Title

Edge-Preserving PET Image Reconstruction Using Trust Optimization Transfer

Journal

IEEE Transactions on Medical Imaging, 34(4)

ISSN

0278-0062

Authors

Wang, Guobao
Qi, Jinyi

Publication Date

2015-04-01

DOI

10.1109/tmi.2014.2371392

Peer reviewed

HHS Public Access

Author manuscript

IEEE Trans Med Imaging. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

IEEE Trans Med Imaging. 2015 April ; 34(4): 930–939. doi:10.1109/TMI.2014.2371392.

Edge-Preserving PET Image Reconstruction Using Trust Optimization Transfer

Guobao Wang and **Jinyi Qi**

Department of Biomedical Engineering, University of California, Davis, CA, USA

Abstract

Iterative image reconstruction for positron emission tomography (PET) can improve image quality by using spatial regularization. The most commonly used quadratic penalty often over-smoothes sharp edges and fine features in reconstructed images, while non-quadratic penalties can preserve edges and achieve higher contrast recovery. Existing optimization algorithms such as the expectation maximization (EM) and preconditioned conjugate gradient (PCG) algorithms work well for the quadratic penalty, but are less efficient for high-curvature or non-smooth edge-preserving regularizations. This paper proposes a new algorithm to accelerate edge-preserving image reconstruction by using two strategies: trust surrogate and optimization transfer descent. Trust surrogate approximates the original penalty by a smoother function at each iteration, but guarantees the algorithm to descend monotonically; Optimization transfer descent accelerates a conventional optimization transfer algorithm by using conjugate gradient and line search. Results of computer simulations and real 3D data show that the proposed algorithm converges much faster than the conventional EM and PCG for smooth edge-preserving regularization and can also be more efficient than the current state-of-art algorithms for the non-smooth ℓ_1 regularization.

Keywords

PET; image reconstruction; edge-preserving regularization; optimization transfer; optimization algorithm

I. INTRODUCTION

Iterative image reconstruction methods can accurately model the system response and noise statistics in positron emission tomography (PET). They have been increasingly used to improve image quality. The maximum likelihood (ML) method [1], [2] reconstructs an image from projections by maximizing the log likelihood of PET data and can be elegantly solved by the expectation maximization (EM) algorithm [3], [4]. However, a true maximum likelihood solution can be very noisy. Common ways to stabilize the image estimation are either terminating the iteration before convergence or using a penalty function to encourage spatially smooth images [5]. Both early termination of the EM algorithm and using the

quadratic penalty function tend to over-smooth sharp edges and fine features in reconstructed images. In comparison, penalized likelihood (PL) methods using a non-quadratic regularization can preserve edges and achieve higher contrast recovery for small targets [9], [10], [33].

Existing optimization transfer (OT) algorithms, either based on the EM algorithm [3], [11], [12] or the separable quadratic surrogates [13], [14], work well for PL reconstruction with quadratic regularization. However, when applied to a high-curvature nonquadratic penalty function, these algorithms can be very inefficient (as shown in the simulation studies in Section V) and they are not applicable to non-smooth regularization because the penalty function is not differentiable at zero. While the conventional preconditioned conjugate gradient (PCG) algorithm [15], [16] can be faster than the OT algorithms, it still suffers the same problem of slow convergence because the preconditioners are often borrowed from the EM algorithm and contain no information of the regularization. Incorporating the regularization information and using a non-diagonal preconditioner can further accelerate the PCG algorithm [17]. Iterative coordinate descent (ICD) algorithms can achieve fast convergence but usually present challenges to parallel implementation [18]-[20].

Recently, there has been growing interests in developing algorithms for non-smooth regularization due to the emerging area of compressive sensing. The alternating direction method of multipliers (ADMM) has been developed for Poisson image deconvolution under different names (PIDAL [22] and PIDSplit [23]) and has been applied to PET and SPECT image reconstruction [24], [25]. ADMM-type algorithms can be very fast, but are not guaranteed to converge monotonically. The ADMM for Poisson data involves Lagrangian penalty parameters that have to be tuned for fast convergence, which is a nontrivial task. The Chambolle-Pock (CP) algorithm [26]-[28], a kind of primal-dual optimization method, can handle non-smooth penalties and other constraints very easily. However, the CP algorithm can be slow when applied to spatially variant inverse problems and has no guarantee on monotonic convergence. The SPIRAL [29] is another new algorithm that utilizes the fast iterative shrinkage and thresholding algorithm (FISTA). It is fast in its non-monotonic implementation, but can be slow if monotonicity is enforced.

In this paper, we develop a new trust optimization transfer algorithm that achieves fast monotonic convergence for edge-preserving PET image reconstruction. We first introduce the optimization transfer descent (OTD) concept by exploring the descent nature of the original optimization transfer [21] for minimization problems. Conjugate gradient and line search [30, p.120] are then incorporated in the OTD for acceleration. The OTD can be viewed as a PCG algorithm with an implicitly defined preconditioner which contains information of both the likelihood term and regularization term. It is therefore expected to achieve faster convergence than the conventional PCG algorithm [15], [16]. To extend the OTD algorithm to high-curvature and non-smooth penalties, we adopt the trust surrogate concept that has been used in the Levenberg-Marquardt algorithm [34] and trust region method [30, p.65, p.262] for nonlinear optimization. We approximate the original objective function by a smooth surrogate and solve the smooth surrogate by the OTD algorithm. The resulting trust optimization transfer algorithm can solve edge-preserving image

reconstruction very efficiently. It is nearly free of parameter tuning and guarantees to descend monotonically.

This paper is organized as follows. We briefly introduce the penalized likelihood PET image reconstruction in Section II. We then present the optimization transfer descent method in Section III and the trust optimization transfer in Section IV. In Section V, we perform simulation studies and compare the proposed algorithm with other state-of-art algorithms. An example application of the proposed algorithm to real 3D PET data is given in Section VI. Conclusions are drawn in Section VII.

II. PENALIZED LIKELIHOOD PET RECONSTRUCTION

PET data $\boldsymbol{y} = \{y_i\}$ can be well modeled as a collection of independent Poisson random variables with the log likelihood function as

$$L(y|\ \bar{y}\ (x))=\sum_{i=1}^{n_i} y_i \log \bar{y}_i(x) - \bar{y}_i(x) \quad (1)$$

where the expected data $\bar{y}\ (x)$ is related to the unknown image $\boldsymbol{x}$ through an affine transform [5]

$$y-(x)=Px+r \quad (2)$$

where $P=\{p_{ij}\}\epsilon IR^{n_i \times n_j}$ is the system matrix with p_{ij} denoting the probability of detecting an event originated at pixel j by detector pair i, $\boldsymbol{r}$ accounts for background events such as randoms and scatters. n_i is the total number of detector pairs and n_j is the total number of pixels in image.

Penalized likelihood (PL) reconstruction (or equivalently maximum *a posteriori*, MAP [6]-[8]) estimates the unknown image by minimizing a penalized negative likelihood function

$$\hat{x}=\text{argmin}_{x\geq 0}\phi(x), \phi(x)=-L(y|\ \bar{y}\ (x))+\beta U(x) \quad (3)$$

where $U(\boldsymbol{x})$ is an image roughness penalty that can be measured based on the intensity difference between neighboring pixels [5], [7],

$$U(x)=\sum_{j=1}^{n_j}\sum\nolimits_{k\epsilon N_j} w_{jk}\,\Psi_\delta(x_j-x_k) \quad (4)$$

where $\psi_\delta(t)$ is the penalty function and δ is a parameter that controls the smoothness of the penalty function, w_{jk} is the weighting factor related to the distance between pixel j and pixel k in the neighborhood N_j, and β controls the strength of the regularization.

In contrast to the quadratic penalty function $\psi_\delta(t) = t^2/2$ which may oversmooth edges, the non-smooth ℓ_1 (absolute value function)

$$\psi_\delta(t)=|t| \quad (5)$$

can preserve edges, but it is not differentiable at zero. Smooth functions that approximate the ℓ_1 function but are differentiable at zero have been proposed. Here we refer to them as smooth ℓ_1 functions. One example is the Fair function [11]

$$\psi_\delta(t)=\delta\left(\frac{|t|}{\delta}-\log\left(1+\frac{|t|}{\delta}\right)\right) \quad (6)$$

which has a continuous second-order derivative. Other examples include the hyperbola function $\sqrt{t^2+\delta^2}$ and the Huber function [31]. A common feature of the smooth ℓ_1 functions is that they approximate the quadratic function when $|t| \ll \delta$ and approach the nonsmooth ℓ_1 for $|t| \gg \delta$.

