Published in final edited form as:

IEEE Trans Med Imaging. 2017 December ; 36(12): 2588–2599. doi:10.1109/TMI.2017.2766185.

Low-dose CBCT Reconstruction Using Hessian Schatten Penalties

Liang Liu,

Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China,
School of Automation, Huazhong University of Science and Technology, Wuhan, China

Xinxin Li,

Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China,
School of Automation, Huazhong University of Science and Technology, Wuhan, China

Kai Xiang,

Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China,
School of Automation, Huazhong University of Science and Technology, Wuhan, China

Jing Wang, and

Department of Radiation Oncology, University of Texas Southwestern Medical Center, Texas,
USA

Shan Tan [Member, IEEE]

Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China,
School of Automation, Huazhong University of Science and Technology, Wuhan, China

Abstract

Cone-beam computed tomography (CBCT) has been widely used in radiation therapy. For accurate patient setup and treatment target localization, it is important to obtain high-quality reconstruction images. The total variation (TV) penalty has shown state-of-the-art performance in suppressing noise and preserving edges for statistical iterative image reconstruction, but it sometimes leads to the so-called staircase effect. In this study, we proposed to use a new family of penalties — the Hessian Schatten (HS) penalties — for CBCT reconstruction. Consisting of the second-order derivatives, the HS penalties are able to reflect the smooth intensity transitions of the underlying image without introducing the staircase effect. We discussed and compared the behaviors of several convex HS penalties with orders 1, 2, and $+\infty$ for CBCT reconstruction. We used the majorization-minimization approach with a primal-dual formulation for the corresponding optimization problem. Experiments on two digital phantoms and two physical phantoms demonstrated the proposed penalty family's outstanding performance over TV in suppressing the staircase effect, and the HS penalty with order 1 had the best performance among the HS penalties tested.

Index Terms

CBCT; image reconstruction; TV; staircase effect; Hessian Schatten penalty

I. Introduction

Cone-beam computed tomography (CBCT) provides volumetric information for patient setup and target localization during radiation therapy. Repeated CBCT scans during a treatment course expose patients to extra imaging radiation doses [1], [2]. Lowering the mAs (tube current exposure time product) level in CBCT acquisition can reduce the imaging dose and associated risks to patients [3]. However, the quality of the reconstructed low-mAs image can be dramatically degraded due to excessive noise [4], [5].

Statistical iterative reconstruction (SIR) algorithms have been used to improve the quality of low-dose CBCT imaging [6], [7]. Commonly used penalty terms in SIR include the isotropic quadratic penalty [8], the Huber penalty [9], and the total variation (TV) penalty [10]–[12].

Originally introduced for image de-noising by Rudin *et al.* [10], the TV penalty has shown state-of-the-art performance in preserving edges and suppressing noise for CBCT imaging [10]–[24]. Various TV-based CBCT reconstruction algorithms have been studied and evaluated in the last three decades [10]–[24]. Sidky and Pan proposed to minimize the TV penalty subject to a non-negative constraint for sparse-view projection reconstruction, where the TV objective function was minimized by steepest descent with an adaptive step size, and the non-negative constraint was enforced by projection onto convex sets (POCS) [13], [18], [19]. Han *et al.* studied the selection of parameter values in the constrained TV minimization-based reconstruction and showed, with both visualization and quantitative assessments, that the reconstructions were greatly improved, relative to the clinical FDK reconstruction, in terms of the devised utility metrics [20]. Bian *et al.* showed that the TV-based iterative algorithm was more robust to different data conditions, such as number of views and exposure levels, than the FDK algorithm for CBCT reconstruction [21]. Hashemi *et al.* proposed a simultaneous deblurring and iterative reconstruction to explicitly account for image unsharpness caused by different factors in the CBCT reconstruction formulation when using the TV penalty [14]. Huang and Hsiao proposed an iterative reconstruction method to accelerate the ordered-subsets reconstruction with a power factor, in combination with the TV minimization method [15].

The major limitation of the TV penalty is that it sometimes causes the staircase effect in reconstructed images [25], [26], particularly at low-dose levels [16]. TV favors penalizing the first-order derivatives, *i.e.*, minimizing the intensity difference between neighboring pixels. Therefore, it tends to produce piecewise constant reconstructed images, even when the underlying images are not necessarily piecewise constant.

To avoid the staircase effect due to the TV penalty, Yan *et al.* considered the anisotropic edge property among neighboring image voxels for TV-POCS implementation, where the associated weights were expressed as an exponential function and can be adjusted adaptively by the local image-intensity gradient to preserve the edge details for sparse-view low-dose

CT image reconstruction [26]. Several alternative penalty terms with higher-order derivatives have been proposed in other fields, such as image restoration [27] and microscopy imaging deconvolution [28]. These penalty terms do not directly penalize the intensity difference between neighboring pixels. Recently, Sun *et al.* replaced the TV penalty with a Hessian penalty in a CBCT SIR algorithm [29]. The Hessian penalty was constructed using the second-order derivative operations and was defined as the Frobenius norm of the Hessian matrix of the image. They used the majorization-minimization (MM) method for the associated optimization problem. In their algorithm, the objective function was upper-bounded by a sequence of quadratic majorizers, which can be minimized directly using the Gauss-Seidel update strategy [29]. The proposed Hessian penalty proved successful in suppressing the TV penalty's staircase effect. However, like most higher-order penalties, the Hessian penalty tends to slightly blur the edges of the reconstructed image. Shi *et al.* [30] proposed a new penalty combining the TV and Hessian penalties in a structure-adaptive way to reconstruct CBCT images without introducing extra parameters. The proposed penalty can automatically adjust the weight parameters between TV and Hessian according to local image structures to suppress the staircase effect and preserve edges simultaneously.

The Hessian Schatten (HS) penalties are a novel family of second-degree penalties proposed recently by Lefkimmiatis *et al.* to avoid the staircase effect for ill-posed linear inverse problems [31]. The penalties in this family are invariant, convex, and non-quadratic, and they are defined as the Schatten norm of the Hessian matrix of the underlying image at each image point [31]. The Schatten norm is a p -norm of the vector of the singular values of a matrix, including many popular matrix norms, such as the nuclear/trace norm ($p = 1$), the Frobenius norm ($p = 2$), and the spectral/operator norm ($p = \infty$). The HS family generalizes the Hessian penalty and provides a more flexible family of second-order penalties. The Hessian penalty employed by Sun *et al.* [29] is a special case of the HS penalty with $p = 2$. Hu *et al.* showed that the HS penalty with $p = 1$ is nearly proportional to the higher degree total variation (HDTV) penalty with order 2 [32]. The HS family provides a natural way to incorporate image local geometry structure information. Specifically, an HS penalty can be interpreted as a scalar measurement of the curvature at a local image surface patch, and the eigenvalues of the Hessian matrix correspond to the principal curvatures, which can be used to measure the surface flatness in different directions [31].

The selection of the p value can affect the regularized solution when using the HS penalties for ill-posed linear inverse problems. Lefkimmiatis *et al.* showed that, for most cases, the HS penalty with $p = 1$ gives better results than the HS penalties with $p = 2$ and $p = \infty$ in sparse image restoration and microscopy image deburring [31]. Similar phenomena have been observed in both the compressed sensing problem [33] and the rank minimization problem for data recovery in computer vision. It is well-known that finding the sparsest representation (the l_0 norm minimization problem) in compressed sensing [34] and minimizing the data rank for recovering noisy/missing input data are both NP -hard [35], [36]. The l_1 norm minimization is known to be the best convex approximation of the l_0 norm minimization problem in compressed sensing among all p 's in $(0, 2)$ [34]. Note that the Schatten norm with $p = 0$ (quasinorm) measures the rank of a matrix and thus provides a natural way to approximate the rank minimization problem. Kong *et al.* [37] showed that the

Schatten norm gave better approximation results to the rank minimization problem at smaller p values for noisy data recovery.

In this study, we employed the Hessian Schatten (HS) penalty family for 3D CBCT image reconstruction. The purpose of this study was two-fold. First, we designed a new SIR algorithm based on the HS penalty family for CBCT reconstruction to avoid the staircase effect [38]. Second, we compared the behavior and performance of HS penalties with different orders ($p = 1$, $p = 2$, and $p = \infty$) for CBCT reconstruction. We note that the HS penalty is convex with $p \geq 1$ and non-convex (quasinorm) with $0 < p < 1$.

