Noise properties of motion-compensated tomographic image reconstruction methods

Se Young Chun, *Member, IEEE,* and Jeffrey A. Fessler, *Fellow, IEEE*

***Abstract*—Motion-compensated image reconstruction (MCIR) methods incorporate motion models to improve image quality in the presence of motion. MCIR methods differ in terms of how they use motion information and they have been well-studied separately. However, there have been less theoretical comparisions of different MCIR methods. This paper compares the theoretical noise properties of three popular MCIR methods assuming known nonrigid motion.**

We show the relationship among three MCIR methods - motion-compensated temporal regularization (MTR), the parametric motion model (PMM), and post-reconstruction motion correction (PMC) - for penalized weighted least square cases. These analyses show that PMM and MTR are *matrix-weighted* sums of all registered image frames, while PMC is a *scalar-weighted* sum.

We further investigate the noise properties of MCIR methods with Poisson models and quadratic regularizers by deriving accurate and fast variance prediction formulas using an "analytical approach". These theoretical noise analyses show that the variances of PMM and MTR are lower than or comparable to the variance of PMC due to the statistical weighting. These analyses also facilitate comparisons of the noise properties of different MCIR methods, including the effects of different quadratic regularizers, the influence of the motion through its Jacobian determinant, and the effect of assuming that total activity is preserved. 2D PET simulations demonstrate the theoretical results.

***Index Terms*—motion-compensated image reconstruction, noise properties, quadratic regularization, nonrigid motion.**

I. Introduction

MOTION-COMPENSATED image reconstruction (MCIR) methods have been actively studied for various imaging modalities. MCIR methods can provide high signal-to-noise ratio (SNR) images (or low radiation dose images) *and* reduce motion artifacts [1]–[14]. Gating methods implicitly use motion information (*i.e.*, no explicit motion estimation required) for motion correction, but yield low SNR images due to insufficient measurements (or require longer acquisition to collect enough measurements) [15], [16]. In contrast, MCIR methods use explicit motion information (*i.e.*, motion estimation obtained jointly or separately) to correct for motion artifacts and to produce high SNR images with all collected data.

This work was supported in part by NIH/NCI grant 1P01 CA87634.

Se Young Chun is with the Department of EECS and Radiology, the University of Michigan, Ann Arbor, MI 48109, USA. (e-mail: delight@umich.edu)

Jeffrey A. Fessler is with the Department of EECS, the University of Michigan, Ann Arbor, MI 48109, USA. (e-mail: fessler@umich.edu).

This paper analyzes three popular MCIR methods that differ in their way of incorporating motion information: post-reconstruction motion correction (PMC) [1]–[3], motion-compensated temporal regularization (MTR) [4], [5], and the parametric motion model (PMM) [6]–[14]. Each MCIR method has been well-studied separately, but there has been less theoretical research on comparing different MCIR methods. There are some empirical comparisons between PMC and PMM [17], [18], and between MTR and PMM [19]. Asma *et al.* compared PMC and PMM theoretically in terms of their mean and covariance by using a discrete Fourier transform (DFT) based approximation [20]. However, the analytical comparison was limited to the unregularized case and the empirical comparison was performed for the regularized case.

Theoretical noise analyses of MCIR methods can be useful for regularizer design and for performance comparisons. Noise prediction methods include matrix-based approaches [21], DFT methods [22], and an "analytical approach" that is much faster [23]. We extend this analytical approach to MCIR, and investigate the noise properties of PMC, PMM, and MTR with quadratic regularizers *theoretically*, assuming known nonrigid motion. This assumption is applicable to some multi-modal medical imaging systems such as PET-CT [7], [8], [10] and PET-MR [14]. These analyses provide fast variance prediction for MCIR methods and may also provide some insight into unknown motion cases. These noise analyses not only facilitate theoretical comparisons of the performance of different MCIR methods, but also help one understand the influence of the motion (through its Jacobian determinant) and the effect of assuming that the total activity is preserved.

This paper is organized as follows. Section II reviews the basic models and the estimators of the MCIR methods [24]: PMC, PMM, and MTR. Section III shows the similarity and difference between three MCIR estimators in penalized weighted least square (PWLS) cases. It shows that MTR and PMM are essentially the Fisher information-based *matrix-weighted* sum of all registered image frames, while PMC is the *scalar-weighted* sum. Section IV derives fast variance prediction formulas for PMC and PMM with Poisson likelihoods and general quadratic regularizers. Section V compares the theoretical noise properties of MCIR methods. Section VI illustrates the theories by 2D PET simulations with digital phantoms for given affine and nonrigid motions.

II. MCIR models and methods

This section reviews MCIR models that were also described in [24] and derives the PWLS estimator for each model.

Although we focus on PWLS for simplicity, the general conclusions are also applicable to penalized-likelihood estimation based on Poisson models [25]. We consider three MCIR methods: PMC [1]–[3], PMM [6]–[12], [26], [27], and MTR [4], [5], [19], [28]. We treat the nonrigid motion information as predetermined (known) and focus on how the motion models affect noise propagation from the measurements into the reconstructed image. In practice, errors in the motion models lead to further variability in the image.

A. Review of basic MCIR models

1) Measurement model: MCIR methods are needed when the time-varying object $f(\vec{x}, t)$ has non-negligible motion during an acquisition interval where $\vec{x} \in \mathbb{R}^d$ denotes spatial coordinate and t denotes time. Often one can use gating or temporal binning to group the measurements into M sets, called "frames" here. Let $\boldsymbol{y}_m$ denote the vector of measurements associated with the mth frame. We assume the time varying object $f(\vec{x}, t)$ is approximately motionless during the acquisition of each $\boldsymbol{y}_m$. Let t_m denote the time associated with the mth frame, and let $\boldsymbol{f}_m = (f(\vec{x}_1, t_m), \ldots, f(\vec{x}_N, t_m))$ denote a spatial discretization of the object $f(\cdot, t_m)$ where $\vec{x}_j$ denotes the center of the jth voxel for $j = 1, \ldots, N$, and N denotes the number of voxels. We assume that the measurements are related to the object linearly as follows:

$$\boldsymbol{y}_m = \boldsymbol{A}_m \boldsymbol{f}_m + \boldsymbol{\epsilon}_m, \qquad m = 1, \ldots, M, \tag{1}$$

where $\boldsymbol{A}_m$ denotes the system model for the mth frame, $\boldsymbol{\epsilon}_m$ denotes noise, and M is the number of gates or frames. We allow the system model $\boldsymbol{A}_m$ to possibly differ for each frame.

2) Warp model: For a given spatial transformation $T_{m,n} : \mathbb{R}^d \to \mathbb{R}^d$, define a warp operator $\mathring{\mathcal{T}}_{m,n}$ as follows:

$$\begin{aligned} f(\vec{x}, t_m) &= (\mathring{\mathcal{T}}_{m,n} f)(\vec{x}, t_n) \\ &\triangleq |\nabla T_{m,n}(\vec{x})|^p f(T_{m,n}(\vec{x}), t_n), \end{aligned} \tag{2}$$

where the total activity is preserved when $p = 1$. We discretize the warp $\mathring{\mathcal{T}}_{m,n}$ to define a $N \times N$ matrix relating the image $\boldsymbol{f}_n$ to the image $\boldsymbol{f}_m$ as follows:

$$\boldsymbol{f}_m = \mathring{\boldsymbol{T}}_{m,n} \boldsymbol{f}_n, \quad n, m = 1, \cdots, M. \tag{3}$$

For applications with periodic motion, we can additionally define $\boldsymbol{f}_{M+1} \triangleq \boldsymbol{f}_1$ and $\mathring{\boldsymbol{T}}_{M+1,M} \triangleq \mathring{\boldsymbol{T}}_{1,M}$. The matrix $\mathring{\boldsymbol{T}}_{m,n}$ can be implemented with any interpolation method; we used a B-spline based image warp [29]. Let $|\nabla T_{m,n}(\vec{x})|$ denote the determinant of the Jacobian matrix of a transform $T_{m,n}(\vec{x})$ for a warp $\mathring{\boldsymbol{T}}_{m,n}$. Throughout we assume the warps $\mathring{\boldsymbol{T}}_{m,n}$ (or equivalently $T_{m,n}$ or $\mathring{\mathcal{T}}_{m,n}$) are known. We also assume that invertibility, symmetry, and transitivity properties hold for $\mathring{\boldsymbol{T}}_{m,n}$ [24].

B. Single gated reconstruction (SGR)

Often one can reconstruct each image $\hat{\boldsymbol{f}}_m$ from the corresponding measurement $\boldsymbol{y}_m$ based on the model (1) and some prior knowledge (*e.g.*, a smoothness prior). A single gated (frame) reconstruction (SGR) can be obtained as follows:

$$\hat{\boldsymbol{f}}_m \triangleq \underset{\boldsymbol{f}_m}{\operatorname{argmin}} \, \mathrm{L}_m\left(\boldsymbol{y}_m, \boldsymbol{A}_m \boldsymbol{f}_m\right) + \eta \mathrm{R}_m(\boldsymbol{f}_m) \tag{4}$$

where $m = 1, \cdots, M$, L_m is a negative likelihood function derived from (1), R_m is a spatial regularizer, and η is a spatial regularization parameter.

For the PWLS case, *i.e.*, $\mathrm{L}_m(\boldsymbol{y}_m, \boldsymbol{A}_m \boldsymbol{f}_m) \triangleq \|\boldsymbol{y}_m - \boldsymbol{A}_m \boldsymbol{f}_m\|^2_{\boldsymbol{W}_m}/2$ where $\boldsymbol{W}_m$ is a weight matrix that usually approximates the inverse of the covariance of $\boldsymbol{y}_m$, one can obtain a closed form estimator $\hat{\boldsymbol{f}}_m$ as follows:

$$\hat{\boldsymbol{f}}_m = [\boldsymbol{F}_m + \eta \boldsymbol{R}_m]^{-1} \boldsymbol{A}_m' \boldsymbol{W}_m \boldsymbol{y}_m \tag{5}$$

where the Fisher information matrix for the mth frame is $\boldsymbol{F}_m \triangleq \boldsymbol{A}_m' \boldsymbol{W}_m \boldsymbol{A}_m$, "$'$" denotes matrix transpose, and $\boldsymbol{R}_m$ is the Hessian matrix of a quadratic regularizer R_m.

C. Post-reconstruction motion correction (PMC)

Once the frames $\hat{\boldsymbol{f}}_1, \ldots, \hat{\boldsymbol{f}}_M$ are reconstructed individually from (4), one can improve SNR by averaging all reconstructed images. Using the motion information to map each image $\boldsymbol{f}_m$ to a single image's coordinates can reduce motion artifacts. Without loss of generality, we chose $\boldsymbol{f}_1$ as our reference image. Using (3) and (4), a natural definition for the (scalar-weighted) PMC estimator is the following motion-compensated average:

$$\hat{\boldsymbol{f}}_{\mathrm{PMC}} \triangleq \sum_{m=1}^{M} \alpha_m \mathring{\boldsymbol{T}}_{1,m} \hat{\boldsymbol{f}}_m \tag{6}$$

where $\sum_{m=1}^{M} \alpha_m = 1$. One choice is $\alpha_m = 1/M$ for all m (unweighted PMC). Another option is $\alpha_m = \tau_m / \sum_{m'=1}^{M} \tau_{m'}$ where τ_m is the acquisition time (or the number of counts) for the mth frame (scalar-weighted PMC). For the PWLS case, there is an explicit form for $\hat{\boldsymbol{f}}_{\mathrm{PMC}}$ using (3), (5), and (6):

$$\hat{\boldsymbol{f}}_{\mathrm{PMC}} = \sum_{m=1}^{M} \alpha_m [\mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m]^{-1} \mathring{\boldsymbol{T}}_{m,1}' \boldsymbol{A}_m' \boldsymbol{W}_m \boldsymbol{y}_m \tag{7}$$

where $\mathring{\boldsymbol{F}}_m \triangleq \mathring{\boldsymbol{T}}_{m,1}' \boldsymbol{F}_m \mathring{\boldsymbol{T}}_{m,1}$ and $\mathring{\boldsymbol{R}}_m \triangleq \mathring{\boldsymbol{T}}_{m,1}' \boldsymbol{R}_m \mathring{\boldsymbol{T}}_{m,1}$ are essentially Hessian matrices for the mth frame in the coordinates of the first (reference) frame.

