Convex Temporal Regularizers in Cardiac C-arm CT

Oliver Taubmann*, Günter Lauritsch†, Gregor Krings‡, Andreas Maier*
*Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg
and Erlangen Graduate School in Advanced Optical Technologies, Germany
†Siemens Healthcare GmbH, Forchheim, Germany
‡University Medical Center Utrecht, Netherlands

E-mail: oliver.taubmann@fau.de

***Abstract*—Dynamic cardiac imaging typically requires the use of gating. In the case of computed tomography (CT), this results in an angular undersampling that renders reconstruction difficult. Prior work has shown that incorporating information from the full scan, i. e. from all cardiac phases, can be beneficial in this regard, for instance by regularization. This paper compares three convex temporal regularizers for 4-D cardiac C-arm CT in both a numerical phantom and clinical patient data. Regularizations based on the nuclear norm, temporal total variation as well as a tight-frame wavelet transform are studied. While all of them improve reconstruction quality notably, the former turns out to be the least effective. The latter two yield comparable results at near-optimal parameterization. However, temporal total variation appears to be more forgiving w. r. t. over-regularization.**

I. Introduction

In electrocardiogram (ECG) gated CT imaging, reconstruction is impeded by missing data due to angular undersampling. Many existing approaches alleviate this problem by directly or indirectly incorporating information from the full acquisition, i. e. from all cardiac phases, when reconstructing an individual phase. For instance, a regular reconstruction from all available data, despite exhibiting motion artifacts, can serve as a prior for regularization [1], or it can be used to identify and remove the object-dependent artifact patterns [2].

In interventional cardiac C-arm CT, motion-compensated reconstruction has been employed to be able to use all projection images by correcting for non-rigid motion—estimated from intermediate reconstructions—during back-projection [3], [4]. For dynamic (4-D) imaging, i. e. reconstruction of multiple phases, it is especially helpful to make use of redundancy observed in the temporal dimension of the reconstructed images [4], [5]. In the context of iterative algebraic reconstruction techniques (ART), this can be achieved conveniently by extending regularization to the temporal domain [6]–[10].

In this work, we investigate the use of temporal regularizers for 4-D reconstruction of the left ventricle using angiographic C-arm devices. In particular, we focus on convex priors as they can readily be incorporated in most optimization methods used in tomographic reconstruction due to their convergence properties. The convergence behavior of three distinct regularizers is analyzed empirically and compared in a numerical phantom. Visual results are also presented for clinical patient data.

II. Materials and Methods

A. Gated Iterative Reconstruction

With a rectangular gating window, a subset of the available projection data is selected for each cardiac phase $t \in \{1, 2, \ldots, N_{\text{phases}}\}$ to be reconstructed. We denote the current solution in image space for each t as $\boldsymbol{I}^t$ and forego iteration indices for the sake of readability. We initialize with zeros, $\boldsymbol{I}^t = \boldsymbol{0}$, and in each iteration perform the following sequence of steps: **(i)** A simultaneous ART (SART) update step, consisting of 8 subiterations, on all $\boldsymbol{I}^t$ based on their respective data determined by ECG gating. In a subiteration, each projection image corresponds to one subset update applied simultaneously with a relaxation factor of 0.8 and subsequent enforcing of non-negativity. **(ii)** A reduction of the spatial total variation (TV) of all $\boldsymbol{I}^t$, applied with a relaxation factor of 0.5. Spatial TV minimization is carried out in analogy to the description in section II-C. **(iii)** One of three temporal regularizers described in the following sections.

A total of 30 iterations is performed. We choose this simple reconstruction scheme to focus on the effect of each temporal regularizer. For spatial regularization, TV is selected exemplarily due to its wide-spread and successful use in compressed sensing reconstruction, but it is not the primary subject of our study.