III. OPTIMIZATION TRANSFER DESCENT

A. Optimization Transfer

Optimization transfer (OT) algorithms (a.k.a. majorization-minimization) are popular for regularized image reconstruction. The basic idea of OT is to construct a surrogate function $Q(\boldsymbol{x}; \boldsymbol{x}^n)$ of the image $\boldsymbol{x}$ at the nth iteration which majorizes the original objective function $\Phi(\boldsymbol{x})$ by satisfying the following two conditions [21]:

$$Q(x;x^n)-Q(x^n;x^n)\geq\phi(x)-\phi(x^n) \quad (7)$$

$$\nabla Q(x;x^n)\big|_{x=x^n}=\nabla\phi(x)\big|_{x=x^n} \quad (8)$$

where ∇ denotes the gradient with respect to $\boldsymbol{x}$. Equation (8) assumes that $\Phi(\boldsymbol{x})$ is differentiable. Then the minimization of $\Phi(\boldsymbol{x})$ is transferred into minimization of $Q(\boldsymbol{x}; \boldsymbol{x}^n)$

$$x_{\mathrm{OT}}^{n+1}=\operatorname{argmin}_{x\geq 0} Q(x;x^n) \quad (9)$$

By design the surrogate function $Q(\boldsymbol{x}; \boldsymbol{x}^n)$ is usually easier to optimize than the original objective function. The new update x_{OT}^{n+1} guarantees a decrease in the original objective function,

$$\phi(x_{\mathrm{OT}}^{n+1})<\phi(x^n) \quad (10)$$

unless $\boldsymbol{x}^n$ is a stationary point of the objective function [21].

By setting $x^{n+1}=x_{\mathrm{OT}}^{n+1}$, the majorization-minimization procedure guarantees monotonic convergence. The well-known expectation maximization (EM) algorithm [1] is a special case of optimization transfer [21]. For the Poisson log-likelihood function of PET data, the EM-based surrogate [1], [21] can be written as

$$y_i=g(\mathbf{x}_i)+\epsilon_i, \quad (11)$$

where p_j, the sensitivity image value at pixel j, is

$$p_j=\sum_i p_{ij} \quad (12)$$

and the intermediate EM update $x_{j,\mathrm{EM}}^{n+1}$ is given by

$$x_{j,\mathrm{EM}}^{n+1}=\frac{x_j^n}{p_j}s_j^n, s_j^n=\sum_{i=1}^{n_i}p_{ij}\frac{y_i}{\bar{y}_i(x^n)} \quad (13)$$

where $s^n=\{sjn\}j=1n_j$ is the back projection of the ratio sinogram.

For the regularization term, we consider penalty functions that satisfy [31]-[33]

$$\Psi_\delta(t)\leq\frac{1}{2}\hat{\mathrm{a}}\Im\mathfrak{t}^{\Psi_\delta}(t^n)t^2+\left[\Psi_\delta(t^n)-\frac{1}{2}\hat{\mathrm{a}}\Im\mathfrak{t}^{\Psi_\delta}(t^n)(t^n)2\right] \quad (14)$$

where $\hat{\mathrm{a}}\Im\mathfrak{t}^{\Psi_\delta}(t)$ is the half-quadratic weight function [31],

$$\hat{\mathrm{a}}\Im\mathfrak{t}^{\Psi_\delta}(t)=\dot{\Psi}_\delta(t)/t \quad (15)$$

with $\dot{\Psi}_\delta(t)$ denoting the first-order derivative of the penalty function $\psi_\delta(t)$. Most convex penalty functions (e.g. Huber, Fair, hyperbola functions) used in image reconstruction belong to this class, excluding the nonsmooth penalty $|t|$ because of its nondifferentiability at zero. The Fair penalty function, for example, has

$$\hat{\mathrm{a}}\Im\mathfrak{t}^{\Psi_\delta}(t)=1/(|t|+\delta) \quad (16)$$

By replacing t with $(x_j - x_k)$, we can get the surrogate function $Q_U^a(x;x^n)$ for the penalty function $U(\boldsymbol{x})$ at iteration n:

$$Q_U^a(x;x^n)=\frac{1}{2}\sum_{j=1}^{n_j}\sum_{k\epsilon N_j}w_{jk}w_{jk}^{\Psi_\delta}(x^n)(x_j-x_k)2 \quad (17)$$

where $w_{jk}^{\Psi_\delta}(x)=$âΩţ$^{\Psi_\delta}(x_j-x_k)$ and a constant term that is only a function of $\boldsymbol{x}^n$ but independent of $\boldsymbol{x}$ is omitted. By using De Pierro's decoupling rule [12], the surrogate function $Q_U^a(x;x^n)$ can be further majorized by a separable surrogate function $Q_U^b(x;x^n)$[33] [33]

$$Q_U^b(x;x^n)=\frac{1}{2}\sum_{j=1}^{n_j}w_{j,\mathrm{reg}}^{n+1}(x_j-x_{j,\mathrm{reg}}^{n+1})2 \quad (18)$$

where the pixel-wise weight $w_{j,reg}^n$ is

$$w_{j,\mathrm{reg}}^{n+1}=4\sum\nolimits_{k\epsilon N_j}w_{jk}w_{jk}^{\Psi_\delta}(x^n) \quad (19)$$

and the intermediate edge-preserving smooth image $x_{j,\mathrm{reg}}^{n+1}$ is calculated by

$$x_{j,\mathrm{reg}}^{n+1}=\frac{2}{w_{j,\mathrm{reg}}^{n+1}}\sum\nolimits_{k\epsilon N_j}w_{jk}w_{jk}^{\Psi_\delta}(x^n)(x_k^n+x_j^n) \quad (20)$$

Solving the Karush-Kuhn-Tucker condition for minimizing the surrogate $(x;x^n)+\beta Q_U^b(x;x^n)$, we get the OT image update equation at iteration (n + 1)

$$x_{j,\mathrm{OT}}^{n+1}=\frac{2x_{j,\mathrm{EM}}^{n+1}}{\sqrt{(b_j^{n+1})2+4\beta_j^{n+1}x_{j\mathrm{EM}}^{n+1}}+b_j^{n+1}} \quad (21)$$

where $\beta_j^{n+1}=\beta w_{j,\mathrm{reg}}^{n+1}/p_j$ and $b_j^{n+1}=1-\beta_j^{n+1}x_{j,\mathrm{reg}}^{n+1}$.

B. Acceleration Using Conjugate Direction and Line Search

We observe from Eq. (10) that the OT direction at iteration (n + 1),

$$d_{\mathrm{OT}}^{n+1}=x_{\mathrm{OT}}^{n+1}-x^n \quad (22)$$

is a descent direction. This motivated us to use the conjugate direction method [16] to achieve a more aggressive update on $\boldsymbol{x}$:

$$x^{n+1}=x^n+\alpha^{n+1}a^{n+1} \quad (23)$$

where α^{n+1} is a step size and $\boldsymbol{a}^{n+1}$ is the conjugate direction of d_{OT}^{n+1} that is given by

$$a^{n+1}=a^n+\gamma^{n+1}d_{\mathrm{OT}}^{n+1} \quad (24)$$

with $a^1 = d^1_{\mathrm{OT}}$. The parameter γ^n is determined by the Polak-Ribiere form [16]

$$\gamma^{n+1} = \frac{(g^{n+1} - g^n)\, T a^n}{(g^n)\, T a^n} \quad (25)$$

where the superscript "T" denotes matrix or vector transpose and the gradient vector $g^{n+1} = \left\{ g_j^{n+1} \right\} j = 1 n_j$ is computed by

$$g_j^{n+1} = \frac{\partial}{\partial x_j} \phi(x) \big|_{x = x^n} = (p_j - s_j^n) + \beta w_{j,\mathrm{reg}}^{n+1} (x_j^n - x_{j,\mathrm{reg}}^{n+1}) \quad (26)$$

The step size α^{n+1} in (23) is determined by a line search

$$\alpha^{n+1} = \mathrm{argmin}_{\alpha > 0} \phi(x^n + \alpha a^{n+1}) \quad (27)$$

which can be solved by the modified Newton-Raphson algorithm in [17].

The nonnegativity constraint is enforced in the same way as that in the PCG [15]. If any negative values exist in $\boldsymbol{x}^{n+1}$, a second line search is performed to enforce the nonnegativity constraint on $\boldsymbol{x}$:

$$\alpha_2^{n+1} = \mathrm{argmin}_{0 \le \alpha \le 1} \phi(x^n + \alpha(\left[x^{n+1} \right]_+ - x^n)) \quad (28)$$

where $[\cdot]_+$ sets any negative values to zeros. Then the new image estimate at $(n + 1)$ is given by

$$x_{\mathrm{new}}^{n+1} = x^n + \alpha_2^{n+1} (\left[x^{n+1} \right]_+ - x^n) \quad (29)$$

To distinguish the new update in Eq. (23) or Eq. (29) from the original OT update, we refer to it as the optimization transfer descent (OTD). The OTD algorithm moves more aggressively than the OT does, while still guaranteeing to descend monotonically. The OT direction d_{OT}^{n+1} can be treated as a preconditioned negative gradient direction, where the preconditioner is defined implicitly. Comparing with the conventional PCG algorithm that uses an EM-based preconditioner, the implicitly defined OT preconditioner contains information from both the likelihood term and the penalty term. Therefore, the proposed OTD is expected to be faster than the PCG using the EM preconditioner for PL image reconstruction.