D. Parametric motion model (PMM)

Without loss of generality, we assume that $\boldsymbol{f}_1$ is our reference image frame for the PMM approach. Combining the measurement model (1) with the warp (3) yields a new measurement model that depends only on the image $\boldsymbol{f}_1$ instead of the all images $\boldsymbol{f}_1, \cdots, \boldsymbol{f}_M$ (*i.e.*, *parameterizing* all images with $\boldsymbol{f}_1$):

$$\boldsymbol{y}_m = \boldsymbol{A}_m \mathring{\boldsymbol{T}}_{m,1} \boldsymbol{f}_1 + \boldsymbol{\epsilon}_m, \quad m = 1, \ldots, M.$$

Stacking up these models yields the overall model

$$\boldsymbol{y}_{\mathrm{c}} = \boldsymbol{A}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} \boldsymbol{f}_1 + \boldsymbol{\epsilon}_{\mathrm{c}}, \tag{8}$$

where the components are each stacked accordingly:

$$\begin{aligned} \boldsymbol{y}_{\mathrm{c}} &\triangleq [\boldsymbol{y}_1', \cdots, \boldsymbol{y}_M']', \\ \boldsymbol{A}_{\mathrm{d}} &\triangleq \operatorname{diag}\{\boldsymbol{A}_1, \cdots, \boldsymbol{A}_M\}, \\ \mathring{\boldsymbol{T}}_{\mathrm{c}} &\triangleq [\boldsymbol{I}, \mathring{\boldsymbol{T}}_{2,1}', \cdots, \mathring{\boldsymbol{T}}_{M,1}']', \text{and} \\ \boldsymbol{\epsilon}_{\mathrm{c}} &\triangleq [\boldsymbol{\epsilon}_1', \cdots, \boldsymbol{\epsilon}_M']'. \end{aligned} \tag{9}$$

The PMM estimator for the measurement model (8) with a spatial regularizer is

$$\hat{\boldsymbol{f}}_{\mathrm{PMM}} \triangleq \operatorname*{argmin}_{\boldsymbol{f}_1} \mathrm{L}(\boldsymbol{y}_{\mathrm{c}}, \boldsymbol{A}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} \boldsymbol{f}_1) + \eta \mathrm{R}_{\mathrm{PMM}}(\boldsymbol{f}_1) \tag{10}$$

where L is a negative likelihood function and $\mathrm{R}_{\mathrm{PMM}}$ is a spatial regularizer.

For the PWLS data fidelity function $\mathrm{L}(\boldsymbol{y}_{\mathrm{c}}, \boldsymbol{A}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} \boldsymbol{f}_1) \triangleq \|\boldsymbol{y}_{\mathrm{c}} - \boldsymbol{A}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} \boldsymbol{f}_1\|^2_{\boldsymbol{W}_{\mathrm{d}}}/2$ where $\boldsymbol{W}_{\mathrm{d}} \triangleq \mathrm{diag}\{\boldsymbol{W}_1, \cdots, \boldsymbol{W}_M\}$ is a diagonal matrix, the PMM estimator is

$$\hat{\boldsymbol{f}}_{\mathrm{PMM}} = [\mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{F}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} + \eta \boldsymbol{R}_{\mathrm{PMM}}]^{-1} \mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{A}'_{\mathrm{d}} \boldsymbol{W}_{\mathrm{d}} \boldsymbol{y}_{\mathrm{c}} \tag{11}$$

where $\boldsymbol{F}_{\mathrm{d}} \triangleq \boldsymbol{A}'_{\mathrm{d}} \boldsymbol{W}_{\mathrm{d}} \boldsymbol{A}_{\mathrm{d}} = \mathrm{diag}\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_M\}$ is a block-diagonal matrix, and $\boldsymbol{R}_{\mathrm{PMM}}$ is the Hessian matrix of a quadratic regularizer $\mathrm{R}_{\mathrm{PMM}}$. Since $\mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{F}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}} = \sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m$, we can rewrite the PMM estimator in (11) as

$$\hat{\boldsymbol{f}}_{\mathrm{PMM}} = \left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \boldsymbol{R}_{\mathrm{PMM}}\right]^{-1} \sum_{m=1}^{M} \mathring{\boldsymbol{T}}'_{m,1} \boldsymbol{A}'_m \boldsymbol{W}_m \boldsymbol{y}_m. \tag{12}$$

E. Motion-compensated temporal regularization (MTR)

The MTR method incorporates the motion information that matches two adjacent images into a temporal regularization term [4], [5]:

$$\frac{1}{2} \|\boldsymbol{f}_{m+1} - \mathring{\boldsymbol{T}}_{m+1,m} \boldsymbol{f}_m\|^2_2. \tag{13}$$

for $m = 1, \cdots, M-1$. This penalty is added to the cost function in (4) for all m to define the MTR cost function.

Equations (4) for all m and (13) can be represented in a simpler vector-matrix notation. First, stack up (1) for all m as follows:

$$\boldsymbol{y}_{\mathrm{c}} = \boldsymbol{A}_{\mathrm{d}} \boldsymbol{f}_{\mathrm{c}} + \boldsymbol{\epsilon}_{\mathrm{c}}, \tag{14}$$

where $\boldsymbol{f}_{\mathrm{c}} = [\boldsymbol{f}'_1, \cdots, \boldsymbol{f}'_M]'$ and $\boldsymbol{A}_{\mathrm{d}}$, $\boldsymbol{\epsilon}_{\mathrm{c}}$ are defined in (9). Then, the MTR estimator based on (13), (14), and a spatial regularizer is

$$\hat{\boldsymbol{f}}_{\mathrm{c}} \triangleq \operatorname*{argmin}_{\boldsymbol{f}_{\mathrm{c}}} \mathrm{L}(\boldsymbol{y}_{\mathrm{c}}, \boldsymbol{A}_{\mathrm{d}} \boldsymbol{f}_{\mathrm{c}}) + \eta \mathrm{R}(\boldsymbol{f}_{\mathrm{c}}) + \frac{\zeta}{2} \|\boldsymbol{T}_{\mathrm{time}} \boldsymbol{f}_{\mathrm{c}}\|^2_2 \tag{15}$$

where L is a negative likelihood function from the noise model of (14), R is a spatial regularizer, ζ is a temporal regularization parameter, and the temporal differencing matrix is

$$\boldsymbol{T}_{\mathrm{time}} \triangleq \begin{bmatrix} -\mathring{\boldsymbol{T}}_{2,1} & \boldsymbol{I} & & \\ & \ddots & \ddots & \\ & & -\mathring{\boldsymbol{T}}_{M,M-1} & \boldsymbol{I} \end{bmatrix}. \tag{16}$$

We may also modify $\boldsymbol{T}_{\mathrm{time}}$ for periodic (or pseudo-periodic) image sequences by adding a row corresponding to the term $\boldsymbol{f}_1 - \mathring{\boldsymbol{T}}_{1,M} \boldsymbol{f}_M$. Note that unlike the PMM method that estimates one frame $\hat{\boldsymbol{f}}_1$, MTR estimates all image frames $\hat{\boldsymbol{f}}_{\mathrm{c}}$. The MTR estimate of $\boldsymbol{f}_1$ (reference image) is

$$\hat{\boldsymbol{f}}_{\mathrm{MTR}} \triangleq [\boldsymbol{I} \quad \boldsymbol{0} \quad \cdots \quad \boldsymbol{0}]\, \hat{\boldsymbol{f}}_{\mathrm{c}}. \tag{17}$$

For the PWLS case, the solution to (15) is

$$\hat{\boldsymbol{f}}_{\mathrm{c}} = [\boldsymbol{F}_{\mathrm{d}} + \eta \boldsymbol{R}_{\mathrm{d}} + \zeta \boldsymbol{R}_{\mathrm{time}}]^{-1} \boldsymbol{A}_{\mathrm{d}} \boldsymbol{W}_{\mathrm{d}} \boldsymbol{y}_{\mathrm{c}}, \tag{18}$$

where $\boldsymbol{R}_{\mathrm{d}} \triangleq \mathrm{diag}\{\boldsymbol{R}_1, \cdots, \boldsymbol{R}_M\}$ and $\boldsymbol{R}_{\mathrm{time}} \triangleq \boldsymbol{T}'_{\mathrm{time}} \boldsymbol{T}_{\mathrm{time}}$.

III. Relationship between MCIR estimators

In this section, we investigate the relationship among PWLS MCIR estimators in (5), (7), (12), and (18). Considering PWLS estimators helps show the similarity and differences among MCIR methods more clearly than estimators for Poisson likelihoods. Although the observations in this section focus on PWLS estimators, similar results can be obtained for the mean and variance of MCIR estimators with Poisson likelihood models [25]. The next section analyzes the variance of these MCIR methods.

A. Properties of MTR estimator for $\zeta \to 0$ and $\zeta \to \infty$

The temporal regularization term (13) in (15) will increase the correlation between the estimators $\hat{\boldsymbol{f}}_i$ and $\hat{\boldsymbol{f}}_j$ for $i \neq j$ as ζ is increased. Even though (18) provides the exact relationship between the PWLS MTR estimator and ζ, this form itself may not be informative in terms of comparing it with other MCIR methods. So, we investigate the limiting behavior of the PWLS MTR estimator as $\zeta \to 0$ and as $\zeta \to \infty$. This provides insights for comparisons with PMM and PMC.

It is straightforward to determine the limit of $\hat{\boldsymbol{f}}_{\mathrm{c}}$ in (18) as $\zeta \to 0$ because

$$\boldsymbol{F}_{\mathrm{d}} + \eta \boldsymbol{R}_{\mathrm{d}} + \zeta \boldsymbol{R}_{\mathrm{time}} \to \boldsymbol{F}_{\mathrm{d}} + \eta \boldsymbol{R}_{\mathrm{d}} \triangleq \boldsymbol{G}_{\mathrm{MTR}} \tag{19}$$

where $\boldsymbol{G}_{\mathrm{MTR}}$ is a block-diagonal matrix, *i.e.*, $\boldsymbol{G}_{\mathrm{MTR}} = \mathrm{diag}\{(\boldsymbol{F}_m + \eta \boldsymbol{R}_m)\}_{m=1}^{M}$. Therefore, as $\zeta \to 0$, the PWLS MTR estimator $\hat{\boldsymbol{f}}_{\mathrm{c}}$ approaches

$$\hat{\boldsymbol{f}}_{\mathrm{c}} \to \boldsymbol{G}^{-1}_{\mathrm{MTR}} \boldsymbol{A}'_{\mathrm{d}} \boldsymbol{W}_{\mathrm{d}} \boldsymbol{y}_{\mathrm{c}} = \begin{bmatrix} \hat{\boldsymbol{f}}'_1 & \cdots & \hat{\boldsymbol{f}}'_M \end{bmatrix}' \tag{20}$$

where $\hat{\boldsymbol{f}}_m$ are defined in (5). Thus, by (17), $\hat{\boldsymbol{f}}_{\mathrm{MTR}} \to \hat{\boldsymbol{f}}_1$ as $\zeta \to 0$. In other words, as $\zeta \to 0$, the PWLS MTR estimator for each frame approaches the PWLS SGR estimator (5).

