \mu) - \beta U(\mu)]$$  \hspace{1cm} (29)$$

where $U(\mu)$ denotes a penalty, and $\beta > 0$ is a scalar control parameter which allows one to tune the MAP (or penalized ML (pML)) estimation for a specific noise-resolution tradeoff. When $\beta$ goes to zero, the reconstructed image from the MAP estimation approaches the ML estimation.

From the Tikhonov regularization point of view, the MAP estimation can be considered as an objective function consisting of two terms: a data-fidelity term (e.g., the log-likelihood) modeling the statistics of projection measurements, and a regularization term (e.g., the log-prior) incorporating prior knowledge or expected properties of the image to be reconstructed.

The log-likelihood functions in Eqs. (18) (20) (22) (24) are concave functions of $\mu$, therefore, the resulting objective functions in Eq. (29) would be concave if and only if $U(\mu)$ is a convex function of $\mu$. The log-likelihood function in Eq. (21) is not concave in $\mu$ for $\sigma^2 > 0$, so the corresponding objective function in Eq. (29) would not be concave in $\mu$ anyway. Global maximum can be found for concave objective functions, while only local maximum can be achieved for others.

In summary, the SIR of low-dose CT can be considered to estimate the attenuation map by maximizing the MAP (or pML) criterion with a non-negativity constraint (using the calibrated transmitted photon counts):

$$\mu^* = \arg\max_{\mu \geq 0} [L(N | \mu) - \beta U(\mu)]$$  \hspace{1cm} (30a)$$

or directly minimizing the penalized WLS (PWLS) criterion (using the calibrated line-integrals):
\[
\mu^* = \arg \max_{\mu \geq 0} \left[ L(y \mid \mu) - \beta U(\mu) \right] = \arg \min_{\mu \geq 0} \left[ \frac{1}{2} (y - A\mu)^T \Lambda(y - A\mu) + \beta U(\mu) \right] 
\]

where \( L(N \mid \mu) \) is defined in Eqs. (18) (21), and \( L(y \mid \mu) \) is defined in Eqs. (20) (22) (24).

4. Regularization strategies

Figure 7 gives an overview of regularizations that can be employed in SIR for low-dose CT. Their conceptual and mathematical bases will be illustrated explicitly in this section. Without loss of generality, we assume 2D configuration for the regularizations, while extension from 2D to 3D is straightforward (in 3D presentation, voxels would be used, instead of pixels).
4.1. Spatially independent priors

The simplest form of priors assumes statistical independence among different pixels, that is, there is no coupling among the pixels. Thus, this family of priors takes the form (Qi and Leahy 2006):

\[ P(\mu) = \frac{1}{Z_i} \prod_j \exp[-\beta U(\mu_j)] = \frac{1}{Z_i} \exp[-\beta \sum_j U(\mu_j)] = \frac{1}{Z_i} \exp[-\beta U(\mu)] \]  \hspace{1cm} (31)

where \( Z_i \) is a normalizing constant.

One spatially independent prior leads to the identity norm which is given as (Fessler 2006):

\[ U(\mu) = \sum_j \mu_j^2 / 2 \]  \hspace{1cm} (32)

The second type of independent prior is based on the entropy criterion (Liang et al 1989; Nunez and Llacer 1990), whose corresponding energy function \( U(\mu) \) can be described as:

\[ U(\mu) = \sum_j \mu_j \ln \mu_j \]  \hspace{1cm} (33)

Basically, these two priors can keep pixel values from "blowing up", but they cannot explicitly enforce smoothness in the image (Fessler 2006).

The third type of independent prior is the Gaussian prior (Levitan and Herman 1987), whose energy function has the form:

\[ U(\mu) = \sum_j \frac{(\mu_j - \bar{\mu}_j)^2}{2\sigma_j^2} \]  \hspace{1cm} (34)

where \( \bar{\mu}_j \) and \( \sigma_j^2 \) are the mean and variance respectively, and when \( \bar{\mu}_j = 0 \) it reduces to Eq. (32).

Similarly, the Gamma prior (Lange et al 1987) allows only non-negative image values and can be a more natural model for an image:

\[ U(\mu) = \sum_j \rho(\mu_j, \bar{\mu}_j, \sigma_j) \]  \hspace{1cm} (35)

where \( \rho(\mu_j, \bar{\mu}_j, \sigma_j) \) is a Gamma PDF.

Basically, the Gaussian and Gamma priors encourage the pixel values to be close to the mean image. Thus, the determination of the mean image has a significant effect on the reconstructed image (Qi and Leahy 2006). Some researchers investigated ways to estimate the mean image during the reconstruction using either the median or the mean of neighboring pixels. However, in these cases, the priors are no longer truly independent, which will be discussed in details in Section 4.2.3.
4.2. Markov random field (MRF) model-based priors

Another form of priors assumes that the attenuation maps are locally smooth, that is, the neighboring pixels tend to have similar values. One simple mathematical model that can describe this property is the MRF model, also known as Gibbs distribution (Geman and Geman 1984):

\[ P(\mu) = \frac{1}{Z_2} \exp[-\beta U(\mu)] \]  

(36)

where \( Z_2 \) is a normalizing constant, and the Gibbs energy \( U(\mu) \) is a weighted sum of potential functions (Wernick and Aarsvold 2004):

\[ U(\mu) = \sum_{j} \sum_{m \in W_j} w_{jm} \phi(\mu_j, \mu_m, \mu_{m_1}, \ldots) \]  

(37)

where \( W_j \) represents the MRF window of the \( j \)th pixel, and pixels indexed by \( m_1, m_2, m_3, \ldots \), are the neighboring pixels within the MRF window; \( w_{jm} \) denotes the weighting coefficient (indicating interaction degree) among the pixels; and \( \phi \) denotes a positive potential function.

4.2.1. Pair-wise Gibbs priors

The most commonly used Gibbs energy \( U(\mu) \) in image reconstruction exploits the potential function that is related to the intensity difference between two neighboring pixels, and the potential function does not vary across the image. Accordingly, the pair-wise Gibbs energy is given as:

\[ U(\mu) = \sum_{j} \sum_{m \in W_j} w_{jm} \phi(\mu_j - \mu_m) \]  

(38)

There are many possibilities for the MRF window \( W_j \) and weighting coefficient \( w_{jm} \), but usually the MRF window consists of a small local neighborhood (typically 8 neighbors in 2D case) and the weighting coefficient is determined by the inverse of the Euclidean distance between the two pixels. That is, for pixel \( j \) and pixel \( m \) with 2D coordinates \((j_x, j_y)\) and \((m_x, m_y)\), the weighting coefficient \( w_{jm} \) is usually given by:

\[ w_{jm} = 1/d_{jm} = 1/\sqrt{(j_x - m_x)^2 + (j_y - m_y)^2} \]  

(39)

That is, \( w_{jm}=1 \) for the four horizontal and vertical neighboring pixels, \( w_{jm}=1/\sqrt{2} \) for the four diagonal neighboring pixels and \( w_{jm}=0 \) otherwise:

\[
\begin{bmatrix}
1/\sqrt{2} & 1 & 1/\sqrt{2} \\
1 & 1 & 1 \\
1/\sqrt{2} & 1 & 1/\sqrt{2}
\end{bmatrix}
\]  

or

\[
\begin{bmatrix}
1/\sqrt{2} & 1 & 1/\sqrt{2} \\
1 & 1 & 1 \\
1/\sqrt{2} & 1 & 1/\sqrt{2}
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 1/\sqrt{2} & 1 \\
1/\sqrt{2} & 1 & 1/\sqrt{2} \\
1 & 1 & 1
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 1/\sqrt{2} & 1 \\
1/\sqrt{2} & 1 & 1/\sqrt{2} \\
1 & 1 & 1
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 1/\sqrt{2} & 1 \\
1/\sqrt{2} & 1 & 1/\sqrt{2} \\
1 & 1 & 1
\end{bmatrix}
\]
Generally, the potential function $\phi$ in Eq. (38) increases monotonically with increasing intensity differences between the neighboring pixels. Therefore, for larger intensity differences, the energy function $U(\mu)$ increases, which in turn reduces the prior probability of the image $P(\mu)$. Because of this, the MAP solution discourages images that are too rough, and therefore smoothes the image.

The choice of potential function is very critical since it strongly determines the smoothness properties of the MAP estimate. One common choice of $\phi$ in image reconstruction is the quadratic function, $\phi(\Delta) = \Delta^2 / 2$, and in this case, Eq. (38) becomes:

$$U(\mu) = \sum_{j} \sum_{m \in R_j} w_{jm} \frac{1}{2} (\mu_j - \mu_m)^2$$

which corresponds to the Gaussian MRF (GMRF) prior that has been widely used for SIR.

A major drawback of the GMRF prior is that it can excessively penalize the differences between neighboring pixels when $\mu_j$ and $\mu_m$ fall across a discontinuous boundary in the image, thus may lead to over smoothing of edges and fine structures in the reconstructed image. To mitigate this issue, some researchers replaced the quadratic potential function with non-quadratic functions that increase less rapidly for sufficiently large differences. In this way, the corresponding priors are expected to remove noise while retaining sharp edges in the reconstructed image.

Generally, the non-quadratic potential functions can be divided into two groups: non-convex (Geman and McClure 1985; Blake and Zisserman 1987; Hebert and Leahy 1989; Geman and Reynolds 1992; Lalush and Tsui 1993) and convex potential functions (Huber 1981; Besag 1989; Green 1990; Lange 1990; Stevenson and Delp 1990; Bouman and Sauer 1993; Charbonnier et al 1994; Thibault et al 2007).

1) Non-convex potential functions

Table 2 lists the commonly used non-convex potential functions and Figure 8 illustrates their corresponding curves with comparison to the quadratic potential function. Each of these potential functions has a region that its influence function $\phi(\Delta)$ is decreasing (i.e., $\phi(\Delta)$ is locally concave), and that is why $\phi(\Delta)$ is a non-convex function of $\Delta$.

Since these potential functions are non-convex, the resulting MRF priors in Eq. (38) are non-convex. According to analysis in Section 3.4, the resulting objective functions in Eq. (29) would not be concave, and would only have local solutions. These non-convex potential functions based MRF
priors may preserve edges better than the GMRF model, but they may also generate discontinuous estimates at image boundaries which can be considered as undesirable artifacts.

Table 2. The commonly used non-convex potential functions

| Potential function | Reference | Comment |
|--------------------|-----------|---------|
| \( \phi(\Delta) = \min \left\{ \frac{\Delta^2}{2}, \frac{\delta^2}{2} \right\} \) | Blake and Zisserman 1987 | weak spring |
| \( \phi(\Delta) = \frac{1}{1 + (\Delta / \delta)^2} = \frac{\Delta^2}{\Delta^2 + \delta^2} \) | Geman and McClure 1985 | |
| \( \phi(\Delta) = \frac{|\Delta|}{|\Delta| + \delta} \) | Geman and Reynolds 1992 | |
| \( \phi(\Delta) = \ln \left( 1 + \frac{\Delta^2}{\delta^2} \right) \) | Hebert and Leahy 1989 | |
| \( \phi(\Delta) = \frac{\delta}{2\omega} \ln \left[ 1 - 2\omega^2(1 - \frac{\Delta^2}{\delta^2}) + \frac{2\omega}{\delta^2} \sqrt{\omega^2(\delta^4 + \Delta^4) + \delta^2\Delta^2(1 - 2\omega^2)} \right] \) | Lalush and Tsui 1993 | |

Figure 8. Plot of commonly used non-convex potential functions.