IV. TRUST OPTIMIZATION TRANSFER

The OTD algorithm developed in the previous section is not directly applicable to the non-smooth ℓ_1 penalty function and can be slow when δ in a smooth ℓ_1 is too small. To

accelerate the algorithm for high-curvature edge-preserving penalties and to extend the OTD algorithm to the non-smooth ℓ_1 penalty, we borrow the trust surrogate concept from the classic Levenberg-Marquardt [34] and trust region methods [30, p.65, p.262] for nonlinear optimization.

At iteration n, the original objective function $\Phi(\boldsymbol{x})$ is approximated by a surrogate function S:

$$S_\sigma(x)=S^L(x)+\beta S^U_\sigma(x) \quad (30)$$

where $S^L(\boldsymbol{x})$ is the surrogate of the likelihood term and $_\sigma(\boldsymbol{x})$ is the smooth approximation of the penalty term, with σ being the damping parameter [34]. Note that $S_\sigma(\boldsymbol{x})$ is not an optimization transfer surrogate and does not have to satisfy the two conditions in (7) and (8). Fig. 1 illustrates the difference between an optimization transfer surrogate and a trust surrogate.

In the trust optimization transfer, a new estimate is obtained by minimizing the surrogate

$$\hat{\mu}(\sigma)=\mathrm{argmin}_{x\geq 0} S_\sigma(x) \quad (31)$$

which can be solved by the OTD algorithm. If

$$\phi(\hat{\mu}(\sigma))\leq\phi(x^n) \quad (32)$$

then the associated surrogate $S_\sigma(\boldsymbol{x})$ is a so called trust surrogate. The image estimate is then updated by

$$x^{n+1}=\hat{\mu}(\sigma) \quad (33)$$

If (32) is not satisfied, a new value of σ will be tested until a trust surrogate is found.

To apply OTD, we use

$$S^L(x)=-L(y|\ \bar{y}\ (x)) \quad (34)$$

$$S^U_{\sigma^n}(x)=\sum_{j=1}^{n_j}\sum\nolimits_{k\epsilon N_j} w_{jk}\,\Psi_{\sigma^n}(x_j-x_k) \quad (35)$$

where σ^n, the value of σ at iteration n, is greater than or equal to δ ($\delta = 0$ for non-smooth ℓ_1). Fig. 2 shows the Fair function $\psi_\delta(t)$ with different δ values. All the functions approach $|t|$ when $t \gg \delta$, but a larger δ value provides a smoother penalty function near $t = 0$, which usually results in faster convergence in the optimization than a smaller δ does. The proposed trust surrogate method shares a similar spirit to the continuation scheme in optimization, but

utilizes a trust mechanism to guarantee monotonicity. For efficient computation, we do not solve the minimization problem in (31) completely. Instead, only one iteration of OTD is used and we check the monotonicity in $\Phi(\boldsymbol{x})$.

A. Search Rule

In order to determine σ^n at each iteration, we define ρ as the ratio between the change $\Delta\Phi$ in the original cost function $\Phi(\boldsymbol{x})$ and the change ΔS in the surrogate function $S_\sigma(\boldsymbol{x}$ caused by $\hat{\boldsymbol{\mu}}(\sigma^n)$,

$$\rho(\hat{\mu}(\sigma^n))=\frac{\triangle\phi}{\triangle S}=\frac{\phi(\hat{\mu}(\sigma^n))-\phi(x^n)}{S_{\sigma^n}(\hat{\mu}(\sigma^n))-S^n_\sigma(x^n)} \quad (36)$$

If $\rho(\hat{\mu}(\sigma^n))\geq 0$, then $\Phi\,\hat{\boldsymbol{\mu}}(\sigma^n) \leq \Phi(\boldsymbol{x}^n)$. Note that the objective function values involved in calculation of $\rho\,\pi$ be rapidly evaluated by reusing the projections that have already been calculated in the OTD algorithm, so no addition forward projection is required.

When $\rho(\hat{\mu}(\sigma^n))$ is large, the value of σ^n is trusted and will be used in the next iteration. To prevent too many iterations being spent on the same value of σ with only an insignifica decrease in $\Phi(\boldsymbol{x})$, we measure the relative change in $\Phi(\boldsymbol{x})$ by

$$\nu=\frac{\phi(x^{n+1})-\phi(x^n)}{\phi(x^{n+1})-\phi(x^{n_{\text{start}}(\sigma)})} \quad (37)$$

where $n_{\text{start}}(\sigma)$ denotes the index of the first iteration at whi the current σ is used. The rule for determining σ^{n+1} is

$$\sigma^{n+1}=\begin{cases}\max(\delta,\sigma^n/3), & \rho\leq 0 \text{ or } n-n_{\text{start}}(\sigma^n)>50\\ \sigma^n, & \rho>0,\nu\geq 0.01/\rho\\ \max(\delta,\sigma^n/3), & \rho>0,\nu<0.01/\rho\end{cases} \quad (38)$$

Basically if $\rho > 0$ and ν is greater than a threshold, the curre value of σ will be used again; otherwise, σ will be decreas by a factor of 3. The threshold $0.01/\rho$ allows more iteratio to be taken for a σ value that results in a large ρ. When ν too small, the σ value shall be reduced in the next step even if $\rho > 0$.

B. Initial σ Value

The initial value of σ can be critical for the convergen speed. A large σ results in fast convergence for the surroga optimization, while a σ closer to δ provides better approxim tion of the original objective function. We empirically find th $\sigma_{\text{init}} = 0.01 \max(\boldsymbol{x})$ or $\sigma_{\text{init}} = 0.1\text{mean}(\boldsymbol{x})$ is a good initi value if an estimate of $\boldsymbol{x}$ is known before reconstruction. A alternative is to determine σ_{init} by the least square estima of a uniform image from the PET data :

$$\sigma_{\text{init}}=0.1q^T(y-r)/(q^Tq) \quad (39)$$

where $\boldsymbol{q} = \boldsymbol{P}\mathbf{1}_{\text{obj}}$ and $\mathbf{1}_{\text{obj}}$ denotes the support mask region of the object. The σ_{init} value from (39) roughly equals to 10% of the average image intensity.

C. Convergence Discussion

Because it is a conjugate gradient algorithm with an implicitly defined preconditioner, the OTD algorithm shares the same convergence property with the PCG algorithm [15], [16] for smooth ℓ_1 optimization with a fixed penalty parameter δ. The trust OTD algorithm uses a sequence of surrogate penalty parameters σ^n in different iterations to accelerate convergence, but σ^n will eventually reach the value of δ after a finite number of iterations according to (38). Due to the use of the trust mechanism, the trust OTD algorithm is also guaranteed to descend monotonically. For nonsmooth ℓ_1 optimization where $\delta = 0$, the surrogate penalty parameter σ can never reach the exact value $\delta = 0$. However, the proposed algorithm is guaranteed to descend monotonically to an approximate ℓ_1 solution defined by the objective function with a δ value close to the machine precision. We found that the difference between the solution of $\delta = 0$ and that of a very small positive δ is negligible in PET image reconstruction (see Fig. 13 in Section V.D).

A pseudo-code of the trust optimization transfer algorithm is given in Algorithm 1. In the algorithm table, we restrict the maximum number of iterations to MaxIter. Alternatively, a stopping rule based on the changes in gradient or image estimate can be used to terminate the iteration.

V. SIMULATION STUDIES

Computer simulation was conducted to compare the proposed algorithm with several existing algorithms. We simulated a PET emission image using a 2D brain phantom shown in Fig. 3(a). A real CT image shown in Fig. 3(b) was used to generate the attenuation factors. The PET image in a 128×128 grid was first forward projected to obtain a noise-free sinogram. A 20% uniform background was added to simulate mean randoms and scatters. Independent Poisson noise was then introduced, resulting in a total of 200k coincidence events. The noisy sinogram was reconstructed using a smooth ℓ_1 penalty via the Fair function with $\delta = 10^0, 10^{-2}, 10^{-6}$, and the non-smooth ℓ_1 penalty. The first-order neighborhood was used and the regularization parameter β was set to $\beta = 2^{-6}$ to achieve a low mean squared error in the reconstructed images.

We compared the proposed algorithm with the De Pierro's EM (DEM) [12] using Eq. (21), PCG [15], [16], ADMM [22], SPIRAL [29], and the Chambolle-Pock (CP) algorithm [27], [28]. The DEM and PCG algorithms were applied only to the smooth ℓ_1 regularizations, because they are not applicable to the non-smooth ℓ_1. The ADMM [22] with two different sets of tuning parameters (ADMM1 and ADMM2) were implemented for both the smooth and non-smooth penalties. The empirical rule used in [35] for parameter tuning was modified here to guide the selection of the three parameters. ADMM1 was tuned to converge as fast as possible in a non-monotonic fashion and ADMM2 was tuned to

converge as fast as possible in a nearly monotonic way. The reconstruction step in the ADMM was solved by a PCG algorithm with two sub-iterations. The initial σ value in the proposed algorithm was calculated by Eq. (39). The SPIRAL algorithm [29], downloaded from the authors' website, was used with its default parameter setting for the non-smooth ℓ_1 penalty. The Chambolle-Pock (CP) algorithm was implemented with preconditioning [27], [28]. All reconstructions start from a same all-one uniform initial image. All of algorithms were implemented in MATLAB on a Apple MacBook Pro with a 2.6 GHz dual-core Intel Core i5 processor.