As $\zeta \to \infty$, $\hat{\boldsymbol{f}}_{\mathrm{c}}$ has more interesting limiting behavior. The following theorem is proven in Appendix A.

Theorem 1: As $\zeta \to \infty$, the MTR estimator $\hat{\boldsymbol{f}}_{\mathrm{c}}$ becomes

$$\begin{aligned} \hat{\boldsymbol{f}}_{\mathrm{c}} \quad &\to \quad \mathring{\boldsymbol{T}}_{\mathrm{c}} [\mathring{\boldsymbol{F}}_{\mathrm{c}} + \eta \mathring{\boldsymbol{R}}_{\mathrm{c}}]^{-1} \mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{A}'_{\mathrm{d}} \boldsymbol{W}_{\mathrm{d}} \boldsymbol{y}_{\mathrm{c}} \qquad (21) \\ &= \quad \mathring{\boldsymbol{T}}_{\mathrm{c}} \left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m\right]^{-1} \sum_{m=1}^{M} \mathring{\boldsymbol{T}}'_{m,1} \boldsymbol{A}'_m \boldsymbol{W}_m \boldsymbol{y}_m \end{aligned}$$

where $\mathring{\boldsymbol{T}}_{m,1} = \mathring{\boldsymbol{T}}_{m,m-1} \cdots \mathring{\boldsymbol{T}}_{2,1}$, $\mathring{\boldsymbol{T}}_{\mathrm{c}}$ is defined in (9), $\mathring{\boldsymbol{F}}_{\mathrm{c}} \triangleq \mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{F}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}}$, and $\mathring{\boldsymbol{R}}_{\mathrm{c}} \triangleq \mathring{\boldsymbol{T}}_{\mathrm{c}}{}' \boldsymbol{R}_{\mathrm{d}} \mathring{\boldsymbol{T}}_{\mathrm{c}}$.

B. Equivalence of MTR and PMM estimators

Equation (21) in Theorem 1 and (12) show that the PWLS estimators of PMM and MTR ($\zeta \to \infty$) are remarkably similar. In particular, if we choose a PMM regularizer with

$$\boldsymbol{R}_{\mathrm{PMM}} = \sum_{m=1}^{M} \mathring{\boldsymbol{R}}_m, \tag{22}$$

then the analysis leading to (21) with (17) shows that

$$\hat{\boldsymbol{f}}_{\mathrm{MTR}} \to \hat{\boldsymbol{f}}_{\mathrm{PMM}} \quad \text{as} \quad \zeta \to \infty. \tag{23}$$

In other words, $\hat{\boldsymbol{f}}_{\mathrm{c}} \to \mathring{\boldsymbol{T}}_{\mathrm{c}} \hat{\boldsymbol{f}}_{\mathrm{PMM}}$ as $\zeta \to \infty$. Therefore, assuming some mild conditions on motion and spatial regularizers, the PWLS estimators of PMM and MTR with sufficiently large ζ will be approximately the same, and thus so will the mean and covariance. For the Poisson likelihood, one can show that the mean and covariance of the MTR estimator will approach the mean and covariance of the PMM estimator as ζ increases. We will show the covariance case for the Poisson likelihood in the next section. The mean case with the Poisson likelihood can be shown by consulting [24] and using Appendix A.

C. Difference between PMC and PMM estimators

Using (5), (7), and (22), we rewrite the PWLS PMM estimator (12) as follows:

$$\begin{aligned} \hat{\boldsymbol{f}}_{\text{PMM}} &= \sum_{m=1}^{M} \boldsymbol{\Gamma}_m (\mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m)^{-1} \mathring{\boldsymbol{T}}_{m,1}' \boldsymbol{A}_m' \boldsymbol{W}_m \boldsymbol{y}_m \\ &= \sum_{m=1}^{M} \boldsymbol{\Gamma}_m \mathring{\boldsymbol{T}}_{1,m} \hat{\boldsymbol{f}}_m, \end{aligned} \tag{24}$$

where the weighting matrices are given by

$$\boldsymbol{\Gamma}_m \triangleq \left[\sum_{l=1}^{M} (\mathring{\boldsymbol{F}}_l + \eta \mathring{\boldsymbol{R}}_l) \right]^{-1} (\mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m). \tag{25}$$

Comparing the PWLS PMM estimator (24) and the PWLS PMC estimator (6), we see that the PWLS PMC estimator is a *scalar-weighted* average of the motion corrected PWLS SGR estimators of all frames whereas the PWLS PMM estimator is a *matrix-weighted* average of the motion corrected PWLS estimators. The PWLS MTR estimator (with proper motion and regularizers) approaches the same *matrix-weighted* average of the motion corrected estimators (24) as $\zeta \to \infty$.

The weights $\boldsymbol{\Gamma}_m$ in (25) are calculated using the Fisher information matrices $\mathring{\boldsymbol{F}}_m$. This implies that the PWLS PMM estimator (and the PWLS MTR estimator with $\zeta \to \infty$) automatically assigns different weights to the estimate $\hat{\boldsymbol{f}}_m$ depending on factors such as noise (Fisher information matrix $\boldsymbol{F}_m$) and motion $\mathring{\boldsymbol{T}}_{m,1}$. For the Poisson likelihood case, the next section shows the benefit of this matrix-weighted average (24) by investigating the noise properties of MCIR methods using an "analytical approach" extended from [24] and [23].

IV. Noise properties of MCIR

This section analyzes the noise properties of different MCIR methods. The analysis applies both to PWLS estimators and to maximum a posteriori (MAP) estimators based on Poisson likelihoods. Since the analysis is based on a first-order approximation of the gradient of the likelihood, the accuracy of the analysis for Poisson likelihoods will decrease as the number of counts per frame decreases as shown in [25]. For simplicity, we focus on 2D PET with a few assumptions. We consider an ideal tomography system, *i.e.*, we ignore detector blur. We also assume that $\boldsymbol{A}_m = \boldsymbol{D}_m \boldsymbol{A}_0$ for all m. The (unitless) elements of $\boldsymbol{A}_0$ describe the probability that an emission from the jth pixel is recorded by the ith detector in the absence of attenuation or scatter and for an ideal detector. The ith element of the diagonal matrix $\boldsymbol{D}_m$ has units of time and includes the detector efficiency, the patient-dependent attenuation along the ith ray, and the acquisition time τ_m for the mth frame.

We assume known attenuation map (*i.e.*, $\boldsymbol{D}_m$ is given), which is the usual assumption for PET-CT [30] or PET-MR [31]. We still allow the warp $\mathring{\boldsymbol{T}}_{m,1}$ to differ for each m. We assume that the given nonrigid motion is locally affine [24]. We also assume that the measurements $\boldsymbol{y}_m$ for all m are independent, *i.e.*, $\mathsf{Cov}\{\hat{\boldsymbol{f}}_m, \hat{\boldsymbol{f}}_n\} = 0$ for all $m \neq n$.

We use an "analytical approach" to derive approximate variances for SGR and MCIR methods. This appproach provides fast variance prediction methods [23] compared to the DFT-based variance approximations or numerical simulations.

A. Single gated reconstruction (SGR)

If L in (4) is a negative Poisson log-likelihood function (*i.e.*, $\mathrm{L}(\boldsymbol{y}, \boldsymbol{u}) \triangleq \sum_i \boldsymbol{u}_i - \boldsymbol{y}_i \log \boldsymbol{u}_i$), then one can approximate the covariance of the SGR estimator $\hat{\boldsymbol{f}}_m$ of (4) by [25]:

$$\mathsf{Cov}\{\hat{\boldsymbol{f}}_m\} \approx [\boldsymbol{F}_m + \eta \boldsymbol{R}_m]^{-1} \boldsymbol{F}_m [\boldsymbol{F}_m + \eta \boldsymbol{R}_m]^{-1} \tag{26}$$

where $\boldsymbol{F}_m \triangleq \boldsymbol{A}_0' \boldsymbol{D}_m' \boldsymbol{W}_m \boldsymbol{D}_m \boldsymbol{A}_0$, $\boldsymbol{W}_m \triangleq \boldsymbol{D}\left(1/[\bar{\boldsymbol{y}}_m(\boldsymbol{f}_m)]_i\right)$ is a diagonal matrix, $\bar{\boldsymbol{y}}_m$ is the mean of $\boldsymbol{y}_m$, the Hessian of the regularizer is $\boldsymbol{R}_m \triangleq \nabla^2 \mathrm{R}(\check{\boldsymbol{f}}_m)$, and $\check{\boldsymbol{f}}_m \triangleq \hat{\boldsymbol{f}}_m(\bar{\boldsymbol{y}}_m(\boldsymbol{f}_m))$.

To study (26) using the "analytical approach" of [23], we focus on a first-order difference quadratic regularizer:

$$\boldsymbol{f}_m' \boldsymbol{R}_m \boldsymbol{f}_m \triangleq \sum_j \sum_{l=1}^{L} r_{l,m}^j ((c_l ** f_m)[\vec{n}_j])^2, \tag{27}$$

where $**$ denotes 2D convolution, $r_{l,m}^j$ is a non-negative regularization weight (*e.g.*, regularization designs for uniform and/or isotropic spatial resolution [24], [32]), $f_m[\vec{n}_j]$ denotes the 2D array corresponding to the lexicographically ordered vector $\boldsymbol{f}_m$, j is the lexicographic index of the pixel at 2D coordinates $\vec{n}_j$, and

$$c_l[\vec{n}_j] = \frac{1}{\|\vec{m}_l\|_2} (\delta_2[\vec{n}_j] - \delta_2[\vec{n}_j - \vec{m}_l]), \tag{28}$$

where $\{\vec{m}_l\}$ denote the spatial offsets of the jth pixel's neighbors and $\delta_2[\vec{n}_j]$ denotes the 2D Kronecker impulse. We used the usual 8-pixel 2D neighborhood with $L = 4$ and $\{\vec{m}_l\}_{l=1}^4 = \{(1,0), (0,1), (1,1), (1,-1)\}$.