One major challenge when using HS penalties for 3D CBCT reconstruction is that members in this penalty family are not smooth and differentiable, which prohibits direct use of a gradient-based method to optimize the objective function. Furthermore, the optimization method proposed by Sun *et al.* [29] cannot be directly generalized to the HS penalties with $p = 1$ and $p = \infty$, as they are both non-quadratic. In this study, we adopted a primal-dual formulation originally proposed by Lefkimmiatis *et al.* [31] for 3D CBCT reconstruction with HS penalties of orders 1, 2, and ∞ . The original formulation was proposed in 2D [31], and extending it to 3D is not trivial. For example, in this study, we deduced the numerical formulation of the conjugate Hessian operator in 3D and proved that the dual objective function using the HS penalty for 3D CBCT reconstruction is Lipschitz continuous. We also derived the projection operator onto the l_q unit-ball ($1/p + 1/q = 1$) using an efficient algorithm in 3D [39].

The HS family tends to penalize the second-order derivatives, thus yielding smooth reconstructions, rather than the piecewise constant reconstructions that the TV penalty yields. Therefore, the reconstructed image can more faithfully reflect the real intensity transitions of the underlying image. We compared experiments conducted from simulated projection data and real projection data to demonstrate the effectiveness of the HS penalty family. Our experimental results showed that all the tested HS penalties significantly suppressed the staircase effect, and the HS penalty with $p = 1$ (nuclear/trace norm) performed best in most cases, closely followed by the HS with $p = 2$ (Frobenius norm).

II. Mathematical Modeling

The X-ray CT projection is the line integration along the X-ray path l of the tissue attenuation, which can be calculated as the logarithm transform of the ratio of the incident and detected intensities [7]:

$$v = \ln \frac{I_0}{I} = \int_l u(x, y, z) dl, \quad (1)$$

where v is the projection, and u is the attenuation coefficient. The noise of the X-ray CT projection approximately follows a Gaussian distribution [6], [7]. The noise variance σ_i^2 associated with the projection v_i at the i th detector bin, can be determined as follows [7]:

$$\sigma_i^2 = \exp(\bar{v}_i) / I_{i0}, \quad (2)$$

where I_{i0} is the incident photon number, and $\bar{v}_i$ is the mean value of the projection data. The CBCT reconstruction seeks the optimal solution by minimizing the objective function [40]:

$$\Phi(u) = \frac{1}{2} (v - Au)^T \Sigma^{-1} (v - Au) + \tau R(u), \quad (3)$$

where $\tau \geq 0$ is the regularization parameter, A is the system projection matrix, and Σ is a diagonal matrix with its i th element σ_i^2 . (Generally, the mean projection $\bar{v}_i$ is not available in practice. To calculate σ_i^2 , $\bar{v}_i$ is commonly replaced with the projection v_i , without considering the statistical variation).

A. Hessian Schatten Penalty

In this study, the HS penalty constructed from the second-order derivatives is defined as [31]:

$$R(u) = \|Hu\|_{1,p} = \sum_{i=1}^N \|Hu_i\|_{S_p} \quad \forall p \geq 1, \quad (4)$$

where

$$Hu_i = \begin{pmatrix} u_{xx}^i & u_{xy}^i & u_{xz}^i \\ u_{yx}^i & u_{yy}^i & u_{yz}^i \\ u_{zx}^i & u_{zy}^i & u_{zz}^i \end{pmatrix}, \quad (5)$$

H refers to the Hessian operator, N stands for the number of image pixels, $\|\cdot\|_{S_p}$ denotes the Schatten norm of order p (S_p norm) of a matrix, and $\|\cdot\|_{1,p}$ stands for the mixed $l_1 - S_p$ norm. The S_p norm of an $m \times n$ matrix $X \in \mathbb{R}^{m \times n}$ is defined as [41]:

$$\|X\|_{S_p} = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^p(X) \right)^{1/p}, \quad (6)$$

where $\sigma_i(X)$ denotes the i th singular value of X . Note that for $0 < p < 1$, $\|X\|_{S_p}$ is a quasinorm. Intuitively, when $p = 0$, the S_p norm measures the rank of the matrix X . The Schatten norm provides an approximation to the rank minimization problem.

After substituting the penalty term in the objective function (3) with the HS penalty defined in Eq.(4), we can rewrite the objective function (3) as:

$$\begin{aligned}\Phi(u) &= \frac{1}{2} (v - Au)^T \sum^{-1} (v - Au) + \tau \|Hu\|_{1,p} \\ &= \frac{1}{2} (v - Au)^T \sum^{-1} (v - Au) + \tau \sum_{i=1}^N \|Hu_i\|_{s_p}.\end{aligned}\quad (7)$$

In this study, we focused on three convex HS penalties with $p = 1, 2$, and ∞ .

B. Calculating the Projection Matrix A

To implement a forward projector more efficiently, we adopted the separable footprint (SF) projectors [42]. One of the SF projectors, SF-TR, approximates the voxel footprint functions as 2D separable functions and uses a trapezoid function in the transaxial direction and rectangular functions in the axial direction.

III. Optimizing the Objective Function

A. Reconstruction Process

Our aim in this study is to minimize the objective function (7), in which the HS penalty term is convex but non-smooth.

The first step of the optimization process was to use the MM approach to generate a sequence of upper-bounded functions (9) of the objective function (7). In step 2, we used a primal-dual method to convert the minimization of an upper-bounded function (primal problem) to the maximization of the dual objective function (21) (dual problem). Note that the dual objective function (21) is convex and smooth. In step 3, we solved the dual problem using the Nesterov algorithm [43]. In step 4, we obtained the minimizer of the upper-bounded function from the result of the dual problem through a projection operation (22). To speed up the convergence in each iteration of the MM method, we adopted MFISTA, an extended Nesterov algorithm, to handle a convex non-smooth minimization problem proposed by Amir Beck and Marc Teboulle [44]. Fig. 1 summarizes the optimization process.

B. MM Approach

The main idea of the majorization-minimization (MM) approach is to find the solution by minimizing a sequence of surrogate functions that upper-bound the original objective function (7) [45]. The original objective function is convex, and the solution obtained by minimizing these surrogate majorizer functions will converge to the minimization of the original objective function (7).

To construct the surrogate majorizers of the objective function (7), we defined a function that measures the distance between u and u_t :

$$d(u, u_t) = \frac{1}{2} (u - u_t)^T \left(\alpha I - A^T \sum^{-1} A \right) (u - u_t), \quad (8)$$

where t represents the number of iterations. Note that, when $\alpha = \|A^T \Sigma^{-1} A\|$, $\alpha I - A^T \Sigma^{-1} A$ is positive definite, and thus $\mathcal{D}(u, u_t)$ is non-negative. We obtained the upper-bounded functions (ignoring the constant part) by adding the distance function (8) to the objective function (7):

$$\begin{aligned} Q(u|u_t) &= \Phi(u) + d(u, u_t) = \frac{\alpha}{2} \|u - z\|_2^2 + \tau \|Hu\|_{1,p} \\ &= \frac{1}{2} \|u - z\|_2^2 + \tau \|Hu\|_{1,p} = \frac{1}{2} \|u - z\|_2^2 + \tau \sum_{i=1}^N \|Hu_i\|_{s_p}, \end{aligned} \quad (9)$$

where $z = u_t + \alpha^{-1} A^T \Sigma^{-1} (v - Au_t)$. In the iterative process of minimizing $Q(u|u_t)$, the value of the objective function $\Phi(u)$ kept monotonically non-increasing (Appendix A). If $\Phi(u)$ is convex, u_t will converge to the minimizer of $\Phi(u)$. We minimized the objective function (7) by iteratively minimizing the function (9) through the MM method to reconstruct the underlying image.