2) Convex potential functions

IRIS Laboratory, Department of Radiology, Stony Brook University
Alternatively, researchers proposed to use convex potential functions to replace the quadratic function. Table 3 lists several convex potential functions, and Figure 9 illustrates their corresponding curves with comparison to the quadratic potential function. For each of these potential functions, the influence function \( \phi(\Delta) \) is non-decreasing, and therefore, \( \phi(\Delta) \) is a convex function of \( \Delta \). Also, it should be noted that for the region \( |\Delta| < \delta \), \( \phi(\Delta) \) approximates the quadratic function (e.g., the Taylor expansion), while for the region \( |\Delta| > \delta \), \( \phi(\Delta) \) is always smaller than the quadratic function (e.g., difference method). The parameter \( \delta \) clearly delineates between the "non-edge" and "edge" regions, and is often referred as the "edge threshold" or "transition point".

### Table 3. The commonly used convex potential functions

| Potential function \( \phi(\Delta) \) | Reference | Comment |
|--------------------------------------|----------|---------|
| \( \phi(\Delta) = \begin{cases} \frac{\Delta^2}{2} & |\Delta| \leq \delta \\ \delta |\Delta| - \frac{\Delta^2}{2} & |\Delta| > \delta \end{cases} \) | Stevenson and Delp 1990 | Huber |
| \( \phi(\Delta) = \delta^2 \log \cosh(\Delta / \delta) = \delta^2 \ln \frac{e^{\Delta / \delta} + e^{-\Delta / \delta}}{2} \) | Green 1990 | |
| \( \phi(\Delta) = \delta^2 (|\Delta| / \delta - \ln(1 + |\Delta| / \delta)) \) | Lange 1990 | |
| \( \phi(\Delta) = \delta^2 (\sqrt{1 + (\Delta / \delta)^2} - 1) \) | Charbonnier et al 1994 | Hyperbola |

![Figure 9. Plot of commonly used convex potential functions.](image-url)
Another family of convex potential functions is corresponding to the q-generalized Gaussian MRF prior (q-GGMRF) (Thibault et al. 2007), which can be described as:

$$\phi(\Delta) = \frac{|\Delta|^p}{1 + |\Delta / \delta|^p q} (1 \leq q \leq p \leq 2)$$  \hspace{1cm} (42)

By giving specific parameter values, it can become:

\[
\begin{align*}
\phi(\Delta) &= \begin{cases} 
\Delta^2 & (q = p = 2, \text{Gaussian prior, Herman 1980}) \\
|\Delta| & (q = p = 1, \text{median pixel prior, Besag 1989}) \\
|\Delta|^p & (1 < q = p \leq 2, \text{generalized Gaussian MRF, Bouman and Sauer 1993}) \\
\frac{\Delta^2}{1 + |\Delta / \delta|^p} & (q = 1, p = 2, \text{approximate Huber prior, Thibault et al. 2007}) \\
\frac{|\Delta|^p}{1 + |\Delta / \delta|^p q} & (1 \leq q < p \leq 2, q\text{-generalized Gaussian MRF, Thibault et al. 2007})
\end{cases}
\end{align*}
\]  \hspace{1cm} (43)

**Figure 10** shows several representative curves of these convex potential functions.

**Figure 10.** Plot of convex potential functions corresponding to the q-GGMRF priors.

Since these potential functions are convex, the resulting MRF priors in Eq. (38) are convex, and the corresponding objective functions in Eq. (29) would be concave, which results in stable MAP estimates. These convex potential functions based MRF priors can preserve edges better than the GMRF model. However, one limitation with them is that the reconstruction results may be sensitive to the choice of "transition point" (or edge threshold) that controls the shape of the functional form,
since the edges in an image are unlikely to be determined by a single value (Bouman and Sauer 1993).

4.2.2. Adjusting weighting coefficient

Inspired by the anisotropic diffusion filter, (Wang et al. 2008a; Ouyang et al. 2011) proposed to utilize anisotropic weighting coefficients while retaining the quadratic-form potential function. Then, the prior takes the following form:

\[ U(\mu) = \sum_{j \in \Omega} \sum_{m} w_{jm} \frac{1}{2}(\mu_j - \mu_m)^2 \]  \hspace{1cm} (44)

where \( w_{jm} = \frac{1}{d_{jm}} \cdot \exp\left(\frac{-(\mu_j - \mu_m)^2}{\varepsilon^2}\right) \) or \( w_{jm} = \frac{1}{d_{jm}} \cdot \frac{1}{(\mu_j - \mu_m)^2 + \varepsilon^2} \).

In this formulation, the weighting coefficient is smaller if the intensity difference between a neighbor and the concerned pixel is larger, since the coupling between two such neighbors is weaker. By this way, the prior discourages the equivalence between neighbors if the gradient between them is large, and the edges or boundaries in the image will be better preserved. The parameter \( \varepsilon \) can be set empirically, or to the value at 90% of the histogram of the gradient magnitude of the image to be reconstructed (Perona and Malik 1990). For the weighting coefficient determination, there could be other choices as long as it satisfies the behavior of weighting similar to those defined in Eq. (44).

4.2.3. Median-based priors

1) Median root prior (MRP)

In MRP, intensity differences among neighboring pixels are not penalized. Instead, the penalty is set according to how much the central pixel differs from the local median. Mathematically, the MRP can be described as (Alenius and Ruotsalainen 1997; Alenius et al. 1998):

\[ U(\mu) = \sum_{j} \frac{(\mu_j - \text{median}(\mu_j))^2}{\text{median}(\mu_j)} \]  \hspace{1cm} (45)

where \( \text{median}(\mu_j) \) is the local median. Therefore, no penalty is applied when the image is locally monotonic, and only non-monotonic local changes among neighboring pixels are penalized.

Although the MRP captures significant edges while encouraging preservation of locally monotonic regions, it is a heuristic empirical method and not convex in theory.

2) Convex median prior (MP)
(Hsiao et al 2003) proposed a class of convex MP which depends on two vector variables \( \mathbf{\mu} \in \mathbb{R}^{J \times 1} \) and \( \mathbf{f} \in \mathbb{R}^{J \times 1} \). The vector \( \mathbf{f} \) is an "auxiliary" field having components in register with the components of \( \mathbf{\mu} \) and is also to be estimated. The new MP is expressed as:

\[
U(\mathbf{\mu}) = \sum_{j} \sum_{w \in W_j} |\mu_j - f_w| \tag{46}
\]

where the potential function, \( \phi(\Delta) = |\Delta| \) due to the connection between the median and the absolute value function (Press et al 1992). However, since the absolute value function is not differentiable at zero, a more practical medical prior is used to replace the absolute value function with the following approximations (Hsiao et al 2003; Fessler 2006):

\[
|\Delta| \approx \sqrt{\Delta^2 + \tau^2} - \tau \\
\approx \tau \log \cosh(\Delta / \tau) \\
\approx \tau (\sqrt{1+(\Delta / \tau)^2} - 1) \tag{47}
\]

Figure 11 illustrates the curves of the absolute value function and its differentiable approximations. And it is noted that the smaller the parameter \( \tau \), the more precise of the approximations. The authors have proved that above prior is convex for any \( \phi(\Delta) \) that is convex, therefore, the new MP is essentially convex.

Figure 11. Plot of the absolute value function and its differentiable approximations.
4.3. Compressed sensing (CS)-based regularizations

According to the classic image reconstruction theory, the sampling rate of view angles must satisfy the Nyquist sampling theorem in order to reconstruct an image without aliasing artifacts. This theorem does not assume any prior knowledge of the imaging object. Recently, a new signal reconstruction theory, CS, has been rigorously formulated to accurately reconstruct a signal from much fewer samples than that is required by the Nyquist sampling theorem (Candès et al 2006; Donoho 2006). The main idea of CS is that most signals are sparse in appropriate orthonormal systems, that is, a majority of their coefficients are close or equal to zero. Researchers tried to apply this theory to accurately reconstruct CT images at a much lower angular-sampling rate than the Nyquist sampling, but the CT images are generally not sparse in their original pixel representation (Chen et al 2008a). Fortunately, one can apply a sparsifying transform $\psi$ to increase the image sparsity. For instance, the discrete gradient transform and wavelet transforms are commonly used for this purpose. Then, in the sparsified image, drastically fewer image pixels have significant image values. So the sparsified version of the image can be first reconstructed, followed by a de-sparsifying transform $\psi^{-1}$ to obtain the target image. In practice, there is no need of explicit form for the de-sparsifying transform, only the sparsifying transform is needed for image reconstruction (Chen et al 2008a).

Mathematically, the CS method reconstructs an image via the $L_p$ norm ($0 \leq p < 2$) minimization. Herein, for a vector $\theta$, $\|\theta\|_0$ represents the $L_0$ norm of vector $\theta$ which counts the number of nonzero components of $\theta$, and $\|\theta\|_p (p > 0)$ denotes the $L_p$ norm of vector $\theta$ which is defined as:

$$\|\theta\|_p = \left(\sum_j |\theta_j|^p\right)^{1/p} \rightarrow \|\theta\|_p = \sum_j |\theta_j|^p$$

(48)

It should be noted that $\|\theta\|_p$ is not actually a norm when $0 \leq p < 1$ because it is not sub-additive, yet we still refer it as norm following convention (sort of abuse of terminology).

Then, the CS image reconstruction can be implemented by solving the following constrained minimization problem ($0 \leq p < 2$):

$$\mu^* = \arg\min_{\mu \geq 0} \|\psi\mu\|_p, \text{ s.t. } \|y - A\mu\|_2^2 \leq \xi$$

(49)
where $\xi > 0$ is the error tolerance parameter which denotes the inconsistency in acquired projection data due to noise or other imperfections.

Eq. (49) is equivalent to the following unconstrained minimization problem ($0 \leq p < 2$):

$$
\mu^* = \arg \min_{\mu \geq 0} \left[ \|y - A\mu\|_2^2 + \beta \|\psi \mu\|_p \right]
$$

(50)

or one can include the projection statistics in the measurement model, that is:

$$
\mu^* = \arg \min_{\mu \geq 0} \left[ (y - A\mu)^T \Lambda (y - A\mu) + \beta \|\psi \mu\|_p \right]
$$

(51)

where the sparsity constraint $\|\psi \mu\|_p$ acts as a regularization term in SIR.

Mathematically,

$$\|0\|_p \geq \|0\|_{p+\kappa}, \text{ for any } p \geq 0 \text{ and } \kappa \geq 0$$

(52)

Therefore, the $L_0$ norm minimization would probably require fewest measurements for exact image reconstruction. Unfortunately, the $L_p$ norm is non-convex for $0 \leq p < 1$ and its minimization can be utterly intractable to solve. The $L_p$ norm is convex for $p \geq 1$, among which the $L_1$ norm favors the highest degree of sparsity according to Eq. (52) and hence as a proxy to $L_0$ norm and frequently used for sparse-view image reconstruction.