To compare the convergence rate of different algorithms, we plotted the normalized objective function, which is defined as

$$C^n = [\phi(x^n) - \phi^*] / \left[\phi(x^0) - \phi^*\right] \quad (40)$$

where Φ^* denotes a reference value of objective function, and the image mean squared error (MSE),

$$\mathrm{MSE}(x^n) = 10\log_{10}\left(\left|\left|x^n - x_{\mathrm{true}}\right|\right|^2 / \left|\left|x_{\mathrm{true}}\right|\right|^2\right) \quad (41)$$

as functions of iteration number and CPU time.

A. Comparison for Smooth ℓ_1 Regularization

Fig. 4 shows the images reconstructed by the DEM, PCG and proposed algorithm using the smooth ℓ_1 penalty with two small δ values: $\delta = 10^{-2}$, 10^{-6}. The DEM result at 100 iterations has poor image quality due to slow convergence. The reason for the noisy appearance is that with a uniform initial image, a small value of δ results in a very large value of the edge-preserving weight $w_{j,\mathrm{reg}}^{n+1}$ in (19) through the half-quadratic function (16). As a result, false edges are often identified and preserved in the subsequent iterations. We observed that the DEM algorithm had to be run for thousands of iterations to get a good image estimate for these small δ values with a uniform initial image. Such problem does not appear when the DEM algorithm is initialized with an ML EM reconstructed image (see also Section V.F). Similarly, when δ in the smooth ℓ_1 function decreases from 10^{-2} to 10^{-6}, the PCG also becomes less efficient, resulting in a substantial increase in the image MSE. In comparison, the proposed algorithm is stable as δ changes.

Fig. 5 shows the cost function and image MSE as functions of iteration number and CPU time for PL reconstruction using different smooth ℓ_1 algorithms with a relatively large $\delta = 1$. The PCG image estimate at iteration 2000 was used as the reference for calculating the normalized cost function. All the algorithms, including the DEM, converge relatively quickly and the proposed algorithm is the fastest in terms of the cost function. The PCG is even faster than ADMM1 and ADMM2 in decreasing the cost function and also converges faster than ADMM2 in terms of image MSE.

Fig. 6 shows the comparison results for smooth ℓ_1 with $\delta = 10^{-6}$. For such a small value of δ, both the DEM and PCG exhibit slow convergence and result in high image MSE. The ADMM reconstructions converge very fast but the faster version, ADMM1, behaves non-monotonically. The proposed algorithm converges monotonically and is as fast as ADMM1.

B. Comparison for non-smooth ℓ_1

Different algorithms for the PL reconstruction with the non-smooth ℓ_1 penalty are compared. The reconstructed images at a constant CPU time of 10 seconds are shown in Fig. 7. The convergence plots of the cost function and image MSE are shown in Fig. 8. The image estimate by the CP algorithm with 5000 iterations was used as the reference for calculating the normalized cost function.

The CP algorithm decreases the cost function monotonically but its convergence speed is slow. The SPIRAL converges quickly and achieves a very low cost function value but the CPU time per iteration, especially at later iterations, is greater than that of other algorithms. Both ADMM1 and SPIRAL display non-monotonic behavior in the cost function value. While SPIRAL can be run in the monotonic mode, it is slower than its non-monotonic implementation and may become extremely slow at later iterations. ADMM2 is slower than ADMM1 but converges almost monotonically. Again, the proposed algorithm is the fastest among all algorithms in terms of minimizing the cost function and is as fast as ADMM1 in terms of MSE convergence.

C. Quadratic regularization

The trust optimization transfer algorithm can also be applied to the PL reconstruction with a quadratic regularization, in which case the algorithm reduces to the OTD algorithm. We performed a study to compare the performance of different algorithms for the quadratic regularization. The regularization parameter was set to $\beta = 2^{-8}$ to obtain an image with a low MSE. Fig. 9 shows the plots of the normalized cost function and image MSE as functions of iteration number and CPU time for different algorithms. The PCG result at iteration 2000 was used as the reference for calculating the normalized cost function. For the quadratic regularization, the DEM and PCG converge quickly and are faster than the two ADMM algorithms in terms of minimizing the cost function. The proposed algorithm converges fastest, indicating the proposed algorithm is applicable to a wide range of cost functions.

D. Effect of Initial σ Value

Fig. 10(a) shows the change in σ as a function of iteration in one reconstruction by the proposed smooth ℓ_1 algorithm with a uniform image initial. The value of δ was 10^{-2}. Different initial σ_{init} values, ranging from 0.2 to 20, are compared. Starting from an initial value σ_{init}, σ stays at the same value if the σ value at previous iteration is trusted, and drops down by a factor of 3 when the value becomes untrusted as iteration continues. After reaching the value of δ, σ is set to δ in all later iterations. The plot of a fixed σ value equal to δ is also included for comparison. The values of the cost function versus iteration number are plotted in Fig. 10(b). Note that the value of $\sigma_{\text{init}} = 2.4$ is the one calculated automatically using (39) for the data, which appears to be a good choice. Comparing with the case of

fixing σ at the value of δ, Fig. 10(b) shows the trust surrogate steps can substantially speed up the convergence of the algorithm with a proper initial σ value.

The evolution of σ and the effect of σ_{init} on convergence in the proposed algorithm for non-smooth ℓ_1 are similar to the results shown here. A difference is that σ in the non-smooth ℓ_1 algorithm never reach the exact value $\delta = 0$ but instead approaches the smallest value defined by the machine precision.

E. Effect of the initial image

Instead of a uniform initial image, an image estimate by the EM or ordered subsets (OS) EM can also be used as the initial image. We investigated the effect of the initial image on the convergence speed. Fig. 11 compares the smooth ℓ_1 ($\delta = 0.01$) optimizations with a uniform initial image and with an EM image initialization. The EM image was reconstructed using 20 iterations of the ML EM algorithm, which provides nearly the lowest MSE among all EM iterations. The $\sigma_{\text{init}} = 2.4$ in the proposed algorithm was the same as that in the case of uniform initialization. The cost function values in Fig. 11 (b) are shifted by the difference between the uniform initial image and the EM initial image for easy comparison. The results indicate that the EM initial image accelerates convergence speed at earlier iterations, in particular for slow-converging algorithms (e.g. DEM). The effect of σ_{init} in the proposed algorithm is shown in Fig. 12(b). Different values of σ_{init} result in similar convergence rates that are faster than a constant σ fixed at $\delta = 0.01$, indicating that an EM-based initialization can be more robust to σ_{init} than a uniform initialization.

F. Smooth ℓ_1 *versus non-smooth* ℓ_1

Fig. 13 shows the MSE values of whole image and tumor region only of the PL reconstruction as a function of the penalty parameter δ. The smooth ℓ_1 reconstructions were solved using the proposed algorithm with 200 iterations. The regularization parameter β was fixed at 2^{-6}. The MSE of the non-smooth ℓ_1 reconstruction obtained by the CP algorithm at iteration 5000 is also shown in the figure and marked by "x". Fig. 14(a)-(d) show the reconstructed images by the smooth ℓ_1 with four different δ values, $\delta = 10, 1, 0.1, 0.001$. Fig. 14(e) and Fig. 14(f) show the images reconstructed by the non-smooth ℓ_1 using the proposed algorithm and using the CP algorithm with 5000 iterations, respectively.

These results indicate that the non-smooth ℓ_1 regularization is not mandatory for edge-preserving image reconstruction to get the best MSE performance. A smooth ℓ_1 with a small δ can provide reconstructed images with equal quality. With a careful selection of δ value ($\delta = 1$ in this simulation), the smooth ℓ_1 can achieve a lower MSE than the non-smooth ℓ_1 regularization does. While it is generally challenging to choose the δ value for the best image MSE, Fig. 13 shows that a wide range of δ values exist and can be chosen to get a comparable MSE performance. Compared with the traditional DEM and PCG algorithms which only work well for large δ values, the proposed algorithm can provide consistent image quality over the wide range of δ values, as shown in Fig. 4, and hence reduces the burden of choosing the δ parameter.

VI. APPLICATION TO REAL 3D DATA

We have applied the proposed algorithm to a 3D nonhuman primate brain data acquired on a microPET P4 scanner. The radiotracer is ^{11}C-SCH 23390 which binds to dopamine D1 receptors in the brain. The total number of prompt events is 3.6 million and number of delayed events is 1.3 million. Randoms were pre-corrected by the delayed window technique. The estimated attenuation factors from a transmission scan and a scatter sinogram estimated by the single-scatter simulation method were incorporated into the forward model of the PL reconstruction.

