Simultaneous activity and attenuation estimation in TOF-PET with TV-constrained nonconvex optimization

Zhimei Ren, Emil Y. Sidky, *Member, IEEE*, Rina Foygel Barber, Chien-Min Kao, *Senior Member, IEEE*, and Xiaochuan Pan, *Fellow, IEEE*

Abstract— An alternating direction method of multipliers (ADMM) framework is developed for nonsmooth biconvex optimization for inverse problems in imaging. In particular, the simultaneous estimation of activity and attenuation (SAA) problem in time-of-flight positron emission tomography (TOF-PET) has such a structure when maximum likelihood estimation (MLE) is employed. The ADMM framework is applied to MLE for SAA in TOF-PET, resulting in the ADMM-SAA algorithm. This algorithm is extended by imposing total variation (TV) constraints on both the activity and attenuation map, resulting in the ADMM-TVSAA algorithm. The performance of this algorithm is illustrated using the penalized maximum likelihood activity and attenuation estimation (P-MLAA) algorithm as a reference. Additional results on step-size tuning and on the use of unconstrained ADMM-SAA are presented in the previous arXiv submission: arXiv:2303.17042v1.

Index Terms— Image reconstruction in TOF-PET, simultaneous activity/attenuation estimation, large-scale nonconvex optimization, alternating direction method of multipliers

I. INTRODUCTION

NUCLEAR medicine imaging modalities such as single-photon emission computed tomography (SPECT) and positron emission tomography (PET) require the input of a gamma ray attenuation map for quantitatively accurate imaging. The combination of nuclear medicine imaging with other image modalities such as X-ray computed tomography (CT) [1], [2] or magnetic resonance imaging (MRI) [3] provides a means for estimating the necessary attenuation map. There are, however, challenges in the separate attenuation map estimation. Use of CT-based attenuation maps requires extrapolation of the photon attenuation map from the diagnostic X-ray

Z. Ren is with the Dept. of Statistics and Data Science, University of Pennsylvania (zren@wharton.upenn.edu).
R. F. Barber is with the Dept. of Statistics, University of Chicago (rina@uchicago.edu).
E. Y. Sidky, C.-M. Kao, and X. Pan are with the Dept. of Radiology, University of Chicago (sidky@uchicago.edu, ckao95@uchicago.edu, xpan@uchicago.edu).

energy range to 511 keV and registration of the PET and CT imaging, which can be particularly difficult in the presence of motion [4]. The use of MRI to estimate a synthetic CT image is further complicated by the fact that bone and air have similar gray values in MRI while bone has a significantly higher attenuation coefficient for gamma rays.

To avoid a separate measurement for obtaining the gamma ray attenuation map, a long-standing inverse problem of interest has been to simultaneously estimate the attenuation and activity distributions from emission data alone [5], [6]. To address simultaneous activity and attenuation (SAA) estimation, Nuyts *et al.* [6] use maximum likelihood to invert the algebraic SAA model, and they find that accurate activity distributions can be recovered by appropriately regularizing the attenuation map. The regularization involves the use of Gibbs and intensity priors on the attenuation distribution that encourage local smoothness and clustering of values around known attenuation values for tissues in the scanned subject. Another interesting result for the SAA problem is obtained in considering time-of-flight positron emission tomography (TOF-PET) [7]. Defrise *et al.* [8] exploit an analytic range condition [9], [10] for the continuous TOF-PET model and obtain a uniqueness result that the attenuation factor and activity can be determined up to a multiplicative constant. Returning to the SAA algebraic model for TOF-PET, a comprehensive study of this inverse problem using maximum likelihood estimation is presented in Rezaei *et al.* [11], where it is found that the activity and attenuation maps can be recovered if the timing resolution of the TOF measurements is sufficiently high and if support constraints are exploited. We note an intriguing extension of the SAA problem where the background radiation from Lutetium-176, present in PET scintillators composed of either lutetium oxyorthosilicate (LSO) or lutetium-yttrium orthosilicate (LYSO), is exploited to provide additional information on the subject's attenuation map without the need for a separate scan [12]. Also, in the context of PET/MRI, anatomical information from standard MRI protocols can be used as a prior to inform the SAA estimation without the need for dedicated pulse sequences need for MR-based attenuation correction [13].

In this work, we seek to build off of Ref. [11] and develop an image reconstruction framework for the SAA problem in TOF-PET that can incorporate nonsmooth, convex constraints in the maximum likelihood estimation. Such constraints can help

to achieve stable inversion of the SAA estimation problem. Of particular interest, here, is the use of total variation (TV) constraints on both activity and attenuation distributions. We have previously exploited such constraints in the context of nuclear medicine imaging; in Refs. [14] and [15] TV constraints are exploited to enable sparse-data sampling configurations in SPECT and PET, respectively. In Ref. [16], a similar methodology is used for image reconstruction in low-count list-mode TOF-PET.

The image reconstruction algorithms developed in Refs. [14]–[16] are all instances of a general primal-dual (PD) solver for nonsmooth convex optimization developed by Chambolle and Pock [17], [18]. The optimization problem posed by applying TV-constraints to the SAA estimation problem, however, is nonsmooth and nonconvex. In our recent work, we develop a framework for such problems in imaging, where the optimization can be split into convex terms plus differentiable terms that are possibly nonconvex [19]. This framework is based on the alternating direction method of multipliers (ADMM) [20] in a way that is closely related to the PD algorithm. This framework has been successfully applied to the nonsmooth and nonconvex optimization problem that arises in spectral computed tomography (CT) when the spectral response of the measurement is included in the data model [21]. Here, we modify this framework to address biconvex optimization and apply it to the SAA estimation problem with convex constraints. The SAA data model and imaging problem are specified in Sec. II, where we then develop an ADMM algorithm to solve the associated optimization problem. As the focus of this work is mainly on the SAA inverse problem, we conduct a number of studies on noiseless TOF-PET data in Sec. III that explore the range of TOF-PET parameters that allow exact recovery of activity and attenuation factors. Also presented in this section are results with noisy data that demonstrate the stability of the proposed algorithm. In Sec. V the results are discussed and the conclusions of the work are given.

II. IMAGE RECONSTRUCTION MODEL AND ALGORITHMS

In presenting the SAA algorithm TOF-PET, we consider a two dimensional (2D) simulation where the lines-of-response (LORs) are organized in parallel-ray fashion and are specified in the same way that the 2D Radon transform is parameterized. For the TOF-PET model, the Radon transform is modified by including weighted line-integration that accounts for TOF information that helps to localize the positron-electron annihilation along a given LOR. After specifying the TOF-PET data model, the MLAA algorithm from Rezaei *et al.* [11] is briefly summarized. We then present the nonconvex ADMM algorithm that performs SAA estimation with nonsmooth convex constraints.

A. TOF-PET modelling

The measurement model for the mean data in TOF-PET is

$$c_{i\ell} = \exp\left[-P_\ell^\top \mu\right] \cdot T_{i\ell}^\top \lambda, \qquad (1)$$

where λ and μ are the unknown activity and attenuation maps, respectively; $T_{i\ell}$ is the TOF sensitivity image for TOF window i, LOR ℓ; P_ℓ is the X-ray projection matrix sensitivity image for LOR ℓ. For defining the TOF projection matrix T, the TOF window sensitivity along the LOR is specified as

$$w_i(t) = \exp[-(t - t_i)^2/(2\sigma_{\text{TOF}})],$$

where the sampling along the LOR is half of the full-width-half-maximum (FWHM) of this Gaussian distribution

$$\Delta t = t_{i+1} - t_i = \text{FWHM}/2 = \sqrt{2\log 2} \cdot \sigma_{\text{TOF}}.$$

For this work, scatter coincidences and random events are not considered.

B. Imaging model based on nonconvex optimization

We consider performing SAA using likelihood maximization, where the measured coincidence count data are assumed to follow a multivariate, mutually independent Poisson distribution

$$C_{i\ell} \sim \text{Poisson}(c_{i\ell}).$$

Equivalently, this estimation is performed by minimization of the negative log-likelihood,

$$l(\lambda, \mu) = \sum_{i\ell} \left\{ c_{i\ell} - C_{i\ell} \cdot \log c_{i\ell} \right\} = \qquad (2)$$

$$\sum_{i\ell} \left\{ \exp(-P_\ell^\top \mu) \cdot T_{i\ell}^\top \lambda - C_{i\ell} \cdot (-P_\ell^\top \mu + \log(T_{i\ell}^\top \lambda)) \right\}.$$

The optimization problem of interest is

$$\lambda, \mu = \arg\min_{\lambda,\mu} \left\{ l(\lambda, \mu) \mid \mathbf{1}^\top \lambda = N_{\text{total}}, \ \lambda, \mu \geq 0 \right\}, \quad (3)$$

where l is the negative log-likelihood in Eq. (2); $\mathbf{1}$ is a vector of size λ with unit entries so that $\mathbf{1}^\top \lambda$ is equivalent to summation over λ; and N_{total} is the total number of annihilations. The constraint on the total number of annihilations is used to overcome the constant ambiguity in the SAA estimation problem [8]. This constraint is enforced in this work instead of the object support constraint investigated in Rezaei *et al.* [11].

C. Summary of MLAA

To solve this imaging model, Rezaei *et al.* [11] developed the MLAA algorithm. For completeness, we write the MLAA update steps including a minor modification in Eq. (6) that accommodates the constraint on the total number of annihilations:

$$a_\ell = \exp\left[-\sum_k P_{\ell k}\mu_k\right] \quad \forall \ell, \qquad (4)$$

$$\lambda_k \leftarrow \frac{\lambda_k}{\sum_{i\ell} a_\ell T_{i\ell k}} \sum_{i\ell} \left\{ T_{i\ell k} \left(\frac{C_{i\ell}}{\sum_{k'} T_{i\ell k'}\lambda_{k'}} \right) \right\} \quad \forall k, \quad (5)$$

$$\lambda \leftarrow \lambda \left(\frac{N_{\text{total}}}{\sum_k \lambda_k} \right), \qquad (6)$$

$$\mu_k \leftarrow \mu_k + \frac{\sum_{i\ell k'} P_{\ell k} (a_\ell T_{i\ell k'}\lambda_{k'} - C_{i\ell})}{\sum_{i\ell k'} P_{\ell k'} P_{\ell k} a_\ell \sum_{k''} T_{i\ell k''}\lambda_{k''}} \quad \forall k, \qquad (7)$$

$$\mu_k \leftarrow \text{pos}(\mu_k) \quad \forall k. \qquad (8)$$

The MLAA algorithm essentially alternates between updating λ with a Poisson likelihood EM step and μ with a Poisson transmission likelihood optimization step. In this MLAA implementation the extra update step in Eq. (6) enforces the constraint on the total number of annihilations, and Eq. (8) performs non-negativity projection, where negative values of μ are set to zero. For MLAA the activity λ should have a strictly positive initialization, and this quantity will remain non-negative during the iteration.