C. Primal-Dual Method

The fidelity term in (9) is continuously differentiable, while the penalty term is convex but still non-smooth. Because of the relation between the original norm and its dual norm [31], [46], we wrote the equivalent form of the penalty term in (9) as

$$\|Hu\|_{1,p} = \max_{\Omega \in B_{\infty,q}} \langle \Omega, Hu \rangle_{R^N \times 3 \times 3}, \quad (10)$$

where $1/p + 1/q = 1$, and $B_{\infty,q}$ denotes a unit-norm ball defined as

$$B_{\infty,q} = \left\{ \Omega = [\Omega_1^T, \Omega_2^T, \dots, \Omega_N^T]^T \in C^{N \times 3 \times 3} : \|\Omega_n\|_{s_q} \leq 1, \forall n=1, 2, \dots, N \right\}. \quad (11)$$

Let H^* represent the conjugate of the Hessian operator H , i.e.,

$$\langle \Omega, Hu \rangle_{R^N \times 3 \times 3} = \langle H^* \Omega, u \rangle_{R^N}. \quad (12)$$

Lefkimmiatis *et al.* deduced the numerical implementation of the conjugate operator H^* of the Hessian operator H in 2D [31]. In our study of 3D CBCT construction, where $\Omega \in R^{N \times 3 \times 3}$, the conjugate operator H^* was quite different than in 2D [31]. To obtain H^* in 3D, we rewrote (12) in the following form:

$$\sum_{i=1}^N \text{tr}(\Omega_i (Hu)_i) = \sum_{i=1}^N u_i (H^* \Omega)_i. \quad (13)$$

where $\text{tr}(\cdot)$ denotes the matrix trace. By expanding the left hand side of (13), we obtained

$$\begin{aligned} & \sum_{i=1}^N \text{tr}(\Omega_i (Hu)_i) \\ &= \sum_{i=1}^N u_i \left(\left(\Delta_{xx}^{(1,1)} \Omega_i^{(1,1)} \right)_i + \left(\Delta_{yy}^{(2,2)} \Omega_i^{(2,2)} \right)_i + \left(\Delta_{zz}^{(3,3)} \Omega_i^{(3,3)} \right)_i \right. \\ & \quad + \left(\Delta_{xy}^{(1,2)} \left(\Omega_i^{(1,2)} + \Omega_i^{(2,1)} \right) \right)_i + \left(\Delta_{xz}^{(1,3)} \left(\Omega_i^{(1,3)} + \Omega_i^{(3,1)} \right) \right)_i \\ & \quad \left. + \left(\Delta_{yz}^{(2,3)} \left(\Omega_i^{(2,3)} + \Omega_i^{(3,2)} \right) \right)_i \right). \end{aligned} \quad (14)$$

where Δ_{xx} , Δ_{yy} , Δ_{zz} , Δ_{xy} , Δ_{xz} and Δ_{yz} denote the forward difference operators. By comparing the right hand side of (14) with the right hand side of (13), we obtained the conjugate operator H^* of the Hessian operator H for 3D CBCT reconstruction:

$$\begin{aligned} (H^* \Omega)_i &= \left(\Delta_{xx}^* \Omega^{(1,1)} \right)_i + \left(\Delta_{yy}^* \Omega^{(2,2)} \right)_i \\ &+ \left(\Delta_{xy}^* \left(\Omega^{(1,2)} + \Omega^{(2,1)} \right) \right)_i \\ &+ \left(\Delta_{zz}^* \Omega^{(3,3)} \right)_i + \left(\Delta_{xz}^* \left(\Omega^{(1,3)} + \Omega^{(3,1)} \right) \right)_i \\ &+ \left(\Delta_{yz}^* \left(\Omega^{(2,3)} + \Omega^{(3,2)} \right) \right)_i, \end{aligned} \quad (15)$$

where Δ_{xx}^* , Δ_{yy}^* , Δ_{zz}^* , Δ_{xy}^* , Δ_{xz}^* and Δ_{yz}^* denote the backward difference operators.

We rewrote the penalty term in (9) as:

$$\|Hu\|_{1,p} = \max_{\Omega \in B_{\infty,q}} \langle \Omega, Hu \rangle_{R^N \times 3 \times 3} = \max_{\Omega \in B_{\infty,q}} \langle H^* \Omega, u \rangle_{R^N}. \quad (16)$$

Replacing the penalty term with its equivalent form (16), the objective function (9) can be written as:

$$\hat{u} = \arg \min_{u \in R^N} \frac{1}{2} \|u - z\|_2^2 + \tau \max_{\Omega \in B_{\infty,q}} \langle H^* \Omega, u \rangle. \quad (17)$$

Using (17), the reconstruction problem became the following min-max problem:

$$\min_{u \in R^N} \max_{\Omega \in B_{\infty, q}} \frac{1}{2} \|u - z\|_2^2 + \tau \langle H^* \Omega, u \rangle. \quad (18)$$

Let $f(u, \Omega) = \frac{1}{2} \|u - z\|_2^2 + \tau \langle H^* \Omega, u \rangle$. Note that $f(u, \Omega)$ is convex in u and concave in Ω , and the order of the minimum and maximum of (18) does not affect the solution [31]. We defined the primal objective function $s(u)$ as

$$s(u) = \max_{\Omega \in B_{\infty, q}} f(u, \Omega) = \frac{1}{2} \|u - z\|_2^2 + \tau \|Hu\|_{1,p}, \quad (19)$$

and the dual objective function as

$$g(\Omega) = \max_{u \in R^N} f(u, \Omega) = \min_{u \in R^N} \frac{1}{2} \|u - z\|_2^2 + \tau \langle H^* \Omega, u \rangle. \quad (20)$$

Letting the gradient $\frac{\partial}{\partial u} \left(\frac{1}{2} \|u - z\|_2^2 + \tau \langle H^* \Omega, u \rangle \right) = u - z + \tau H^* \Omega = 0$, we obtained $u = P_{R^N}(z - \tau H^* \Omega)$, where P_{R^N} was the projection operator onto the set R^N . It followed that

$$\begin{aligned} g(\Omega) &= \min_{u \in R^N} \frac{1}{2} \|u - (z - \tau H^* \Omega)\|_2^2 + \frac{1}{2} \|z\|_2^2 - \frac{1}{2} \|z - \tau H^* \Omega\|_2^2 \\ &= \frac{1}{2} \|P_{R^N}(z - \tau H^* \Omega) - (z - \tau H^* \Omega)\|_2^2 + \frac{1}{2} \|z\|_2^2 - \frac{1}{2} \|z - \tau H^* \Omega\|_2^2. \end{aligned} \quad (21)$$

We can calculate the optimal solution $\hat{u}$ of the primal objective function (19) or (9), as long as we have obtained the optimal solution $\hat{\Omega}$ of the dual objective function (21) using the relation of the saddle-value $(\hat{u}, \hat{\Omega})$:

$$\hat{u} = P_{R^N}(z - \tau H^* \hat{\Omega}). \quad (22)$$

The next step is to calculate the optimal solution $\hat{\Omega}$ by optimizing the dual objective function (21).

D. Nesterov Algorithm

Nesterov *et al.* introduced a method that achieves a greater convergence rate than the theoretical convergence rate of the classical gradient projection method [43]. We utilized the Nesterov algorithm [47] to optimize the dual objective function (21).

Let P_E denote the projection onto a set E . For the problem of $\min f(x)$ where $x \in E$, if ∇f is Lipschitz continuous with a Lipschitz constant L , the Nesterov algorithm can be summarized as follows [43]:

1. $x_t = P_E(y_{t-1} + \nabla f(y_{t-1})/L),$
2. $k_t = \left(1 + \sqrt{1 + 4k_{t-1}^2}\right) / 2,$
3. $y_t = x_{t-1} + ((k_{t-1} - 1)/k_t) (x_t - x_{t-1}).$

To utilize the Nesterov algorithm, we first proved that $\nabla g(\Omega)$ is Lipschitz continuous for 3D CBCT reconstruction in Lemma 1 (Appendix B). The dual objective function $g(\Omega)$ is smooth and differentiable in Ω . The gradient of $g(\Omega)$ is

$$\nabla g(\Omega) = \tau H P_{R^N} (z - \tau H^* \Omega). \quad (23)$$

Lemma 1—For 3D CBCT reconstruction, the dual objective function (21) is Lipschitz continuous, namely,

$$\|\nabla g(\Omega) - \nabla g(\tilde{\Omega})\| \leq \tau^2 \|H\|^2 \|\Omega - \tilde{\Omega}\|_{R^{N \times 3 \times 3}} \leq 144\tau^2 \quad (24)$$

To optimize (21) using the Nesterov algorithm, we used the reconstruction image generated by FDK as the initial value.

E. Orthogonal Projections

To construct the penalty term of the objective function, we selected the three commonly used Schatten norms: the Schatten nuclear norm ($p = 1$), the Schatten Frobenius norm ($p = 2$), and the Schatten spectral norm ($p = \infty$). To optimize the dual objective function (21) using the Nesterov algorithm, we needed to calculate the orthogonal projection of a matrix $X \in R^{3 \times 3}$ onto the set $B_{S_q} = \{X \in R^{3 \times 3} : \|X\|_{S_q} \leq 1\}$. (Note that $1/p + 1/q = 1$. Please see (11)). For a matrix X with singular value decomposition, $U[\text{diag}(\sigma(X))]V^T$, its orthogonal projection to B_{S_q} can be expressed as [31]:

$$P_{B_{S_q}}(X) = U \left[\text{diag} \left(P_{B_q}(\sigma(X)) \right) \right] V^H, \quad (25)$$

where $\sigma(X)$ denotes the singular value of the matrix X , $\text{diag}(\cdot)$ maps a vector to a diagonal matrix, and P_{B_q} is the orthogonal projection onto the l_q unit-norm ball

$B_q = \left\{ x \in R_+^{min(m,n)} : \|x\|_q \leq 1 \right\}$. According to (25), the orthogonal projection of a matrix in $R^{3 \times 3}$ onto the S_q norm ball converts to compute the orthogonal projection of its singulars onto the corresponding l_q unit-norm ball.