The projection data were reconstructed on 128 × 128 × 65 grid using both smooth ℓ_1 and nonsmooth ℓ_1 regularizations. The initial value of σ in the trust OT was automatically set by Eq. (39). The regularization parameter β was chosen empirically to give a visually pleasing image. The convergence speed of the cost function of the proposed algorithm in comparison with those of the PCG, ADMM, and CP algorithms is shown in Fig. 15(a) for the smooth ℓ_1 regularization and in Fig. 15(b) for non-smooth ℓ_1 regularization. The CPU time was measured on a Linux PC with dual 2.3 GHz quad-core CPUs running MATLAB and a parallel implementation of the forward and back projectors. Each point on the curves represents 10 iterations. The three Lagrangian penalty parameters in the ADMM were manually tuned to make the algorithm as fast as possible. The SPIRAL algorithm was not compared here because the downloaded program is not applicable to 3D regularization. The results show that the proposed algorithm is faster than PCG and ADMM for the smooth ℓ_1 regularization and is also faster than the ADMM and CP for the non-smooth ℓ_1 regularization.

We note that choosing the proper parameters for ADMM is nontrivial and very challenging in practice. In the 2D simulation in Section IV, the three tuning parameters in ADMM were selected by using the empirical rule in [35] as a guide. However, the same approach failed to choose a good parameter set for the 3D data and the parameters had to be manually tuned based on trial and error. In comparison, the proposed algorithm is nearly free of parameter tuning, consistently performing well in both 2D and 3D cases, and is therefore easier to use.

VII. CONCLUSION

We have proposed a trust optimization transfer algorithm for edge-preserving PET image reconstruction. The fast convergence of the proposed algorithm is demonstrated using simulated data and real 3D PET data. The results show that the proposed algorithm is faster than the traditional DEM and PCG algorithms for smooth ℓ_1 regularization. Compared with the emerging algorithms such as the ADMM, SPIRAL and CP algorithms for non-smooth ℓ_1 regularization, the proposed algorithm is guaranteed to descend monotonically and its convergence speed is at least comparable to that of non-monotonic ADMM and SPIRAL.

This paper mainly focuses on the development of a fast convergent optimization algorithm for edge-preserving image reconstruction. While the anisotropic regularization function in (4) often suffers from blocky artifacts, new regularization functions have been proposed to reduce such artifacts (e.g. [36]-[40]). We also note that a commercial product of edge-

preserving PET image reconstruction has recently been made available on clinical PET scanners [41]. These new developments will likely promote more applications of edge-preserving PET image reconstruction in both clinic and research in the future.

Acknowledgments

This work was supported in part by NIH grant R01EB00194 and by an investigator-sponsored research grant from GE Healthcare.

References

[1]. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B. 1977; 39(1):1–38.

[2]. Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging. 1982; 1(2):113–122. [PubMed: 18238264]

[3]. Lange K, Carson RE. EM reconstruction algorithms for emission and transmission tomography. Journal of Computer Assisted Tomography. 1984; 8(2):306–316. [PubMed: 6608535]

[4]. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Transactions on Medical Imaging. 1994; 13(4):601–609. [PubMed: 18218538]

[5]. Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography. Physics in Medicine and Biology. 2006; 51(15):R541–578. [PubMed: 16861768]

[6]. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984; 6(6):721–741. [PubMed: 22499653]

[7]. Geman, S.; McClure, D. Bayesian image analysis: an application to single photon emission tomography. Proceedings of the Statistical Computing Section of the American Statistical Association; Washington, DC, USA. 1985. p. 12-18.

[8]. Kao CM, Pan X, Chen CT, Wong WH. Image restoration and reconstruction with a Bayesian approach. Medical Physics. 1998; 25(5):600–613. [PubMed: 9608469]

[9]. Alenius S S, Ruotsalainen U. Bayesian image reconstruction for emission tomography based on median root prior. Europian Journal of Nuclear Medicine. 1997; 24(3):258–265.

[10]. Bouman C, Sauer K. A generalized Gaussian image model for edge-preserving MAP estimation. IEEE Transactions on Image Processing. 2002; 2(3):296–310. [PubMed: 18296219]

[11]. Lange K. Convergence of EM image reconstruction algorithms with Gibbs smoothing. IEEE Trans Med Imaging. 1990; 9(4):439–446. [PubMed: 18222791]

[12]. De Pierro AR. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Transactions on Medical Imaging. 1995; 14(1):132–137. [PubMed: 18215817]

[13]. Fessler JA, Erdogan H. A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction. IEEE Nuclear Science Syposium and Medical Imaging Conference. 1998; 2:1132–5. 1998.

[14]. Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. IEEE Transactions on Medical Imaging. 2003; 22(5):613–626. [PubMed: 12846430]

[15]. Kaufman L. Implementing and accelerating the EM algorithm for positron emission tomography. IEEE Transations on Medical Imaging. 1987; 6(1):37–51.

[16]. Mumcuoglu EU, Leahy R, Cherry SR, Zhou ZY. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. IEEE Transactions on Medical Imaging. 1994; 13(4):687–701. [PubMed: 18218547]

[17]. Fessler JA, Booth SD. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. IEEE Transactions on Image Processing. 1998; 8(5):688–99. [PubMed: 18267484]

[18]. Bouman CA, Sauer K. A unified approach to statistical tomography using coordinate descent optimization. IEEE Transactions on Image Processing. 1996; 5(3):480–92. [PubMed: 18285133]

[19]. Fessler JA, Ficaro EP, Clinthorne NH, Lange K. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. IEEE Transactions on Medical Imaging. 1997; 16(2):166–75. [PubMed: 9101326]

[20]. Yu Z, Thibault JB, Bouman CA, Sauer KD, Hsieh J. Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization. IEEE Transactions on Image Processing. 2011; 20(1):161–75. [PubMed: 20643609]

[21]. Lange K, Hunter DR, Yang I. Optimization transfer using surrogate objective functions. Journal of Computational and Graphical Statistics. 2000; 9(1):1–20.

[22]. Figueiredo MAT, Bioucas-Dias JM. Restoration of Poissonian images using alternating direction optimization. IEEE Transactions on Image Processing. 2010; 19(12):3133–3145. [PubMed: 20833604]

[23]. Setzer S, Steidl G, Teuber T. Deblurring Poissonian images by split Bregman techniques. Journal of Visual Communication and Image Representation. 2010; 21(3):193–199.

[24]. Lingenfelter DJ, Fessler JA. Augmented Lagrangian methods for penalized likelihood reconstruction in emission tomography. IEEE Nuclear Science Symposium Conference Record (NSS/MIC). 2010:3288–3291. 2010.

[25]. Chun SY, Dewaraja YK, Fessler JA. Alternating direction method of multiplier for tomography with nonlocal regularizers. IEEE Transactions on Medical Imaging. 2014; 33(10):1960–1968. [PubMed: 25291351]

[26]. Chambolle A, Pock T. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. Journal of Mathematical Imaging and Vision. 2011; 40(1):120–145.

[27]. Pock T, Chambolle A. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. 2011 International Conference on Computer Vision, Barcelona, Spain. 2011; 40(1): 1762–1769.

[28]. Sidky EY, Jørgensen JH, Pan X. Convex optimization problem prototyping with the Chambolle-Pock algorithm for image reconstruction in computed tomography. Physics in Medicine and Biology. 2012; 57(10):30653091.

[29]. Harmany ZT, Marcia RF, Willett RM. This is SPIRAL-TAP: sparse Poisson intensity reconstruction algorithms – theory and practice. IEEE Transactions on Image Processing. 2012; 21(3):1084–1096. [PubMed: 21926022]

[30]. Nocedal, J.; J. Wright, S. Numerical Optimization. Springer; 1999.

[31]. Huber, PJ. Robust Statistics. John Wiley & Sons; New York, NY, USA: 1981.

[32]. Erdogan H, Fessler JA. Monotonic algorithms for transmission tomography. IEEE Transactions on Medical Imaging. Apr; 1999 18(9):801–814. [PubMed: 10571385]

[33]. Wang G, Qi J. Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization. IEEE Transactions on Medical Imaging. 2012; 31(12):2194–204. [PubMed: 22875244]

[34]. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics. 1963; 11(2):431–441.

[35]. Ramani S, Fessler JA. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. IEEE Transactions on Medical Imaging. 2012; 31(3):677–88. [PubMed: 22084046]

[36]. Bredies K, Kunisch K, Pock T. Total Generalized Variation. SIAM Journal on Imaging Sciences. 2010; 3(3):492–526.