For a polar coordinate (ρ, φ) in the frequency domain, we can represent the variance of (26) at the jth voxel in an analytical form as follows [23]:

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_m\} \approx \int_0^{2\pi} \int_0^{\rho_{\max}} P_{\mathrm{SGR}}^j(\rho, \varphi) \rho \,\mathrm{d}\rho \,\mathrm{d}\varphi \tag{29}$$

where $\rho_{\max} \triangleq 1/2/\Delta$, Δ is the pixel spatial sampling distance, and the local power spectrum $P_{\mathrm{SGR}}^j(\rho, \varphi)$ at the jth pixel, which is the Fourier transform of the jth column of the covariance in (26) (see also p. 220 of [33]), is

$$P_{\mathrm{SGR}}^j(\rho, \varphi) \triangleq \frac{\bar{w}_m(\varphi; \vec{x}_j)/\rho/\tilde{\Delta}}{\left(\bar{w}_m(\varphi; \vec{x}_j)/\rho/\tilde{\Delta} + \eta (2\pi\rho)^2 Q_m^j(\varphi)\right)^2} \tag{30}$$

where the angular component of the local frequency response of the regularizer (27) is

$$Q_m^j(\varphi) \triangleq \sum_{l=1}^{L} r_{l,m}^j \cos^2(\varphi - \varphi_l) \quad (31)$$

and $\varphi_l = \angle \vec{m}_l$. For a standard quadratic regularizer, $Q_m^j(\varphi) = w_0$ where w_0 is a constant. The analytical forms of $\boldsymbol{F}_m$ and $\boldsymbol{R}_m$ at the jth voxel are $\bar{w}_m(\varphi; \vec{x}_j)/\rho$ and $(2\pi\rho)^2 Q_m^j(\varphi)$ (see [23], [32]) where

$$\bar{w}_m(\varphi; \vec{x}_j) \triangleq \frac{\sum_{i \in \mathcal{I}_\varphi} a_{ij}^2 \tilde{w}_{m,i}}{\sum_{i \in \mathcal{I}_\varphi} a_{ij}^2}, \quad (32)$$

$\mathcal{I}_\varphi$ is the set of rays at the angle φ, $a_{ij} \triangleq [\boldsymbol{A}_0]_{ij}$, $\tilde{w}_{m,i} \triangleq [\boldsymbol{D}_m \boldsymbol{W}_m \boldsymbol{D}_m]_{ii}$, $\tilde{\Delta} \triangleq \Delta^2 \Delta_r \Delta_\varphi$, Δ_r is a detector sampling interval, and Δ_φ is an angular sampling interval. For fast computation, one can approximate $\bar{w}_m(\varphi; \vec{x}_j) \approx \tilde{w}_m(\vec{x}_j \cdot (\cos\varphi, \sin\varphi), \varphi)$ where $\tilde{w}_m(r_i, \varphi_i) \triangleq \tilde{w}_{m,i}$. One can further simplify the local variance $\mathsf{Var}_j\{\hat{\boldsymbol{f}}_m\}$ in (29) by calculating the intergral (29) with respect to ρ as follows [23]:

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_m\} \approx \int_0^\pi \frac{2/3}{\frac{\bar{w}_m(\varphi; \vec{x}_j)}{\tilde{\Delta}\rho_{\max}^3} + \eta 4\pi^2 Q_m^j(\varphi)} \,\mathrm{d}\varphi, \quad (33)$$

where $P_{\mathrm{SGR}}^j(\rho, \varphi + \pi) = P_{\mathrm{SGR}}^j(\rho, \varphi)$. The variance of the SGR estimator for the Poisson likelihood depends on the measurement statistics $\bar{w}_m$, the sampling distances Δ, Δ_r, Δ_φ, and the regularization parameter η. One can also obtain the local autocovariance of the SGR estimator at the jth pixel by taking an inverse Fourier transform (FT) of the local power spectrum $P_{\mathrm{SGR}}^j(\rho, \varphi)$ in (30).

B. Post-reconstruction motion correction (PMC)

Assuming that the measurements $\boldsymbol{y}_m$ for each frame are statistically independent and the reconstruction algorithm uses the Poisson likelihood, the covariance of the PMC estimator (6) is approximately

$$\begin{aligned} \mathsf{Cov}\{\hat{\boldsymbol{f}}_{\mathrm{PMC}}\} &= \sum_{m=1}^{M} \alpha_m^2 \mathring{\boldsymbol{T}}_{1,m} \mathsf{Cov}\{\hat{\boldsymbol{f}}_m\} \mathring{\boldsymbol{T}}_{1,m}' \quad (34) \\ &\approx \sum_{m=1}^{M} \alpha_m^2 [\mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m]^{-1} \mathring{\boldsymbol{F}}_m [\mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m]^{-1}. \end{aligned}$$

We can derive the analytical forms of $\mathring{\boldsymbol{F}}_m$ and $\mathring{\boldsymbol{R}}_m$ (the quadratic regularizer (27)) in the frequency domain as follows (see Appendix B in [24]):

$$\begin{aligned} \mathring{\boldsymbol{F}}_m &: \frac{\bar{w}_m(\tilde{\varphi}; \vec{x}_k) |\nabla T_{m,1}(\vec{x}_j)'|^{2p-1}}{\rho \|\nabla T_{m,1}(\vec{x}_j)'(\cos\varphi, \sin\varphi)'\|_2} \quad (35) \\ \mathring{\boldsymbol{R}}_m &: (2\pi\rho)^2 \|\nabla T_{m,1}(\vec{x}_j)'(\cos\varphi, \sin\varphi)'\|_2^2 \quad (36) \\ &\quad \cdot |\nabla T_{m,1}(\vec{x}_j)'|^{2p-1} Q_m^k(\tilde{\varphi}) \end{aligned}$$

where $\vec{x}_k$ is the closest pixel to $T_{m,1}^{-1}(\vec{x}_j)$ and $\tilde{\varphi} \triangleq \angle \nabla T_{m,1}(\vec{x}_j)'(\cos\varphi, \sin\varphi)'$. Therefore, by using analytical forms, we approximate the variance of $\hat{\boldsymbol{f}}_{\mathrm{PMC}}$ at the jth voxel:

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMC}}\} \approx \int_0^{2\pi} \int_0^{\rho_{\max}} P_{\mathrm{PMC}}^j(\rho, \varphi) \rho \,\mathrm{d}\rho \,\mathrm{d}\varphi \quad (37)$$

where the local power spectrum, $P_{\mathrm{PMC}}^j(\rho, \varphi)$, at the jth pixel is given by

$$\sum_{m=1}^{M} \frac{\alpha_m^2 \bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)/\rho/\tilde{\Delta}}{\left(\bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)/\rho/\tilde{\Delta} + \eta(2\pi\rho)^2 Q_m^k(\tilde{\varphi}) t_{\mathrm{R_m}}^j(\varphi)\right)^2}$$

where the following factors arise from the Fisher information matrix $\boldsymbol{F}_m$ and the Hessian of the regularizer $\boldsymbol{R}_m$ respectively due to motion compensation

$$\begin{aligned} t_{\mathrm{F_m}}^j(\varphi) &\triangleq \frac{|\nabla T_{m,1}(\vec{x}_j)'|^{2p-1}}{\|\nabla T_{m,1}(\vec{x}_j)'(\cos\varphi, \sin\varphi)'\|_2} \quad (38) \\ t_{\mathrm{R_m}}^j(\varphi) &\triangleq \|\nabla T_{m,1}(\vec{x}_j)'(\cos\varphi, \sin\varphi)'\|_2^2 |\nabla T_{m,1}(\vec{x}_j)'|^{2p-1}. \end{aligned}$$

For rigid motion, $t_{\mathrm{F_m}}^j(\varphi) = t_{\mathrm{R_m}}^j(\varphi) = 1$ whereas for nonrigid motion such as (isotropic or anisotropic) scaling, $t_{\mathrm{F_m}}^j(\varphi)$ and $t_{\mathrm{R_m}}^j(\varphi)$ usually differ from 1. By integrating, we simplify the local variance $\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMC}}\}$ in (37) further as follows:

$$\int_0^\pi \sum_{m=1}^{M} \frac{2\alpha_m^2/3}{\frac{\bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)}{\tilde{\Delta}\rho_{\max}^3} + \eta 4\pi^2 Q_m^k(\tilde{\varphi}) t_{\mathrm{R_m}}^j(\varphi)} \,\mathrm{d}\varphi. \quad (39)$$

Note that the variance of the PMC estimator depends on the motion through $t_{\mathrm{F_m}}^j(\varphi)$ and $t_{\mathrm{R_m}}^j(\varphi)$ terms. One can also obtain the local autocovariance of the PMC estimator by taking an inverse FT of $P_{\mathrm{PMC}}^j(\rho, \varphi)$.

C. Parametric motion model

For the PMM estimator (10) with the Poisson likelihood, the covariance of the PMM estimator, $\mathsf{Cov}\{\hat{\boldsymbol{f}}_{\mathrm{PMM}}\}$, can be approximated using the matrix-based methods of [25] as

$$\left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \boldsymbol{R}_{\mathrm{PMM}}\right]^{-1} \sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m \left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \boldsymbol{R}_{\mathrm{PMM}}\right]^{-1}. \quad (40)$$

Using the analytical forms in (35) and (36), the variance of the PMM estimator at the jth pixel is approximately

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMM}}\} \approx \int_0^{2\pi} \int_0^{\rho_{\max}} P_{\mathrm{PMM}}^j(\rho, \varphi) \rho \,\mathrm{d}\rho \,\mathrm{d}\varphi, \quad (41)$$

where the local power spectrum, $P_{\mathrm{PMM}}^j(\rho, \varphi)$, at the jth pixel for (40) is defined as follows:

$$\frac{\sum_{m=1}^{M} \bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)/\rho/\tilde{\Delta}}{\left(\sum_{m=1}^{M} \bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)/\rho/\tilde{\Delta} + \eta(2\pi\rho)^2 Q_{\mathrm{PMM}}^j(\varphi)\right)^2},$$

where $Q_{\mathrm{PMM}}^j(\varphi) \triangleq \sum_{l=1}^{L} r_l^j \cos^2(\varphi - \varphi_l)$. Integrating $P_{\mathrm{PMM}}^j(\rho, \varphi)$ over ρ simplifies the local variance $\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMM}}\}$ in (41) to

$$\int_0^\pi \frac{2/3}{\sum_{m=1}^{M} \frac{\bar{w}_m(\tilde{\varphi}; \vec{x}_k) t_{\mathrm{F_m}}^j(\varphi)}{\tilde{\Delta}\rho_{\max}^3} + \eta 4\pi^2 Q_{\mathrm{PMM}}^j(\varphi)} \,\mathrm{d}\varphi. \quad (42)$$

Like the PMC case, the noise depends on the given motion. The local covariance of the PMM estimator can be approximated with an inverse FT of $P_{\mathrm{PMM}}^j(\rho, \varphi)$.

The covariance of the PMM estimator with the regularizer (22) will be approximately

$$\left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m\right]^{-1} \sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m \left[\sum_{m=1}^{M} \mathring{\boldsymbol{F}}_m + \eta \mathring{\boldsymbol{R}}_m\right]^{-1} \quad (43)$$

and with the same procedure as above, the variance of the PMM estimator at the jth pixel for (43) is approximately

$$\int_0^{\pi} \frac{2/3}{\sum_{m=1}^{M} \frac{\bar{w}_m(\tilde{\varphi};\vec{x}_k) t^j_{\mathrm{F_m}}(\varphi)}{\tilde{\Delta}\rho^3_{\max}} + \eta 4\pi^2 Q^k_m(\tilde{\varphi}) t^j_{\mathrm{R_m}}(\varphi)} \,\mathrm{d}\varphi. \quad (44)$$

One can evaluate (33), (39), (42), and (44) using a simple back projection (*i.e.*, approximate integral by sum over projection angle φ) to predict variance for every image pixel.

D. Motion-compensated temporal regularization

From (15) with the Poisson likelihood, the covariance matrix of the MTR estimator $\hat{\boldsymbol{f}}_{\mathrm{c}}$ is approximately

$$\mathsf{Cov}\{\hat{\boldsymbol{f}}_{\mathrm{c}}\} \approx [\boldsymbol{G}_{\mathrm{MTR}} + \zeta \boldsymbol{R}_{\mathrm{time}}]^{-1} \boldsymbol{F}_{\mathrm{d}} [\boldsymbol{G}_{\mathrm{MTR}} + \zeta \boldsymbol{R}_{\mathrm{time}}]^{-1} \quad (45)$$

where $\boldsymbol{G}_{\mathrm{MTR}} \triangleq \boldsymbol{F}_{\mathrm{d}} + \eta \boldsymbol{R}_{\mathrm{d}}$. Section III showed that the PWLS MTR estimator converges to the PWLS SGR and PMM estimators as $\zeta \to 0$ and $\zeta \to \infty$, respectively. For the estimators with the Poisson likelihood, one can show that the covariance of the MTR estimator (45) "approximately" converges to the covariance of the SGR estimator and the PMM estimator as $\zeta \to 0$ and $\zeta \to \infty$, respectively, using (64) in Appendix A. Therefore, the local variance of the MTR estimator at the jth pixel will approach the SGR result (33) approximately as $\zeta \to 0$ and will approach the PMM result (44) approximately as $\zeta \to \infty$.