B. Temporal Rank-Reducing Regularization

The $\boldsymbol{I}^t$ are not independent. In fact, they show the same object in different states with the underlying variation governed by a small number of intrinsic components. Albeit a simplification, the assumption that few linear components may explain most of the dynamics reasonably well can serve as a global measure of consistency along the temporal domain [9]. In terms of linear algebra, this can be expressed as the rank of a matrix, i. e. the number of linearly independent rows or columns it contains. The observed rank is expected to be higher than it should ideally be due to the influence of artifacts and noise which also cause variations in the images. To mitigate this influence in our current image estimates $\boldsymbol{I} = [\boldsymbol{I}^1, \boldsymbol{I}^2, \ldots, \boldsymbol{I}^{N_{\text{phases}}}] \in \mathbb{R}^{N_{\text{voxels}} \times N_{\text{phases}}}$, arranged column-wise in a matrix, we find the solution to the proximal operator,

$$\text{prox}_{\|\|_*, \lambda_{\text{NN}}}(\boldsymbol{I}) = \underset{\boldsymbol{I}'}{\text{argmin}} \frac{1}{2} \|\boldsymbol{I} - \boldsymbol{I}'\|_{\text{F}}^2 + \lambda_{\text{NN}} \|\boldsymbol{I}'\|_*, \qquad (1)$$

where $\|\boldsymbol{I}'\|_*$ denotes the nuclear norm, i. e. the sum of the singular values of $\boldsymbol{I}'$, and $\|\cdot\|_\mathrm{F}$ the Frobenius norm used to calculate the squared Euclidean distance of the images. λ_NN controls the strength of the regularization. While using the rank function itself would result in an NP-hard reconstruction problem, the nuclear norm as its best convex approximation over the unit ball of matrices yields a proximal operator with an analytical solution that consists in a soft-thresholding of the singular values [11],

$$\boldsymbol{I} \leftarrow \boldsymbol{U}(\boldsymbol{\Sigma} - \mathbf{1} \cdot \lambda_\mathrm{NN})_+ \boldsymbol{V}^\top, \tag{2}$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the singular value decomposition (SVD) of $\boldsymbol{I}$, $(\cdot)_+$ the element-wise computation of $\max(\cdot, 0)$, and $\mathbf{1}$ the identity matrix the size of $\boldsymbol{\Sigma}$. Regarding memory efficiency, note that with $N_\mathrm{phases} \ll N_\mathrm{voxels}$, the required dimensions of the involved matrices reduce to $N_\mathrm{voxels} \times N_\mathrm{phases}$ for $\boldsymbol{U}$, and $N_\mathrm{phases} \times N_\mathrm{phases}$ for $\boldsymbol{\Sigma}$ and $\boldsymbol{V}^\top$.

C. Temporal Total Variation Regularization

In contrast, temporal total variation (tTV) [6]–[8] is a more local measure of consistency. Similar to the well-known spatial TV approach, it is based on the idealization that medical images are "cartoons" exhibiting large homogeneous regions for, e. g., certain tissue types, separated by sharp edges at organ boundaries. Considering such a model and disregarding the effects of limited resolution, in the case of motion, the gray values of affected voxels should "jump" when an edge is displaced across them, i. e. instantly change the organ they are affiliated with, and stay constant otherwise. More precisely, while traditional TV postulates sparsity in the domain of the spatial image gradient, the sparsifying transform used for tTV is the temporal gradient, estimated by forward differences,

$$\mathcal{D}_t \boldsymbol{I} = [\boldsymbol{I}^2 - \boldsymbol{I}^1, \boldsymbol{I}^3 - \boldsymbol{I}^2, \ldots, \boldsymbol{I}^1 - \boldsymbol{I}^{N_\mathrm{phases}}]. \tag{3}$$

We minimize the tTV norm of the current image estimates by iteratively descending along its negative gradient with respect to the image voxels,

$$-\frac{\partial}{\partial \boldsymbol{I}} \|\boldsymbol{I}\|_\mathrm{tTV} = -\frac{\partial}{\partial \boldsymbol{I}} \|\operatorname{vec}(\mathcal{D}_t \boldsymbol{I})\|_1, \tag{4}$$

using a backtracking line search to find a suitable step length. $\operatorname{vec}(\boldsymbol{I}) = [\boldsymbol{I}^{1\top}, \ldots, \boldsymbol{I}^{N_\mathrm{phases}\top}]^\top$ is a vectorization operator. The non-differentiability of $|x|$ at the origin is avoided by "corner rounding," i. e. its derivative is approximated as $\frac{x}{|x|+\varepsilon}$, with $\varepsilon = 10^{-4}$. After a fixed number of gradient descent iterations is performed (10 in our experiments), the resulting tTV-reduced volumes $\boldsymbol{I}_\mathrm{tTV}$ are used to update the current images, $\boldsymbol{I} \leftarrow \boldsymbol{I} + \lambda_\mathrm{tTV}(\boldsymbol{I}_\mathrm{tTV} - \boldsymbol{I})$, with the relaxation factor λ_tTV controlling the regularizer strength. The convexity of the L^1-norm guarantees that after a convex combination ($0 < \lambda_\mathrm{tTV} < 1$) with $\boldsymbol{I}_\mathrm{tTV}$, $\boldsymbol{I}$ will be smaller in terms of the tTV norm than before [12]. Note that, in analogy to the relationship between the nuclear norm and the rank operator, in a d-dimensional space, L^1 is the convex envelope of the L^0-quasi-norm on $[-1, 1]^d$ [13]. While L^0 is a better measure of sparsity, its use would again render the overall reconstruction problem NP-hard.