In the case of $q = 1$ (corresponding to the Schatten spectral norm ($p = \infty$)), for 3D CBCT image reconstruction, the orthogonal projection of a matrix $X \in \mathbb{R}^{3 \times 3}$ onto the B_{S_1} unit-norm ball is quite different than that in 2D [31]. In this study, we adopted an efficient algorithm [39] for the projection. The orthogonal projection for $q = 1$ in 3D can be given as (Appendix C):

$$P_{B_{S_1}}(\Omega_n) = U [\text{diag}(S_\gamma(\sigma(\Omega_n)))] V^H, \quad (26)$$

where $S_\gamma(x) = \max(x - \gamma, 0)$ with a threshold

$$\gamma = \begin{cases} 0, & \text{if } \sigma_1(\Omega_i) + \sigma_2(\Omega_i) + \sigma_3(\Omega_i) \leq 1, \\ \frac{(\sigma_1(\Omega_i) + \sigma_2(\Omega_i) + \sigma_3(\Omega_i) - 1)}{3}, & \text{if } (\sigma_1(\Omega_i) + \sigma_2(\Omega_i) + \sigma_3(\Omega_i)) > 1 \text{ and } \sigma_1(\Omega_i) + \sigma_2(\Omega_i) - 2\sigma_3(\Omega_i) < 1 \\ \frac{(\sigma_1(\Omega_i) + \sigma_2(\Omega_i) - 1)}{2}, & \text{if } (\sigma_1(\Omega_i) + \sigma_2(\Omega_i) + \sigma_3(\Omega_i)) > 1 \text{ and } \sigma_1(\Omega_i) + \sigma_2(\Omega_i) - 2\sigma_3(\Omega_i) \geq 1 \text{ and } \sigma_1(\Omega_i) - \sigma_2(\Omega_i) < 1 \\ \max(\sigma_1(\Omega_i), \sigma_2(\Omega_i), \sigma_3(\Omega_i)) - 1, & \text{if } (\sigma_1(\Omega_i) + \sigma_2(\Omega_i) + \sigma_3(\Omega_i)) > 1 \text{ and } \sigma_1(\Omega_i) + \sigma_2(\Omega_i) - 2\sigma_3(\Omega_i) \geq 1 \text{ and } \sigma_1(\Omega_i) - \sigma_2(\Omega_i) \geq 1 \end{cases} \quad (27)$$

In the case of $q = 2$ (corresponding to the Schatten Frobenius norm ($p = 2$)), finding the projection of a matrix $X \in \mathbb{R}^{3 \times 3}$ onto the B_{S_2} unit-norm ball requires finding a point in the ball that has the shortest distance to the matrix. If the matrix lies inside the B_{S_2} unit-norm ball, the projection is simply itself. Otherwise, the projection is the normalization of the matrix itself. Therefore, we had

$$P_{B_2}(\Omega_i) = \begin{cases} \Omega_i, & \text{if } \|\Omega_i\|_F \leq 1 \\ \frac{\Omega_i}{\|\Omega_i\|_F}, & \text{if } \|\Omega_i\|_F > 1 \end{cases}, \quad i = 1, \dots, N. \quad (28)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

In the case of $q = \infty$ (corresponding to the Schatten nuclear norm ($p = 1$)), the orthogonal projection of a matrix $X \in \mathbb{R}^{3 \times 3}$ onto the B_{S_∞} unit-norm ball is similar to the projection onto the B_{S_2} unit-norm ball. The projection can be divided into two cases: inside the B_{S_∞} unit-norm ball and outside the B_{S_∞} unit-norm ball. Different locations correspond to different projection results. Projection onto the B_{S_∞} unit-norm ball can be given as

$$P_{B_{S_\infty}}(\Omega_i) = U [\text{diag}(\min(\sigma(\Omega_i), 1))] V^H, \quad (29)$$

where $\sigma(\Omega_i)$ denotes the singular value of Ω_i , and the $\min(\cdot)$ operation compares each of the three singular values with 1 and takes the smaller one.

The pseudo-code of the proposed CBCT reconstruction algorithm using the HS penalties is shown in Appendix D.

IV. Materials and Evaluation

Experiments were conducted on two computer simulation phantoms and two physical phantoms. We used the peak signal to noise ratio (PSNR), the improvement signal to noise ratio (ISNR), the contrast-to-noise ratio (CNR), the structural similarity (SSIM), and the full width at half maximum (FWHM) to quantify the quality of the reconstructed images using different penalty terms [29].

A. Synthetic Phantom

We used two computer simulation phantoms in this study: a Compressed Sensing (CS) phantom [48] and a modified Shepp-Logan phantom. For the CS phantom, we used 360 projections, each containing 800×200 pixels, and the pixel size was $0.776 \times 0.776 \text{ mm}^2$. For the modified Shepp-Logan phantom, we used 360 projections, each containing 500×500 pixels, and the pixel size was $0.776 \times 0.776 \text{ mm}^2$. For both phantoms, the source-to-axial distance was 100 cm, and the source-to-detector distance was 150 cm. The incident photon number for the CS phantom was set to be 1×10^3 and 5×10^3 to simulate low-dose and high-dose protocols, respectively. The incident photon number for the Shepp-Logan phantom was set to be 1×10^4 , and the noise level was fixed. For both phantoms, the reconstructed image had $350 \times 350 \times 16$ pixels with a pixel size of $0.776 \times 0.776 \times 0.776 \text{ mm}^3$.

B. Physical Phantom

We used two physical phantoms in this study: a commercial calibration phantom CatPhan 600 (The Phantom Laboratory, Inc., Salem, NY) and an anthropomorphic head phantom. The projections were acquired by the ExactArms (kV source/detector arms) of a Trilogy treatment system (Varian Medical Systems, Palo Alto, CA). The number of projections for a full 360° rotation was 634. The dimensions of each acquired projection were $397 \times 298 \text{ mm}^2$, containing 1024×768 pixels. For each phantom, the X-ray tube current was set at 10 mA (low-dose) and 80 mA (high-dose) during projection data acquisitions. In both phantom studies, the tube voltage was set at 125 kVp, and the duration of the X-ray pulse at each projection scan was 10 ms. The projection data were acquired in full-fan mode with a bowtie filter. The source-to-axial distance was 100 cm, and the source-to-detector distance was 150 cm. The sizes of the reconstructed image were $350 \times 350 \times 16$ and $550 \times 550 \times 32$ for the CatPhan 600 and anthropomorphic head phantoms, respectively. The sizes of the reconstructed image were $0.776 \times 0.776 \times 0.776 \text{ mm}^3$ and $0.388 \times 0.388 \times 0.388 \text{ mm}^3$, respectively.

V. Experimental Results

A. Parameters Setting

CBCT reconstruction always involves a trade-off between image resolution and noise level. Most evaluation criteria used in this work were related to the noise level. To compare the

performance of different penalties, we adjusted the parameter τ in Eq. (3) for each phantom by trial and error so that the reconstruction results had the same noise level in each comparison group [49].

To determine the value of α in Eq.(9), we calculated the value of $\|A^T \Sigma^{-1} A\|$ for each tested phantom and found that setting α to be 2×10^{10} was enough to assure that $\alpha I - A^T \Sigma^{-1} A$ is positive definite for all phantoms in our experiment.

In all iterative reconstruction algorithms, the attenuation coefficient u was initialized using the FDK reconstruction u_0 .

B. Convergence Analysis

The range of the values of objective functions using different penalties can differ significantly, even when using exactly the same regularization parameter τ . To make these objective functions “comparable” in the same plot, we used a normalized objective cost to indicate the convergence of the reconstruction algorithms. The normalized objective cost was defined as the ratio of the objective function value of the reconstruction at the k th iteration and that of the initial reconstruction (using FDK), *i.e.* $\Phi(u_k)/\Phi(u_0)$.

To distinguish between different HS penalty terms, we referred to them as HS_p , with p denoting the order of the Schatten norm. HS_1 , HS_F and HS_∞ denoted Hessian nuclear ($p = 1$), Hessian Frobenius ($p = 2$), and Hessian spectral norms ($p = \infty$), respectively. Fig. 2 shows the convergence curves of reconstruction methods with TV, HS_∞ , HS_1 , and HS_F for the CatPhan 600 phantom. Note that the reconstruction method used for TV is the MM method proposed by Sun *et al.* [29], while the method used for the three HS penalties (HS_∞ , HS_1 , and HS_F) was the one developed in this study. For all penalties, good convergence was achieved after 30 iterations. Fig. 2 shows that all penalties had a similar or close convergence speed.