Obtaining an analytical form for the variance of MTR with any ζ seems challenging due to the complicated structure of $\boldsymbol{T}_{\mathrm{time}}$ matrix. However, from (45) one can show that the covariance of the MTR decreases as ζ increases. We can also intuitively expect that high ζ value will increase the correlation between estimated image frames, which will reduce the variance of MTR. We evaluate this intuition empirically in Section VI.

V. PERFORMANCE COMPARISONS IN MCIR

This section presents theoretical comparisons of the noise properties of SGR and MCIR methods with the Poisson likelihood.

A. Comparing noise properties between PMC and PMM

As discussed in Section III-C, the PMC estimator is a *scalar-weighted* average of the motion corrected estimators of all frames, whereas the PMM estimator is a *matrix-weighted* average using the weight in (25). This difference led to the different variances of the PMC estimator (39) and of PMM (44) (and the variance of MTR for $\zeta \to \infty$). By matching the spatial resolutions of PMM and PMC using the regularizer (22) for PMM (see [24]), we can also compare the variance of PMC and PMM *theoretically*.

For $v^j_m(\varphi) \geq 0$, one can show that

$$\frac{1}{\sum_{m=1}^{M} v^j_m(\varphi)} \leq \sum_{m=1}^{M} \frac{\alpha_m^2}{v^j_m(\varphi)} \quad (46)$$

using the Cauchy-Schwarz inequality [20] and $\sum_{m=1}^{M} \alpha_m = 1$. If we set

$$v^j_m(\varphi) \triangleq \bar{w}_m(\tilde{\varphi};\vec{x}_k) t^j_{\mathrm{F_m}}(\varphi)/\tilde{\Delta}/\rho^3_{\max} + \eta 4\pi^2 t^j_{\mathrm{R_m}}(\varphi) Q^k_m(\tilde{\varphi}),$$

then (39), (44), and (46) show that

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMM}}\} \leq \mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMC}}\} \quad (47)$$

for the regularized PMC and PMM. Equality holds when all v^j_m are the same for all m. This inequality is consistent with the empirical observations in [20]. Therefore, PMM (and MTR with sufficiently large ζ) is preferable over PMC in terms of noise variance.

B. Comparing noise properties of SGR for three regularizers

Because of the interactions between the likelihood and regularizer, spatial resolution will be anisotropic and non-uniform if one uses a standard regularizer [21], *i.e.*, $Q^j_m(\varphi) = w_0$ in (30), which we call SGR-S. There has been some research on regularizers that provide approximately uniform and/or isotropic spatial resolution [21], [32], [34]. This section analyzes the effect of such regularizers on the noise properties of SGR.

The certainty-based quadratic regularizer proposed in [21] can provide approximately uniform (but still anisotropic) spatial resolution. In this case, $r^j_{l,m}$ in (27) is designed to approximately satisfy

$$Q^j_m(\varphi) \approx \frac{1}{\pi} \int_0^{\pi} \bar{w}_m(\varphi';\vec{x}_j) \,\mathrm{d}\varphi', \quad (48)$$

and we call the estimation SGR-C. Alternatively, one can design $\{r^j_{l,m}\}$ to approximately satisfy

$$Q^j_m(\varphi) \approx \bar{w}_m(\varphi;\vec{x}_j) \quad (49)$$

so that the spatial resolution will be approximately uniform and isotropic [32], [35], which we call SGR-P. From (33), one can show the relationship between the variances of SGR-S and SGR-C as follows:

$$\begin{aligned} &\mathsf{Var}_j\{\hat{\boldsymbol{f}}^{\mathrm{SGR-S}}_m\} \leq \mathsf{Var}_j\{\hat{\boldsymbol{f}}^{\mathrm{SGR-C}}_m\} \qquad w_0 \geq Q^j_m(\varphi) \\ &\mathsf{Var}_j\{\hat{\boldsymbol{f}}^{\mathrm{SGR-S}}_m\} > \mathsf{Var}_j\{\hat{\boldsymbol{f}}^{\mathrm{SGR-C}}_m\} \qquad o.w. \end{aligned} \quad (50)$$

The same relationship holds between SGR-S and SGR-P. The variance of SGR-S can be larger or smaller than the variance of SGR-C and SGR-P for each location (jth pixel).

There is a more interesting relationship between the variances of SGR-C and SGR-P. In both (48) and (49), $\bar{w}_m(\varphi;\vec{x}_j) \approx Q^j_m(\varphi)$ [21], [32], and substituting this further approximation into (33) yields the following simplified variance approximation:

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_m\} \approx \left(\frac{2/3}{1/\tilde{\Delta}/\rho^3_{\max} + \eta 4\pi^2}\right) \int_0^{\pi} \frac{1}{Q^j_m(\varphi)} \,\mathrm{d}\varphi. \quad (51)$$

This approximation becomes increasingly accurate as $\rho_{\max}$ and/or η increase. In our simulations, using (49) in (51) significantly reduced the accuracy of (51) because small differences in (49) became large differences in (51) due to their reciprocal relationship. Using (48) and (49) to achieve approximately uniform and/or isotropic spatial resolution will increase the effect of the measurement statistics $\bar{w}_m$ on the estimator variance (51) compared to (33). This tendency was empirically observed in [21]. Using the Cauchy-Schwarz inequality, one can show that the variance approximation in (51) satisfies

$$\mathsf{Var}_j\{\hat{\boldsymbol{f}}_m^{\mathrm{SGR-C}}\} \leq \mathsf{Var}_j\{\hat{\boldsymbol{f}}_m^{\mathrm{SGR-P}}\}. \tag{52}$$

This inequality is verified empirically in Section VI-B. Evidently, imposing more properties on the spatial resolution such as isotropy requires sacrificing the noise performance, which shows the spatial resolution-noise trade-off.

C. Comparing noise properties between SGR and MCIR

If there is no motion between image frames and $\bar{w}_m = \bar{w}_1$ for all m, then (33), (39), and (44) yield $\mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMC}}\} = \mathsf{Var}_j\{\hat{\boldsymbol{f}}_{\mathrm{PMM}}\} = \mathsf{Var}_j\{\hat{\boldsymbol{f}}_m\}/M$, as expected since PMC and PMM used M times more counts than SGR. The MTR variance $\mathsf{Var}\{\hat{\boldsymbol{f}}_{\mathrm{MTR}}\}$ with very high ζ also yields approximately the same variance as PMM and PMC in this case.

However, this $1/M$ relationship between MCIR and SGR variances may not hold exactly when there is motion between image frames. For example, if there is locally isotropic scaling motion between frames as follows:

$$\nabla T_{m,1}(\vec{x}_j) = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} \tag{53}$$

where $s > 0$, then $t^j_{\mathrm{F_m}}(\varphi) = s^{4p-3}$ and $t^j_{\mathrm{R_m}}(\varphi) = s^{4p}$ in (38). For PMC, if we design the regularizer to achieve isotropic resolution by using

$$t^j_{\mathrm{R_m}}(\varphi) Q^k_m(\tilde{\varphi}) \approx \bar{w}_m(\tilde{\varphi}; \vec{x}_k) t^j_{\mathrm{F_m}}(\varphi), \tag{54}$$

and if $\rho_{\max}$ and/or η are relatively large, then the variance of the PMC estimator at the jth pixel in (39) approximately reduces to

$$\left(\frac{2/3/M^2}{1/\tilde{\Delta}/\rho^3_{\max} + \eta 4\pi^2}\right) \sum_{m=1}^{M} \int_0^\pi \frac{1}{t^j_{\mathrm{R_m}}(\varphi) Q^k_m(\tilde{\varphi})} \,\mathrm{d}\varphi. \tag{55}$$

Comparing with (51), the variance of PMC (55) will be approximately $1/M/s^{4p}$ times the variance of SGR for $M \gg 1$. The variance of PMM (44) will have a similar relationship with the variance of SGR. If the total activity is preserved (*i.e.*, $p = 1$), then local expansion ($s < 1$) will increase the variance and local shrinkage ($s > 1$) will decrease the variance. Intuitively, if the same amount of total activity produces the same number of Poisson counts, the expanded area that contains the same total activity will have larger image area to estimate, *i.e.*, effectively more parameters. Thus, the expanded area will lead to higher estimator variance. For regularizers other than (54), the variance of PMC will also be affected by motion through $t^j_{\mathrm{F_m}}(\varphi)$ and $t^j_{\mathrm{R_m}}(\varphi)$ terms.

D. Total activity preserving condition for MCIR

The total activity preserving condition (2) is important for accurate motion modeling and it also affects the spatial resolution [24] and noise properties of MCIR. Using the example in Section V-C, we analyze the influence of motion on the noise, focusing on PMC and PMM. (MTR with sufficiently large ζ will have approximately the same noise properties as PMM.)

If one uses standard quadratic regularizers for PMC and PMM (*e.g.*, $Q^j_m(\varphi) = w_0$ and $Q^j_{\mathrm{PMM}}(\varphi) = M w_0$ in (31)), then the variance of the PMC estimator in (39) reduces to

$$\int_0^\pi \sum_{m=1}^{M} \frac{2\alpha_m^2/3}{\bar{w}_m(\tilde{\varphi}; \vec{x}_k) s^{4p-3}/\tilde{\Delta}/\rho^3_{\max} + \eta 4\pi^2 w_0 s^{4p}} \,\mathrm{d}\varphi \tag{56}$$

since $t^j_{\mathrm{F_m}}(\varphi) = s^{4p-3}$ and $t^j_{\mathrm{R_m}}(\varphi) = s^{4p}$ when (53) holds. The variance of the PMM estimator in (42) reduces to

$$\int_0^\pi \frac{2/3}{\sum_{m=1}^{M} \left(\bar{w}_m(\tilde{\varphi}; \vec{x}_k) s^{4p-3}/\tilde{\Delta}/\rho^3_{\max} + \eta 4\pi^2 w_0\right)} \,\mathrm{d}\varphi. \tag{57}$$

When $s = 1$ (*e.g.*, rigid motion), the variance is not affected by motion. However, when $s \neq 1$, the variance of PMC will be always affected by motion, whether the total activity is preserved or not, due to the s^{4p-3} and s^{4p} terms in (56). However, when $\rho_{\max}$ and/or η are relatively large, the variance of PMM may be less affected by motion when $p = 1$ than when $p = 0$ since (57) only contains s, which is relatively closer to 1 than s^4 or s^{-3}. Since the regularizer in PMM does not involve the motion warp, there is no $t^j_{\mathrm{R_m}}(\varphi)$ term in the variance (42) of PMM. Thus, when we use the total activity preserving condition $p = 1$ with the standard regularizers, the variance of PMM may be less affected by motion than the variance of PMC.

When one designs the spatial regularizers (*i.e.*, determine $r^j_{l,m}$ in (27)) to achieve approximately uniform and/or isotropic spatial resolution for the MCIR methods [24], as shown in Section V-C, the variances of PMC and PMM will be affected by motion with the factor of s^{4p-3}. Thus, the variance of PMC and PMM will be less affected by motion when $p = 1$ than when $p = 0$. Note that the analyses above assumed that both measurement model and reconstruction model follow the same condition. One could generalize these analyses to consider the effects of motion model mismatch.