Early stopping of the iteration is the primary means of performing regularization with MLAA, but explicit regularization can also be included with the use of Gibbs smoothing [22]–[24]. In this work, we develop a framework for SAA which can include nonsmooth regularization.

D. ADMM for nonsmooth and biconvex optimization

The general convex optimization problem that ADMM solves takes the form

$$\min_{x,y} \left\{ f(x) + g(y) \mid Ax + By = c \right\},$$

where f and g are convex and possibly non-smooth functions; A and B are linear operators; x, y and c are vectors. The steps of the ADMM algorithm are

$$x \leftarrow \arg\min_{x'} \left\{ f(x') + u^\top Ax' \right.$$
$$\left. + \tfrac{1}{2}\|Ax' + By - c\|_\Sigma^2 + \tfrac{1}{2}\|x' - x\|_{H_f}^2 \right\} \quad (9)$$
$$y \leftarrow \arg\min_{y'} \left\{ g(y') + u^\top By' \right.$$
$$\left. + \tfrac{1}{2}\|Ax + By' - c\|_\Sigma^2 + \tfrac{1}{2}\|y' - y\|_{H_g}^2 \right\} \quad (10)$$
$$u \leftarrow u + \Sigma(Ax + By - c), \quad (11)$$

where Σ, H_f, and H_g are symmetric positive definite, and $\|v\|_M^2 \equiv v^\top M v$ for any symmetric positive definite matrix M. Because optimizing the TOF-PET likelihood for SAA is a non-convex optimization problem, the ADMM algorithm does not directly apply. One strategy to adapt ADMM to SAA is to base the ADMM steps on a series of successive convex approximations as developed by Chun *et al.* [25] using the separable quadratic surrogates (SQS) method. In the present work, we develop an alternative form of ADMM that directly applies to SAA, exploiting the biconvex structure of the TOF-PET likelihood function; i.e. fixing either λ or μ, the likelihood is a convex function in the other variable.

The ADMM algorithm can be modified to accommodate a biconvex function, and we consider the case that only g is a biconvex function

$$g(y) = g(y_1, y_2),$$

where y is the concatenation of y_1 and y_2; and $g(y_1, \cdot)$ and $g(\cdot, y_2)$ are convex functions for fixed y_1 and y_2, respectively. To accommodate the biconvexity of g, the second update equation, Eq. (10), is replaced by an inner iteration with the following update equations

$$y_1 \leftarrow \arg\min_{y_1'} \left\{ g(y_1', y_2) + u^\top B(y_1', y_2) \right. \quad (12)$$
$$\left. + \tfrac{1}{2}\|Ax + B(y_1', y_2) - c\|_\Sigma^2 + \tfrac{1}{2}\|(y_1', y_2) - (y_1, y_2)\|_{H_g}^2 \right\}$$
$$y_2 \leftarrow \arg\min_{y_2'} \left\{ g(y_1, y_2') + u^\top B(y_1, y_2') \right. \quad (13)$$
$$\left. + \tfrac{1}{2}\|Ax + B(y_1, y_2') - c\|_\Sigma^2 + \tfrac{1}{2}\|(y_1, y_2') - (y_1, y_2)\|_{H_g}^2 \right\}.$$

The inner loop consists of alternating between Eqs. (12) and (13) for a predetermined number of iterations N_y, where $N_y \geq 1$. After the inner loop is completed, the ADMM iteration continues with Eq. (11) after the following assignment

$$y = (y_1, \ y_2).$$

This inner loop, specified in Eqs. (12) and (13), is computationally efficient if multiplication by the matrix B is efficient; this is the case in our application because we consider $B = I$ where I is the identity matrix. Note that multiplication by A is not performed within this inner iteration because the matrix A only appears in the term Ax which is computed before entering the inner loop.

E. ADMM for large-scale tomographic image reconstruction

For the large-scale optimization problems that arise in tomographic image reconstruction, the update step in Eq. (9) can be problematic because of the term Ax, which appears in the minimization over x. The matrix A usually contains the system matrix for the imaging model, and computation of Ax can be expensive particularly for 3D imaging; thus numerical solution of Eq. (9) may not be feasible. This "expensive inner loop" problem can be circumvented by linearization, i.e. by including the additional term $\tfrac{1}{2}\|x' - x\|_{H_f}^2$ in Eq. (9) [19], [26], resulting in an algorithm closely related to the primal-dual (PD) algorithm of Chambolle and Pock [17], [18]. Considering only scalar step size parameters, i.e.

$$\Sigma = \sigma I,$$

the metric H_f in Eq. (9) is set to

$$H_f = I/\tau - \sigma A^\top A. \quad (14)$$

This choice cancels the Ax' term in Eq. (9), and the requirement that H_f be positive definite yields a constraint on the step sizes σ and τ. In the context of the image reconstruction problem, we also have

$$H_g = 0; \quad B = -I; \quad c = 0.$$

The ADMM generic optimization problem becomes

$$\min_{x,y} \left\{ f(x) + g(y) \mid Ax - y = 0 \right\}, \quad (15)$$

and the algorithm for convex optimization is then specified by the following update equations

$$x \leftarrow \arg\min_{x'}\Big\{ f(x') + {x'}^\top A^\top(u + \sigma(Ax - y))$$

$$+ \tfrac{1}{2\tau}\|x' - x\|^2 \Big\} \tag{16}$$

$$y \leftarrow \arg\min_{y'}\Big\{ g(y') - u^\top y' + \tfrac{\sigma}{2}\|Ax - y'\|^2 \Big\} \tag{17}$$

$$u \leftarrow u + \sigma(Ax - y). \tag{18}$$

Aside from minor details, this set of update equations is equivalent to the PD algorithm, but as a starting point to modify the update steps for non-convex optimization, this form is more convenient because both f and g functions appear directly in the updates. In contrast, the PD algorithm dualizes g and the convex conjugate $g^\star$ is needed. If it is desired to apply PD to non-convex g, figuring out what to put in place of $g^\star$, while possible [27], adds another layer of complication to the algorithm development.

The modification of the linearized ADMM updates for addressing the case where g is biconvex replaces Eq. (17) with inner loop update equations

$$y_1 = \arg\min_{y_1'}\Big\{ g(y_1', y_2) - u^\top(y_1', y_2) \tag{19}$$

$$+ \tfrac{\sigma}{2}\|Ax - (y_1', y_2)\|^2 \Big\}$$

$$y_2 = \arg\min_{y_2'}\Big\{ g(y_1, y_2') - u^\top(y_1, y_2') \tag{20}$$

$$+ \tfrac{\sigma}{2}\|Ax - (y_1, y_2')\|^2 \Big\}.$$

Convergence of this modified ADMM algorithm for biconvex functions is not theoretically guaranteed and thus convergence is demonstrated empirically.

F. ADMM for SAA in TOF-PET

The instantiation of ADMM for SAA estimation by minimization of the negative log-likelihood is covered here in detail. The optimization problem of interest, restated from Eq. (3), is

$$\lambda, \mu = \arg\min_{\lambda,\mu}\left\{ l(\lambda, \mu) \mid \mathbf{1}^\top\lambda = N_{\text{total}}, \ \lambda \geq 0, \ \mu \geq 0 \right\}. \tag{21}$$

In this sub-section, we map this optimization problem on to the ADMM algorithm, derive the x-update and biconvex y-updates, and provide the pseudo-code for SAA estimation.

To map the optimization problem in Eq. (21) onto the generic ADMM optimization in Eq. (15), the primal, splitting, and dual variables x, y, and u, are respectively assigned as

$$x = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}, \quad y = \begin{pmatrix} y_\lambda \\ y_\mu \end{pmatrix}, \quad u = \begin{pmatrix} u_\lambda \\ u_\mu \end{pmatrix}.$$

The linear system A is assigned as

$$A = \begin{pmatrix} T & 0 \\ 0 & P \end{pmatrix}.$$

The convex function f is used to represent the non-negativity constraints and the constraint on the total number of annihilations by setting

$$f(\lambda, \mu) = \delta(\mathbf{1}^\top\lambda = N_{\text{total}}) + \delta(\lambda \geq 0) + \delta(\mu \geq 0), \tag{22}$$

where δ is the convex indicator function, which is zero if the conditional argument is true and infinity otherwise. The biconvex function g accounts for the negative log-likelihood objective function in Eq. (21)

$$g(y_\lambda, y_\mu) = L(y_\lambda, y_\mu), \tag{23}$$

$$L(y_\lambda, y_\mu) = \sum_{i\ell}\Big\{ \exp(-y_{\mu,\ell}) \cdot y_{\lambda, i\ell}$$

$$- C_{i\ell} \cdot (-y_{\mu,\ell} + \log(y_{\lambda, i\ell})) \Big\},$$

where

$$l(\lambda, \mu) = L(T\lambda, P\mu).$$

Parametrization of the step sizes: Step size selection is a critical issue for first-order, large-scale optimization algorithms. There can be much flexibility in the step size selection, and it is important to select a minimal set of free parameters that are effective for algorithm efficiency but not too cumbersome in the tuning procedure. Because the system matrix A for SAA is block-diagonal, a slight generalization of the ADMM linearization is considered. The metric H_f is written as

$$H_f = \begin{pmatrix} H_\lambda & 0 \\ 0 & H_\mu \end{pmatrix},$$

$$H_\lambda = \frac{I}{\tau_\lambda} - \sigma_\lambda T^\top T,$$

$$H_\mu = \frac{I}{\tau_\mu} - \sigma_\mu P^\top P,$$

and the step size parameters are chosen according to

$$\sigma_\lambda\tau_\lambda = 1/\|T\|_2^2, \quad \sigma_\mu\tau_\mu = 1/\|P\|_2^2,$$

where $\|M\|_2$ is the largest singular value of the matrix M. With four step size parameters and two equality constraints, there are two free step size parameters. Specifically, the step size ratios, ρ_λ and ρ_μ, are chosen to be the free parameters that need to be tuned:

$$\sigma_\lambda = \rho_\lambda/\|T\|_2, \quad \tau_\lambda = 1/(\rho_\lambda\|T\|_2), \tag{24}$$

$$\sigma_\mu = \rho_\mu/\|P\|_2, \quad \tau_\mu = 1/(\rho_\mu\|P\|_2). \tag{25}$$

Tuning of ρ_λ and ρ_μ is a necessary step any time the T or P matrices are changed due to, for example, a change in scan configuration or sampling pattern.