Note that there was a sub-optimization problem (the dual objective function (21). Please see also step 3 in Fig. 1) in the reconstruction process. This dual objective function is smooth and differentiable over Ω . To solve this sub-optimization problem, we selected the Nesterov algorithm [47], rather than the classical gradient projection method. Our experiments showed that, for this sub-optimization problem, the Nesterov algorithm can converge in about 20 iterations, while the classical gradient decent method needs at least 80 iterations to converge. This is consistent with the convergence analysis in the literature [43].

C. CS Phantom

To demonstrate the superiority of our approach over the TV penalty, we visually compared the reconstructed images using the proposed HS penalties with those using the TV penalty for the CS phantom in the low-dose protocol (with incident photon number 1×10^3). Fig. 3(a) displays a slice of the original CS phantom, in which the blue rectangle is a region of interest (ROI) over the octahedron with smoothly changed intensity, and the red rectangles were chosen to calculate the noise level (calculated as the average standard deviation (std) of these three rectangles). We also studied the resolution in the reconstructed image via FWHM along the yellow line (Fig. 3(a)). To calculate the FWHM value, the image gradient along the

yellow line was fitted using a Gaussian function, then the standard deviation of the fitted Gaussian function, δ , was used to calculate the FWHM as $2\sqrt{2\ln 2}\delta$ [29]. The FDK reconstruction result was very noisy, and details such as the small circles (light and dark grey ones in the upper-left part in the image) became hard to perceive (Fig. 3(b)). By comparison, the SIR reconstruction results using various penalties were less noisy and exhibited significantly better image quality (Fig. 3(c–f)).

Regions with smoothly changed intensity, such as the octahedron in the upper-right corner and the sphere at the lower-left corner, exhibited many small and unnatural constant intensity artifacts in the image reconstructed using the TV penalty, due to the staircase effect (Fig. 3(c)). The HS penalty family yielded more natural results (Fig. 3(d–f)), thus indicating its good ability to suppress the staircase effect. A zoom-in view of the ROI over the octahedron from different algorithms illustrates these differences even more clearly (Fig. 4).

We quantitatively compared the reconstruction quality in terms of PSNR, ISNR, SSIM, and FWHM for the proposed HS penalties together with the TV penalty and the FDK reconstruction (Table I). SSIM was used to measure the structural similarity between two images in a local region (an 11×11 window, in this study), with a higher value indicating greater similarity [29], [50]. As the window moved pixel-by-pixel over the reconstructed image and the reference image, we obtained an SSIM map. In practice, we used a single Mean-SSIM (MSSIM) value to evaluate the overall image quality by averaging the SSIM values over a pre-defined region. In Table I, we used original CS image as the reference image to calculate the SSIM. For both the low-dose and high-dose protocols, all reconstructions and the simulation of the digital CS phantom itself were repeated eight times, and the evaluation indexes were expressed as mean $\pm$ std.

Table I shows that the SIR algorithms using TV and the HS penalty family all had much better reconstruction quality than FDK, according to all evaluation indexes. In fact, the evaluation indexes of the SIR reconstruction algorithms at the low-dose protocol were better than the FDK at the high-dose protocol. At the low-dose protocol, the HS_1 penalty gave the best result, in terms of PSNR, ISNR, and SSIM. The HS_∞ and HS_F penalty had similar reconstruction quality as the TV penalty, all slightly lower than the HS_1 penalty. At the high-dose protocol, the HS_1 penalty had the best result in terms of PSNR, ISNR, and SSIM, with the HS_F penalty following behind. All three HS penalties produced better reconstruction quality than the TV penalty, in terms of PSNR, ISNR, and SSIM.

The TV penalty had the smallest FWHM value, indicating that it produced the reconstruction image with the sharpest edges (thus, the highest resolution). The FDK results at both protocols were too noisy to calculate the FWHM in Table I. Among the HS penalties, the HS_1 penalty had the smallest FWHM, indicating a better edge preserving ability. These numerical observations are consistent with the visual inspection of Fig. 3, where the HS penalty family tended to slightly blur image edges, while the TV penalty preserved edges better.

D. Shepp-Logan phantom

The reconstruction results using different penalties for the modified Shepp-Logan phantom also showed that the proposed HS penalties can avoid the staircase effect (Fig. 5). The blue rectangle in Fig. 5(a) denotes the ROI, which is magnified and displayed in the lower-right corner of Fig. 5(a); corresponding ROIs are magnified and displayed in the lower-right corners of Figs. 5(b–f). The FDK result looks very noisy, especially in the ROI (Fig. 5(b)), and the reconstructed result using the TV penalty shows an obvious staircase effect in the magnified region (Fig. 5(c)). The ROIs reconstructed by the HS penalties (Figs. 5(d–f)) had higher visual quality than those reconstructed by either FDK or TV.

The intensity curves along the yellow line in Fig. 5(a) show that the TV-based reconstruction (Fig. 6(c)) had several staircase artifacts, but the reconstructed results using the HS penalties (Figs. 6(d–f)) were much smoother. The profiles displayed in Fig. 6 confirmed the HS penalties' ability to avoid the staircase effect. To quantitatively analyze the penalty terms' performance in preserving edges during CBCT reconstruction, we studied the FWHM of the reconstructed images along the red line in Fig. 5(a). Different penalties exhibited clear differences in sharpness along the red line. The FWHM corresponding to TV, HS_{∞} , HS_1 , and HS_F was 2.03, 3.14, 3.48, and 3.49, respectively. The TV penalty has the smallest FWHM value, indicating that the TV penalty produced the reconstructed image with the sharpest edges, and the HS penalties produced reconstructed images with edges that were less sharp.

E. CatPhan 600 Phantom

The reconstructed CatPhan 600 images show that the low-dose protocol (10mA/10ms) results using TV (Fig. 7(c)), HS_{∞} (Fig. 7(d)), HS_1 (Fig. 7(e)), and HS_F (Fig. 7(f)) penalties had significantly better visual quality than the low-dose protocol FDK result (Fig. 7(b)). To quantitatively compare the penalties' abilities to suppress noise, the blue rectangle in Fig. 7(a) was chosen to calculate the noise level. The black arrows indicate the regions selected to calculate CNR values.

Table II lists CNRs of the four ROIs from different algorithms. The high-dose FDK result had a higher CNR value than the low-dose FDK result. The TV and HS reconstructions had higher CNR values than the high-dose and low-dose FDK results. The HS penalties and the TV penalty had comparable CNR values, indicating that the HS penalties and the TV penalty can comparably suppress noise in the reconstructed images. The HS_1 penalty ($p = 1$) had a slightly higher CNR values than the other two HS penalties in most cases.

F. Anthropomorphic Head Phantom

Fig. 8 displays the reconstructed anthropomorphic head phantom images. The black rectangle in Fig. 8(a) denotes the ROI selected to calculate the noise level. The blue rectangle represents the region we selected to calculate SSIM between the reconstructed image and the reference image. We used the high-dose FDK result as the reference image to see if the HS penalties could produce comparable results. The red rectangle in Fig. 8(a) is magnified and displayed in the lower-right corner of Fig. 8(a); corresponding regions are magnified and displayed in the lower-right corner of Figs. 8(b–f). Again, the TV result (Fig.

8(c)) has piecewise constant artifacts, while the HS penalty family results (Figs. 8(d–f)) look more natural but slightly blurred.

MSSIM values calculated between the targeted image and the reference image in the blue ROI are listed in Table III. We also show the SSIM maps in Fig. 9. The value of each pixel in the SSIM map measures the structural similarity between the targeted image and the reference image in the 11×11 window regions centered on each voxel.

The MSSIM values in the blue ROI corresponding to the low-dose FDK, TV, HS_{∞} , HS_1 , and HS_F are 0.74, 0.79, 0.86, 0.93, and 0.92, respectively (Table III). The three HS penalties had higher MSSIM values than the TV penalty, indicating that the HS penalties were better able to preserve local structures than TV. According to the MSSIM values, the HS_1 penalty had the best reconstruction result, while the HS_F penalty followed rather closely. The differences between different penalties' ability to preserve local structures are better demonstrated in Fig. 9, where the SSIM map resulting from the HS_1 penalty looks the whitest, indicating the best performance in preserving local structures over the selected ROI. The HS_F penalty showed the second best ability to preserve local structures.