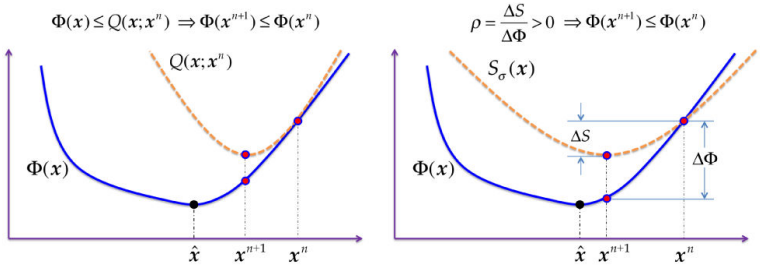
[37]. Lefkimmiatis S, Bourquard A, Unser M. Hessian-Based Norm Regularization for Image Restoration With Biomedical Applications. IEEE Trans. Image Processing. 2012; 21(3):983–995.

[38]. Papafitsoros K, Schnlieb CB. A Combined First and Second Order Variational Approach for Image Reconstruction. Journal of Mathematical Imaging and Vision. 2014; 48(2):308–338.

[39]. Wang, GB.; Qi, J. Penalized likelihood PET image reconstruction using isotropic total variation. IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 13); Seoul, South Korea. 2013. (abstract only)

[40]. Niu S, Gao Y, Bian Z, Huang J, Chen W, Yu G, Liang Z, Ma J. Sparse-view x-ray CT reconstruction via total generalized variation regularization. Physics in Medicine and Biology. 2014; 59(12):2997. [PubMed: 24842150]

[41]. Ross S. Q.Clear. GE Healthcare, White Paper. 2014:1–9.



(a) Optimization transfer (OT) (b) Trust OT

Fig. 1.

Conceptual illustration of the surrogate functions constructed at iteration n for estimating a new image $\boldsymbol{x}^{n+1}$. (a) An optimization transfer surrogate $Q(\boldsymbol{x}; \boldsymbol{x}^n)$ is designed to be always above the original objective function $\Phi(\boldsymbol{x})$. (b) A trust surrogate $S_\sigma(\boldsymbol{x})$ approximate $\Phi(\boldsymbol{x})$ by only requiring its minimum point $\boldsymbol{x}^{n+1}$ to decrease $\Phi(\boldsymbol{x})$. The minimum point $\boldsymbol{x}^{n+1}$ of the surrogate functions in both cases decreases $\Phi(\boldsymbol{x})$ monotonically.

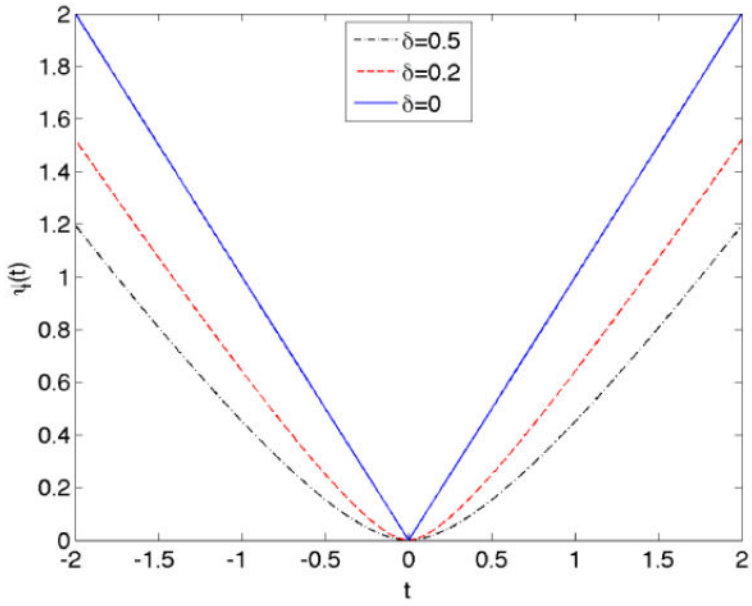


Fig. 2.

Edge-preserving function $\psi_\delta(t)$ with different δ values. A larger δ value usually results in faster convergence in the optimization than a smaller δ does, so the proposed trust surrogate algorithm can be viewed as using a continuation mechanism to accelerate convergence.

Set maximum iteration number `MaxIter`, regularization parameter β and penalty parameter δ;
Initialize image estimate $\boldsymbol{x}^0$, $\rho = 1$; Calculate σ_{init} using (39) and set $\sigma^0 = \sigma_{\text{init}}$;
Perform the initial forward projection to get $\bar{\boldsymbol{y}}(\boldsymbol{x}^0)$ using (2);
for $n = 0$ to `MaxIter-1` **do**
    Edge-preserving image smoothing with the penalty parameter σ^n to get the smooth image $\boldsymbol{x}_{\text{reg}}^{n+1}$ using (20);
    If $\rho > 0$, obtain $\boldsymbol{x}_{\text{EM}}^{n+1}$ using (13);
    Compute the OT image update $\boldsymbol{x}_{\text{OT}}^{n+1}$ by fusing $\boldsymbol{x}_{\text{EM}}^{n+1}$ and $\boldsymbol{x}_{\text{reg}}^{n+1}$ pixel-by-pixel using (21);
    Compute the OT descent direction $\boldsymbol{d}_{\text{OT}}^{n+1}$ using (22) and the conjugate direction $\boldsymbol{a}^{n+1}$ using (24);
    If $x_j^n < \text{mean}(\boldsymbol{x}^n)/10^9$ and $a_j^{n+1} < 0$, then truncate the direction: $a_j^{n+1} = 0$;
    If $\frac{(\boldsymbol{a}^{n+1})^T \boldsymbol{g}^{n+1}}{||\boldsymbol{a}^{n+1}|| \cdot ||\boldsymbol{g}^{n+1}||} < 0.001$, then reset the search direction to the gradient: $\boldsymbol{a}^{n+1} = \boldsymbol{g}^{n+1}$;
    Forward project $\boldsymbol{a}^{n+1}$ to get $\boldsymbol{y}_a^{n+1} = \boldsymbol{P}\boldsymbol{a}^{n+1}$ and estimate the step size $\alpha^{n+1} = \arg\min_\alpha S_{\sigma^n}(\boldsymbol{x}^n + \alpha \boldsymbol{a}^{n+1})$;
    Compute the OTD update $\hat{\boldsymbol{\mu}}(\sigma^n) = \boldsymbol{x}^n + \alpha^{n+1}\boldsymbol{a}^{n+1}$;
    If any negative values exist in $\hat{\boldsymbol{\mu}}(\sigma^n)$, update $\boldsymbol{a}^{n+1} = [\hat{\boldsymbol{\mu}}(\sigma^n)]_+ - \boldsymbol{x}^n$ and $\boldsymbol{y}_a^{n+1} = \boldsymbol{P}\boldsymbol{a}^{n+1}$. A second line search is then performed using Eq. (28) to get a new α^{n+1} and $\hat{\boldsymbol{\mu}}(\sigma^n)$;
    Update the projection $\bar{\boldsymbol{y}}(\hat{\boldsymbol{\mu}}(\sigma^n)) = \bar{\boldsymbol{y}}(\boldsymbol{x}^n) + \alpha^{n+1}\boldsymbol{y}_a^{n+1}$;
    Calculate ρ using (36);
    If $\rho > 0$, update the image estimate $\boldsymbol{x}^{n+1} = \hat{\boldsymbol{\mu}}(\sigma^n)$ and projection $\bar{\boldsymbol{y}}(\boldsymbol{x}^{n+1}) = \bar{\boldsymbol{y}}(\hat{\boldsymbol{\mu}}(\sigma^n))$; otherwise, no update on the image and projection, i.e. $\boldsymbol{x}^{n+1} = \boldsymbol{x}^n$ and $\bar{\boldsymbol{y}}(\boldsymbol{x}^{n+1}) = \bar{\boldsymbol{y}}(\boldsymbol{x}^n)$;
    Calculate the change rate ν using (37) and determine σ^{n+1} using (38);
    If $\sigma^{n+1} < \sigma^n$, record $n_{\text{start}}(\sigma^{n+1}) = n + 1$.
end for
return the image estimate $\hat{\boldsymbol{x}}^{n+1}$

Algorithm 1.
The trust optimization transfer (TOT) algorithm for edge-preserving PET image reconstruction

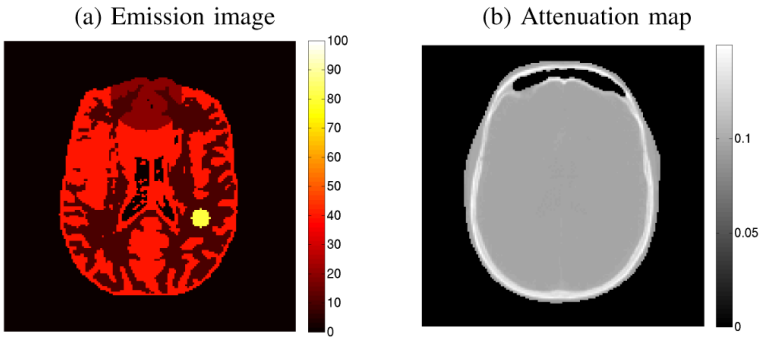


Fig. 3.

(a) The simulated PET emission image and (b) the attenuation map from a real CT image.