The x-update: For the SAA problem in TOF-PET the x-update in Eq. (16) splits into two optimization problems

$$\lambda \leftarrow \arg\min_{\lambda'}\Big\{ {\lambda'}^\top T^\top(u_\lambda + \sigma_\lambda(T\lambda - y_\lambda)) \tag{26}$$

$$+ \tfrac{1}{2\tau_\lambda}\|\lambda' - \lambda\|^2 \mid \mathbf{1}^\top\lambda' = N_{\text{total}}, \lambda' \geq 0 \Big\},$$

$$\mu \leftarrow \arg\min_{\mu'}\Big\{ {\mu'}^\top P^\top(u_\mu + \sigma_\mu(P\mu - y_\mu)) \tag{27}$$

$$+ \tfrac{1}{2\tau_\mu}\|\mu' - \mu\|^2 \mid \mu' \geq 0 \Big\},$$

where the indicator terms of the convex function f from Eq. (22) are incorporated as constraints in the λ- and μ-update equations. The optimization problem for the μ-update in Eq. (27) is solved by setting the gradient of the objective function to zero and solving for μ', followed by a non-negativity projection to enforce the constraint on μ_I

$$\mu \leftarrow \mu - \tau_\mu P^\top \bar{u}_\mu, \qquad (28)$$
$$\bar{u}_\mu = u_\mu + \sigma_\mu(P\mu - y_\mu),$$
$$\mu \leftarrow \mathrm{pos}(\mu),$$

where the function $\mathrm{pos}(\cdot)$ thresholds negative components of the argument to zero.

For the λ optimization problem in Eq. (26), completing the square in the quadratic objective function, rescaling the objective function, and ignoring the λ'-independent term yields

$$\lambda \leftarrow \arg\min_{\lambda'}\left\{ \tfrac{1}{2}\left\| \lambda' - (\lambda - \tau_\lambda T^\top \bar{u}_\lambda) \right\|^2 \right. \qquad (29)$$
$$\left. \mid \mathbf{1}^\top \lambda' = N_{\mathrm{total}},\ \lambda' \geq 0 \right\},$$
$$\bar{u}_\lambda = u_\lambda + \sigma_\lambda(T\lambda - y_\lambda).$$

This optimization problem is now in the form of a projection onto the positive simplex, for which an efficient algorithm is developed by Duchi *et al.* [28]. Computationally, the updates in λ and μ are the most expensive steps in the ADMM algorithm because they involve forward- and back-projection of μ and $\bar{u}_\mu$, respectively, in addition to TOF forward- and back-projection of λ and $\bar{u}_\lambda$, respectively.

The biconvex y-updates: The g function in Eq. (23) is biconvex in that it is convex in y_λ if y_μ is fixed and *vice versa*. Splitting up the g function over the two update equations in Eqs. (19) and (20) yields

$$y_\lambda = \arg\min_{y'_\lambda}\left\{ \sum_{i\ell} \left(\exp(-y_{\mu,\ell}) \cdot y'_{\lambda,i\ell} - C_{i\ell} \cdot \log(y'_{\lambda,i\ell}) \right) \right.$$
$$\left. - u_\lambda^\top y'_\lambda + \frac{\sigma_\lambda}{2}\|y'_\lambda - T\lambda\|^2 \mid y'_\lambda \geq 0 \right\}, \qquad (30)$$

and

$$y_\mu = \arg\min_{y'_\mu}\left\{ \sum_{i\ell} \left(\exp(-y'_{\mu,\ell}) \cdot y_{\lambda,i\ell} + C_{i\ell} \cdot y'_{\mu,\ell} \right) \right.$$
$$\left. - u_\mu^\top y'_\mu + \frac{\sigma_\mu}{2}\|y'_\mu - P\mu\|^2 \mid y'_\mu \geq 0 \right\}, \qquad (31)$$

noting that the $\exp(-y_\mu) \cdot y_\lambda$ term is the only one that mixes the y_λ and y_μ variables and is therefore common to both minimization problems. In order for the biconvex alternation to converge it is necessary to introduce the non-negativity constraints on y'_λ and y'_μ. Physically, these constraints are redundant with the non-negativity constraints imposed on λ and μ; if these physical constraints are not used, it is still necessary to impose non-negativity constraints on $T\lambda$ and $P\mu$.

The minimization problems for the y-update are both separable over the components of y'_λ and y'_μ. The minimization over y'_λ in Eq. (30) is solved analytically by setting the gradient of the objective function to zero, yielding a quadratic equation when $C_{i\ell} > 0$,

$$\sigma_\lambda y^2_{\lambda,i\ell} - b_{i\ell} y_{\lambda,i\ell} - C_{i\ell} = 0,$$
$$\text{where } b_{i\ell} = u_{\lambda,i\ell} + \sigma_\lambda T_{i\ell}^\top \lambda - \exp(-y_{\mu,\ell}),$$

and a linear equation when $C_{i\ell} = 0$,

$$\sigma_\lambda y_{\lambda,i\ell} - b_{i\ell} = 0.$$

For the $C_{i\ell} > 0$ case, the non-negativity constraint on y'_λ is respected by selecting the non-negative root of the corresponding quadratic equation, and for the $C_{i\ell} = 0$ case, the non-negativity constraint on y'_λ yields an update,

$$y_{\lambda,i\ell} = \max(b_{i\ell}/\sigma_\lambda, 0).$$

Both the linear and quadratic cases can be merged into the following update equation for $y_{\lambda,i\ell}$,

$$y_{\lambda,i\ell} = \left(b_{i\ell} + \sqrt{b_{i\ell}^2 + 4\sigma_\lambda C_{i\ell}} \right)/(2\sigma_\lambda). \qquad (32)$$

Solving the minimization over y'_μ in Eq. (31) is more involved because setting the gradient of the objective function to zero results in a transcendental equation, which requires the use of a numerical solver. The objective function is convex in y'_μ and its derivatives are easily computed analytically. Thus Newton's algorithm can be applied to obtain an efficient and accurate solution to Eq. (31). Both the first and second derivatives of the objective function are needed for Newton's algorithm. Defining ψ to be the objective function of Eq. (31)

$$\psi(y'_\mu) = \sum_{i\ell} \left(\exp(-y'_{\mu,\ell}) \cdot y_{\lambda,i\ell} + C_{i\ell} \cdot y'_{\mu,\ell} \right)$$
$$- u_\mu^\top y'_\mu + \frac{\sigma_\mu}{2}\|y'_\mu - P\mu\|^2,$$

the first derivative of ψ is

$$\frac{\partial\psi(y'_\mu)}{\partial y'_{\mu,\ell}} = -\exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell}$$
$$+ C_\ell - u_\mu + \sigma_\mu(y'_{\mu,\ell} - P_\ell^\top \mu), \quad (33)$$

where

$$y_{\lambda,\ell} = \sum_i y_{\lambda,i\ell}, \quad C_\ell = \sum_i C_{i\ell}.$$

The second derivative of ψ is

$$\frac{\partial^2\psi(y'_\mu)}{\partial y'^2_{\mu,\ell}} = \exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell} + \sigma_\mu, \qquad (34)$$

which is strictly positive. Thus Newton's algorithm can be applied without any difficulties with the following update equation

$$y'_{\mu,\ell} \leftarrow y'_{\mu,\ell} - \frac{\partial\psi(y'_\mu)}{\partial y'_{\mu,\ell}}\left(\frac{\partial^2\psi(y'_\mu)}{\partial y'^2_{\mu,\ell}} \right)^{-1}. \qquad (35)$$

There is also the non-negativity constraint in Eq. (31), and this can be accounted for by thresholding negative values of $y'_{\mu,\ell}$ to zero after the Newton iteration is completed.

The proposed y-update involves two additional levels of iteration. The first additional level of iteration involves alternating between solving Eqs. (30) and (31). In the second additional level of iteration Eq. (31) is solved with the Newton iteration in Eq. (35). Nevertheless, these additional nested iterations do not negatively impact the efficiency of the overall algorithm because all of the iterations for the y-update separate over the

components of y. The complete y-update computation takes less effort than computing $T\lambda$, the TOF data of an estimate of the activity map, λ. This is one of the useful aspects of the powerful splitting technique that ADMM exploits.

Algorithm 1 ADMM pseudocode for SAA estimation with biconvex optimization. Variables λ, μ, y_λ, y_μ, $\bar{y}_\lambda$, $\bar{y}_\mu$, u_λ, and u_μ are initialized to zero. Step size ratio parameters ρ_λ and ρ_μ are chosen, and step size parameters σ_λ, σ_μ, τ_λ, and τ_μ are determined according to Eqs. (24) and (25).