D. Temporal Tight-Frame Regularization

Another common choice are sparsifying transforms based on wavelet decompositions, a big advantage of which is their invertibility. Tight frames (TF) are redundant wavelet systems popular in image restoration problems as they provide very sparse representations of piecewise smooth functions [10], [14], [15]. We use the piecewise linear TF basis,

$$\boldsymbol{h}_0 = \frac{1}{4}[1, 2, 1],\ \boldsymbol{h}_1 = \frac{\sqrt{2}}{4}[1, 0, -1],\ \boldsymbol{h}_2 = \frac{1}{4}[-1, 2, -1], \tag{5}$$

where $\boldsymbol{h}_0$ is a low-pass kernel while $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ correspond to first and second order difference operators, respectively [10]. By component-wise convolution along the temporal dimension t with periodic boundary conditions, denoted by $*_t$, we compute coefficient images $\boldsymbol{C} = \{\boldsymbol{C}_i : \boldsymbol{C}_i = \boldsymbol{h}_i *_t \boldsymbol{I}\}$. If $\boldsymbol{W}$ is the operator performing this decomposition, $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{I}$, its adjoint is $\boldsymbol{W}^T = \boldsymbol{W}^{-1}$, where $\boldsymbol{W}^\top \boldsymbol{C} = \sum_{i=0}^{2} \boldsymbol{h}_i \star_t \boldsymbol{C}_i = \boldsymbol{I}$ and $\star_t$ denotes correlation in analogy to $*_t$. Sparsity is enhanced by reducing the summed up high-pass component energies [10], $\|\boldsymbol{C}\|_\mathrm{TF} = \|\operatorname{vec}(\sum_{i=1}^{2}(\boldsymbol{C}_i)^2)^{\frac{1}{2}}\|_1$, where $(\cdot)^2$ and $(\cdot)^{\frac{1}{2}}$ are to be understood element-wise. To this end, we find,

$$\operatorname{prox}_{\|\|_\mathrm{TF}, \lambda_\mathrm{TF}}(\boldsymbol{C}) = \operatorname*{argmin}_{\boldsymbol{C}'} \sum_{i=0}^{2} \frac{1}{2}\|\boldsymbol{C}'_i - \boldsymbol{C}_i\|_\mathrm{F}^2 + \lambda_\mathrm{TF}\|\boldsymbol{C}'\|_\mathrm{TF}, \tag{6}$$

which is used in the update step, $\boldsymbol{I} \leftarrow \boldsymbol{W}^\top \operatorname{prox}_{\|\|_\mathrm{TF}, \lambda_\mathrm{TF}}(\boldsymbol{W}\boldsymbol{I})$, and realized by element-wise vector shrinkage [14],

$$\boldsymbol{C}^t_{i,j} \leftarrow \begin{cases} \boldsymbol{C}^t_{i,j} & \text{if } i = 0, \\ \boldsymbol{C}^t_{i,j} \cdot \left(1 - \lambda_\mathrm{TF} \cdot (\sum_{i=1}^{2}(\boldsymbol{C}_{i,j})^2)^{-\frac{1}{2}}\right)_+ & \text{else,} \end{cases} \tag{7}$$

with $\boldsymbol{C}^t_{i,j}$ the j-th element of coefficient image $\boldsymbol{C}^t_i$ and λ_TF the regularization parameter.