VI. Simulation Results

The analyses in this paper apply to nonrigid motions that are approximately locally affine [24]. We performed PET simulations with two digital phantoms: one is a simple phantom with global affine motion between frames and the other is the XCAT phantom [36] with non-affine nonrigid motion that we modeled using B-splines [37].

A. Simulation setting

Two digital phantoms were used, each with four frames of 160×160 pixels with 3.4 mm pixel width. Sinograms were generated using a PET scanner geometry with 400 detector samples, 1.9 mm spacing, 220 angular views, and 1.9 mm strip width. We used 300K, 500K, 200K, 200K mean true coincidences for each frame (1.2M total) with 10% random coincidences. Simple uniform attenuation maps were used for the first simulation and no attenuation was used for the second.

We investigated SGR, PMC, and PMM by comparing analytical standard deviation (SD) with empirical SD from 500 Poisson noise realizations. We used spatial regularizers (with regularization parameter $\eta = 10^4$) that provide approximately uniform (SGR-C, PMC-C, PMM-C) and uniform/isotropic (SGR-P, PMC-P, PMM-P) spatial resolutions, respectively [20], [21], [24], [32]. We also studied the noise properties of MTR empirically with various ζ values. The spatial resolutions of SGR, PMC, PMM and MTR were all matched to each other using the regularization designs in [24]. All images were reconstructed using a L-BFGS-B (quasi-Newton) algorithm with non-negativity constraints [38], [39].

B. Simple phantom with affine motion

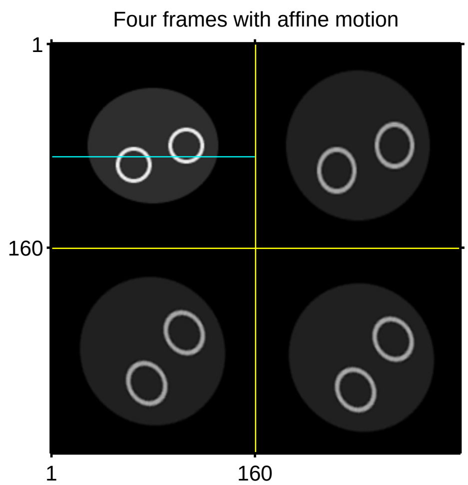

Fig. 1. Four true images with anisotropic scaling, rotation and translation. Total activity is preserved.

We used a simple digital phantom with known affine motion (anisotropic scaling between frame 1 and 2, rotation between frame 2 and 3, and translation between frame 3 and 4) as shown in Fig. 1. The total activity is preserved between frames.

Fig. 2 displays profiles through the variance image and shows that our analytical equation for SGR in (33) (and (51)) provides accurate noise predictions. (The location of the profile is indicated in Fig. 1 as a horizontal line). The analytical SD of SGR with quadratic regularizers (A-SGR-C and A-SGR-P) matches well with the empirical SD of SGR from 500 noise realizations (E-SGR-C and E-SGR-P). Fig. 2 also shows that the variance of SGR-C is lower than the variance of SGR-P as shown in (52) (in this case, η was fairly large). This analytical and empirical agreement of SGR does not hold well near the boundary of and outside the object because of the non-negativity constraint and because the “locally shift invariant” approximation is less accurate there. We observed similar results for a constant quadratic regularizer (not shown).

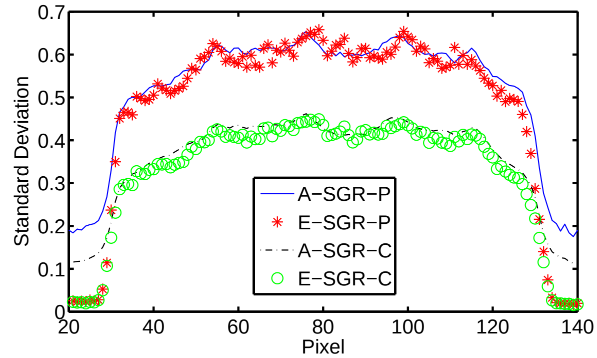

Fig. 2. Analytical SD of SGR (A-SGR-P, A-SGR-C) matches well with empirical SD of SGR (E-SGR-P, E-SGR-C), respectively. SD of SGR-P (with regularizer that approximately uniform and isotropic spatial resolution) is higher than SD of SGR-C (with regularizer that approximately uniform spatial resolution), which is consistent with theoretical comparison.

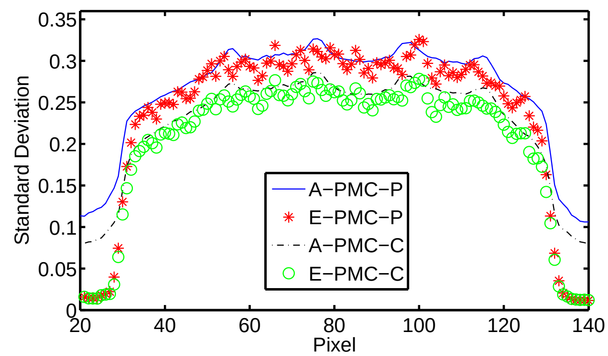

Fig. 3. Analytical SD of PMC (A-PMC-P, A-PMC-C) matches well with empirical SD of PMC (E-PMC-P, E-PMC-C), respectively.

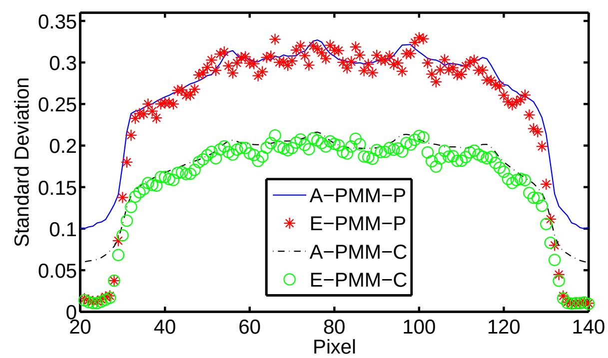

Fig. 4. Analytical SD of PMM (A-PMM-P, A-PMM-C) matches well with empirical SD of PMM (E-PMM-P, E-PMM-C), respectively.

Fig. 3 shows that our analytical variance prediction for PMC (A-PMC-C and A-PMC-P) in (39) agrees with the empirical variance of PMC (E-PMC-C and E-PMC-P). Fig. 4 also shows that the analytical variance formula for PMM in (44) predicts the empirical variance of PMM well.

Fig. 5 confirms the theoretical noise comparison between PMC and PMM shown in (47). As shown in Fig. 5, the SD of unweighted PMC was generally lower than the SD of PMM. However, the difference between the SD of PMM and the SD of scalar-weighted PMC (using weights that account for the number of counts per frame) was very small. Using the spatial regularizer for PMM as proposed in (22) that matches to PMC, the full-width-half-maximum (FWHM) of PMC (2.30 ± 0.13 pixels) was slightly larger than the FWHM of PMM (2.19 ± 0.05 pixels). Our target FWHM was 2.19 ± 0.01 pixels. This small discrepancy was because our analysis assumed perfect interpolations for warps, whereas the actual interpolations induce slight blurring. For PMC, the warp is applied *after* the reconstruction, thus the FWHM was slightly larger than the target FWHM. We observed that the SD of scalar-weighted PMC was *slightly* lower than the SD of PMM empirically, due to it being slightly blurred more.

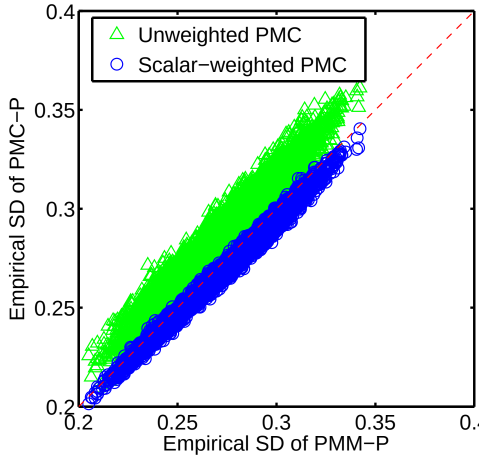

Fig. 5. If the spatial resolutions are matched, the SD of PMC is higher than or comparable to the SD of PMM, depending on the choice of weights α_m.

Section V-C showed that if we combine M image frames with the motion (53), then the variance of MCIR would not be $1/M$ of the variance of SGR due to motion effects. In other words, as shown in Fig. 6, the SD of PMC will not be $1/2$ of the SD of SGR (4 frames), but will be approximately $1/2/|\boldsymbol{J}|$ of the SD of SGR where $\boldsymbol{J} \triangleq \nabla T_{m,1}$ and $m \neq 1$. This example confirms that the variance of MCIR methods depend on the Jacobian determinant of the transformation T.

Fig. 7 shows that the empirical variance of MTR approaches the analytical variance of SGR if $\zeta \to 0$ and to the analytical variance of PMM if $\zeta \to \infty$ as shown in Section IV-D.

We also repeated the reconstructions and noise predictions using motion parameters that were translated by 1 pixel (3.4 mm) away from their true values. We examined the empirical and predicted noise standard deviations for all pixels within two pixels of the outer boundary of the object. For PMC-C the maximum (mean) percent error between the predicted and empirical SD increased from 16.5% (3.2%) without motion error to 17.0% (3.3%) with motion error. For PMM-C the maximum (mean) percent errors were 16.0% (3.7%) and 15.0% (3.8%) without and with motion error, respectively.

C. XCAT phantom with nonrigid motion

We used the XCAT digital phantom [36] to generate 4 volumes with respiratory and cardiac motion and selected one

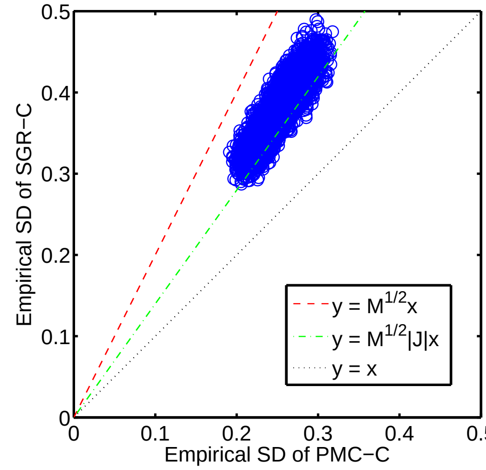

Fig. 6. Empirical SD of SGR vs. empirical SD of PMC (MCIR) with 4 frames ($M = 4$). The SD of PMC will be affected by *both* the number of frames and the motion (Jacobian determinant of transformation $|\boldsymbol{J}|$).

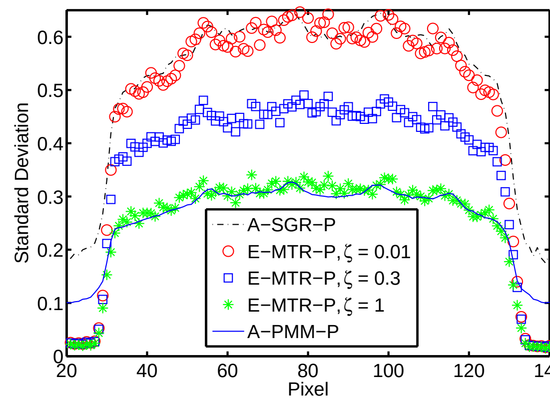

Fig. 7. Empirical SD of MTR with different ζ. As $\zeta \to 0$, the SD of MTR approaches the analytical SD of SGR. As $\zeta \to \infty$, the SD of MTR approaches the analytical SD of PMM.

slice per each volume (same location) for a 2D simulation. After estimating transformations between frames for all MCIR methods consistently (see [24] for details), we used them as the true motion, leading to the true images shown in Fig. 8. Thus, there is no motion model mismatch in this experiment.