1: **for** $k \leftarrow 1, N_{\text{iter}}$ **do**
2:      $\bar{\lambda} = T^\top(u_\lambda + \sigma_\lambda(\bar{y}_\lambda - y_\lambda))$
3:      $\lambda \leftarrow \mathcal{P}_{\text{simplex}(N_{\text{total}})}\left(\lambda - \tau_\lambda\bar{\lambda}\right)$
4:      $\bar{y}_\lambda = T\lambda$
5:      $\bar{\mu} = P^\top(u_\mu + \sigma_\mu(\bar{y}_\mu - y_\mu))$
6:      $\mu \leftarrow \text{pos}(\mu - \tau_\mu\bar{\mu})$
7:      $\bar{y}_\mu = P\mu$
8:      $v_{\mu,\ell} = \sum_i C_{i\ell} - (u_{\mu,\ell} - \sigma_\mu\bar{y}_{\mu,\ell}) \;\; \forall\ell$
9:      **for** $k' \leftarrow 1, N_y$ **do**   ▷ Biconvex alternation loop
10:          $b_{i\ell} = u_{\lambda,i\ell} + \sigma_\lambda\bar{y}_{\lambda,i\ell} - \exp(-y_{\mu,\ell}) \;\; \forall i,\ell$
11:          $y_{\lambda,i\ell} = \left(b_{i\ell} + \sqrt{b_{i\ell}^2 + 4\sigma_\lambda C_{i\ell}}\right)/(2\sigma_\lambda) \;\; \forall i,\ell$
12:          $y_{\lambda,\ell} = \sum_i y_{\lambda,i\ell} \;\; \forall\ell$
13:          $y'_{\mu,\ell} = 0 \;\; \forall\ell$ ▷ Initialize Newton iteration
14:          **for** $k'' \leftarrow 1, N_{\text{newt}}$ **do** ▷ Loop for solving Eq. (31)
15:              $\psi_\ell^{(1)} = -\exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell} + \sigma_\mu y'_{\mu,\ell} + v_{\mu,\ell}$
16:              $\psi_\ell^{(2)} = \exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell} + \sigma_\mu \;\; \forall\ell$
17:              $y'_{\mu,\ell} \leftarrow y'_{\mu,\ell} - \psi_\ell^{(1)} \cdot \left(\psi_\ell^{(2)}\right)^{-1} \;\; \forall\ell$
18:              $y'_{\mu,\ell} \leftarrow \text{pos}(y'_{\mu,\ell}) \;\; \forall\ell$ ▷ Nonneg., Eq. (31)
19:          **end for**
20:          $y_{\mu,\ell} = y'_{\mu,\ell} \;\; \forall\ell$
21:      **end for**
22:      $u_\lambda \leftarrow u_\lambda + \sigma_\lambda(\bar{y}_\lambda - y_\lambda)$
23:      $u_\mu \leftarrow u_\mu + \sigma_\mu(\bar{y}_\mu - y_\mu)$
24: **end for**

ADMM pseudocode for SAA estimation: The x-, y-, and u-update equations are assembled into a complete pseudocode given in Algorithm 1. The expensive projection and back-projection computations are collected in as few lines as possible, and their results stored, to avoid unnecessary repetition of these burdensome operations. The simplex projection at line 3 is the optimization problem defined in Eq. (29); efficient computer code for implementing this projection is available from Duchi *et al.* [28]. The first derivative computation from Eq. (33) is performed at lines 8 and 15, where line 8 collects all terms that are not dependent on y_λ or y'_μ. The function $\text{pos}(\cdot)$ in lines 6 and 18 returns the argument if it is non-negative, otherwise it returns zero. For the results presented in this work, we only consider zero initialization for all of the algorithm variables. The choice of step size ratios ρ_λ and ρ_μ will impact the convergence rate of the algorithm, and these parameters must be tuned for optimal performance.

G. ADMM for TV-constrained SAA in TOF-PET

The proposed ADMM framework for solving SAA estimation in TOF-PET allows for great flexibility in imposing convex constraints in the imaging optimization problem. Accordingly, we augment the total annihilation count and nonnegativity constraints in Eq. (21) with additional total variation constraints on the activity and attenuation maps

$$\lambda, \mu = \underset{\lambda,\mu}{\arg\min}\{l(\lambda,\mu) \mid \|\lambda\|_{\text{TV}} \leq \gamma_\lambda, \;\; \|\mu\|_{\text{TV}} \leq \gamma_\mu,$$
$$\mathbf{1}^\top\lambda = N_{\text{total}}, \;\; \lambda \geq 0, \;\; \mu \geq 0\}, \quad (36)$$

where $\|\cdot\|_{\text{TV}}$ is the isotropic TV seminorm; γ_λ and γ_μ are the TV constraint values for the activity and attenuation maps, respectively. The additional TV constraints exploit gradient sparsity in in both the activity and attenuation that potentially improves accurate estimation of their corresponding images.

Because the novel aspect of this work is the treatment of the biconvex log-likelihood term, which is explained in detail in Sec. II-F, the ADMM instance for this optimization problem is covered in the Appendix. The ADMM algorithm for TV-constrained SAA estimation (ADMM-TVSAA) is also designed so that it makes use of the same step size ratio parameters as discussed for Algorithm 1. Because of the additional constraints, the TV constraint values γ_λ and γ_μ become additional parameters of the algorithm.

H. Step size scaling of ADMM for SAA

When tuning the step size parameters ρ_λ and ρ_μ in ADMM-SAA or ADMM-TVSAA, it is important to account for the fact that the the optimal settings for maximum algorithm efficiency change with scaling of the coincidence data C. A practical consequence is that optimally efficient ρ-values would depend on, for example, collection time of the TOF-PET system. In the following results presented in Sec. III, this issue is addressed by normalizing the data C with the factor $\text{size}(C)/\|C\|_2$. The choice of data normalization is arbitrary because of the following scaling relationship. If we replace the coincidence data C by aC, the same ADMM iterates, up to a scaling, can be obtained by adjusting the constraint and step size parameters as follows:

$$N_{\text{total}} \to aN_{\text{total}}, \;\; \gamma_\lambda \to a\gamma_\lambda,$$
$$\sigma_\lambda \to \sigma_\lambda/a, \;\; \tau_\lambda \to a\tau_\lambda, \;\; \sigma_\mu \to a\sigma_\mu, \;\; \tau_\mu \to \tau_\mu/a.$$

With this scaling of the algorithm parameters, the algorithm variables transform as follows:

$$\lambda \to a\lambda, \;\; y_\lambda \to ay_\lambda, \;\; u_\lambda \to u_\lambda,$$
$$\mu \to \mu, \;\; y_\mu \to y_\mu, \;\; u_\mu \to au_\mu.$$

The scaling transformations can be verified by making the appropriate substitutions into Eqs. (18), (26), (27), (30), and (31), or into the update equations of Algorithm 1.

I. Huber-penalized MLAA

Use of ADMM allows for nonsmooth terms in the optimization such as use of the TV-norms and complex constraints
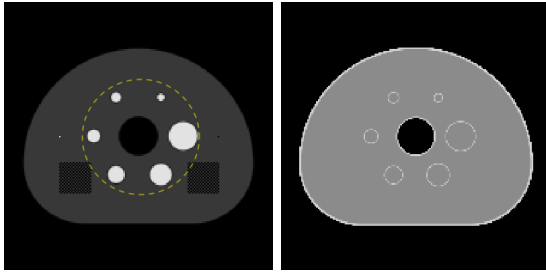
Fig. 1. (left) Slice number 40 from the University of Washington Digital Reference Object: activity image in arbitrary units, and (right) attenuation map displayed in the gray scale window $[0.075, 0.115]$ cm^{-1}. The dashed circle in the activity image indicates the activity distribution used for the investigation of SAA with interior data.
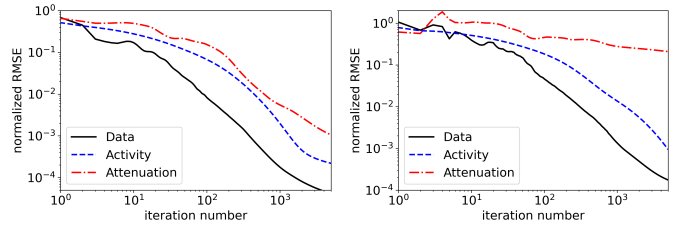

Fig. 2. Convergence of ADMM-TVSAA with noiseless TOF-PET data for the case of the full activity distribution (left) and the interior activity distribution (right). The data RMSE is normalized to the mean value of the TOF-PET data, and the activity/attenuation RMSEs are normalized to the mean values of their respective images.

on the activity and attenuation. The closest comparison from the literature involves a smooth objective function with Huber penalties on the activity and attenuation. Accordingly comparison results are obtained with the penalized-MLAA (P-MLAA) algorithm presented in Mehranian *et al.* [24]. The P-MLAA algorithm implemented here addresses the following optimization problem Huber penalties

$$\lambda, \mu = \arg\min_{\lambda, \mu} \{ l(\lambda, \mu) + \beta H_\delta(\lambda) + \gamma H_\delta(\mu) \mid$$
$$\mathbf{1}^\top \lambda = N_{\text{total}}, \ \lambda \geq 0, \ \mu \geq 0 \}, \quad (37)$$

where $H_\delta(\cdot)$ is the Huber penalty with smoothing parameter δ, see for example Nuyts *et al.* [6] for the definition of this penalty function. The P-MLAA algorithm replaces Eq. (5) with the "one step late" update equation developed by Green [30] and introduces a Gibbs prior into Eq. (7) in the manner developed in Nuyts *et al.* [6].

III. RESULTS WITH A 2D TOF-PET SIMULATION

The results demonstrating the ADMM-SAA algorithm are all derived from a 2D simulation using the digital reference object (DRO) shown in Fig. 1 [31]. This digital phantom is cropped to 176x176 image array with physical dimension 30x30 cm^2. The LORs are arranged in a 2D parallel-beam geometry with 176 views covering a π radian arc, and 176 parallel rays being measured per view with a spacing of 0.17 cm ($30/176 \approx 0.17$). The TOF FWHM is taken to be 4.5 cm, which corresponds to a timing resolution of approximately 300 picoseconds. The spacing between TOF window samples is 2.25 cm, and a total of 17 TOF samples are taken per LOR. For the image reconstruction, both the attenuation and activity images are represented on a 176x176 grid. The purpose of the presented results is to demonstrate usage of the ADMM-SAA algorithm and the impact of the TV constraints on the reconstruction of the activity and attenuation.