III. Experiments

A. Data

For evaluation, a numerical phantom data set was used [16]. Projection images were generated [17] using a polychromatic spectrum, discretized in energy bins 5 keV wide from 10 keV to 90 keV (peak energy), and a time-current product of 2.5 mAs per X-ray pulse. For bones and bone marrow, material properties match the mass attenuation coefficients found in the NIST X-ray table[1]. Contrasted blood in the left ventricle and aorta was simulated as iodine-based contrast agent (Ultravist-150, Bayer AG, Leverkusen, Germany) mixed with equal parts water. All other structures were modeled with the absorption behavior of water for modified densities. The acquisition protocol is identical to that of the clinical data sets described below. Additionally, a set of projection images with complete angular sampling for the end-diastolic phase (static phantom) was generated to reconstruct a ground truth image using SART with spatial TV regularization (relaxation factor 1.0) as described above. A heart rate of 120 bpm was simulated, resulting in 10 observed heartbeats.

[1] http://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html

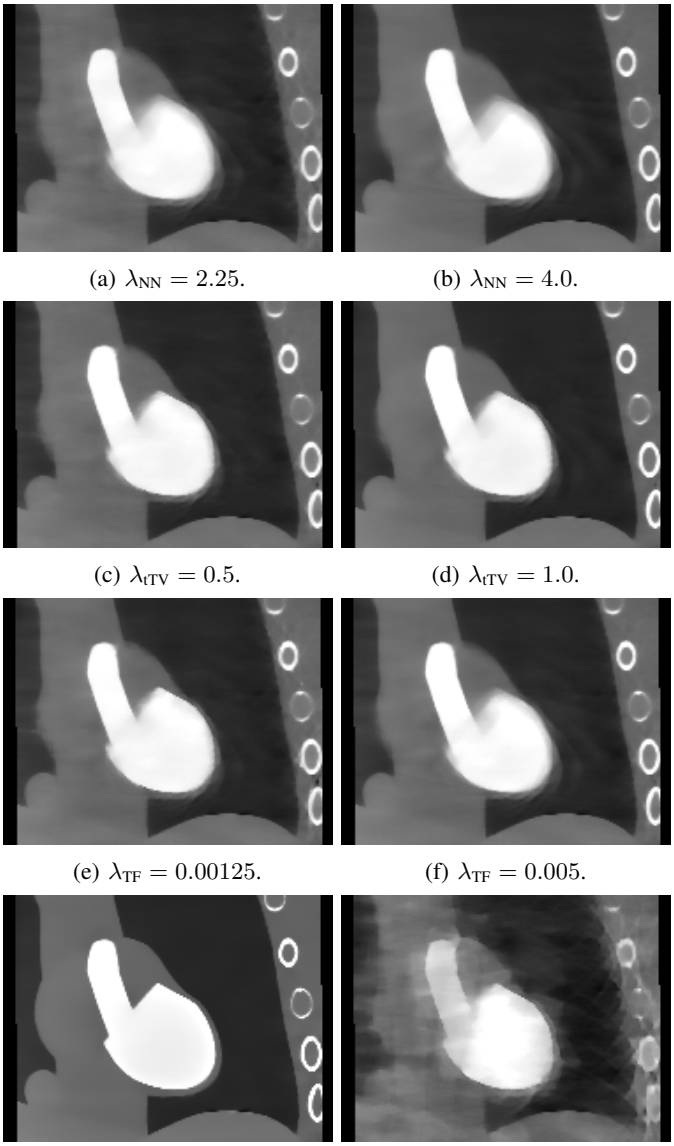
A clinical patient data set was acquired with an Artis zee biplane (Siemens Healthcare GmbH, Forchheim, Germany). 133 projection images were captured at approx. 30 Hz with an angular increment of 1.5° during one 5 s long rotation of the C-arm. The isotropic pixel resolution was 0.31 mm/pixel (0.21 in isocenter), the detector size 960×960 pixels. The heart was paced through external stimulation to 140 bpm, resulting in 13 heartbeats. The gating windows cover 12.5 % of the heart cycle each and use all data without overlap, such that $N_{\text{phases}} = 8$. Contrast agent was administered in the left ventricle (42 ml, undiluted) and right atrium (50 ml, 60 % dilution). All images were reconstructed on a grid of 256^3 voxels covering a volume of $(25.6\,\text{cm})^3$.

B. Evaluation

For the phantom study, the root mean square error (RMSE) to the ground truth is calculated over a region of interest (ROI) tightly enclosing the heart to assess convergence. In order to study the sensitivity of the results w. r. t. regularizer strength, a 1-D grid search is conducted for each $\lambda_\diamond$, $\diamond \in \{\text{NN, tTV, TF}\}$, such that we obtain sets of parameter values covering an "interesting" range, i. e. near-optimal as well as under- and over-regularized settings. Visual results are shown for both the phantom and the clinical data set.