VI. Discussion

The penalty term in SIR reflects the prior information of the underlying image. Different penalty terms can be constructed based on different prior information. The TV penalty has demonstrated its superiority in preserving edges in image processing and CBCT reconstruction. However, the TV penalty can sometimes lead to the staircase effect. Our experiments demonstrated the staircase effect in the CS phantom (Figs. 3–4(c)), the Shepp-Logan phantom (Figs. 5–6(c)), and the anthropomorphic head phantom (Fig. 8(c)). The over-sharpening of regions that are not piecewise constant may wipe out some important details in the underlying image, leading to a false interpretation. To overcome the staircase effect and improve the quality of the reconstructed image, we proposed in this study to use the HS penalty family, instead of the TV penalty, for CBCT reconstruction. Our experiments showed that the HS penalty family performed well in avoiding the staircase effect and preserving image regions with smooth intensity transitions (Figs. 3–6(e, f, g) and Fig. 8(e, f, g)). The MSSIM values and the SSIM maps for the anthropomorphic head phantom also demonstrated this (Fig. 9).

Among the HS penalties tested, the HS_1 penalty yielded the best reconstruction results in most cases, closely followed by the HS_F penalty. The improvement due to the HS_{∞} penalty is less pronounced, perhaps because the HS_{∞} penalty considers only the maximum absolute eigenvalue that satisfies the threshold criteria. In other words, the HS_{∞} penalty failed to include additional information that other eigenvalues might have provided.

The FWHM comparison in the Shepp-Logan phantom indicated that the HS penalty family does not preserve edges as well as the TV penalty. This seems to be a common limitation for the second-order penalties that has been observed in a previous study of the Hessian penalty [29]. Shi *et al.* combined TV and Hessian penalties using a local structure adaptive strategy for CBCT reconstruction [30], where the associated objective function was not convex

anymore and an alternative minimization method was used for the optimization process. A similar strategy can be used in the future to combine the HS penalty family with the TV penalty for 3D CBCT reconstruction to preserve both sharp edges and regions with smooth intensity transitions. However, the optimization method should be re-designed.

In this study, the performance of different penalties for CBCT reconstruction was evaluated under the same noise level by adjusting the regularization parameter [49]. Different strategies can be potentially used for the performance evaluation. For example, Han *et al.* evaluated the constrained TV minimization-based reconstruction with parameters selected to have the highest performance according to different utility metrics characterizing either reconstruction contrast or spatial resolution [20]. Thus the selected optimal parameters are strongly metric-dependent [20]. In this work, we tuned the parameters for different algorithms such that the reconstructed images were at the same noise level and then compared algorithm performance at the matched noise level [49]. Therefore, the parameters we used for TV reconstruction may not be optimal according to different metrics defined in [20]. We will compare the proposed penalties with TV by choosing metric-specific optimal parameters in a future work similar to that in [20]. As the parameters were not adjusted to be optimal according to certain specific evaluation metrics in our experiments, undesired artifacts existed in the reconstructed images in some cases. For example, the HS_1 slice (CS phantom) in Fig. 4(d) has localized diagonal stripe high-frequency artifacts at its lower middle portion. These artifacts can be removed using a larger regularization parameter with a stronger noise suppressing ability but a lower spatial resolution. For the four small dots in Fig. 7, the HS_∞ (Fig. 7(d)) and HS_1 (Fig. 7(e)) penalties tended to reconstruct them to faint line segments. This could be related to the anisotropic spatial resolution property of the used penalties. Similar trend was also observed in the image reconstructed by the TV penalty (Fig. 7(c)). These dots can be better restored by reconstructing images with a higher resolution through adjusting the regularization parameter.

VII. Conclusion

In this study, we proposed to use the HS penalty family for 3D CBCT reconstruction, and we developed an effective algorithm to minimize the objective function using the majorization-minimization approach followed by a primal-dual formulation. Both simulated and physical experiments showed that the HS penalty family can effectively reconstruct the underlying three-dimensional image and preserve regions with smooth intensity transitions without introducing the staircase effect often induced by the TV penalty. Among the three HS penalties ($p = 1$, $p = 2$, and $p = \infty$), the one with $p = 1$ performed the best in most cases, followed by the HS penalty with $p = 2$. However, since it consists of the second-order derivatives, the HS penalty family tends to slightly blur sharp edges in the reconstruction results. A possible remedy is to combine the HS penalty family with the TV penalty in the future.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NNSFC), under Grant Nos. 61375018 and 61672253. J. Wang was supported in part by grants from the Cancer Prevention and Research

Institute of Texas (RP130109 and RP110562-P2), the National Institute of Biomedical Imaging and Bioengineering (R01 EB020366), and a grant from the American Cancer Society (RSG-13-326-01-CCE).

We would like to thank Dr. Jonathan Feinberg for editing the paper.

References

1. Islam MK, Purdie TG, Norrlinger BD, Alasti H, Moseley DJ, Sharpe MB, et al. Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy. *Medical physics*. 2006; 33:1573–1582. [PubMed: 16872065]
2. Wen N, Guan H, Hammoud R, Pradhan D, Nurusev T, Li S, et al. Dose delivered from Varian's CBCT to patients receiving IMRT for prostate cancer. *Physics in medicine and biology*. 2007; 52:2267. [PubMed: 17404468]
3. Lee L, Le QT, Xing L. Retrospective IMRT dose reconstruction based on cone-beam CT and MLC log-file. *International Journal of Radiation Oncology* Biology* Physics*. 2008; 70:634–644.
4. Hsieh J. Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise. *Medical Physics*. 1998; 25:2139–2147. [PubMed: 9829238]
5. Kachelriess M, Watzke O, Kalender WA. Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. *Medical physics*. 2001; 28:475–490. [PubMed: 11339744]
6. Li T, Li X, Wang J, Wen J, Lu H, Hsieh J, et al. Nonlinear sinogram smoothing for low-dose X-ray CT. *Nuclear Science, IEEE Transactions on*. 2004; 51:2505–2513.
7. Wang J, Lu H, Liang Z, Eremina D, Zhang G, Wang S, et al. An experimental study on the noise properties of x-ray CT sinogram data in Radon space. *Physics in medicine and biology*. 2008; 53:3327. [PubMed: 18523346]
8. Fessler J. Penalized weighted least-squares image reconstruction for positron emission tomography. *Medical Imaging, IEEE Transactions on*. 1994; 13:290–300.
9. Chlewicki W, Hermansen F, Hansen S. Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior. *Physics in Medicine and Biology*. 2004; 49:4717. [PubMed: 15566170]
10. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*. 1992; 60:259–268.
11. Sidky EY, Kao C-M, Pan X. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. 2009arXiv preprint arXiv:0904.4495
12. Jørgensen JH, Sidky EY, Pan X. Analysis of discrete-to-discrete imaging models for iterative tomographic image reconstruction and compressive sensing. *IEEE Trans Med Imag*. 2011
13. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in medicine and biology*. 2008; 53:4777. [PubMed: 18701771]
14. Hashemi S, Song WY, Sahgal A, Lee Y, Huynh C, Grouza V, et al. Simultaneous deblurring and iterative reconstruction of CBCT for image guided brain radiosurgery. *Physics in Medicine and Biology*. 2017; 62:2521. [PubMed: 28248652]
15. Huang HM, Hsiao T. Accelerating an Ordered-Subset Low-Dose X-Ray Cone Beam Computed Tomography Image Reconstruction with a Power Factor and Total Variation Minimization. *PloS one*. 2016; 11:e0153421. [PubMed: 27073853]
16. Tang J, Nett BE, Chen GH. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms. *Physics in medicine and biology*. 2009; 54:5781. [PubMed: 19741274]
17. Jia X, Lou Y, Lewis J, Li R, Gu X, Men C, et al. GPU-based fast low-dose cone beam CT reconstruction via total variation. *Journal of X-ray Science and Technology*. 2011; 19:139–154. [PubMed: 21606579]
18. Han X, Bian J, Eaker DR, Kline TL, Sidky EY, Ritman EL, et al. Algorithm-enabled low-dose micro-CT imaging. *IEEE Trans Med Imaging*. Mar.2011 30:606–20. [PubMed: 20977983]