For the following results, the biconvex alternation loop at line 9 of Algorithm 1 is run for $N_y = 100$ iterations, and the Newton solver at line 14 is run for $N_{\text{newt}} = 10$ iterations. With both of these loop settings, Eq. (23) is solved accurately in a numerical sense. Even with the inner loops being executed with such high iteration numbers, the efficiency of the whole biconvex alternation loop is still high, because all of the computations separate across the vector components.

The computational effort for the biconvex alternation loop is $O(N_y \cdot N_{\text{TOF}} \cdot N_{\text{views}} \cdot \sqrt{N_{\text{pix}}})$ (the Newton loop does not increase the order of this loop because it involves the attenuation sinogram only), and by comparison, computing TOF projection, $T\lambda$, is $O(N_{\text{TOF}} \cdot N_{\text{views}} \cdot N_{\text{pix}})$, where N_{pix} is the total number of pixels and N_{views} is the number of projection angles. For the small 176x176 images of this study the biconvex loop is of the same order as TOF projection because $N_y \approx \sqrt{N_{\text{pix}}}$, but as the data and image size increase, TOF projection becomes the more burdensome computation. It is also possible, in practice, to reduce N_y and N_{newt} and work with inexact solution of Eq. (23) but we do not investigate this option in this work.

A. SAA from noiseless data

Image reconstruction is performed on noiseless data using the mean counts as the measured data, and ADMM-TVSAA is employed to study the effectiveness for solving the associated inverse problem for two situations: (1) the full activity distribution from the DRO phantom is used as the test object, and (2) the activity distribution is truncated at the yellow, dashed circle in Fig. 1. The second case is a more challenging inverse problem for SAA because recovery of the full attenuation map is complicated by the fact that only LORs that intersect the non-zero activity provide useful data, and accordingly recovery of the attenuation map has a similar degree of difficulty as the interior problem in tomography.

The convergence results for 5000 iterations of ADMM-TVSAA on the phantom with the full activity distribution and the truncated (or interior) activity distribution are shown in Fig. 2. In both cases, the normalized data RMSE and activity RMSE is observed to steadily converge to zero although the activity RMSE is noted to converge more slowly than the data RMSE. Recovering the true attenuation is clearly more challenging; for the full activity distribution case the attenuation RMSE exhibits convergent behavior all the way to the last computed iteration, but for the truncated activity case the global attenuation RMSE does not demonstrate convergent behavior. For truncated activity, however, the attenuation factor derived from the inaccurate attenuation map must have high degree of accuracy since the activity distribution is nicely recovered.

A series of image estimates are shown for the full activity in Fig. 3 and only the 5000th iteration results are shown
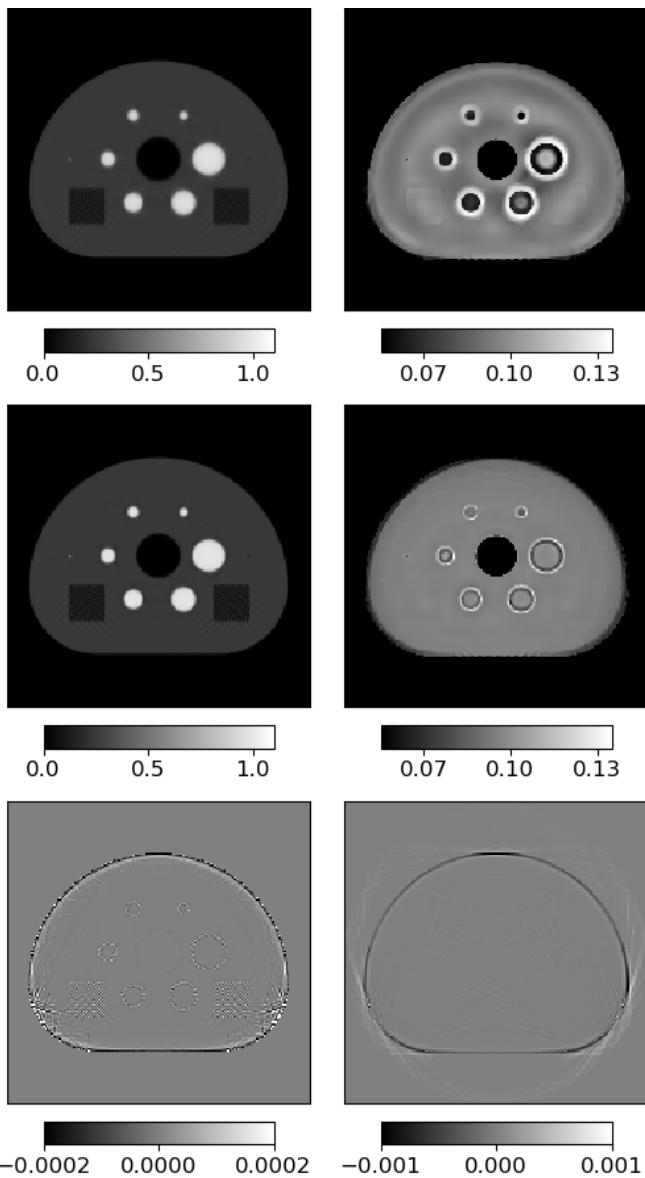
Fig. 3. Reconstructed activity (left column) and attenuation (right column) images from noiseless data with ADMM-TVSAA at 50 (top row), 100 (middle row), and 5000 (bottom row) iterations. Because the result at 5000 iterations is visually indistinguishable from the test phantom the difference from the ground truth is displayed in the bottom row. The activity distribution is normalized to 1.0 for the maximum value.

Fig. 4. For the results with the interior activity distribution we only show iteration 5000 for the activty (left) and attenuation (right). The actual activity/attenuation images are shown in the top row, and the difference from ground truth is shown in the bottom row.

Fig. 5. Normalized standard deviation versus normalized bias of the activity (left) and attenuation (right) images as a function of iteration number computed empirically from 100 noise realizations for MLAA, MLAA with Huber penalties, and TV-constrained SAA. Normalization of bias and standard deviation is achieved by dividing by the mean value of the corresponding ground truth image. The labeled dots indicate the iteration numbers for the respective algorithm curves. For TV-constrained ADMM-SAA, curves are shown for activity and attenuation TV constraints set to $\gamma_\lambda = 1.0$ and $\gamma_\mu = 1.0$, respectively, where the constraint values are given as a fraction of the ground truth TV values. The Huber penalty parameters for P-MLAA are set so that the resulting activity and attenuation images have nearly the same TV values as the ground truth after 500 iterations.

for truncated activity in Fig. 4. The difference images at 5000 iterations shows accurate reconstruction of both activity and attenuation for the full distribution case. The attenuation difference in the bottom panels reveals error at the 1% level for the attenuation map near the edges of the phantom. The panels at the earlier iterations provide a sense of the convergence to the solution; specifically the activity images at 50 or 100 iterations closely resemble the result at 5000 iterations. The attenuation map clearly converges more slowly. For the case of truncated activity shown in Fig. 4, the activity image is accurately recovered but it is clear that the attenuation map is not completely recovered. Interestingly, ADMM-TVSAA does seem to be able to recover the support of the attenuation map even if there is substantial error in the outer portions of the
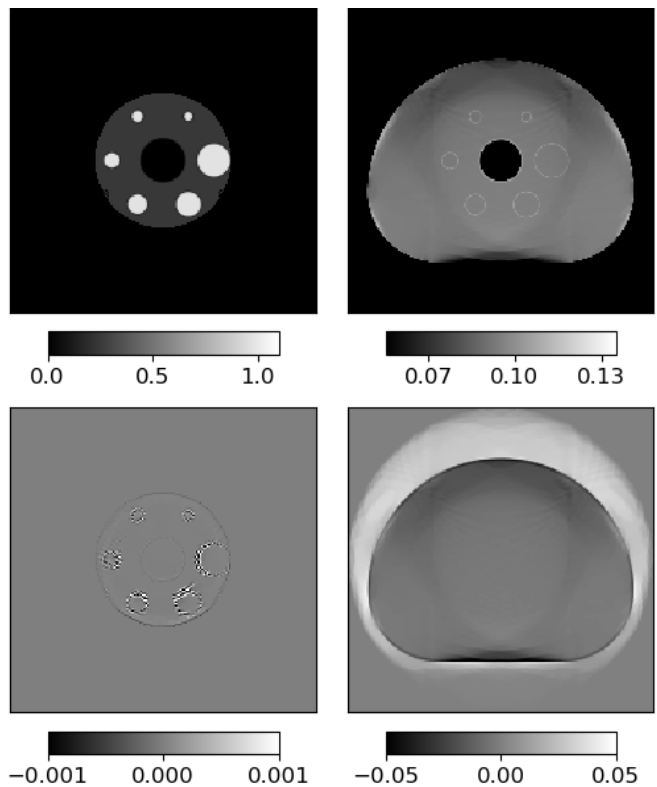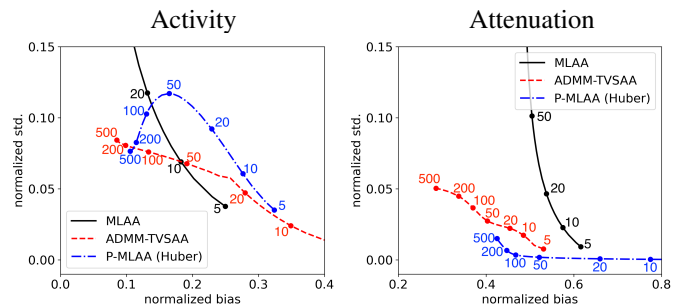
image. Moreover, the central portion of the attenuation image in the location where the activity is non-zero does appear to be reconstructed accurately.