4.3.1. $L_1$ norm regularization

1) Total variation (TV) norm

In CS image reconstruction, discrete gradient transform is one of the commonly used sparsifying transforms, which leads to the TV norm (Sidky et al 2006; Sidky and Pan 2008). In the 2D case, the TV norm can be given as:

$$
\|\nabla \mu\|_p = \|\mu\|_{TV} = \sum_{s,t} \sqrt{(\mu_{s,t} - \mu_{s-1,t})^2 + (\mu_{s,t} - \mu_{s,t-1})^2}
$$

(53)

where $\nabla$ denotes the first-order discrete gradient transform, $s$ and $t$ are the indices of the location of the attenuation coefficients in discretized 2D grid of the imaging object $\mu$.

While the conventional TV regularization has favorable noise and artifacts mitigating properties, it sometimes results in over-smoothness on the edges and small details, due to the piecewise constant assumption. To solve this problem, researchers introduced adaptive weight to the original TV norm. That is, the pixels at the edges are given lower weight so that the smoothing is preferentially
performed on the non-edge parts, and the edges are better preserved. Mathematically, the edge-preserving TV (EPTV) regularization term is given as (Tian et al 2011):

\[
\|\mu\|_{\text{EPTV}} = \sum_{s,t} w_{s,t} \left[ (\mu_{s,t} - \mu_{s-1,t})^2 + (\mu_{s,t} - \mu_{s,t-1})^2 \right]
\]

where \( w_{s,t} = \exp\left( \frac{-(\mu_{s,t} - \mu_{s-1,t})^2 + (\mu_{s,t} - \mu_{s,t-1})^2}{\varepsilon^2} \right) \), and the parameter \( \varepsilon \) in the weight is a scale factor which controls the strength of smoothing for those pixels at edges.

Alternatively, the adaptive-weighted TV (AwTV) regularization term is described as (Liu et al 2012; Liu et al 2014b):

\[
\|\mu\|_{\text{AwTV}} = \sum_{s,t} w_{s,t,1} (\mu_{s,t} - \mu_{s-1,t})^2 + w_{s,t,2} (\mu_{s,t} - \mu_{s,t-1})^2
\]

where \( w_{s,t,1} = \exp\left( \frac{-(\mu_{s,t} - \mu_{s-1,t})^2}{\varepsilon^2} \right) \) and \( w_{s,t,2} = \exp\left( \frac{-(\mu_{s,t} - \mu_{s,t-1})^2}{\varepsilon^2} \right) \).

The experimental results illustrated that the EPTV norm is superior to the TV norm in preserving the low-contrast structures (Tian et al 2011), while the comparisons in (Liu et al 2012) showed that the AwTV norm outperforms both the TV norm and EPTV norm due to the consideration of the anisotropic edge property.

Furthermore, the conventional TV norm has been shown to produce "staircase effects" where intensity ramps are discretized into steps, resulting in blocky or patchy artifacts in the reconstructed images (Tang et al 2009; Do et al 2010; Xu et al 2012; Liu et al 2014a). High-order TV (Do et al 2010), TV-strokes (Liu et al 2014a) and total generalized variation (Niu et al 2014) strategies were investigated for low-dose CT reconstruction to reduce the staircase artifacts without sacrificing edge sharpness.

2) Wavelet transform based L_1 norm

The wavelet transform is another sparsifying transforms frequently employed. For a medical image, one can find an orthonormal basis (or more general, a frame) to make the image sparse in terms of significant transform coefficients. The preliminary investigation of the Haar transform (the simplest version of the wavelet transform) based L_1 norm minimization (Yu and Wang 2010; Garduño et al 2011) showed no obvious benefits than the TV norm minimization from the medical diagnostic point of view, yet it is by no means conclusive due to limited cases. Some other image transformation techniques, such as tight frame (TF) transform (Jia et al 2011; Zhou et al 2013),
curvelet transform (Wu et al 2013), and shearlet transform (Vandeginste et al 2012b), may provide sparser representations to piecewise smooth functions than the traditional wavelets. They were also studied by researchers for the L_1 norm based CT image reconstruction and compared with the TV norm based reconstruction. Both positive and negative results were reported, so this is still an open question for further investigation.

3) Combined L_1 norm

In order to take advantage of the above-mentioned two types L_1 norm, some novel CS-based methods which combine the two sparsity transforms were proposed (Vandeginste et al 2012a; Zhu et al 2013):

\[ \lambda_1 \| \mathbf{u} \|_{TV} + \lambda_2 \| W \mathbf{u} \|, \text{ or } \lambda_1 \| \mathbf{u} \|_{TV} + \lambda_2 \| S \mathbf{u} \|, \]

(56)

where \( W \) and \( S \) represent the wavelet transform and the shearlet transform respectively, and constants \( \lambda_1 \) and \( \lambda_2 \) control the relative weight of the two components.

Experimental results exhibited the advantages of the combined L_1 norm in Eq. (56) over the simple TV norm in terms of noise and streak artifacts suppression, and edge preservation (Vandeginste et al 2012a; Zhu et al 2013).

4) Dose reduction-prior image constrained compressed sensing (DR-PICCS)

The prior image constrained compressed sensing (PICCS) norm was initially proposed by Chen et al to reconstruct dynamic CT images from highly undersampled projection datasets (Chen et al 2008a). Then, it is extended to the CT data acquisitions where the projection datasets are densely sampled but with high level of noise, called the DR-PICCS (Lauzier and Chen 2013). In DR-PICCS applications, the prior image is generated by the spatial filtering (e.g., Gaussian or Diffusion filtering) of a high-noise image reconstructed from the same projection dataset with the FBP method. Then, the low-noise and low-spatial resolution prior image is used together with the original projection data to recover spatial resolution using PICCS. The DR-PICCS norm is incorporated into the SIR framework as the regularization term and is given as:

\[ \| \mathbf{u} \|_{piccs} = \lambda_1 \| \psi_1 (\mathbf{u} - \mathbf{u}_{\text{prior}}) \| + (1 - \lambda) \| \psi_2 \mathbf{u} \| \]

(57)

where \( \mathbf{u}_{\text{prior}} \in \mathbb{R}^{J \times L} \) is the prior image obtained from the strategy motioned above, \( \psi_1 \) and \( \psi_2 \) are sparsifying transforms (e.g., discrete gradient transform \( \nabla \)), and \( \lambda \in [0,1] \) is a weighting parameter which controls the relative influence of the prior image and the CS term. It is noted that when \( \lambda = 0 \), the PICCS norm reduces to the conventional L_1 norm.
4.3.2. L-p norm (0<p<1) based regularization

It has been proved that an exact reconstruction of the sparse signal is possible with fewer measurements by replacing the $L_1$ norm with the L-p norm (0<p<1) (Chartrand 2007). However, instead of minimizing the $L-p$ norm $\|\psi\|_p$, directly, the $p$th power of the L-p norm, $\|\psi\|_p^p$, is often adopted for simplicity of computation. The $\|\psi\|_p^p$ (0<p<1) was used as the regularization term in the SIR framework for low-dose CT (Chun and Talavage 2013). Although there is no guarantee of globally optimal solution due to the non-convexity, simulation results demonstrated good image reconstruction from limited measurements than that with $L_1$ norm.

4.3.3. L_0 norm based regularization

Based on Eq. (52), the L_0 norm minimization requires the fewest measurements for exact image reconstruction. In (Hu et al 2011), the L_0 norm based regularization for SIR was presented by defining the potential function in Eq. (38) as:

$$\phi(\Delta) = \left|\text{sgn}(\Delta)\right|$$

(58)

Since the potential function is not continuous, it is approximated with a set of asymptotic potential functions (Trzasko and Manduca 2009; Hu et al 2011):

$$\phi(\Delta, \tau) = \begin{cases} 
1 - e^{-\left|\Delta\right|^\tau} & \left|\Delta\right| \leq \tau \\
\frac{\left|\Delta\right|}{\left|\Delta\right| + \tau} & \left|\Delta\right| > \tau
\end{cases}$$

$$\frac{1}{\ln(1+1/\tau)} \ln\left(1+\frac{\left|\Delta\right|}{\tau}\right)$$

$$\frac{2}{\pi} \arctan\left(\frac{\left|\Delta\right|}{\tau}\right)$$

(59)

Figure 12 illustrates the potential functions in Eq. (59) with comparison to $\left|\text{sgn}(\Delta)\right|$, and as $\tau \to 0$, the potential functions in Eq. (59) would more closely approximate $\left|\text{sgn}(\Delta)\right|$. Correspondingly, the pseudo-L_0 norm would better approximate the 'true L_0 norm'.

The simulation results showed that the SIR with the pseudo-L_0 norm regularization can provide better reconstructions than those obtained with the $L_1$ or $L_2$ norm regularizations (Hu et al 2011). However, due to the non-convexity of the regularization, a good initial guess should be carefully selected to avoid the solution falling into a local minimum.
4.4. Nonlinear neighborhood filters (NNFs)-based regularizations

Inspired by the success of NNFs for noise reduction in the image processing scenario, researchers reformulated them as regularization terms in SIR. Elad has shown that the bilateral filter has strong origins in the MAP estimation (Elad 2002), and this proof theoretically connects the NNFs to the classical approaches as regularization terms.

Essentially, the NNFs reduce image noise by replacing each pixel with a weighted average of its neighbors according to the similarity. The exponential function converts the similarity to weighting coefficient ($w_{jk}$) which indicates the interaction degree between two pixels. The weighting coefficient is positive and symmetric, i.e., $w_{jk}>0$ and $w_{jk}=w_{kj}$. Although the similarity comparison can be performed between any two pixels within the entire image, for computation efficiency, it is usually limited to a fixed neighboring window area (called search-window (SW)) of the target pixel in practice. The bilateral filter (Tomasi and Manduchi 1998) and nonlocal means (NLM) filter (Buades et al 2005a; Buades et al 2005b) are two of the most popular NNFs.

Mathematically, the NNFs can be described as:

$$\hat{\mu}_{j,\text{NNF}} = \sum_{k \in \text{SW}_j} w_{jk}(\hat{\mu}) \hat{\mu}_k$$ (60)
where \(\tilde{u} \in \mathbb{R}^{J \times 1}\) represents the noisy image to be smoothed, \(SW_j\) denotes the search-window of the \(j\)th image pixel, and \(w_{jk}(\tilde{u})\) is the weighting coefficient for pixel \(k\) to the central pixel \(j\). Although the search-window can be of various shapes and sizes at different positions, it is typically of square shape (e.g., 17×17) and does not vary across the image.

The NNFs differ from each other mainly in the weighting coefficient \(w_{jk}\) computation. For instance, the bilateral filter determines the weighting coefficient according to the spatial proximity and intensity similarity between the central pixel and the neighboring pixel (Tomasi and Manduchi 1998):

\[
22 \quad 22 \quad 22 \quad 22 \quad \exp \left( -\frac{d_{jk}^2}{2\sigma_d^2} \right) \cdot \exp \left( -\frac{(\tilde{\mu}_j - \tilde{\mu}_k)^2}{2\sigma_{\mu}^2} \right)
\]

where \(d_{jk}\) denotes the Euclidean distance between the two pixels \(j\) and \(k\), the parameters \(\sigma_d\) and \(\sigma_{\mu}\) control the spatial and intensity weighting respectively.