As shown in the previous simulation with affine motion, our fast variance predictions for PMC and PMM, which correspond to (39) and (44), work well for the case of nonrigid, non-affine motion as shown in Figs. 9 and 10. There are some areas that match less well than other areas (and compared to the case of affine motion) since there are areas that contain abrupt change of motion so that the local affine approximation does not hold well. Fig. 11 also shows that the empirical SD of MTR approached to the analytical SD of SGR and PMM as $\zeta \to 0$ and $\zeta \to \infty$, respectively.

VII. DISCUSSION

We analyzed the noise properties of three different PWLS MCIR methods for the case of known nonrigid motion.

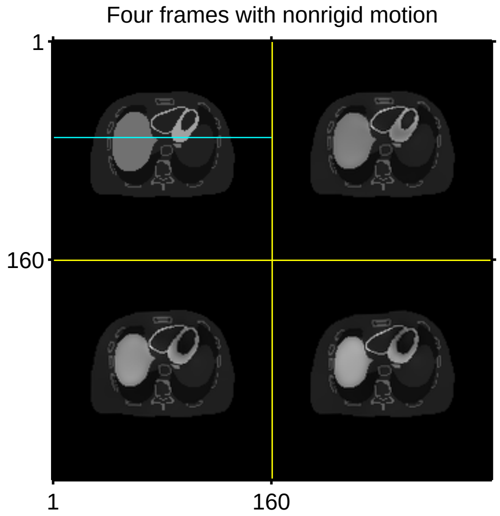

Fig. 8. Four true images with nonrigid motion. Total activity is preserved.

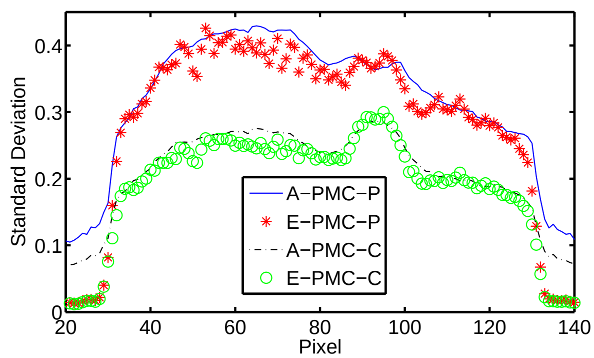

Fig. 9. Analytical SD of PMC (A-PMC-P, A-PMC-C) matches well with empirical SD of PMC (E-PMC-P, E-PMC-C), respectively.

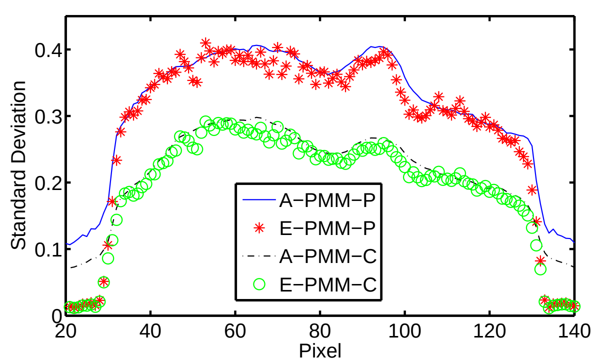

Fig. 10. Analytical SD of PMM (A-PMM-P, A-PMM-C) matches well with empirical SD of PMM (E-PMM-P, E-PMM-C), respectively.

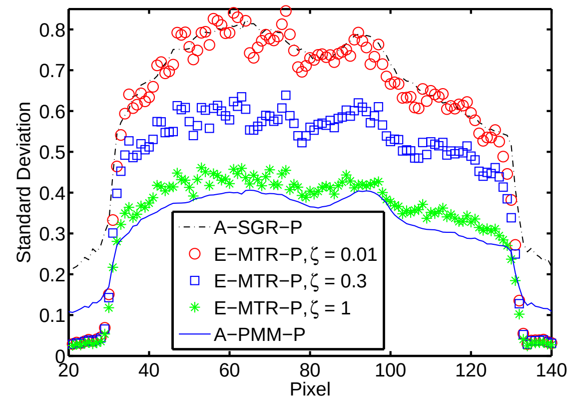

Fig. 11. Empirical SD of MTR with different ζ. As $\zeta \to 0$, the SD of MTR approaches to the analytical SD of SGR. As $\zeta \to \infty$, the SD of MTR approaches to the analytical SD of PMM.

We showed that the PMC is a *scalar-weighted* sum of the motion corrected estimated image frames, whereas the PMM and the MTR with $\zeta \to \infty$ are *matrix-weighted* sum with weights that depend on the Fisher information matrix of each frame. We further investigated the noise properties of three different MCIR methods with Poisson likelihood. We derived approximate variance prediction equations for PMC and PMM and also studied the limiting behavior of the MTR variance as $\zeta \to \infty$ and $\zeta \to 0$. These predictions worked well for digital phantoms with affine motion and non-affine nonrigid motion. Furthermore, as in [23], the variance predictions (33), (39), and (42) require computation time comparable to a back-projection, which is much faster than DFT-based variance prediction methods [20]. However, as the number of counts per frame decreases (due to less total counts or more number of frames), the accuracy of the variance predictions will also decrease since our variance approximations are based on a first-order approximation of the gradient of the likelihood function. [25]. More accurate variance predictions based on higher-order approximations will be challenging.

These analytical variance formulas showed a few interesting relationship between MCIR methods. The variance of SGR-C (using spatial regularizer that approximately provides uniform spatial resolution) is lower than the variance of SGR-P (using spatial regularizer that approximately provides uniform and isotropic spatial resolution). We observed this trend in PMC and PMM as well. The variance of PMM is less than or comparable to the variance of PMC and the gap between them will be larger when the frames have significantly different counts and PMC uses equal scalar weighted sum. When PMC uses proper weights (*e.g.*, normalized scan durations), PMC and PMM empirically had similar variances in our simple phantom simulation with affine motions. The variance of PMM is also less affected by motion than the variance of PMC when the total activity preserving condition is used. The variance of MCIR with M frames may not provide $1/M$ times lower variance than the variance of SGR due to motion. This suggests that one can choose the reference frame to minimize the variance of MCIR methods based on this intuition. Lastly, MTR with very large ζ usually yields images as good as PMM. However, too large ζ can slow convergence of the reconstruction algorithm. When the motion is given, PMM seems to be preferable to PMC and MTR.

This paper has focused on the case of known true motion. In practice motion is never known perfectly and motion errors may introduce further bias and/or variability into MCIR results and motion errors may also degrade the accuracy of noise predictions. Our anecdotal results with motion errors in Section VI-B suggest that the noise predictions are not highly sensitive to small motion errors; in fact the noise predictions seem to be less sensitive to motion errors than were the regularizer designs for MCIR described in [24]. Methods for reducing motion errors will of course improve MCIR results, regularizer designs, and noise prediction accuracy.

This analysis can serve as a starting point for understanding joint estimation of image and motion [12]. Since the Jacobian determinant of estimated deformations affects the noise properties, it is important to enforce correct prior knowledge for local volume changes. Extending this analysis for unknown nonrigid motion will be interesting future work [40]. Our work has been focused on spatial resolution [24] and noise analyses of MCIR methods; it would also be interesting to extend the work to analyze detection performance [41], [42].

APPENDIX A
PROOF OF THEOREM 1

To prove this theorem, we need to treat the null space of $\boldsymbol{R}_{\mathrm{time}}$ carefully. Since the matrix $\boldsymbol{R}_{\mathrm{time}}$ in (18) is symmetric nonnegative definite (*i.e.*, positive semidefinite), it has an orthonormal eigen-decomposition of the form

$$\boldsymbol{R}_{\mathrm{time}} = [\boldsymbol{U}_1 \quad \boldsymbol{U}_0] \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} [\boldsymbol{U}_1 \quad \boldsymbol{U}_0]' \tag{58}$$

where the columns of the matrices $\boldsymbol{U}_1$, $\boldsymbol{U}_0$ are orthonormal and $\boldsymbol{\Sigma}_1 \succ \boldsymbol{0}$, *i.e.*, $\boldsymbol{\Sigma}_1$ is positive definite. The columns of $\boldsymbol{U}_0$ span the null space of $\boldsymbol{R}_{\mathrm{time}}$. From the definition of $\boldsymbol{T}_{\mathrm{time}}$ in (16), it is clear that the null space of $\boldsymbol{R}_{\mathrm{time}}$ consists of images that satisfy the following conditions:

$$\begin{aligned} \boldsymbol{f}_2 &= \mathring{\boldsymbol{T}}_{2,1}\boldsymbol{f}_1 \\ \boldsymbol{f}_3 &= \mathring{\boldsymbol{T}}_{3,2}\mathring{\boldsymbol{T}}_{2,1}\boldsymbol{f}_1 \\ &\vdots \\ \boldsymbol{f}_M &= \mathring{\boldsymbol{T}}_{M,M-1}\cdots\mathring{\boldsymbol{T}}_{2,1}\boldsymbol{f}_1, \end{aligned} \tag{59}$$

for any image $\boldsymbol{f}_1 \in \mathbb{R}^N$. In other words, the $MN \times MN$ matrix $\boldsymbol{R}_{\mathrm{time}}$ has a null space of dimension N. (In contrast, the spatial regularizer $\boldsymbol{C}_{\mathrm{d}}$ usually has a null space only of dimension 1, which is usually formed of constant images.) We rewrite the system of equations (59) as

$$\boldsymbol{f}_{\mathrm{c}} = \mathring{\boldsymbol{T}}_{\mathrm{c}}\boldsymbol{f}_1, \tag{60}$$

where $\mathring{\boldsymbol{T}}_{\mathrm{c}}$ is defined in (9) and $\mathring{\boldsymbol{T}}_{m,1} = \mathring{\boldsymbol{T}}_{m,m-1}\cdots\mathring{\boldsymbol{T}}_{2,1}$. Even if we add a periodic condition $\boldsymbol{f}_1 = \mathring{\boldsymbol{T}}_{1,M}\boldsymbol{f}_M$ to (16), then $\boldsymbol{R}_{\mathrm{time}}$ still has a null space of dimension N provided the transitivity property of the motion model holds. Using (60) we can construct $\boldsymbol{U}_0$ in (58) as follows:

$$\boldsymbol{U}_0 = \mathring{\boldsymbol{T}}_{\mathrm{c}}\boldsymbol{S}, \tag{61}$$

where $\boldsymbol{S} \triangleq \left(\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\mathring{\boldsymbol{T}}_{\mathrm{c}}\right)^{-1/2}$ so that $\boldsymbol{U}_0$ is orthonormal. Note that $\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\mathring{\boldsymbol{T}}_{\mathrm{c}} \succ \boldsymbol{0}$ because $\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\mathring{\boldsymbol{T}}_{\mathrm{c}} = \boldsymbol{I} + \sum_{m=2}^{M}\mathring{\boldsymbol{T}}'_{m,1}\mathring{\boldsymbol{T}}_{m,1}$ and $\boldsymbol{I}$ is positive definite. So, $\boldsymbol{S}$ is invertible.