B. SAA from noisy data

The next set of studies focus on SAA with noisy data and only the case of the full activity distribution is considered.
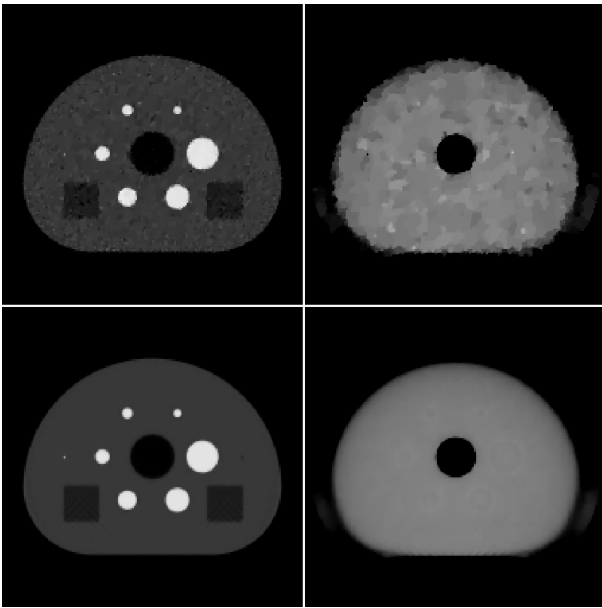
Fig. 6. Reconstructed activity (left column) and attenuation (right column) images from noisy data with ADMM-TVSAA. The top row shows a reconstructed set of images from a single noise realization at 500 iterations, and the bottom row shows the corresponding mean over 100 noise realizations. With the activity distribution normalized to 1.0 for the maximum value, the gray scale for the activity images is $[0, 1.1]$, and the gray scale for the attenuation is $[0.055, 0.135]$.
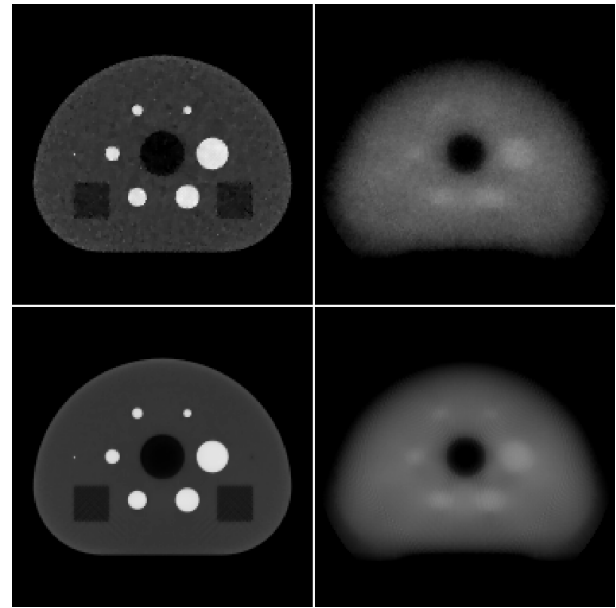
Fig. 7. Same as Fig. 6 except that the Huber-regularized MLAA is used to generate the image iterates.

Noise realizations are obtained by scaling the mean TOF-PET data so that the total number of measured coincidences is 4×10^6; the realization is then obtained by selecting a number of detected coincidences for each time-window sample and LOR, drawn from a Poisson distribution. In demonstrating the use of ADMM-TVSAA, two forms of MLAA provide reference algorithms. The MLAA algorithm described in Sec. II-C is one of the references, where early stopping provides regularization. The P-MLAA algorithm with Huber penalties provides the other reference where the parameters are chosen in such a way that it could conceivably yield similar results as ADMM-TVSAA after 500 iterations. All three algorithms enforce the total annihilation count and non-negativity constraints. The smoothing parameter for P-MLAA's Huber functions are both chosen to be 0.1% of the phantom maximum value, which is much less than the contrast of structures in either the attenuation or activity maps. In this way the Huber penalties approximate the TV-norm accurately. The penalty parameters are tuned so that the phantom TV values are achieved at 500 iterations of P-MLAA.

For a quantitative bias-variance analysis of the activity and attenuation, MLAA, P-MLAA, and ADMM-TVSAA are used to perform SAA on an ensemble of 100 noise realizations of TOF-PET data. The mean and pixel standard deviation are computed and plotted in Fig. 5 as a function of iteration number. The use of the TV-constraints, allows ADMM-TVSAA to achieve activity estimates with low bias and variance as compared with basic maximum-likelihood estimation as implemented with MLAA. Use of explicit Huber penalties with P-MLAA also yields images at 500 iterations that have low bias and variance with respect to MLAA; although the paths

in the bias-variance plot for ADMM-TVSAA and P-MLAA are quite different as a function of iteration number. For the particular parameter settings chosen, the ADMM-TVSAA algorithm achieves slightly lower global bias in the activity image with slightly larger variance as compared to P-MLAA. In the activity bias-variance curves, the proximity of the points at 200 and 500 iterations is an indication that the respective ADMM-TVSAA and P-MLAA are near convergence.

The bias-variance curves for the attenuation image reveal a much different behavior than that of the activity image. Both ADMM-TVSAA and P-MLAA substantially improve on the use of MLAA without explicit penalty terms. The variance of the ADMM-TVSAA result is larger than that of P-MLAA, but ADMM-TVSAA achieves a lower bias. It is also clear that 500 iterations is not sufficient to achieve a converged attenuation map because the there is still quite some separation between the points at 200 and 500 iterations for ADMM-TVSAA and P-MLAA. This difference in the rate of convergence of the activity and attenuation images was also seen in the noiseless results shown in Fig. 2. That there can be such a difference in the convergence rate of the two images is due to the fact that the attenuation image only impacts the activity image through the attenuation factor.

The activity and attenuation images for the present noise studies are shown in Figs. 6 and 7 for ADMM-TVSAA and P-MLAA, respectively. The top row of each figure shows images at 500 iterations for a single noise realization, and the bottom row shows the mean over 100 noise realizations, which reveals the spatial dependence of the image bias. The most notable difference between the two algorithms is in the attenuation images. The ADMM-TVSAA algorithm achieves an attenuation distribution that has the correct support and gray level even if it suffers from noticeable noise artifacts. The result for P-MLAA, however, shows significant bias in the attenuation images and features from the activity distribution
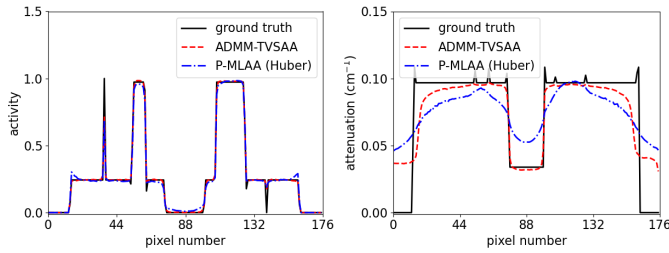
Fig. 8. Line profiles through the middle row of pixels comparing the ground truth with the mean reconstructed activity (left) and attenuation (right) images generated at 500 iterations by ADMM-TVSAA and P-MLAA.
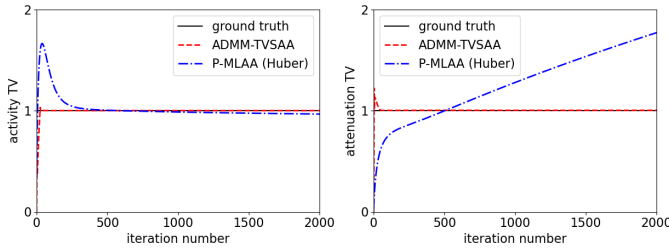

Fig. 9. Evolution of activity (left) and attenuation (right) image TV as a function of iteration number for 2000 iterations of ADMM-TVSAA and P-MLAA. The TV values are normalized so that the ground truth phantom values are 1.0.

clearly bleed through to the attenuation images. The error in the attenuation images, however, may not be critical if the attenuation factors are recovered, and this appears to be the case for both algorithms as the activity images at 500 iterations are accurate. The activity for P-MLAA shows slightly more bias, relative to ADMM-TVSAA, at the phantom border.

The line profiles in Fig. 8 reveal the bias more quantitatively. For ADMM-TVSAA, the mean activity follows the line profile of the ground truth quite closely except for the fact that the sharp features are slightly rounded. The mean attenuation is somewhat accurate except for near the border of the object support where there is significant blurring of the edge. For P-MLAA, the mean activity has a slight increase at the phantom border and a slight filling in of the central cold spot. The attenuation profile for P-MLAA reflects the significant bias that was seen in Fig. 7.

IV. DISCUSSION

The constrained likelihood model for SAA in TOF-PET enables novel numerical investigation into the SAA inverse problem. In particular, for this work, TV-constraints that exploit gradient sparsity are investigated for simultaneous recovery of activity and attenuation images. From the formulation presented in Sec. II-G, given knowledge of the total annihilation count and TV values of the activity and attenuation, the question is whether or not it is possible to recover the underlying activity and attenuation distributions. For the presented noiseless study where the activity and attenuation have the same support, it appears that both distributions can be accurately recovered. For the case where the activity distribution is interior to the attenuation, the former is recovered accurately while the latter is recovered only within

the support of the activity distribution. The study with data noise provides some sense of the stability of the SAA inverse problem when using the TV-constrained likelihood model.

We point out that carrying out inverse problems studies with the penalized likelihood approach shown in Sec. II-I is much more difficult. Aside from the difference between TV and Huber regularization terms, penalized likelihood needs to be investigated in the limit where the penalty parameters approach zero in order to provide the equivalent solution to constrained likelihood when the data are noiseless.

Even with noisy data, it is difficult to use the penalized-likelihood approach for achieving the same solution as the constrained-likelihood model. This is illustrated in the evolution of the activity and attenuation image TV shown in Fig. 9. These results correspond to the noise study that yielded the single noise realization images in Figs. 6 and 7 except that the iteration number is extended to 2000. For ADMM-TVSAA, the image TVs reach their constraint values quickly and maintain these values. For P-MLAA, the desired constraint values are achieved at 500 iterations because a two-dimensional penalty parameter search was performed to achieve this goal. Iterating further with P-MLAA causes the image TV to change. This change is not large for the activity because this image is nearly converged, but for the attenuation the change in TV is rather large. In order to come closer to the desired constraint values with P-MLAA, the penalty parameter search would need to be performed at larger iteration numbers, increasing the computational burden.