The NLM filter calculates the weighting coefficient according to the Euclidean distance of two patches. A patch of a pixel can be defined as a squared region centered at that pixel (e.g., 7×7, called patch-window). Let \(P(\tilde{\mu}_j)\) denote the patch centered at pixel \(j\) and \(P(\tilde{\mu}_k)\) denote the patch centered at pixel \(k\). The similarity between pixel \(j\) and \(k\) depends on the weighted Euclidean distance of their patches, \(\left\| P(\tilde{\mu}_j) - P(\tilde{\mu}_k) \right\|_{L^2}\), which is computed as the distance of two intensity vectors in high dimensional space with a Gaussian kernel (\(c > 0\) is the standard deviation of the Gaussian kernel) to weight the contribution for each dimension. Then the weighting coefficient is calculated as (Buades et al 2005a; Buades et al 2005b):

\[
2 \quad 2 \quad 2 \quad 2 \quad \exp \left( -\frac{\left\| P(\tilde{\mu}_j) - P(\tilde{\mu}_k) \right\|_{L^2}^2}{h^2} \right)
\]

where the filtering parameter \(h\) controls the decay of the exponential function.

Based on the NNFs described above, the NNFs-based regularizations can be given in three general forms.

One general form for the NNFs-based roughness regularization is (Kindermann et al 2005):
where \( \tilde{\mu} \) represents a reference image. The quadratic function is one common choice for the potential function \( \phi \). Essentially, this regularization also determines the weighting coefficient by considering the pixel intensity information, which is similar to that in Eq. (44). That is, the weighting coefficient is smaller if the intensity difference between a neighbor and the concerned pixel is larger, and vice versa.

Another general form for NNFs-based roughness regularization is (Buades et al 2006):

\[
U(\mu) = \sum_j \phi \left( \mu_j - \sum_{k \in SW_j} w_{jk}(\tilde{\mu}) \mu_k \right) 
\]

(64)

This regularization is closely related to regularization in Eq. (63), but different in thought. It assumes that each pixel can be replaced by a weighted average of its neighbors according to the similarity, which is also the original idea of the neighborhood filters.

The third general form for the NNFs-based CS regularization is (Elmoataz et al 2008):

\[
U(\mu) = \frac{1}{p} \sum_j \left( \sum_{k \in SW_j} w_{jk}(\tilde{\mu}) (\mu_j - \mu_k)^2 \right)^{p/2}, \quad 0 < p < 2
\]

(65)

This regularization is inspired by both the neighborhood filter and the CS theory.

It should be noted that all three regularization models stated above depend on the weighting coefficient \( w_{jk}(\tilde{\mu}) \) that is calculated from a reference image \( \tilde{\mu} \). Lou et al (Lou et al 2010) suggested to use the FBP reconstructed image as the reference image, but such a reference image is typically noisy and the resulting regularization may lead to suboptimal reconstruction result.

Therefore, investigators (Chen et al 2008b; Zhang et al 2010; Zhang et al 2014d) tried to modify the above three regularization models by replacing \( w_{jk}(\tilde{\mu}) \) with \( w_{jk}(\mu) \), which is described as:

\[
w_{jk}(\mu) = \frac{\exp \left( -\frac{d_{jk}^2}{2\sigma_d^2} \right) \cdot \exp \left( -\frac{(\mu_j - \mu_k)^2}{2\sigma_\mu^2} \right)}{\sum_{k \in SW_j} \exp \left( -\frac{d_{jk}^2}{2\sigma_d^2} \right) \cdot \exp \left( -\frac{(\mu_j - \mu_k)^2}{2\sigma_\mu^2} \right)} \quad \text{or} \quad w_{jk}(\mu) = \frac{\exp \left( -\frac{\|P(\mu_j) - P(\mu_k)\|_c^2}{h^2} \right)}{\sum_{k \in SW_j} \exp \left( -\frac{\|P(\mu_j) - P(\mu_k)\|_c^2}{h^2} \right)}
\]

(66)

In this way, the regularization models in Eqs. (63)-(65) will be based on generic information and do not require the prior reference image. However, the weighting coefficients in Eq. (66) are computed on the unknown image \( \mu \) and the direct minimization of the resulting objective function can be very
complicated. Instead, an empirical one-step-late (OSL) implementation is usually employed in an iterative approach to the solution, where the weighting coefficients are computed on the current image estimate, and then are assumed to be constants when updating the image (Chen et al. 2008b; Zhang et al. 2010).

While previous neighborhood filters calculate the similarity based on single pixel intensity, the NLM filter calculates the similarity based on the patch distance. So the NLM filter is believed to be more robust than the other neighborhood filters, since the intensity value in a single pixel is usually noisy. Consequently, the NLM filter is the most common choice for the NNFs-based regularizations, and the resulting regularizations are also referred to as the nonlocal regularization. The nonlocal regularization in Eq. (63) has been applied to image reconstruction for PET (Chen et al. 2008b) and low-dose CT (Chen et al. 2009), where the potential function takes the quadratic form $\phi(\Delta) = \frac{1}{2} \Delta^2$.

The performance of the NLM-based regularization in Eq. (64) on low-dose CT image reconstruction was also investigated by (Zhang et al. 2014d), and it was found that using local adaptive filtering parameter $h$ can further improve its performance (Zhang et al. 2015b). And with $p=1$, Eq. (65) becomes the nonlocal TV regularization, which has also been explored for CT image reconstruction (Lou et al. 2010).

4.5. Patch-based roughness regularizations

Traditional regularizations penalize image roughness based on the intensity difference between neighboring pixels, but the pixel intensity differences may not be reliable in differentiating sharp edges from random fluctuation due to noise. To address this issue, (Wang and Qi 2012) proposed patch-based regularizations which utilize neighborhood patches instead of individual pixels to measure the image roughness. Since they compare the similarity between patches, the patch-based roughness regularizations are believed to be more robust in distinguishing real edges from noisy fluctuation.

The patch-based roughness regularizations are defined as (Wang and Qi 2012):

$$U(\mathbf{\mu}) = \sum_j \sum_{k \in W_j} w_{jk} \phi\left(\|P(\mathbf{\mu}_j) - P(\mathbf{\mu}_k)\|_{L^2,c}\right)$$

where $w_{jk} = 1$, or $w_{jk} = 1/d_{jk}$.

The authors proved that the above regularizations are convex for any $\phi(\Delta)$ that is convex. Therefore, all the convex potential functions in Table 3 can be employed.
The conventional pixel-based roughness regularizations in Eq. (38) can be considered as a special case of the patch-based roughness regularizations with patch size equal to one. Also, similar to the strategy in Section 4.2.2, we can also adjust the weighting coefficients by considering the intensity while retaining the quadratic potential function:

\[
U(\mu) = \sum_j \sum_{k \in \mathcal{W}_j} w_j^k \frac{1}{2} \left\| P(\mu_j) - P(\mu_k) \right\|^2_{L_2}
\]  

(68)

where

\[
w_j^k = \exp \left(-\frac{\left\| P(\mu_j) - P(\mu_k) \right\|^2_{L_2}}{\mu^2} \right).
\]

In Eq. (68), the weighting coefficient is smaller if the distance between the patch of a neighboring pixel and the patch of the concerned pixel is larger. By this way, the regularization can better preserve edges and boundaries, which is similar to those in Eq. (67) where the edge-preserving non-quadratic potential functions are employed.

4.6. Dictionary learning (DL)-based regularizations

Motivated by the success of DL-based techniques in image processing, face recognition, and texture classification, Xu et al (Xu et al 2012) proposed to utilize the dictionary-based sparsification as the regularization in SIR for low-dose CT. A dictionary is a redundant basis, whose elements are called atoms and are learned from training images. Then, an object image can be sparsely represented as a linear combination of the atoms. The dictionary-based method processes the object image patch by patch, instead of pixel by pixel as conventional sparse transforms do. Because of the patch-based analysis, it is expected to capture local image features and structures more effectively.

A dictionary is a matrix \( D \in \mathbb{R}^{G \times B} \), whose column \( d_b \in \mathbb{R}^{G \times 1} \) is called an atom. Herein, \( G \) is the number of pixels in a patch, and \( B \) is the number of atoms in the dictionary. Given a training image of \( V \) patches, the DL is to seek a dictionary which makes each patch in the training image be sparsely represented by the atoms in this dictionary. Thus, the DL is to solve:

\[
\min_{D, \alpha} \sum_{v=1}^{V} \left\| \alpha_v \right\|_0 \quad \text{s.t.} \quad \forall v, \left\| F_v \mu - Da_v \right\|_2 \leq \xi
\]  

(69)

where \( \alpha_v \in \mathbb{R}^{B \times 1} \) is the sparse representation of the \( v \)th patch under the trained dictionary \( D \), and \( F_v \in \mathbb{R}^{G \times J} \) represents an operator to extract \( v \)th patch from the image \( \mu \in \mathbb{R}^{J \times 1} \).

Via the Lagrange method, Eq. (69) is equivalent to the following unconstraint forms:
\[
\min_{D, \alpha} \sum_{i=1}^{V} \left( \|a_i\|_0 + \gamma_i \|F_i \mu - Da_i\|_2^2 \right) \text{ or } \min_{D, \alpha} \sum_{i=1}^{V} \left( \|F_i \mu - Da_i\|_2^2 + \gamma_i \|a_i\|_0 \right) \tag{70}
\]

where \(\gamma_i\) is the Lagrange multiplier.

Consequently, the above sparsity constraint in terms of a redundant dictionary can be incorporated into the SIR as a regularization term:

\[
U(\mu) = \sum_{i=1}^{V} \|F_i \mu - Da_i\|_2^2 + \sum_{i=1}^{V} \gamma_i \|a_i\|_0 \tag{71}
\]

Herein, the dictionary is learned from an intermediate image during the iteration process, and is called adaptive dictionary (AD). Alternatively, it can be learned from a pre-specified training image, which is called global dictionary (GD) and will be discussed in Section 4.7.

Additionally, Lu et al (Lu et al 2012) investigated two dictionaries, a transitional dictionary for atom matching and a global dictionary for image updating, to improve the image quality. Also, the aforementioned dictionary is generally assumed to be single-scale, that is, size of all the atoms in the dictionary is the same. Bai et al (Bai et al 2013) proposed to train a multi-scale dictionary which can extract more details and lead to better resolution than that based on single-scale dictionary for CT reconstruction.

4.7. Previous scan-induced regularizations

Repeated CT scans are required in some clinical applications such as perfusion imaging, cardiac dynamic imaging, image-guided intervention, and image-guided radiotherapy (Mayer et al 2000; Jaffray et al 2002; Dharap et al 2005; Lauritsch et al 2006; Zhao et al 2009; Siewerdsen 2011; Schubert et al 2012). In these applications, previous scan/scans can be exploited as prior information due to the similarity among the reconstructed image series of the scans. These scans generally contain the same anatomical structures. While somewhat misalignment or deformation may occur among the image series, they can be mitigated through registration of the image series, or modeling the effects in constructing the objective function. Using the reconstruction from the previous scan/scans to improve the current low-dose scan reconstructions has drawn great interests recently. These efforts also explored the use of the previous scan as regularization for SIR and some examples are given below.