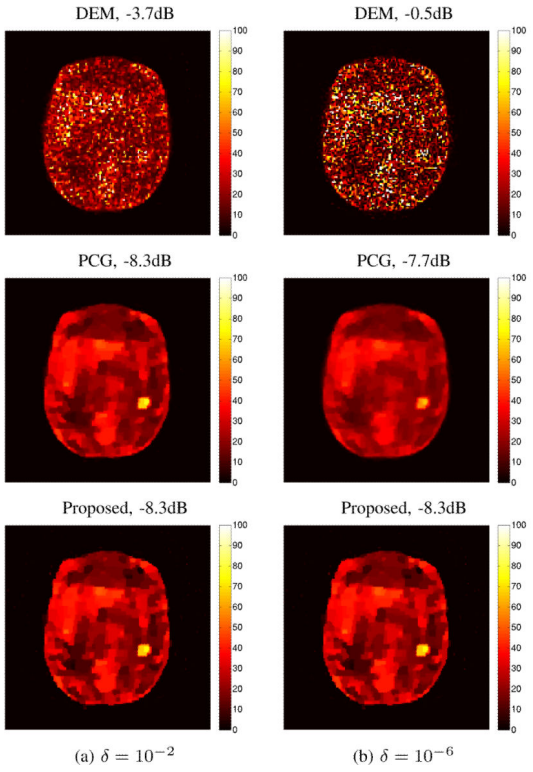
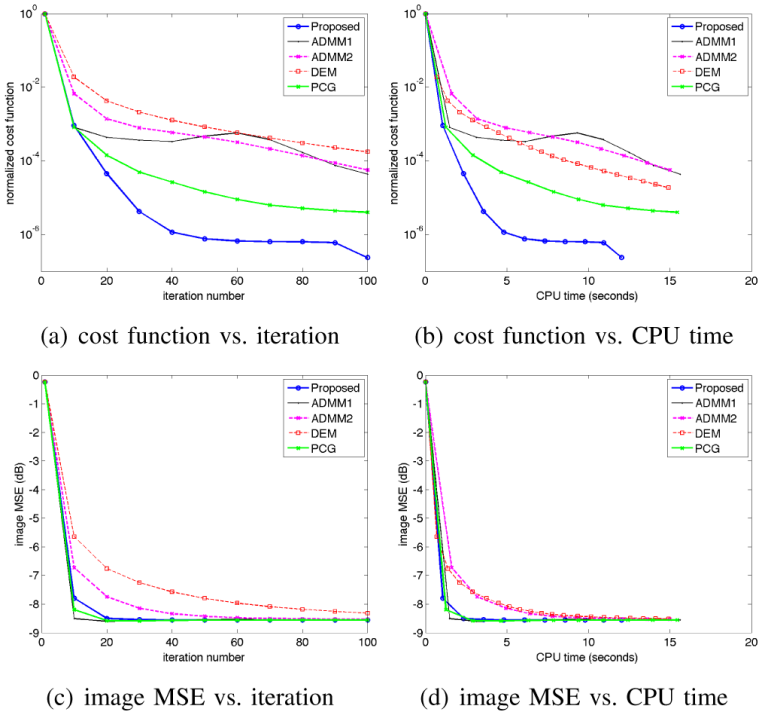


Fig. 4.

PL reconstructions at iteration 100 using (a) the smooth R_1 with $\delta = 10^{-2}$ and (b) smooth R_1 with $\delta = 10^{-6}$ by DEM, PCG and proposed algorithms. The image MSE values are given on top of each image.

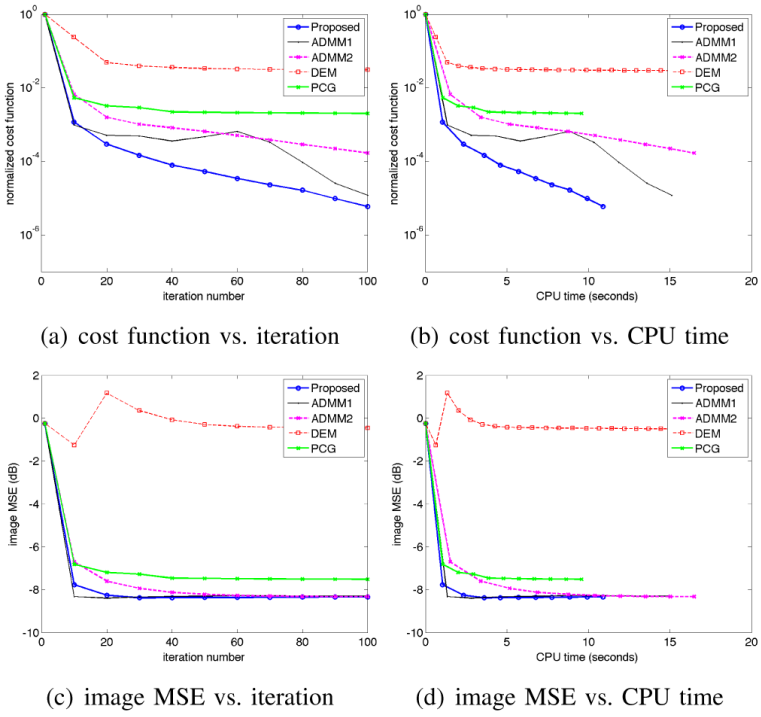


(a) cost function vs. iteration (b) cost function vs. CPU time

(c) image MSE vs. iteration (d) image MSE vs. CPU time

Fig. 5.

Convergence of (a) cost function and (b) image MSE for the smooth ℓ_1 regularization with $\delta = 1$ by different algorithms.



(a) cost function vs. iteration (b) cost function vs. CPU time

(c) image MSE vs. iteration (d) image MSE vs. CPU time

Fig. 6.

Convergence of (a) cost function and (b) image MSE for the smooth ℓ_1 regularization with $\delta = 10^{-6}$ by different algorithms.

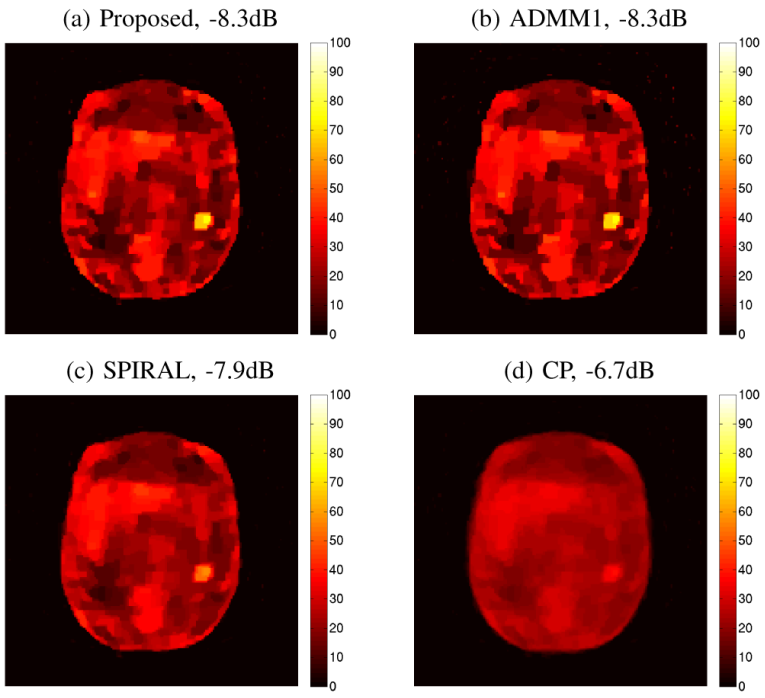
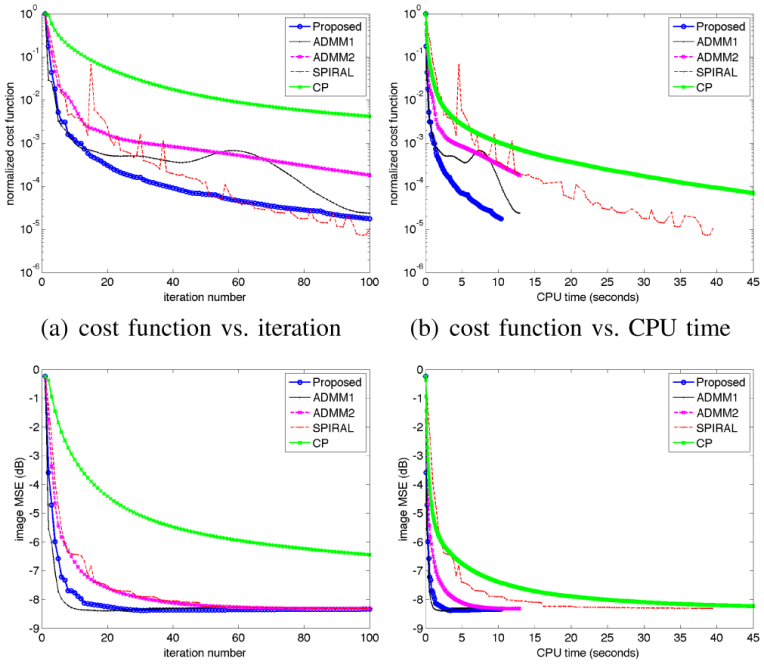


Fig. 7.