So far this discussion has focused on algorithms for investigating the TOF-PET SAA inverse problem, which can be quite different than their use for image reconstruction in real data scenarios. For an actual scan, ground truth is not available and mathematical accuracy of the solution may not correlate well with the imaging task of interest. Parameters of ADMM-TVSAA or P-MLAA need to be optimized on a metric reflecting performance on this imaging task. Also, because accurate solution is not necessarily the goal, early stopping is almost always used in practical applications of iterative image reconstruction. In this sense, iteration number becomes a parameter to tune and the practical difference between these two algorithms is the different paths they take in approaching their respective solutions. We also point out that we have shown results for P-MLAA using Green's one step late algorithm, and use of other solvers such as ADMM with SQS [25] may lead to improved performance of P-MLAA.

The presented study made use of a relatively simple test object that may favor the use of TV constraints. For realistic activity distributions, such constraints can still be employed effectively as is shown in Refs. [15] and [16], where human PET data is reconstructed using such constraints.

V. CONCLUSION

In this work, an ADMM framework is developed that can be applied to nonsmooth and nonconvex optimization problems that arise in imaging. The particular form of nonconvexity addressed is when the optimization problem has a biconvex structure. The imaging problem posed by simultaneous es-

timation of the activity and attenuation (SAA) in time-of-flight positron emission tomography (TOF-PET) has such a structure. Using this ADMM framework, a limited study on the impact of total variation (TV) constraints on the activity and attenuation for the SAA problem is presented. The use of both of these constraints is seen to help stabilize the SAA inverse problem. The shown results are intended to demonstrate the ADMM-TVSAA algorithm and its potential in solving the SAA inverse problem with the use of TV constraints on the activity and attenuation. While we have shown results only for ADMM-TVSAA, the ADMM framework is easily extended to include other nonsmooth, convex terms. Further study varying the test phantom and TOF-PET setup are needed to obtain a more comprehensive picture of the SAA inverse problem.

DATA AVAILABILITY

The implementation of the algorithms, which are presented in this article, and the code, which generates the figures, are available at:
`https://github.com/zhimeir/saa_admm_paper`.

APPENDIX

ADMM for TVSAA in TOF-PET

In this Appendix, the details for the instantiation of ADMM for TV-constraineed SAA estimation (ADMM-TVSAA) is explained. The optimization problem of interest is Eq. (36) from the main text, which we restate here

$$\lambda, \mu = \arg\min_{\lambda,\mu}\{l(\lambda,\mu) \mid \|D\lambda\|_1 \leq \gamma_\lambda,\ \|D\mu\|_1 \leq \gamma_\mu,$$

$$\mathbf{1}^\top\lambda = N_{\text{total}},\ \lambda \geq 0,\ \mu \geq 0\}, \quad \text{(A1)}$$

where D is the discretization of the spatial gradient operator, and $\|Dx\|_1$ is the anisotropic TV of x. To map the optimization problem onto the generic ADMM optimization in Eq. (14), the primal, splitting, and dual variables x, y, and u, are respectively assigned as

$$x = \begin{pmatrix} \lambda \\ \mu \end{pmatrix},\quad y = \begin{pmatrix} y_\lambda \\ z_\lambda \\ y_\mu \\ z_\mu \end{pmatrix},\quad u = \begin{pmatrix} u_\lambda \\ v_\lambda \\ u_\mu \\ v_\mu \end{pmatrix},$$

The linear system A is assigned as

$$A = \begin{pmatrix} T & 0 \\ \nu_\lambda D & 0 \\ 0 & P \\ 0 & \nu_\mu D \end{pmatrix},$$

where

$$\nu_\lambda = \|T\|_2/\|D\|_2,\quad \nu_\mu = \|P\|_2/\|D\|_2, \quad \text{(A2)}$$

are constants that normalize the gradient matrices to the projection matrices.

As with the SAA problem, the convex function f is used to represent the constraint on the total number of annihilations by setting

$$f(\lambda,\mu) = \delta(\mathbf{1}^\top\lambda = N_{\text{total}}) + \delta(\lambda \geq 0) + \delta(\mu \geq 0).$$

The biconvex function g accounts for the remaining terms in Eq. (A1)

$$g(y_\lambda, y_\mu) = L(y_\lambda, y_\mu)+$$
$$\delta(\|z_\lambda\|_1 \leq \nu_\lambda\gamma_\lambda) + \delta(\|z_\mu\|_1 \leq \nu_\mu\gamma_\mu), \quad \text{(A3)}$$

where the biconvex function $L(y_\lambda, y_\mu)$ is defined in Eq. (23) and the TV constraint values have also been scaled to reflect the normalization of D.

Parametrization of the step-sizes: With the modified system matrix A, the metric from Eq. (14) becomes

$$H_f = \begin{pmatrix} H_\lambda & 0 \\ 0 & H_\mu \end{pmatrix},$$

$$H_\lambda = \frac{I}{\tau_\lambda} - \sigma_\lambda(T^\top T + \nu_\lambda^2 D^\top D),$$

$$H_\mu = \frac{I}{\tau_\mu} - \sigma_\mu(P^\top P + \nu_\mu^2 D^\top D),$$

and the step-size parameters are chosen so that

$$\sigma_\lambda\tau_\lambda = (\|T\|_2^2 + \nu_\lambda^2\|D\|_2^2)^{-1},\quad \sigma_\mu\tau_\mu = (\|P\|_2^2 + \nu_\mu^2\|D\|_2^2)^{-1}.$$

As with ADMM-SAA the step size ratios, ρ_λ and ρ_μ, need to be determined by a similar grid search to what is shown in Sec. IIIA. The grid search performed for ADMM-SAA provides a good initial starting point for the step-size parameter search for ADMM-TVSAA when the parameters ν_λ and ν_μ are determined by Eq. (A2). Accordingly, the step size parameters for ADMM-TVSAA are

$$\sigma_\lambda = \rho_\lambda(\|T\|_2^2 + \nu_\lambda^2\|D\|_2^2)^{-1},$$
$$\tau_\lambda = \rho_\lambda^{-1}(\|T\|_2^2 + \nu_\lambda^2\|D\|_2^2)^{-1},$$
$$\sigma_\mu = \rho_\mu(\|P\|_2^2 + \nu_\mu^2\|D\|_2^2)^{-1},$$
$$\tau_\mu = \rho_\mu^{-1}(\|P\|_2^2 + \nu_\mu^2\|D\|_2^2)^{-1}.$$

The y-update: The g function in Eq. (A3) separates into a biconvex function in y_λ and y_μ, and convex functions in z_λ and z_μ. The biconvex terms are treated in exactly the same way as the ADMM-SAA presentation in Sec. IIF. Focusing on the last two convex terms in Eq. (A3) yields the optimization problems

$$z_\lambda = \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{(A4)}$$
$$\arg\min_{z_\lambda'}\left\{\delta(\|z_\lambda'\|_1 \leq \nu_\lambda\gamma_\lambda) - v_\lambda^\top z_\lambda' + \frac{\sigma_\lambda}{2}\|z_\lambda' - \nu_\lambda D\lambda\|^2\right\},$$

and

$$z_\mu = \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{(A5)}$$
$$\arg\min_{z_\mu'}\left\{\delta(\|z_\mu'\|_1 \leq \nu_\mu\gamma_\mu) - v_\mu^\top z_\mu' + \frac{\sigma_\mu}{2}\|z_\mu' - \nu_\mu D\mu\|^2\right\}.$$

Because both of these problems are identical, we focus on Eq. (A4). The objective function consists of a quadratic function and an indicator function that enforces the ℓ_1 constraint on z_λ'. Furthermore, the quadratic function has uniform curvature, i.e. a Hessian matrix that is proportional to the identity matrix. For this special case, the solution to Eq. (A4) is obtained in a two-step process that involves finding the minimizer of the unconstrained quadratic function, then projecting the result to

the closest z'_λ that satisfies the ℓ_1 constraint. The unconstrained minimizer is given by

$$z''_\lambda = v_\lambda/\sigma_\lambda + \nu_\lambda D\lambda,$$

and the solution to Eq. (A4) becomes

$$z'_\lambda = \mathcal{P}_{L1(\nu_\lambda \gamma_\lambda)}(z''_\lambda), \quad L1(r) = \{z \mid \|z\|_1 \le r\},$$

where $\mathcal{P}_{L1(\nu_\lambda \gamma_\lambda)}(\cdot)$ denotes projection onto the ℓ_1-ball of "radius" $\nu_\lambda \gamma_\lambda$. An efficient algorithm for performing this projection is presented in Ref. [28], or it can also be accomplished by vector shrinkage and use of a root finding algorithm to determine the shrinkage parameter to attain an ℓ_1-norm of $\nu_\lambda \gamma_\lambda$. The update equations for z_λ and z_μ are

$$z_\lambda \leftarrow \mathcal{P}_{L1(\nu_\lambda \gamma_\lambda)}(v_\lambda/\sigma_\lambda + \nu_\lambda D\lambda),$$
$$z_\mu \leftarrow \mathcal{P}_{L1(\nu_\mu \gamma_\mu)}(v_\mu/\sigma_\mu + \nu_\mu D\mu).$$

Algorithm 2 ADMM pseudocode for TV-constrained SAA estimation. Variables λ, μ, y_λ, y_μ, z_λ, z_μ, $\bar{y}_\lambda$, $\bar{y}_\mu$, $\bar{z}_\lambda$, $\bar{z}_\mu$, u_λ, u_μ, v_λ, and v_μ are initialized to zero. Step size ratio parameters ρ_λ and ρ_μ are determined in the grid search described in Sec. IIIA.