4.7.1. Via the PICCS method
The PICCS method was initially proposed to reconstruct dynamic CT images from highly undersampled projection datasets (Chen et al 2008a). Then, it was also applied to reconstruct dynamic/perfusion CT images from noisy projection datasets (Lauzier and Chen 2012). The idea of the PICCS method is to take advantage of a prior image to improve the image quality of dynamic CT, wherein the prior image is typically generated by averaging over FBP reconstructions from different time frames of the dynamic CT exam (Chen et al 2008a; Lauzier and Chen 2012). For the prior image, although the dynamic information is lost, the static structures are well kept with greatly reduced streaking artifacts or noise. This strategy is also known as temporal resolution improvement PICCS (TRI-PICCS). Mathematically, the L_1 norm based TRI-PICCS can be given as (Chen et al 2008a; Lauzier and Chen 2012):

$$\lambda \|\psi_1(\mu - \mu_{\text{prior}})\|_1 + (1 - \lambda) \|\psi_2 \mu\|_1$$

where the parameters $\lambda$, $\psi_1$, and $\psi_2$ are defined the same as those in Eq. (57), but the prior image $\mu_{\text{prior}} = (\mu_{\text{prior}, 1}, ..., \mu_{\text{prior}, T})^T$ is reconstructed using FBP from the union of the interleaved dynamic projections, which is blurred with low temporal resolution.

Theoretically, Eq. (72) can be generalized to L-p norm ($0 < p \leq 2$) based PICCS:

$$\lambda \|\psi_1(\mu - \mu_{\text{prior}})\|_p + (1 - \lambda) \|\psi_2 \mu\|_p$$

Similarly, the $p$th power of the L-p norm is employed, and setting $p = 1$ yields the L_1 norm based PICCS norm in Eq. (72). However, when $0 < p < 1$, Eq. (73) is non-convex. The non-convex PICCS (NCPICCS) has been studied by (Ramirez-Giraldo et al 2011), and the simulation results have shown that the NCPICCS allows for image reconstruction from fewer samples than the L_1 norm based PICCS method in Eq. (72).

It should be noted that the prior image in Eqs. (72) and (73) is actually obtained from the current data and is of low quality. In some applications, a previous high quality image reconstructed from the previously acquired scans of the same patient may also be available. This previous high quality image contains rich patient-specific anatomical information and can also be incorporated into the PICCS framework. However, the previous high quality image needs to be registered to the current data.

The first way is to do the image registration before the reconstruction (Nett et al 2009):

$$\lambda \|\psi_1(\mu - \mu_{\text{reg}})\|_p + (1 - \lambda) \|\psi_2 \mu\|_p$$

(0 < p \leq 2)
where $\mu^{\text{preHQ registered}} \in \mathbb{R}^{J \times l}$ represents the previous high quality (preHQ) image registered to the current low-dose image.

The other way is to do the image registration during the reconstruction process (Stayman et al 2011; Stayman et al 2013; Dang et al 2014):

$$
\lambda \left\| \Psi_j \left( \mu - T \left( \mu^{\text{preHQ}} \right) \right) \right\|^p_p + (1 - \lambda) \left\| \Psi_j \mu \right\|^p_p \quad (0 < p \leq 2)
$$

(75)

where $\mu^{\text{preHQ}} \in \mathbb{R}^{J \times l}$ is the previous high quality image, and $T$ denotes a registration transformation operator and one can choose a rigid or deformable registration algorithm.

4.7.2. Via the NLM method

Since a previous high quality image of the same patient exists in some applications, (Zhang et al 2014b) proposed a family of previous normal-dose scan induced NLM (ndiNLM) regularizations for low-dose CT reconstruction, which is given as:

$$
U(\mu) = \sum_j \sum_{k \in \mathcal{S}_j} w_{jk}(\mu - \mu^{\text{preHQ registered}}) \phi(\mu_j - \mu_k^{\text{preHQ registered}})
$$

(76)

where $w_{jk}(\mu - \mu^{\text{preHQ registered}}) = \frac{\exp \left( -\| P(\mu_j) - P(\mu_k^{\text{preHQ registered}}) \|_2^2 / h^2 \right)}{\sum_{k \in \mathcal{S}_j} \exp \left( -\| P(\mu_j) - P(\mu_k^{\text{preHQ registered}}) \|_2^2 / h^2 \right)}$, and $\mu^{\text{preHQ registered}}$ represents the previous high quality image registered to the current low-dose image. It has been shown that the ndiNLM regularizations do not heavily depend on the accuracy of the image registration due to the patch-based search mechanism. Therefore, a rough registration will be sufficient, which makes the regularizations more practical.

Alternatively, this family of regularizations can also be given as the following form according to Section 4.4:

$$
U(\mu) = \sum_j \phi \left( \mu_j - \sum_{k \in \mathcal{S}_j} w_{jk}(\mu - \mu^{\text{preHQ registered}}) \mu_k^{\text{preHQ registered}} \right)
$$

(77)

(Ma et al 2012b) utilized this regularization to reconstruct low-dose perfusion CT images using a pre-contrast normal-dose CT scan, and generated very promising results.

4.7.3. Via global dictionary learning

As mentioned in Section 4.6, the dictionary $\mathbf{D}$ in Eq. (71) can also be learned from a prespecified training image. The training image for global dictionary learning can come from a previous
high quality scan of the same patient. In (Xu et al 2012), it was demonstrated that the SIR-GD performed robustly well with a dictionary learned from a quite different image (e.g., distant slice of the same patient), although sometimes tiny details may be invisible as compared to the SIR-AD.

4.7.4. Via predicted MRF coefficients

More recently, motivated by the work of (Wang et al 2012), two previous high quality image induced MRF regularizations were proposed to utilize the previous high quality image to improve the subsequent low-dose CT scans (Zhang et al 2014a; Zhang et al 2015a):

\[
U(\mu) = \sum_j \sum_{n \in \Omega} w^\text{preHQ\_predict}_{jm} \frac{1}{2} (\mu_j - \mu_m)^2
\]

where the weighting coefficients \( w^\text{preHQ\_predict}_{jm} \) are no longer global constant but are predicted from the previous high quality image, either pixel-specific (Zhang et al 2014a) or region-specific (Zhang et al 2015a).

Essentially, the traditional regularizations generally utilize a small MRF window size (e.g., \( W=3\times3 \)), which limit the spectral description to a crude low-pass model. By increasing MRF window size (e.g., \( \Omega=7\times7 \)), the penalty term increases the number of degrees of freedom in spectral description, and can include the high frequency information (Wang et al 2012). In this way, the high frequency components could be better preserved. The experimental results demonstrated that the regularization in Eq. (78) is superior in terms of noise reduction and the image edge/detail/contrast preservation (Zhang et al 2014a; Zhang et al 2015a).

5. Discussions and conclusions

The SIR methods for CT have been explored for radiation dose reduction in the past decades. Each of the SIR methods consists of two basic components of (1) construction of an adequate objective function and (2) utilization of a suitable algorithm for maximization or minimization of the objective function.

5.1. Construction of objective function

In construction of the objective function, the statistical modeling of the projection measurements for the data fidelity term is a prerequisite. The statistical phenomena in CT projection measurements are so complicated (Siewerdsen et al 1997; Nuyts et al 2013) that it is rarely practical to have an exact statistical model and likelihood function. Instead, the statistical models in Table 1 are
considered as good approximations, and have been widely employed in the CT image reconstruction. Besides, Xu and Tsui (Xu and Tsui 2007) proposed to use a subclass of the exponential dispersion (ED) models to approximate the compound Poisson statistics. While its manipulation is numerically tractable, the resulting log-likelihood function is not concave in $\mu$, and only local maximum can be reached for the resulting objective function (Xu and Tsui 2007). They further consider to include the electronic noise modeling as in Eq. (9), but the likelihood function of the sum of the ED model and a Gaussian noise model is analytically intractable (Xu and Tsui 2009). These statistical models were not explicitly reviewed in this work, but interested readers can refer to their papers for details.

In addition to the statistical model, a good regularization term is also a key to solve the ill-posed inverse problems including low-dose CT reconstruction. In practice, effectively suppressing the noise and streak artifacts while preserving the edges/details/contrasts are the two major concerns when designing a regularization term. Many regularization terms have been proposed and studied for low-dose CT reconstruction in the past decades. Figure 7 gives an overview of the regularizations which were, or potentially can be, employed in SIR for low-dose CT, and Section 4 illustrates their conceptual and mathematical bases. However, the regularizations reviewed herein are by no means comprehensive or conclusive, and more desirable regularizations based on prior knowledge may be proposed in the future.

While the statistical models in Table 1 can reasonably describe the statistical properties of the low-dose CT projection data, the regularization for faithful image reconstruction remains an open question. There are many regularization models to impose local smoothness on the image, however, their diagnostic values are still not well documented. In theory, the regularization shall incorporate the prior knowledge of the image to be reconstructed. Unfortunately, the diagnostic information is not known in prior and actually need to be from the current scan. However, for some specific clinical tasks, such image-guided interventions, prior knowledge is available to achieve the task.

Furthermore, the hyper-parameter $\beta$ in Eq. (29) controls the tradeoff between the data fidelity term and the regularization term. A larger $\beta$ value produces a more smoothed reconstruction with lower noise but also lower resolution, and vice versa. The noise-resolution tradeoff curve, or bias-variance tradeoff curve, can indicate the influence of $\beta$ value selection on reconstructed image quality. In practice, a series of $\beta$ values can be tested for a specific SIR method with specific projection data, and after the images are reconstructed, visual inspection and quantitative measurements are used to determine the optimal $\beta$ value. This scheme can also be considered as a
process of trial and error. But generally, the selection of $\beta$ value for SIR methods is still an open question and is considered as one of their drawbacks.

5.2. Optimization of objective function

This paper reviews the formulation of the objective function for SIR in great detail, but it does not include the optimization algorithm of the objective function, which is the other crucial component of SIR methods. Efficient algorithms for objective function optimization are important to avoid local maximum or minimum for the optimal solution. An iterative algorithm is typically employed because the objective function usually does not have a closed-form solution, or even if it has closed-form solution it is still impractical to directly invert the system matrix due to the large size (Fessler 2006). Several types of iterative algorithms have been proposed for solving objective functions of SIR (Fessler 2000; Fessler 2006), including the steepest descent (SD) (or gradient descent), conjugate gradient (CG), expectation maximization (EM), iterative coordinate descent (ICD), separable parabolic surrogates (SPS) and so on. The selection of an iterative algorithm depends on several factors such as the form of the objective function, convergence rate, parallelization ability, and reconstruction accuracy (Fessler 2000). There are also some variants of these iterative algorithms which attempt to accelerate the reconstruction, such as the ordered subsets (OS) (Hudson and Larkin 1994; Kamphuis and Beekman 1998; Erdogan and Fessler 1999; Lee 2000; Beekman and Kamphuis 2001), the space-alternating generalized EM (SAGE) (Fessler and Hero 1995), group coordinate descent (Fessler et al 1997), block-based ICD (Benson et al 2010), and alternating direction method of multipliers (ADMM) (Ramani and Fessler 2012) approaches. Specially, for the non-concave objective functions in Eq. (29), special strategies may need to be adopted to avoid the solution falling into local maximum or minimum.