19. Ritschl L, Bergner F, Fleischmann C, Kachelrieß M. Improved total variation-based CT image reconstruction applied to clinical data. *Physics in medicine and biology*. 2011; 56:1545. [PubMed: 21325707]
20. Han X, Pearson E, Pelizzari C, Al-Hallaq H, Sidky EY, Bian JG, et al. Algorithm-enabled exploration of image-quality potential of cone-beam CT in image-guided radiation therapy. *Physics in Medicine and Biology*. Jun 21. 2015 60:4601–4633. [PubMed: 26020490]
21. Bian J, Sharp GC, Park Y-K, Ouyang J, Bortfeld T, El Fakhri G. Investigation of cone-beam CT image quality trade-off for image-guided radiation therapy. *Physics in medicine and biology*. 2016; 61:3317. [PubMed: 27032676]
22. Gao H. Fused analytical and iterative reconstruction (AIR) via modified proximal forward-backward splitting: a FDK-based iterative image reconstruction example for CBCT. *Physics in Medicine and Biology*. Oct 7. 2016 61:7187–7204. [PubMed: 27649259]
23. Zhang Y, Tehrani JN, Wang J. A Biomechanical Modeling Guided CBCT Estimation Technique. *IEEE Transactions on Medical Imaging*. 2017; 36:641–652. [PubMed: 27831866]
24. Chen Y, Shi L, Feng Q, Yang J, Shu H, Luo L, et al. Artifact Suppressed Dictionary Learning for Low-Dose CT Image Processing. *IEEE Transactions on Medical Imaging*. 2014; 33:2271–2292. [PubMed: 25029378]
25. Chan T, Marquina A, Mulet P. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*. 2000; 22:503–516.
26. Yan L, Jianhua M, Yi F, Zhengrong L. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in Medicine & Biology*. 2012; 57:7923. [PubMed: 23154621]
27. Lefkimmiatis S., Bourquard, A., Unser, M. Hessian-based regularization for 3-D microscopy image restoration. *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*; 2012; p. 1731-1734.
28. Arigovindan M, Fung JC, Elnatan D, Mennella V, Chan Y-HM, Pollard M, et al. High-resolution restoration of 3D structures from widefield images with extreme low signal-to-noise-ratio. *Proceedings of the National Academy of Sciences*. 2013; 110:17344–17349.
29. Sun T, Sun N, Wang J, Tan S. Iterative CBCT reconstruction using Hessian penalty. *Physics in medicine and biology*. 2015; 60:1965. [PubMed: 25674998]
30. Shi Q, Sun N, Sun T, Wang J, Tan S. Structure-adaptive CBCT reconstruction using weighted total variation and Hessian penalties. *Biomedical Optics Express*. 2016; 7:3299–3322. [PubMed: 27699100]
31. Lefkimmiatis S, Ward JP, Unser M. Hessian Schatten-Norm Regularization for Linear Inverse Problems. *IEEE Transactions on Image Processing*. 2013; 22:1873–1888. [PubMed: 23303692]
32. Hu Y, Ongie G, Ramani S, Jacob M. Generalized higher degree total variation (HDTV) regularization. *IEEE Transactions on Image Processing*. 2014; 23:2423–2435. [PubMed: 24710832]
33. Candès EJ, Wakin MB. An introduction to compressive sampling. *IEEE signal processing magazine*. 2008; 25:21–30.
34. Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*. 2008; 14:877–905.
35. Nie F, Huang H, Ding CH. Low-Rank Matrix Recovery via Efficient Schatten p -Norm Minimization. *AAAI*. 2012
36. Xie Y, Qu Y, Tao D, Wu W, Yuan Q, Zhang W. Hyperspectral Image Restoration via Iteratively Regularized Weighted Schatten-Norm Minimization. *IEEE Transactions on Geoscience and Remote Sensing*. 2016; 54:4642–4659.
37. Kong, D., Zhang, M., Ding, C. Minimal shrinkage for noisy data recovery using Schatten- p norm objective. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; 2013; p. 177-193.
38. Li X, Wang J, Tan S. Hessian Schatten-norm regularization for CBCT image reconstruction using fast iterative shrinkage-thresholding algorithm. *SPIE Medical Imaging*. 2015

39. Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T. Efficient projections onto the l_1 -ball for learning in high dimensions. Proceedings of the 25th international conference on Machine learning; 2008; p. 272-279.
40. Wang J, Li T, Xing L. Iterative image reconstruction for CBCT using edge-preserving prior. Medical physics. 2009; 36:252–260. [PubMed: 19235393]
41. Horn, RA., Johnson, CR. Matrix analysis. Cambridge university press; 2012.
42. Long Y, Fessler JA, Balter JM. 3D forward and back-projection for X-ray CT using separable footprints. Medical Imaging, IEEE Transactions on. 2010; 29:1839–1850.
43. Nesterov Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Soviet Mathematics Doklady. 1983:372–376.
44. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences. 2009; 2:183–202.
45. Figueiredo MA, Bioucas-Dias JM, Nowak RD. Majorization–minimization algorithms for wavelet-based image restoration. Image Processing, IEEE Transactions on. 2007; 16:2980–2991.
46. Sra, S. Machine learning and knowledge discovery in databases. Springer; 2011. Fast projections onto l_1 , q -norm balls for grouped feature selection; p. 305-317.
47. Rockafellar, RT. Convex analysis. Princeton University Press; 1970.
48. Smith, D. Nov 11. 2010 “<http://www.mathworks.com/matlabcentral/fileexchange/29364-compressed-sensing-mri-phantom--v1-1->”
49. Wang J, Li T, Xing L. Iterative image reconstruction for CBCT using edge-preserving prior. Medical physics. 2008; 36:252–260.
50. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on. 2004; 13:600–612.

Appendix

A. Monotonicity of the MM Approach

It follows from Eq. (9) that

$$Q(u|u_t) \geq \Phi(u), \quad \forall u,$$

and

$$Q(u_t|u_t) = \Phi(u_t).$$

Let u_{t+1} denote the minimizer of $Q(u|u_t)$, i.e., $u_{t+1} = \arg \min_u Q(u|u_t)$. We had

$$\Phi(u_{t+1}) \leq Q(u_{t+1}|u_t) \leq Q(u_t|u_t) = \Phi(u_t).$$

In other words, the value of the objective function $\Phi(u)$ kept monotonically non-increasing in the MM iterative process.

B. Proof of Lemma 1

For any pair of variables $\Omega, \tilde{\Omega} \in \mathbb{R}^{N \times 3 \times 3}$, we have

$$\begin{aligned}
& \|\nabla g(\Omega) - \nabla g(\tilde{\Omega})\| = \|\tau H \{P_{R^N}(z - \tau H^* \Omega) - P_{R^N}(z - \tau H^* \tilde{\Omega})\}\|_{R^N \times 3 \times 3} \\
& \leq \|\tau H P_{R^N}(\tau H^* \Omega - \tau H^* \tilde{\Omega})\|_{R^N \times 3 \times 3} \\
& \leq \tau^2 \|H\| \|H^*\| \|\Omega - \tilde{\Omega}\|_{R^N \times 3 \times 3} \\
& = \tau^2 \|H\|^2 \|\Omega - \tilde{\Omega}\|_{R^N \times 3 \times 3}.
\end{aligned}$$

The Lipschitz constant of $\nabla g(\Omega)$ can be $\tau^2 \|H\|^2$ or larger than $\tau^2 \|H\|^2$. To compute an upper bound of $\|H\|$, we utilized the relation that $\|H\|^2 = \|H^* H\|$:

$$\|H^* H u\|_{R^N} = \left\| \left(\Delta_{xx}^* \Delta_{xx} + \Delta_{yy}^* \Delta_{yy} + \Delta_{zz}^* \Delta_{zz} + 2\Delta_{xy}^* \Delta_{xy} + 2\Delta_{xz}^* \Delta_{xz} + \Delta_{yz}^* \Delta_{yz} \right) u \right\|_{R^N} \leq \left(\|\Delta_{xx}\|^2 + \|\Delta_{yy}\|^2 + \|\Delta_{zz}\|^2 + 2\|\Delta_{xy}\|^2 + 2\|\Delta_{xz}\|^2 + 2\|\Delta_{yz}\|^2 \right) \|u\|_{R^N}$$

Without any loss of generalization, we assumed that all pixels of the CBCT image u had been normalized to the range of $[0, 1]$. From the definition of the second-order differential operator, each of $\|\Delta_{xx}\|$, $\|\Delta_{yy}\|$, $\|\Delta_{zz}\|$, $\|\Delta_{xy}\|$, $\|\Delta_{xz}\|$, and $\|\Delta_{yz}\|$ is equal to or smaller than 4. Therefore, an upper bound of the Lipschitz constant of $\nabla g(\Omega)$ can be $L = \tau^2 \|H\|^2 = 144 \tau^2$ for 3D CBCT reconstruction.