1: **for** $k \leftarrow 1, N_{\text{iter}}$ **do**
2: $\quad \bar{\lambda}_1 = T^\top(u_\lambda + \sigma_\lambda(\bar{y}_\lambda - y_\lambda))$
3: $\quad \bar{\lambda}_2 = \nu_\lambda D^\top(v_\lambda + \sigma_\lambda(\bar{z}_\lambda - z_\lambda))$
4: $\quad \lambda \leftarrow \mathcal{P}_{\text{simplex}(N_{\text{total}})}(\lambda - \tau_\lambda(\bar{\lambda}_1 + \bar{\lambda}_2))$
5: $\quad \bar{y}_\lambda = T\lambda$
6: $\quad \bar{z}_\lambda = \nu_\lambda D\lambda$
7: $\quad \bar{\mu}_1 = P^\top(u_\mu + \sigma_\mu(\bar{y}_\mu - y_\mu))$
8: $\quad \bar{\mu}_2 = \nu_\mu D^\top(v_\mu + \sigma_\mu(\bar{z}_\mu - z_\mu))$
9: $\quad \mu \leftarrow \mu - \tau_\mu(\bar{\mu}_1 + \bar{\mu}_2)$
10: $\quad \bar{y}_\mu = P\mu$
11: $\quad \bar{z}_\mu = \nu_\mu D\mu$
12: $\quad v_{\mu,\ell} = \sum_i C_{i\ell} - (u_{\mu,\ell} - \sigma_\mu \bar{y}_{\mu,\ell}) \ \ \forall \ell$
13: $\quad$ **for** $k' \leftarrow 1, N_y$ **do**
14: $\quad\quad b_{i\ell} = u_{\lambda,i\ell} + \sigma_\lambda \bar{y}_{\lambda,i\ell} - \exp(-y_{\mu,\ell}) \ \ \forall i,\ell$
15: $\quad\quad y_{\lambda,i\ell} = \left(b_{i\ell} + \sqrt{b_{i\ell}^2 + 4\sigma_\lambda C_{i\ell}}\right)/(2\sigma_\lambda) \ \ \forall i,\ell$
16: $\quad\quad y_{\lambda,\ell} = \sum_i y_{\lambda,i\ell} \ \ \forall \ell$
17: $\quad\quad y'_{\mu,\ell} = 0 \ \ \forall \ell$
18: $\quad\quad$ **for** $k'' \leftarrow 1, N_{\text{newt}}$ **do**
19: $\quad\quad\quad \psi_\ell^{(1)} = -\exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell} + \sigma_\mu y'_{\mu,\ell} + v_{\mu,\ell}$
20: $\quad\quad\quad \psi_\ell^{(2)} = \exp(-y'_{\mu,\ell}) \cdot y_{\lambda,\ell} + \sigma_\mu \ \ \forall \ell$
21: $\quad\quad\quad y'_{\mu,\ell} \leftarrow y'_{\mu,\ell} - \psi_\ell^{(1)} \cdot \left(\psi_\ell^{(2)}\right)^{-1} \ \ \forall \ell$
22: $\quad\quad\quad y'_{\mu,\ell} \leftarrow \text{pos}(y'_{\mu,\ell}) \ \ \forall \ell$
23: $\quad\quad$ **end for**
24: $\quad\quad y_{\mu,\ell} = y'_{\mu,\ell} \ \ \forall \ell$
25: $\quad$ **end for**
26: $\quad z_\lambda \leftarrow \mathcal{P}_{L1(\nu_\lambda \gamma_\lambda)}(v_\lambda/\sigma_\lambda + \bar{z}_\lambda)$
27: $\quad z_\mu \leftarrow \mathcal{P}_{L1(\nu_\mu \gamma_\mu)}(v_\mu/\sigma_\mu + \bar{z}_\mu)$
28: $\quad u_\lambda \leftarrow u_\lambda + \sigma_\lambda(\bar{y}_\lambda - y_\lambda)$
29: $\quad u_\mu \leftarrow u_\mu + \sigma_\mu(\bar{y}_\mu - y_\mu)$
30: $\quad v_\lambda \leftarrow v_\lambda + \sigma_\lambda(\bar{z}_\lambda - z_\lambda)$
31: $\quad v_\mu \leftarrow v_\mu + \sigma_\mu(\bar{z}_\mu - z_\mu)$
32: **end for**

ADMM pseudocode for TV-constrained SAA estimation: The derivations in this appendix yield the additional steps that are needed to formulate the pseudocode shown in Algorithm 2 from the ADMM-SAA pseudocode in Algorithm 1.

REFERENCES

[1] P. E. Kinahan, D. W. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3D PET/CT scanner," *Med. Phys.*, vol. 25, pp. 2046–2053, 1998.

[2] T. Xia, A. M. Alessio, B. De Man, R. Manjeshwar, E. Asma, and P. E. Kinahan, "Ultra-low dose CT attenuation correction for PET/CT," *Phys. Med. Biol.*, vol. 57, pp. 309–328, 2011.

[3] N. Burgos, M. J. Cardoso, K. Thielemans, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, R. Ahmed, C. J.Mahoney, J. M. Schott, J. S. Duncan, D. Atkinson, S. R. Arridge, B. F. Hutton, and S. Ourselin, "Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies," *IEEE Trans. Med. Imaging*, vol. 33, pp. 2332–2341, 2014.

[4] M. M. Osman, C. Cohade, Y. Nakamoto, and R. L. Wahl, "Respiratory motion artifacts on PET emission images obtained using CT attenuation correction on PET-CT," *Euro. J. Nuc. Med. Mol. Imaging*, vol. 30, pp. 603–606, 2003.

[5] F. Natterer and H. Herzog, "Attenuation correction in positron emission tomography," *Math. Method. Appl. Sci.*, vol. 15, pp. 321–330, 1992.

[6] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Trans. Med. Imaging*, vol. 18, pp. 393–403, 1999.

[7] T. K. Lewellen, "Time-of-flight PET," in *Semin. Nucl. Med.*, 1998, vol. 28, pp. 268–275.

[8] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Phys. Med. Biol.*, vol. 57, pp. 885–899, 2012.

[9] M. Defrise, V. Panin, C. Michel, and M. E. Casey, "Continuous and discrete data rebinning in time-of-flight PET," *IEEE Trans. Med. Imaging*, vol. 27, pp. 1310–1322, 2008.

[10] V. Y. Panin, M. Defrise, and M. E. Casey, "Restoration of fine azimuthal sampling of measured TOF projection data," in *2010 IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, 2011, pp. 3079–3084.

[11] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson, and J. Nuyts, "Simultaneous reconstruction of activity and attenuation in time-of-flight PET," *IEEE Trans Med. Imaging*, vol. 31, pp. 2224–2233, 2012.

[12] L. Cheng, T. Ma, X. Zhang, Q. Peng, Y. Liu, and J. Qi, "Maximum likelihood activity and attenuation estimation using both emission and transmission data with application to utilization of Lu-176 background radiation in TOF PET," *Med. Phys.*, vol. 47, pp. 1067–1082, 2020.

[13] A. Salomon, A. Goedicke, B. Schweizer, T. Aach, and V. Schulz, "Simultaneous reconstruction of activity and attenuation for PET/MRI," *IEEE Trans. Med. Imaging*, vol. 30, pp. 804–813, 2010.

[14] P. A. Wolf, J. S. Jørgensen, T. G. Schmidt, and E. Y. Sidky, "Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model," *Phys. Med. Biol.*, vol. 58, pp. 5629–5652, 2013.

[15] Z. Zhang, J. Ye, B. Chen, A. E. Perkins, S. Rose, E. Y. Sidky, C.-M. Kao, D. Xia, C.-H. Tung, and X. Pan, "Investigation of optimization-based reconstruction with an image-total-variation constraint in PET," *Phys. Med. Biol.*, vol. 61, pp. 6055–6084, 2016.

[16] Z. Zhang, S. Rose, J. Ye, A. E. Perkins, B. Chen, C.-M. Kao, E. Y. Sidky, C.-H. Tung, and X. Pan, "Optimization-based reconstruction from low-count, list-mode TOF-PET data," *IEEE Trans. Biomed. Eng.*, vol. 65, pp. 936–946, 2018.

[17] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. lmaging Vis.*, vol. 40, pp. 120–145, 2011.

[18] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, pp. 3065–3091, 2012.

[19] R. F. Barber and E. Y. Sidky, "Convergence for nonconvex ADMM, with applications to CT imaging," *arXiv preprint arXiv:2006.07278*, 2020.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trend. Mach. Learn.*, vol. 3, pp. 1–122, 2011.

[21] T. G. Schmidt, B. A. Sammut, R. F. Barber, X. Pan, and E. Y. Sidky, "Addressing CT metal artifacts using photon-counting detectors and one-step spectral CT image reconstruction," *Med. Phys.*, vol. 49, pp. 3021–3040, 2022.

[22] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imaging*, vol. 9, pp. 439–446, 1990.

[23] T. Heußer, C. M. Rank, Y. Berker, M. T. Freitag, and M. Kachelrieß, "MLAA-based attenuation correction of flexible hardware components in hybrid PET/MR imaging," *EJNMMI physics*, vol. 4, pp. 1–23, 2017.

[24] A. Mehranian, H. Zaidi, and A. J. Reader, "MR-guided joint reconstruction of activity and attenuation in brain PET-MR," *NeuroImage*, vol. 162, pp. 276–288, 2017.

[25] S. Y. Chun, K. Y. Kim, J. S. Lee, and J. A. Fessler, "Joint estimation of activity distribution and attenuation map for TOF-PET using alternating direction method of multiplier," in *Proc. IEEE 13th Int. Symp. Biomed. Imaging (ISBI)*, 2016, pp. 86–89.

[26] H. Nien and J. A. Fessler, "Fast X-ray CT image reconstruction using a linearized augmented Lagrangian method with ordered subsets," *IEEE Trans. Med. Imaging*, vol. 34, pp. 388–399, 2014.

[27] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral CT data," *Phys. Med. Biol.*, vol. 61, pp. 3784–3818, 2016.

[28] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 272–279.

[29] Z. Ren, E. Y. Sidky, R. F. Barber, C.-M. Kao, and X. Pan, "Simultaneous activity and attenuation estimation in TOF-PET with TV-constrained nonconvex optimization," *arXiv preprint arXiv:2303.17042*, 2024.

[30] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, pp. 84–93, 1990.

[31] L. A. Pierce, B. F. Elston, D. A. Clunie, D. Nelson, and P. E. Kinahan, "A digital reference object to analyze calculation accuracy of PET standardized uptake value," *Radiology*, vol. 277, pp. 538–545, 2015.