5.3. Reconstructed CT images by SIR methods

In order to intuitively demonstrate the potential benefits of SIR methods over the conventional FBP method and the influence of regularizations on SIR methods, reconstructed CT images from low-dose acquisitions with clinical patients are illustrated in Figure 13 and Figure 14 (Xu et al 2012), respectively.

The first patient was scanned by a Siemens Sensation16 CT scanner. The X-ray tube voltage was set to be 120kVp and the mAs level was 20mAs, which is considered as ultra low-dose scan in clinic. The scan included 1160 sampling views evenly spanned on a 360° circular orbit. A reconstructed slice of the patient from the 20mAs acquisition by the FBP, SIR-GMRF [Eq. (41)],
SIR-Huber [Eq. (38) and Table 3], SIR-TV [Eq. (53)], SIR-NLS [Eq. (64)], and SIR-adaptiveNLM (Zhang et al 2015b) are shown in Figure 13(a)-(f). It can be observed that all the SIR methods outperform the FBP in terms of noise and streak artifact suppression. However, the SIR-GMRF result is blurred on the edges and fine structures, as indicated in Section 4.2.1. The SIR-Huber shows some gains over the SIR-GMRF on edge preservation, but it cannot generate excellent image quality for all objects with different contrasts due to the "edge threshold" selection. The SIR-TV exhibits slight blocky or patchy artifacts as indicated in Section 4.3.1. The SIR-NLM also produces inferior reconstruction for low-contrast objects in the image because of the spatially-invariant filtering across the entire image. In contrast, the SIR-adaptiveNLM generates superior image in terms of edge/detail/contrast preservation, due to the consideration of local characteristics of image and the incorporation of spatial adaptivity.

![Image](a) ![Image](b)
Figure 13. A reconstructed slice of the patient from 20mAs acquisition: (a) FBP reconstruction; (b) SIR-GMRF reconstruction; (c) SIR-Huber reconstruction; (d) SIR-TV reconstruction; (e) SIR-NLM reconstruction; (f) SIR-adaptiveNLM reconstruction. All the images are displayed with the same window.

The second patient was scanned by a GE Discovery CT750 HD scanner. Although 2200 projections were uniformly acquired on a 360° circular orbit, only 440 views were used for image reconstruction. Therefore, the sinogram data is both noisy and undersampled. Figure 14 shows a reconstructed slice of the patient by the FBP, SIR-TV [Eq. (53)], SIR-GD [Eq. (71)] and SIR-AD [Eq. (71)] respectively. The FBP reconstructed image has significant noise and streak artifacts, while
the three SIR reconstructed images are much better. However, the SIR-TV reconstructed image in
Figure 14(b) suffers from blocky or patchy artifacts slightly, as mentioned in Section 4.3.1. The SIR-
GD and SIR-AD results show good noise suppression and no obvious artifacts.

Figure 14. A reconstructed slice of clinical CT scan: (a) FBP; (b) SIR-TV; (c) SIR-GD; (d) SIR-AD.
The image display window is [-1000, 1000] HU. (Figure reprinted from Xu et al. 2012, Low-dose
X-ray CT reconstruction via dictionary learning, IEEE Trans. Med. Imag., 31(9): 1682-1697)

5.4. Clinical use of SIR methods

Although considerable progress has been made on SIR of CT during the past decades, the
analytical FBP method is still employed by most commercial scanners for image reconstruction (Pan
et al 2009). Due to a strong desire for radiation dose reduction, SIR methods have recently become
an endeavor for major CT vendors, and some prototypes products based on this have been exhibited
in a number of national and international meetings. Since the optimization of objective function for
SIR is routinely performed by iterative algorithms, the computational burden due to multiple re-
projection and back-projection operation cycles in the projection and image domains has always been one big challenge for clinical use. However, software approaches mentioned in Section 5.2 and hardware approaches using the graphical processing unit (GPU) (Xu and Mueller 2005; Jia et al 2010; Jia et al 2011) and cell broad-band engine (Knaup et al 2006; Kachelriess et al 2007) have been investigated to accelerate the iterative procedure and substantially reduce the reconstruction time. With constant improvements in computation technology, SIR methods can move closer to clinical use and may play a dominant role there in the near future.

Acknowledgements

The authors would like to thank Drs. Yan Liu, Hao Han, Gene Gindi and Junhai Wen for helpful discussions, and thank Mr. Michael Salerno for English editing.

References

http://www.imvinfo.com/index.aspx?sec=ct&

Alenius S and Ruotsalainen U (1997). "Bayesian image reconstruction for emission tomography based on median root prior." European Journal of Nuclear Medicine 24(3): 258-265.

Alenius S, Ruotsalainen U and Astola J (1998). "Using local median as the location of the prior distribution in iterative emission tomography image reconstruction." IEEE Transactions on Nuclear Science 45(6): 3097-3104.

Andersen A and Kak A (1984). "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm." Ultrasonic imaging 6(1): 81-94.

Bai T, Mou X, Xu Q and Zhang Y (2013). "Low-dose CT reconstruction based on multiscale dictionary." SPIE Medical Imaging 8668: 86683L-86683L-86688.

Beekman F J and Kamphuis C (2001). "Ordered subset reconstruction for x-ray CT." Phys Med Biol 46(7): 1835.

Beister M, Kolditz D and Kalender W A (2012). "Iterative reconstruction methods in X-ray CT." Physica Medica 28(2): 94-108.

Benson T M, De Man B K, Fu L and Thibault J-B (2010). "Block-based iterative coordinate descent." IEEE Nuclear Science Symposium Conference Record (NSS/MIC): 2856-2859.

Besag J (1989). "Digital Image Processing: Towards Bayesian image analysis." Journal of Applied statistics 16(3): 395-407.
Blake A and Zisserman A (1987). Visual Reconstruction. Cambridge, MA, MIT Press.

Bouman C and Sauer K (1993). "A generalized Gaussian image model for edge-preserving MAP estimation." IEEE Transactions on Image Processing 2(3): 296-310.

Bouman C A and Sauer K (1996). "A unified approach to statistical tomography using coordinate descent optimization." IEEE Transactions on Image Processing 5(3): 480-492.

Brenner D J and Hall E J (2007). "Computed tomography—an increasing source of radiation exposure." New England Journal of Medicine 357(22): 2277-2284.

Buades A, Coll B and Morel J-M (2005a). "A non-local algorithm for image denoising." IEEE Computer Vision and Pattern Recognition (CVPR) 2: 60-65.

Buades A, Coll B and Morel J-M (2005b). "A review of image denoising algorithms, with a new one." Multiscale Modeling & Simulation 4(2): 490-530.

Buades A, Coll B and Morel J-M (2006). "Image enhancement by non-local reverse heat equation." Preprint CMLA 22: 2006.

Candès E J, Romberg J and Tao T (2006). "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." IEEE Transactions on Information Theory 52(2): 489-509.

Charbonnier P, Blanc-Feraud L, Aubert G and Barlaud M (1994). "Two deterministic half-quadratic regularization algorithms for computed imaging." IEEE International Conference on Image Processing (ICIP) 2: 168-172.

Chartrand R (2007). "Exact reconstruction of sparse signals via nonconvex minimization." IEEE Signal Processing Letters 14(10): 707-710.

Chen G-H, Tang J and Leng S (2008a). "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets." Med Phys 35(2): 660-663.

Chen G-H, Thériault-Lauzier P, Tang J, Nett B, Leng S, Zambelli J, et al. (2012). "Time-resolved interventional cardiac C-arm cone-beam CT: an application of the PICCS algorithm." IEEE Transactions on Medical Imaging 31(4): 907-923.

Chen Y, Gao D, Nie C, Luo L, Chen W, Yin X and Lin Y (2009). "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior." Computerized Medical Imaging and Graphics 33(7): 495-500.

Chen Y, Ma J, Feng Q, Luo L, Shi P and Chen W (2008b). "Nonlocal prior Bayesian tomographic reconstruction." Journal of Mathematical Imaging and Vision 30(2): 133-146.
Chun I Y and Talavage T M (2013). "Efficient Compressed Sensing Statistical X-Ray CT Reconstruction from Fewer Measurements." The 12th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine: 30-33.

Dang H, Wang A, Sussman M S, Siewerdsen J and Stayman J (2014). "dPIRPLE: a joint estimation framework for deformable registration and penalized-likelihood CT image reconstruction using prior images." Phys Med Biol 59(17): 4799.

de González A B and Darby S (2004). "Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries." The lancet 363(9406): 345-351.

de González A B, Mahesh M, Kim K-P, Bhargavan M, Lewis R, Mettler F and Land C (2009). "Projected cancer risks from computed tomographic scans performed in the United States in 2007." Archives of Internal Medicine 169(22): 2071-2077.

De Man B and Basu S (2004). "Distance-driven projection and backprojection in three dimensions." Phys Med Biol 49(11): 2463.

Dharap S, Khandkar A, Pandey A and Sharma A (2005). "Repeat CT scan in closed head injury." Injury 36(3): 412-416.

Do S, Karl W C, Kalra M K, Brady T J and Pien H (2010). "Clinical low dose CT image reconstruction using high-order total variation techniques." SPIE Medical Imaging 7622: 76225D-7622D-76228.

Dollinger M B and Staudte R G (1991). "Influence functions of iteratively reweighted least squares estimators." Journal of the American Statistical Association 86(415): 709-716.

Donoho D L (2006). "Compressed sensing." IEEE Transactions on Information Theory 52(4): 1289-1306.

Elad M (2002). "On the origin of the bilateral filter and ways to improve it." IEEE Transactions on Image Processing 11(10): 1141-1151.

Elbakri I A and Fessler J A (2002). "Statistical image reconstruction for polyenergetic X-ray computed tomography." IEEE Transactions on Medical Imaging 21(2): 89-99.

Elbakri I A and Fessler J A (2003). "Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography." SPIE Medical Imaging 5032: 1839-1850.

Elmoataz A, Lezoray O and Bougleux S (2008). "Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing." IEEE Transactions on Image Processing 17(7): 1047-1060.

Erdogan H and Fessler J A (1999). "Ordered subsets algorithms for transmission tomography." Phys Med Biol 44(11): 2835.
Fessler J A (2000). "Statistical image reconstruction methods for transmission tomography." Handbook of Medical Imaging 2: 1-70.

Fessler J A (2006). "Iterative methods for image reconstruction." ISBI Tutorial. Arlington Virginia: April.

Fessler J A, Ficaro E P, Clinthorne N H and Lange K (1997). "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction." IEEE Transactions on Medical Imaging 16(2): 166-175.

Fessler J A and Hero A O (1995). "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms." IEEE Transactions on Image Processing 4(10): 1417-1429.

Garduño E, Herman G and Davidi R (2011). "Reconstruction from a few projections by ℓ1-minimization of the Haar transform." Inverse problems 27(5): 055006.