C. Projection of a Matrix onto the B1 Unit-norm Ball in 3D

Let σ denote the singular value vector and w denote the projection. The projection task can be formally described as the following optimization problem [39]:

$$\min_w \frac{1}{2} \|w - \sigma\|_2^2 \quad s.t. \sum_{i=1}^3 w_i = 1, w_i \geq 0.$$

Using the Lagrangian of the above problem and KKT conditions, we can get $w_i = \max\{\sigma_i -$

$\theta, 0\}$, where $\theta = \frac{1}{M} \left(\sum_{i=1}^M z_i - 1 \right)$ and $M = \max \left\{ z_j - \frac{1}{j} \left(\sum_{r=1}^j z_r - 1 \right) > 0; j=1, 2, 3 \right\}$, with z denoting the vector obtained by sorting σ in a descending order. The orthogonal projection of a matrix onto the B_{s_1} unit-norm ball can be calculated using the operator $\mathcal{S}_{\gamma}(\sigma(\Omega_i)) = \max(\sigma(\Omega_i) - \gamma, 0)$, with the threshold γ defined in Eq. (27).

D. Reconstruction Algorithm Using the Hessian Schatten Penalty

Our task was to estimate the attenuation coefficient u from the projection data v by minimizing the cost function $\Phi(u)$ in Eq. (7). We used the MM approach (Step 1 in Fig. 1), followed by a primal-dual formulation (Step 2 in Fig. 1), where the dual objective function was optimized by the Nesterov algorithm (Step 3 in Fig. 1). In our experiments, u was initialized using the FDK reconstruction $FDK\{v\}$. Let T and Ne be the maximum iteration numbers for the MM approach and the Nesterov algorithm, respectively. In our experiment, we set $K = 100$ and $Ne = 20$. T_1 and T_2 are two small real numbers, and L is the Lipschitz constant (Please see Lemma 1).

Initialization: $u_0 = FDK(v)$, $k_0 = 1$, $k'_0 = 1$, $c_0 = \Phi(u_0)$, $\tilde{\Omega}_0$, and Ω_0 are both all 0's matrices in $R^{N \times 3 \times 3}$.

Iteration:

For $t = 1: T$ (**MM approach**)

$$z = x_{t-1} + \alpha^{-1} A^T \Sigma^{-1} (v - Ax_{t-1})$$

For $n = 1: Ne$ (**Nesterov algorithm**)

$$\Omega_n = P_{B_{sq}}(\tilde{\Omega}_{n-1} + \tau H P_{R^N}(z - \tau H^* \tilde{\Omega}_{n-1})/L),$$

$$k'_n = \left(1 + \sqrt{1 + 4k'^2_{n-1}}\right) / 2,$$

$$\tilde{\Omega}_n = \Omega_{n-1} + \left(\left(k'_{n-1} - 1\right) / k'_n\right) (\Omega_n - \Omega_{n-1}),$$

If $\|\Omega_n - \Omega_{n-1}\|_2 / \|\Omega_{n-1}\|_2 \leq T_1$, Break, End

End

$$s_t = P_{R^N}(z - \tau H^* \tilde{\Omega}_{n-1}), \text{ (Note: Step 4 in Fig. 1)}$$

$$k_t = \left(1 + \sqrt{1 + 4k'^2_{t-1}}\right) / 2,$$

$$c_t = \Phi(s_t),$$

If $c_t > c_{t-1}$, then $u_t = u_{t-1}$, else $u_t = s_{t-1}$, End

$$x_t = u_t + (k_{t-1}/k_t)(s_t - u_t) + (k_{t-1} - 1/k_t)(u_t - u_{t-1}),$$

If $\|u_t - u_{t-1}\|_2 / \|u_{t-1}\|_2 \leq T_2$, Break, End

End

Return $\hat{u} = u_t$

Fig. 1.
Main procedure to optimize the objective function.

Fig. 2. Convergence curves of the reconstruction algorithms on the CatPhan 600 phantom using TV, HS_1 , HS_F and HS_∞ . 'Ratio' is the normalized objective function value.

Fig. 3.

A representative slice of (a) the original CS phantom; and reconstructed images from different reconstructed algorithms with incident photon number 1×10^3 (low-dose protocol): (b) FDK; (c) TV; (d) HS_∞ ; (e) HS_1 ; (f) HS_F .

Fig. 4.

Zoom-in visual inspections of ROIs of images reconstructed by different algorithms with incident photon number 1×10^{-3} (low-dose protocol): (a) the original CS phantom; (b) FDK; (c) TV; (d) HS_I ; (e) HS_2 ; (f) HS_∞ .

Fig. 5. Representative slices from the reconstructed volume using different penalty terms: (a) Original image; (b) FDK; (c) TV; (d) HS_∞ ; (e) HS_1 ; (f) HS_F .

Fig. 6.

Profiles through the ellipsoid object (the yellow line in Fig. 5(a)): (a) Original image; (b) FDK; (c) TV; (d) HS_{∞} ; (e) HS_1 ; and (f) HS_F .

Fig. 7.

Representative slices from the reconstructed CatPhan 600 images. (a) FDK with a high-dose protocol (80mA/10ms); Low-dose protocol (10mA/10ms): (b) FDK, (c) TV, (d) HS_{∞} , (e) HS_1 , (f) HS_F

Fig. 8.

Representative slices of the reconstructed anthropomorphic head phantom image using different penalties: (a) FDK high-dose protocol (80mA/10ms); Low-dose protocol(10mA/10ms); (b) FDK; (c) TV; (d) HS_∞ ; (e) HS_1 ; and (f) HS_F .

Fig. 9.

SSIM maps from the low dose reconstructed image over the reference image (FDK reconstructed image using the high dose protocol): (a) FDK (MSSIM = 0.74); (b) TV (MSSIM = 0.79); (c) HS_∞ (MSSIM = 0.86); (d) HS_1 (MSSIM = 0.93); (e) HS_F (MSSIM = 0.92).

TABLE I
Comparisons Between Images From Different Reconstruction Algorithms for the CS Phantom

Incident photon	Penalty	Noise level ($\times 10^{-4}$)	PSNR	ISNR	MSSIM	FWHM
1×10^3	FDK	27.00 $\pm$ 0.00	18.81 $\pm$ 0.09	0.00 $\pm$ 0.00	0.12 $\pm$ 0.00	N/A
	TV	3.32 $\pm$ 0.03	32.63 $\pm$ 0.18	8.76 $\pm$ 0.01	0.90 $\pm$ 0.00	1.82$\pm$0.05
	HS $_{\infty}$	3.33 $\pm$ 0.00	31.94 $\pm$ 0.01	8.86 $\pm$ 0.00	0.90 $\pm$ 0.00	3.75 $\pm$ 0.17
	HS $_1$	3.27 $\pm$ 0.01	33.75$\pm$0.16	9.53$\pm$0.01	0.93$\pm$0.00	2.22 $\pm$ 0.12
	HS $_F$	3.34 $\pm$ 0.02	32.14 $\pm$ 0.08	8.87 $\pm$ 0.01	0.92 $\pm$ 0.00	3.69 $\pm$ 0.04
5×10^3	FDK	12.00 $\pm$ 0.00	23.26 $\pm$ 0.06	0.00 $\pm$ 0.00	0.38 $\pm$ 0.00	N/A
	TV	1.68 $\pm$ 0.01	36.11 $\pm$ 0.06	7.47 $\pm$ 0.01	0.95 $\pm$ 0.00	1.43$\pm$0.01
	HS $_{\infty}$	1.78 $\pm$ 0.00	37.90 $\pm$ 0.02	8.58 $\pm$ 0.00	0.98 $\pm$ 0.00	2.57 $\pm$ 0.10
	HS $_1$	1.76 $\pm$ 0.00	40.12$\pm$0.11	9.52$\pm$0.02	0.99$\pm$0.00	2.12 $\pm$ 0.02
	HS $_F$	1.77 $\pm$ 0.00	37.89 $\pm$ 0.11	8.49 $\pm$ 0.02	0.98 $\pm$ 0.00	3.04 $\pm$ 0.01

TABLE II

CNRs of Different ROIs of the CatPhan 600 Phantom

	Noise($\times 10^{-4}$)	ROI1	ROI2	ROI3	ROI4
FDK_10mA	35	0.29	0.37	0.45	0.64
FDK_80mA	13	0.58	0.83	0.94	1.03
TV	7.60	0.61	1.16	1.21	1.27
HS _∞	8.25	0.66	1.17	1.14	1.34
HS ₁	7.86	0.68	1.29	1.26	1.42
HS _F	7.87	0.69	1.27	1.20	1.40

TABLE III

MSSIM of the Blue ROI Shown in Fig. 7(a)

Penalty	Noise level($\times 10^{-4}$)	MSSIM
FDK_10mA	17	0.74
FDK_80mA	8.54	N/A
TV	7.49	0.79
HS_{∞}	7.60	0.86
HS_1	7.68	0.93
HS_F	7.57	0.92