PL reconstructions using the non-smooth R_1 by (a) the proposed algorithm, (b) ADMM, (c) SPIRAL and (d) CP, all taking a total of 10 seconds of CPU time. The image MSE values are given on top of each image.

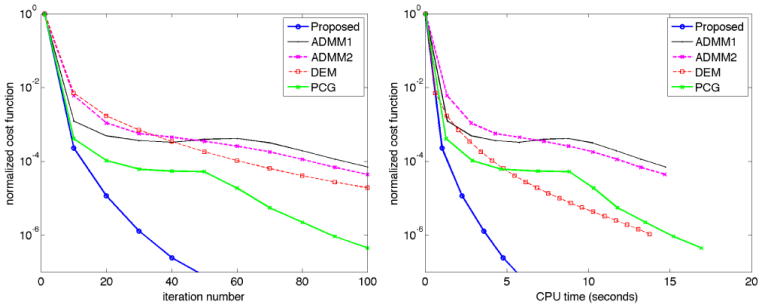


(a) cost function vs. iteration (b) cost function vs. CPU time

(c) image MSE vs. iteration (d) image MSE vs. CPU time

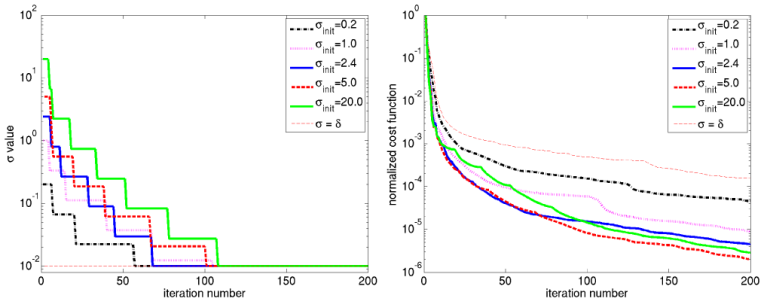
Fig. 8.

Convergence of cost function and image MSE as functions of iteration number and CPU time for the non-smooth R_1 regularization by different algorithms.



(a) cost function vs. iteration (b) cost function vs. CPU time

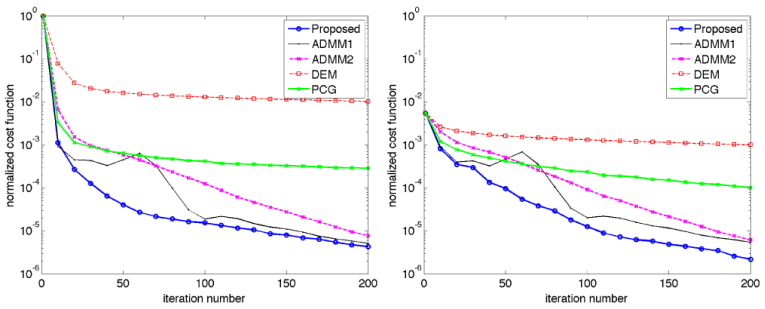
Fig. 9.
Convergence of cost function for the quadratic regularization by different algorithms.



(a) σ value (b) cost function

Fig. 10.

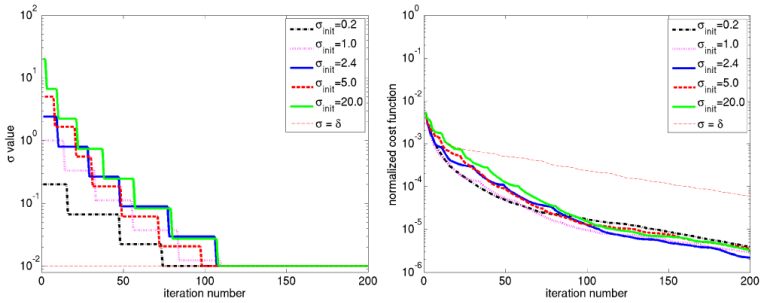
Effect of the initial value σ_{init} on (a) the evolution of σ values and (b) convergence speed of cost function of the proposed algorithm for the smooth R_1 with $\delta = 0.01$ and a uniform image initial. $\sigma_{\text{init}} = 2.4$ was automatically calculated using Eq. (39).



(a) Uniform initial (b) EM initial

Fig. 11.

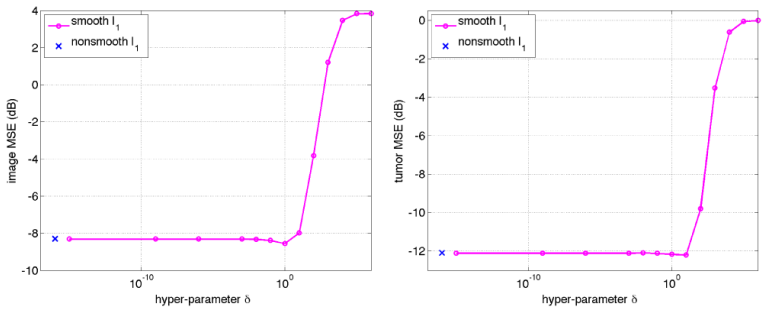
Convergence speed of different smooth R_1 algorithms ($\delta = 0.01$) with different initial estimates: (a) initialized with a uniform image and (b) with an EM reconstruction.



(a) σ value (b) cost function

Fig. 12.

Effect of the initial value σ_{init} in the proposed algorithm with an EM image initial for the smooth R_1 with $\delta = 0.01$. (a) The evolution of σ values vs. iteration number; (b) Plot of the cost function vs. iteration number.



(a) whole image (b) tumor region

Fig. 13.

Effect of the hyper-parameter δ in the smooth R_1 on (a) MSE of the whole image and (b) MSE of the tumor region. The smooth R_1 regularization uses $\beta = 2^{-6}$ and is solved by the proposed algorithm. The MSE by the non-smooth R_1 using the preconditioned Chambolle-Pock algorithm at iteration 5000 is marked by x.

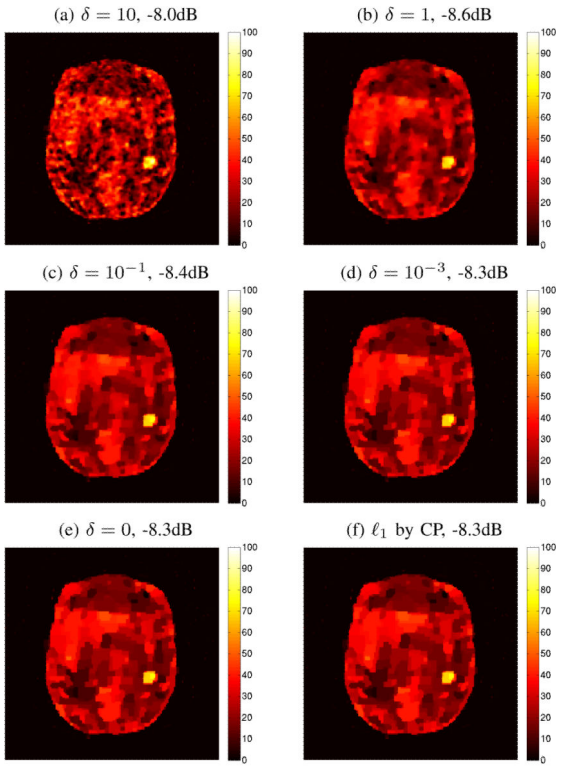
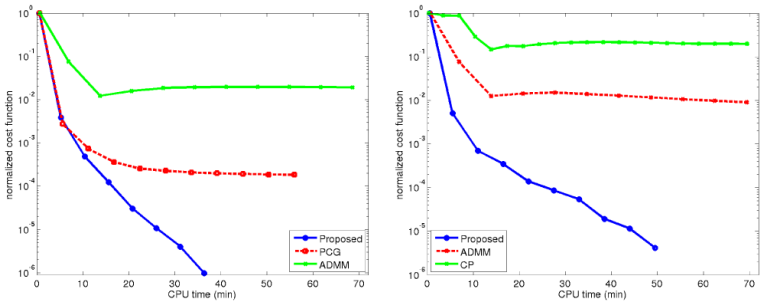


Fig. 14.

Reconstructed images by the smooth R_1 regularization with different δ values from 10 to 0 (a-e) and by the non-smooth R_1 regularization (f). The smooth regularizations were solved using the proposed algorithm with 200 iterations and the non-smooth R_1 regularization was solved using the CP algorithm with 5000 iterations. The image MSE values are given on top of each image.



(a) smooth $\ell_1, \delta = 0.001$ (b) non-smooth ℓ_1

Fig. 15.

Comparison of convergence speed of different algorithms for the PL reconstruction of the 3D brain data using (a) smooth R_1 and (b) non-smooth R_1 regularizations.