Geman D and Reynolds G (1992). "Constrained restoration and the recovery of discontinuities." IEEE Transactions on Pattern Analysis and Machine Intelligence 14(3): 367-383.

Geman S and Geman D (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." IEEE Transactions on Pattern Analysis and Machine Intelligence(6): 721-741.

Geman S and McClure D (1985). "Bayesian image analysis: an application to single photon emission tomography." Statist. Comput. Sect: 12-18.

Gordon R, Bender R and Herman G T (1970). "Algebraic reconstruction techniques (Art) for 3-dimensional electron microscopy and x-ray photography." J. Theor. Biol. 29: 471-481.

Green P J (1984). "Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives." Journal of the Royal Statistical Society B 46(2): 149-192.

Green P J (1990). "Bayesian reconstructions from emission tomography data using a modified EM algorithm." IEEE Transactions on Medical Imaging 9(1): 84-93.

Han G, Liang Z and You J (1999). "A fast ray-tracing technique for TCT and ECT studies." Nuclear Science Symposium, 1999. Conference Record. 1999 IEEE 3: 1515-1518.

Hebert T and Leahy R (1989). "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors." IEEE Transactions on Medical Imaging 8(2): 194-202.

Herman G (1980). Image reconstruction from projections: the fundamentals of computerized tomography, Academic.
Hounsfield G N (1968). Method of and apparatus for examining a body by radiation such as X or gamma radiation. USA, Google Patents.

Hsiao I-T, Rangarajan A and Gindi G (2003). "A new convex edge-preserving median prior with applications to tomography." IEEE Transactions on Medical Imaging 22(5): 580-585.

Hsieh J (2009). "Computed tomography: principles, design, artifacts, and recent advances."

Hu Y, Xie L, Luo L, Nunes J C and Toumoulin C (2011). "L0 constrained sparse reconstruction for multi-slice helical CT reconstruction." Phys Med Biol 56(4): 1173.

Huber P (1981). Robust Statistics. New York, John Wiley & Sons.

Hudson H M and Larkin R S (1994). "Accelerated image reconstruction using ordered subsets of projection data." IEEE Transactions on Medical Imaging 13(4): 601-609.

Jaffray D A, Siewerdsen J H, Wong J W and Martinez A A (2002). "Flat-panel cone-beam computed tomography for image-guided radiation therapy." International Journal of Radiation Oncology Biology Physics 53(5): 1337-1349.

Jia X, Dong B, Lou Y and Jiang S B (2011). "GPU-based iterative cone beam CT reconstruction using tight frame regularization." Phys Med Biol 56(13): 3787-3807.

Jia X, Lou Y, Li R, Song W Y and Jiang S B (2010). "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation." Med Phys 37(4): 1757-1760.

Jiang M and Wang G (2003). "Convergence of the simultaneous algebraic reconstruction technique (SART)." IEEE Transactions on Image Processing 12(8): 957-961.

Kachelriess M, Knaup M and Bockenbach O (2007). "Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware." Med Phys 34(4): 1474-1486.

Kak A C and Slaney M (1988). Principles of Computerized Tomographic Imaging, IEEE Press.

Kamphuis C and Beekman F J (1998). "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm." IEEE Transactions on Medical Imaging 17(6): 1101-1105.

Kindermann S, Osher S and Jones P W (2005). "Deblurring and denoising of images by nonlocal functionals." Multiscale Modeling & Simulation 4(4): 1091-1115.

Knaup M, Kalender W A and Kachelrie M (2006). "Statistical cone-beam CT image reconstruction using the cell broadband engine." IEEE Nuclear Science Symposium Conference Record 5: 2837-2840.

La Riviére P J (2005). "Penalized-likelihood sinogram smoothing for low-dose CT." Med Phys 32(6): 1676-1683.
La Rivière P J, Bian J and Vargas P A (2006). "Penalized-likelihood sinogram restoration for computed tomography." IEEE Transactions on Medical Imaging 25(8): 1022-1036.

La Rivière P J and Billmire D M (2005). "Reduction of noise-induced streak artifacts in X-ray computed tomography through spline-based penalized-likelihood sinogram smoothing." IEEE Transactions on Medical Imaging 24(1): 105-111.

La Rivière P J and Vargas P (2008). "Correction for resolution nonuniformities caused by anode angulation in computed tomography." IEEE Transactions on Medical Imaging 27(9): 1333-1341.

Lalush D and Tsui B (1993). "A generalized Gibbs prior for maximum a posteriori reconstruction in SPECT." Phys Med Biol 38(6): 729.

Lange K (1990). "Convergence of EM image reconstruction algorithms with Gibbs smoothing." IEEE Transactions on Medical Imaging 9(4): 439-446.

Lange K, Bahn M and Little R (1987). "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography." IEEE Transactions on Medical Imaging 6(2): 106-114.

Lange K and Carson R (1984). "EM reconstruction algorithms for emission and transmission tomography." J Comput Assist Tomogr 8(2): 306-316.

Lasio G M, Whiting B R and Williamson J F (2007). "Statistical reconstruction for x-ray computed tomography using energy-integrating detectors." Phys Med Biol 52(8): 2247.

Lauritsch G, Boese J, Wigstrom L, Kemeth H and Fahrig R (2006). "Towards cardiac C-arm computed tomography." Medical Imaging, IEEE Transactions on 25(7): 922-934.

Lauzier P T and Chen G-H (2012). "Characterization of statistical prior image constrained compressed sensing. I. Applications to time-resolved contrast-enhanced CT." Med Phys 39(10): 5930-5948.

Lauzier P T and Chen G-H (2013). "Characterization of statistical prior image constrained compressed sensing (PICCS): II. Application to dose reduction." Med Phys 40(2): 021902.

Lee S-J (2000). "Accelerated coordinate descent methods for Bayesian reconstruction using ordered subsets of projection data." International Symposium on Optical Science and Technology: 170-181.

Levitan E and Herman G T (1987). "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography." IEEE Transactions on Medical Imaging 6(3): 185-192.

Li T, Li X, Wang J, Wen J, Lu H, Hsieh J and Liang Z (2004). "Nonlinear sinogram smoothing for low-dose X-ray CT." IEEE Transactions on Nuclear Science 51(5): 2505-2513.
Liang Z, Jaszczak R and Greer K (1989). "On Bayesian image reconstruction from projections: uniform and nonuniform a priori source information." IEEE Transactions on Medical Imaging 8(3): 227-235.

Liu Y, Fan Y, Lu H and Liang Z (2013). "Improved area-simulating-volume method for X-ray CT reprojection and back-projection operations." IEEE Nuclear science symposium and medical imaging conference (NSS/MIC)

Liu Y, Liang Z, Ma J, Lu H, Wang K, Zhang H and Moore W (2014a). "Total Variation-Stokes Strategy for Sparse-View X-ray CT Image Reconstruction." IEEE Transactions on Medical Imaging 33(3): 749 - 763.

Liu Y, Ma J, Fan Y and Liang Z (2012). "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction." Phys Med Biol 57(23): 7923.

Liu Y, Ma J, Zhang H, Wang J and Liang Z (2014b). "Low-mAs X-ray CT image reconstruction by adaptive-weighted TV-constrained penalized re-weighted least-squares." Journal of X-ray Science and Technology 22(4): 437-457.

Lo S-C (1988). "Strip and line path integrals with a square pixel matrix: A unified theory for computational CT projections." IEEE Transactions on Medical Imaging 7(4): 355-363.

Long Y, Fessler J A and Balter J M (2010). "3D forward and back-projection for X-ray CT using separable footprints." Medical Imaging, IEEE Transactions on 29(11): 1839-1850.

Lou Y, Zhang X, Osher S and Bertozzi A (2010). "Image recovery via nonlocal operators." Journal of Scientific Computing 42(2): 185-197.

Lu H, Hsiao I-T, Li X and Liang Z (2001). "Noise properties of low-dose CT projections and noise treatment by scale transformations." IEEE Nuclear Science Symposium Conference Record 3: 1662-1666.

Lu Y, Zhao J and Wang G (2012). "Few-view image reconstruction with dual dictionaries." Phys Med Biol 57(1): 173.

Ma J, Liang Z, Fan Y, Liu Y, Huang J, Chen W and Lu H (2012a). "Variance analysis of x-ray CT sinograms in the presence of electronic noise background." Med Phys 39(7): 4051-4065.

Ma J, Zhang H, Gao Y, Huang J, Liang Z, Feng Q and Chen W (2012b). "Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior." Phys Med Biol 57(22): 7519.

Macovski A (1983). Medical Imaging Systems. New Jersey, Prentice Hall.
Manduca A, Yu L, Trzasko J D, Khaylova N, Kofler J M, McCollough C M and Fletcher J G (2009). "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT." Med Phys 36(11): 4911-4919.

Matej S and Lewitt R M (1996). "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements." Medical Imaging, IEEE Transactions on 15(1): 68-78.

Mayer T E, Hamann G F, Baranczyk J, Rosengarten B, Klotz E, Wiesmann M, et al. (2000). "Dynamic CT perfusion imaging of acute stroke." American journal of neuroradiology 21(8): 1441-1449.

McCollough C H, Primak A N, Braun N, Kofler J, Yu L and Christner J (2009). "Strategies for reducing radiation dose in CT." Radiologic Clinics of North America 47(1): 27-40.

Mueller K, Yagel R and Wheller J J (1999). "Fast implementations of algebraic methods for three-dimensional reconstruction from cone-beam data." Medical Imaging, IEEE Transactions on 18(6): 538-548.

Nett B, Tang J, Aagaard-Kienitz B, Rowley H and Chen G-H (2009). "Low radiation dose C-arm cone-beam CT based on prior image constrained compressed sensing (PICCS): including compensation for image volume mismatch between multiple data acquisitions." SPIE Medical Imaging 7258: 725803-725801:725803-725812.

Niu S, Gao Y, Bian Z, Huang J, Chen W, Yu G, et al. (2014). "Sparse-view x-ray CT reconstruction via total generalized variation regularization." Phys Med Biol 59(12): 2997.

Nunez J and Llacer J (1990). "A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images." IEEE Transactions on Medical Imaging 9(2): 159-171.

Nuyts J, De Man B, Fessler J A, Zbijewski W and Beekman F J (2013). "Modelling the physics in the iterative reconstruction for transmission computed tomography." Phys Med Biol 58(12): R63.

Ouyang L, Solberg T and Wang J (2011). "Effects of the penalty on the penalized weighted least-squares image reconstruction for low-dose CBCT." Phys Med Biol 56(17): 5535.

Pan X, Sidky E Y and Vannier M (2009). "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" Inverse problems 25(12): 123009.

Perona P and Malik J (1990). "Scale-space and edge detection using anisotropic diffusion." IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7): 629-639.

Peters T (1981). "Algorithms for fast back-and re-projection in computed tomography." Nuclear Science, IEEE Transactions on 28(4): 3641-3647.
Press W H, Flannery B P, Teukolsky S A and Vetterling W T (1992). Numerical recipes: The art of scientific computing, Cambridge university press.

Qi J and Leahy R M (2006). "Iterative reconstruction techniques in emission computed tomography." Phys Med Biol 51(15): R541.