Under the usual assumption that $\boldsymbol{F}_{\mathrm{d}}$ and $\boldsymbol{R}_{\mathrm{d}}$ have disjoint null spaces, one can verify that

$$\boldsymbol{B} \triangleq \boldsymbol{U}_0'\boldsymbol{G}_{\mathrm{MTR}}\boldsymbol{U}_0 \succ \boldsymbol{0}. \tag{62}$$

To proceed, we express $\boldsymbol{G}_{\mathrm{MTR}}$ in (19) as follows:

$$[\boldsymbol{U}_1 \quad \boldsymbol{U}_0]'\boldsymbol{G}_{\mathrm{MTR}}[\boldsymbol{U}_1 \quad \boldsymbol{U}_0] = \begin{bmatrix} \boldsymbol{N} & \boldsymbol{M}' \\ \boldsymbol{M} & \boldsymbol{B} \end{bmatrix}.$$

Thus,

$$[\boldsymbol{G}_{\mathrm{MTR}} + \zeta\boldsymbol{R}_{\mathrm{time}}]^{-1} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{N}+\zeta\boldsymbol{\Sigma}_1 & \boldsymbol{M}' \\ \boldsymbol{M} & \boldsymbol{B} \end{bmatrix}^{-1}\boldsymbol{U}'$$

where $\boldsymbol{U} \triangleq [\boldsymbol{U}_1 \quad \boldsymbol{U}_0]$. By Schur complement [43], we have

$$[\boldsymbol{G}_{\mathrm{MTR}} + \zeta\boldsymbol{R}_{\mathrm{time}}]^{-1} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Delta} & -\boldsymbol{\Delta}\boldsymbol{M}'\boldsymbol{B}^{-1} \\ -\boldsymbol{B}^{-1}\boldsymbol{M}\boldsymbol{\Delta} & \boldsymbol{B}^{-1}+\boldsymbol{B}^{-1}\boldsymbol{M}\boldsymbol{\Delta}\boldsymbol{M}'\boldsymbol{B}^{-1} \end{bmatrix}\boldsymbol{U}' \tag{63}$$

where $\boldsymbol{\Delta} \triangleq [\boldsymbol{N}+\zeta\boldsymbol{\Sigma}_1 - \boldsymbol{M}'\boldsymbol{B}^{-1}\boldsymbol{M}]^{-1}$. Since $\boldsymbol{\Sigma}_1$ is positive definite, $\boldsymbol{\Delta} \to \boldsymbol{0}$ as $\zeta \to \infty$. Thus, by (62)

$$\begin{aligned} [\boldsymbol{G}_{\mathrm{MTR}} + \zeta\boldsymbol{R}_{\mathrm{time}}]^{-1} &\to \boldsymbol{U}_0\boldsymbol{B}^{-1}\boldsymbol{U}_0' \\ &= \mathring{\boldsymbol{T}}_{\mathrm{c}}[\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\boldsymbol{G}_{\mathrm{MTR}}\mathring{\boldsymbol{T}}_{\mathrm{c}}]^{-1}\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'. \end{aligned} \tag{64}$$

Therefore, as $\zeta \to \infty$

$$\hat{\boldsymbol{f}}_{\mathrm{c}} \to \mathring{\boldsymbol{T}}_{\mathrm{c}}[\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\boldsymbol{G}_{\mathrm{MTR}}\mathring{\boldsymbol{T}}_{\mathrm{c}}]^{-1}\mathring{\boldsymbol{T}}_{\mathrm{c}}{}'\boldsymbol{A}_{\mathrm{d}}'\boldsymbol{W}_{\mathrm{d}}\boldsymbol{y}_{\mathrm{c}}. \tag{65}$$

REFERENCES

[1] G. J. Klein, B. W. Reutter, and R. H. Huesman, “Non-rigid summing of gated PET via optical flow,” *IEEE Trans. Nuc. Sci.*, vol. 44, no. 4, pp. 1509–12, Aug. 1997.

[2] M. Dawood, N. Lang, X. Jiang, and K. P. Schafers, “Lung motion correction on respiratory gated 3-D PET/CT images,” *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 476–85, Apr. 2006.

[3] W. Bai and M. Brady, “Regularized B-spline deformable registration for respiratory motion correction in PET images,” *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2719–36, May 2009.

[4] E. Gravier, Y. Yang, and M. Jin, “Tomographic reconstruction of dynamic cardiac image sequences,” *IEEE Trans. Im. Proc.*, vol. 16, no. 4, pp. 932–42, Apr. 2007.

[5] B. A. Mair, D. R. Gilland, and J. Sun, “Estimation of images and nonrigid deformations in gated emission CT,” *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1130–44, Sep. 2006.

[6] M. W. Jacobson and J. A. Fessler, “Joint estimation of image and deformation parameters in motion-corrected PET,” in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, vol. 5, 2003, pp. 3290–4.

[7] F. Qiao, T. Pan, J. W. Clark, and O. R. Mawlawi, “A motion-incorporated reconstruction method for gated PET studies,” *Phys. Med. Biol.*, vol. 51, no. 15, pp. 3769–84, Aug. 2006.

[8] T. Li, B. Thorndyke, E. Schreibmann, Y. Yang, and L. Xing, “Model-based image reconstruction for four-dimensional PET,” *Med. Phys.*, vol. 33, no. 5, pp. 1288–98, May 2006.

[9] K. Taguchi, Z. Sun, W. P. Segars, E. K. Fishman, and B. M. W. Tsui, “Image-domain motion compensated time resolved 4D cardiac CT,” in *Proc. SPIE 6510, Medical Imaging 2007: Phys. Med. Im.*, 2007, p. 651016.

[10] F. Lamare, M. J. Ledesma Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C. Cheze LeRest, A. J. Reader, and D. Visvikis, “List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations,” *Phys. Med. Biol.*, vol. 52, no. 17, pp. 5187–204, Sep. 2007.

[11] F. Odille, N. Cîndea, D. Mandry, C. Pasquier, P.-A. Vuissoz, and J. Felblinger, “Generalized MRI reconstruction including elastic physiological motion and coil sensitivity encoding,” *Mag. Res. Med.*, vol. 59, no. 6, pp. 1401–11, Jun. 2008.

[12] S. Y. Chun and J. A. Fessler, "Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility," in *Proc. SPIE 7258, Medical Imaging 2009: Phys. Med. Im.*, 2009, p. 72580U.

[13] M. Blume, A. Martinez-Möller, A. Keil, N. Navab, and M. Rafecas, "Joint reconstruction of image and motion in gated positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1892–1906, Nov 2010.

[14] B. Guérin, S. Cho, S. Y. Chun, X. Zhu, N. M. Alpert, G. E. Fakhri, T. Reese, and C. Catana, "Nonrigid pet motion compensation in the lower abdomen using simultaneous tagged-mri and pet imaging," *Medical Physics*, vol. 38, no. 6, pp. 3025–3038, 2011.

[15] C. J. Ritchie, J. Hsieh, M. F. Gard, J. D. Godwin, Y. Kim, and C. R. Crawford, "Predictive respiratory gating: a new method to reduce motion artifacts on CT scans," *Radiology*, vol. 190, no. 3, pp. 847–52, Mar. 1994.

[16] S. A. Nehmeh, Y. E. Erdi, C. C. Ling, K. E. Rosenzweig, H. Schoder, S. M. Larson, H. A. Macapinlac, O. D. Squire, and J. L. Humm, "Effect of respiratory gating on quantifying PET images of lung cancer," *J. Nuc. Med.*, vol. 43, no. 7, pp. 876–81, Jul. 2002.

[17] R. Manjeshwar, . X. Tao, E. Asma, and K. Thielemans, "Motion compensated image reconstruction of respiratory gated PET/CT," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, 2006, pp. 674–7.

[18] K. Thielemans, R. Manjeshwar, X. Tao, and E. Asma, "Lesion detectability in motion compensated image reconstruction of respiratory gated pet/ct," in *Nuclear Science Symposium Conference Record, 2006. IEEE*, vol. 6, Nov 2006, pp. 3278–3282.

[19] M. V. W. Zibetti, F. S. V. Bazán, and J. Mayer, "Determining the regularization parameters for super-resolution problems," *Signal Process.*, vol. 88, no. 12, pp. 2890–2901, 2008.

[20] E. Asma, R. Manjeshwar, and K. Thielemans, "Theoretical comparison of motion correction techniques for PET image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, vol. 3, 2006, pp. 1762–7.

[21] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, Sep. 1996.

[22] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, May 2000.

[23] Y. Zhang-O'Connor and J. A. Fessler, "Fast predictions of variance images for fan-beam transmission tomography with quadratic regularization," *IEEE Trans. Med. Imag.*, vol. 26, no. 3, pp. 335–46, Mar. 2007.

[24] S. Y. Chun and J. A. Fessler, "Spatial resolution properties of motion-compensated tomographic image reconstruction methods," *IEEE Transactions on Medical Imaging*, 2012, accepted for publication.

[25] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Im. Proc.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.

[26] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Im. Proc.*, vol. 6, no. 12, pp. 1646–58, Dec. 1997.

[27] R. Fransens, C. Strecha, and L. Van Gool, "Optical flow based super-resolution: A probabilistic approach," *Comp. Vision & Im. Understanding*, vol. 106, no. 1, pp. 106–15, Apr. 2007.

[28] M. Zibetti and J. Mayer, "A robust and computationally efficient simultaneous super-resolution scheme for image sequences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 10, pp. 1288–1300, Oct. 2007.

[29] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I—theory," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 821–33, Feb. 1993.

[30] P. E. Kinahan, D. W. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3d PET/CT scannter," *Med. Phys.*, vol. 25, no. 10, pp. 2046–53, Oct. 1998.

[31] M. Hofmann, B. Pichler, B. Schlkopf, and T. Beyer, "Towards quantitative pet/mri: a review of mr-based attenuation correction techniques," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 0, pp. 93–104, 2009.

[32] H. R. Shi and J. A. Fessler, "Quadratic regularization design for 2D CT," *IEEE Trans. Med. Imag.*, vol. 28, no. 5, pp. 645–56, May 2009.

[33] H. H. Barrett and K. J. Myers, *Foundations of image science*. New York: Wiley, 2003.

[34] J. W. Stayman and J. A. Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 269–84, Mar. 2004.

[35] J. A. Fessler, "Analytical approach to regularization design for isotropic spatial resolution," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, vol. 3, 2003, pp. 2022–6.

[36] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–8, Aug. 2008.

[37] S. Y. Chun and J. A. Fessler, "A simple regularizer for B-spline nonrigid image registration that encourages local invertibility," *IEEE J. Sel. Top. Sig. Proc.*, vol. 3, no. 1, pp. 159–69, Feb. 2009, special Issue on Digital Image Processing Techniques for Oncology.

[38] J. L. Morales and J. Nocedal, "Remark on "algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound constrained optimization"," *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 7:1–7:4, Nov. 2011.

[39] S. Becker, "L-BFGS-B mex wrapper [online]," Available: http://www.mathworks.com/matlabcentral/fileexchange/35104, Feb 2012.

[40] D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1413–1428, June 2006.

[41] P. Bonetto, J. Qi, and R. M. Leahy, "Covariance approximation for fast and accurate computation of channelized Hotelling observer statistics," *IEEE Trans. Nuc. Sci.*, vol. 47, no. 4, pp. 1567–72, Aug. 2000.

[42] Y. Xing, I.-T. Hsiao, and G. Gindi, "Rapid calculation of detectability in Bayesian single photon emission computed tomography," *Phys. Med. Biol.*, vol. 48, no. 22, pp. 3755–74, Nov. 2003.

[43] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Review*, vol. 23, no. 1, pp. 53–60, Jan. 1981.