Ramani S and Fessler J A (2012). "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction." IEEE Transactions on Medical Imaging 31(3): 677-688.

Ramirez-Giraldo J, Trzasko J, Leng S, Yu L, Manduca A and McCollough C (2011). "Nonconvex prior image constrained compressed sensing (NCPICCS): Theory and simulations on perfusion CT." Med Phys 38(4): 2157-2167.

Riddell C and Trousslet Y (2006). "Rectification for cone-beam projection and backprojection." Medical Imaging, IEEE Transactions on 25(7): 950-962.

Rockmore A and Macovski A (1977). "A maximum likelihood approach to transmission image reconstruction from projections." IEEE Transactions on Nuclear Science 24(3): 1929-1935.

Sauer K and Bouman C (1993). "A local update strategy for iterative reconstruction from projections." IEEE Transactions on Image Processing 41(2): 534-548.

Schubert T, Jacob A L, Takes M, Menter T, Gutzeit A and Kos S (2012). "CT-guided percutaneous biopsy of a mass lesion in the upper presacral space: a sacral transneuroforaminal approach." Cardiovascular and interventional radiology: 1-3.

Shepp L A and Vardi Y (1982). "Maximum likelihood reconstruction for emission tomography." IEEE Transactions on Image Processing 1(2): 113-122.

Siddon R L (1985). "Fast calculation of the exact radiological path for a three - dimensional CT array." Med Phys 12(2): 252-255.

Sidky E Y, Kao C-M and Pan X (2006). "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT." Journal of X-ray Science and Technology 14(2): 119-139.

Sidky E Y and Pan X (2008). "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." Phys Med Biol 53(17): 4777.

Siewerdsen J, Antonuk L, El-Mohri Y, Yorkston J, Huang W, Boudry J and Cunningham I (1997). "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology." Med Phys 24(1): 71-89.

Siewerdsen J, Waese A, Moseley D, Richard S and Jaffray D (2004). "Spektr: A computational tool for x-ray spectral analysis and imaging system optimization." Med Phys 31(11): 3057-3067.
Siewerdsen J H (2011). "Cone-beam CT with a flat-panel detector: From image science to image-guided surgery." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 648: S241-S250.

Smith-Bindman R, Lipson J, Marcus R, Kim K-P, Mahesh M, Gould R, et al. (2009). "Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer." Archives of Internal Medicine 169(22): 2078-2086.

Snyder D L, Hammoud A M and White R L (1993). "Image recovery from data acquired with a charge-coupled-device camera." JOSA A 10(5): 1014-1023.

Snyder D L, Helstrom C W, Lanterman A D, Faisal M and White R L (1995). "Compensation for readout noise in CCD images." JOSA A 12(2): 272-283.

Stayman J, Zbijewski W, Otake Y, Uneri A, Schafer S, Lee J, et al. (2011). "Penalized-likelihood reconstruction for sparse data acquisitions with unregistered prior images and compressed sensing penalties." SPIE Medical Imaging: 79611L-79611L-79616.

Stayman J W, Dang H, Ding Y and Siewerdsen J H (2013). "PIRPLE: a penalized-likelihood framework for incorporation of prior images in CT reconstruction." Phys Med Biol 58(21): 7563.

Stevenson R and Delp E (1990). "Fitting Curves with Discontinuities." the 1st Int. Workshop. on Robust Comm. Vision: 127-136.

Tang J, Nett B E and Chen G-H (2009). "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms." Phys Med Biol 54(19): 5781.

Thibault J-B, Bouman C A, Sauer K D and Hsieh J (2006). "A recursive filter for noise reduction in statistical iterative tomographic imaging." SPIE Electronic Imaging 6065: 60650X-60650X-60610.

Thibault J-B, Sauer K D, Bouman C A and Hsieh J (2007). "A three-dimensional statistical approach to improved image quality for multislice helical CT." Med Phys 34(11): 4526-4544.

Tian Z, Jia X, Yuan K, Pan T and Jiang S B (2011). "Low-dose CT reconstruction via edge-preserving total variation regularization." Phys Med Biol 56(18): 5949.

Tomasi C and Manduchi R (1998). "Bilateral filtering for gray and color images." IEEE International Conference on Computer Vision (ICCV): 839-846.

Toth T L, Cesmeli E, Ikhlef A and Horiuchi T (2005). "Image quality and dose optimization using novel x-ray source filters tailored to patient size." SPIE Medical imaging 5745: 283-291.

Trzasko J and Manduca A (2009). "Highly undersampled magnetic resonance image reconstruction via homotopic-minimization." IEEE Transactions on Medical Imaging 28(1): 106-121.
Vandeghinste B, Goossens B, Van Holen R, Vanhove C, Pizurica A, Vandenberghhe S and Staelens S (2012a). "Combined shearlet and TV regularization in sparse-view CT reconstruction." Proc. 2nd International Meeting on Image Formation in X-ray Computed Tomography

Vandeghinste B, Goossens B, Van Holen R, Vanhove C, Pizurica A, Vandenberghhe S and Staelens S (2012b). "Iterative CT reconstruction using shearlet-based regularization." SPIE Medical Imaging 8313: 83133I-83133I-83137.

Wang G and Qi J (2012). "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization." IEEE Transactions on Medical Imaging 31(12): 2194-2204.

Wang J, Li T, Lu H and Liang Z (2006). "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography." IEEE Transactions on Medical Imaging 25(10): 1272-1283.

Wang J, Li T and Xing L (2008a). "Iterative image reconstruction for CBCT using edge-preserving prior." Med Phys 36(1): 252-260.

Wang J, Lu H, Liang Z, Eremina D, Zhang G, Wang S, et al. (2008b). "An experimental study on the noise properties of x-ray CT sinogram data in Radon space." Phys Med Biol 53(12): 3327.

Wang J, Sauer K, Thibault J-B, Yu Z and Bouman C (2012). "Prediction coefficient estimation in Markov random fields for iterative x-ray ct reconstruction." SPIE Medical Imaging: 831444-831444-831449.

Wang J, Wang S, Li L, Fan Y, Lu H and Liang Z (2008c). "Virtual colonoscopy screening with ultra low-dose CT and less-stressful bowel preparation: A computer simulation study." IEEE Transactions on Nuclear Science 55(5): 2566-2575.

Wernick M N and Aarsvold J N (2004). Emission tomography: the fundamentals of PET and SPECT, Academic Press.

Whiting B R (2002). "Signal statistics in x-ray computed tomography." SPIE Medical Imaging 4682: 53-60.

Whiting B R, Massoumzadeh P, Earl O A, O’Sullivan J A, Snyder D L and Williamson J F (2006). "Properties of preprocessed sinogram data in x-ray computed tomography." Med Phys 33(9): 3290-3303.

Wu H, Maier A and Hornegger J (2013). Iterative CT Reconstruction Using Curvelet-Based Regularization. Bildverarbeitung für die Medizin 2013, Springer: 229-234.

Wu M and Fessler J A (2011). "GPU acceleration of 3D forward and backward projection using separable footprints for X-ray CT image reconstruction." Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med: 56-59.

IRIS Laboratory, Department of Radiology, Stony Brook University
Xu F and Mueller K (2005). "Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware." IEEE Transactions on Nuclear Science 52(3): 654-663.

Xu J and Tsui B (2014). "Quantifying the importance of the statistical assumption in statistical x-ray CT image reconstruction." IEEE Transactions on Medical Imaging 33(1): 61-73.

Xu J and Tsui B M (2009). "Electronic noise modeling in statistical iterative reconstruction." IEEE Transactions on Image Processing 18(6): 1228-1238.

Xu J and Tsui B M W (2007). "A compound Poisson maximum likelihood iterative reconstruction algorithm for x-ray CT." Fully three-dimensional Image Reconstruction in Radiology and Nuclear Medicine: 108-111.

Xu Q, Yu H, Mou X, Zhang L, Hsieh J and Wang G (2012). "Low-dose X-ray CT reconstruction via dictionary learning." IEEE Transactions on Medical Imaging 31(9): 1682-1697.

Yavuz M and Fessler J A (1998). "Statistical image reconstruction methods for randoms-precorrected PET scans." Med Image Anal 2(4): 369-378.

Yu H and Wang G (2010). "SART-type image reconstruction from a limited number of projections with the sparsity constraint." Journal of Biomedical Imaging 2010: 3.

Yu L, Liu X, Leng S, Kofler J M, Ramirez-Giraldo J C, Qu M, et al. (2009). "Radiation dose reduction in computed tomography: techniques and future perspective." Imaging in Medicine 1(1): 65-84.

Zeng G and Gullberg G (1993). "A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms." Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.: 1199-1201.

Zhang H, Han H, Liang Z, Liu Y, Moore W, Wang J and Ma J (2015a). "Deriving regional specific MRF coefficients from previous normal-dose CT scan as a priori knowledge for statistical low-dose image reconstruction." IEEE Trans. Medical Imaging (in submission).

Zhang H, Han H, Wang J, Ma J, Liu Y, Moore W and Liang Z (2014a). "Deriving adaptive MRF coefficients from previous normal-dose CT scan for low-dose image reconstruction via penalized weighted least-squares minimization." Med Phys 41(4): 041916.

Zhang H, Huang J, Ma J, Bian Z, Feng Q, Lu H, et al. (2014b). "Iterative Reconstruction for X-Ray Computed Tomography using Prior-Image Induced Nonlocal Regularization." IEEE Transactions on Biomedical Engineering 61(9): 2367 - 2378.

Zhang H, Ma J, Liu Y, Han H, Li L, Wang J and Liang Z (2014c). "Nonlocal means-based regularizations for statistical CT reconstruction." SPIE Medical Imaging 9033: 903337.
Zhang H, Ma J, Wang J, Liu Y, Lu H and Liang Z (2014d). "Statistical image reconstruction for low-dose CT using nonlocal means-based regularization." Computerized Medical Imaging and Graphics 38(6): 423-435.

Zhang H, Ma J, Wang J, Liu Y, Moore W, Lu H and Liang Z (2015b). "Statistical image reconstruction for low-dose CT using nonlocal means-based regularization. Part II: An adaptive approach." Computerized Medical Imaging and Graphics (under review).

Zhang X, Burger M, Bresson X and Osher S (2010). "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction." SIAM Journal on Imaging Sciences 3(3): 253-276.

Zhao B, James L P, Moskowitz C S, Guo P, Ginsberg M S, Lefkowitz R A, et al. (2009). "Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer I." Radiology 252(1): 263-272.

Zhou W, Cai J-F and Gao H (2013). "Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography." Inverse problems 29(12): 125006.

Zhu Z, Wahid K, Babyn P, Cooper D, Pratt I and Carter Y (2013). "Improved Compressed Sensing-Based Algorithm for Sparse-View CT Image Reconstruction." Computational & Mathematical Methods in Medicine 2013: 1-15.

Zhuang W, Gopal S and Hebert T (1994). "Numerical evaluation of methods for computing tomographic projections." Nuclear Science, IEEE Transactions on 41(4): 1660-1665.

Ziegler A, Köhler T, Nielsen T and Proksa R (2006). "Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry." Med Phys 33(12): 4653-4663.