IV. Results

Figs. 1 and 3 show reconstructed image slices of the phantom and the clinical data set, respectively, while the plots in Fig. 2 summarize the quantitative results of the phantom study. While both TF and tTV achieve similar RMSE values at the optimum, the "valley" of the TF curve is much more pronounced (Fig. 2), indicating that it is more sensitive to the choice of $\lambda_\diamond$. This is reflected by the strongly regularized result being blurred in the case of TF (Fig. 1f), but not tTV (Fig. 1d). At near-optimal parameterization, both yield visually comparable images (Figs. 1e and 1c). This also holds for the clinical case when using the same $\lambda_\diamond$ (Figs. 3c and 3b). Rank-based regularization does not achieve quite as low RMSE values and introduces a blur even for the optimal parameter value (Figs. 1a and 3a). All tested settings of temporal regularization outperform spatial TV alone ($\lambda_\diamond = 0$ in Fig. 2, Figs. 1h and 3d). Universally, convergence is faster in the first iterations for a higher $\lambda_\diamond$, but then levels out earlier.



(a) $\lambda_{\text{NN}} = 2.25$. (b) $\lambda_{\text{NN}} = 4.0$.

(c) $\lambda_{\text{tTV}} = 0.5$. (d) $\lambda_{\text{tTV}} = 1.0$.

(e) $\lambda_{\text{TF}} = 0.00125$. (f) $\lambda_{\text{TF}} = 0.005$.

(g) Ground truth reconstruction. (h) No temporal regularization.

Fig. 1: Reconstructed images of the phantom data set in end-diastole. The best results in terms of RMSE achieved for each regularizer are shown on the left (a, c, e), strongly regularized results on the right (b, d, f). Ground truth reconstructed from static data (g) and an image obtained without temporal regularization (h) are shown for comparison. All images are displayed with the same window at $[\text{center}, \text{width}] = [300, 2600]$ HU.

V. Discussion and Conclusion

The difficulty of the reconstruction problem is illustrated well by the poor performance of the variant without temporal regularization; with about 10-13 distinct views available per phase, it features a degree of undersampling that is extremely challenging even in the context of compressed sensing.

Although tTV lacks a simple closed-form solution of its proximal operator, requiring optimization based on, e. g., iterative descent or primal-dual splitting algorithms [13], its comparatively forgiving nature w. r. t. parameterization renders it highly useful. When this property is not critical, TF regularization is a viable alternative due to its straight-forward implementation. It should be noted that TV as a spatial regularizer could put tTV at an advantage due to synergy effects, so it may be worthwhile to test combinations with other spatial regularizers as well. While the results demonstrate that rank-reduction is inferior to the other methods on its own, it could be investigated whether its global approach, which is capable of suppressing local outliers, might complement other regularizers well at early stages of the optimization when used in combination.

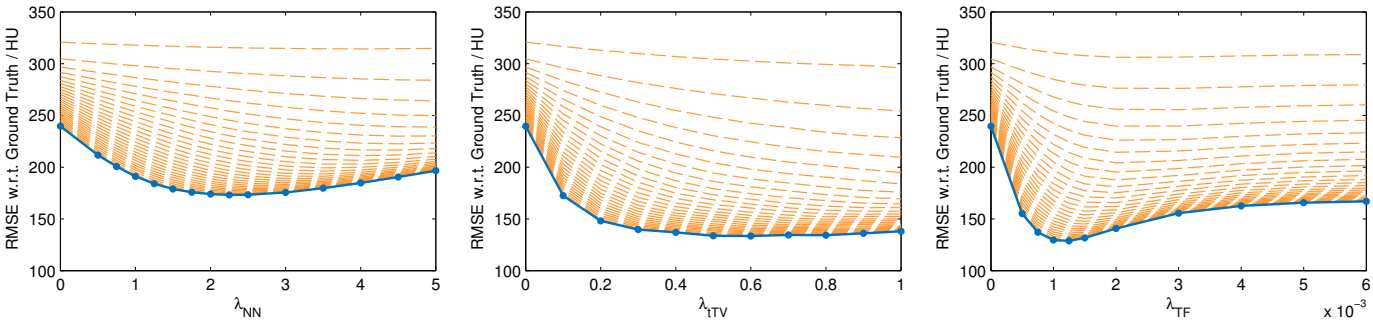
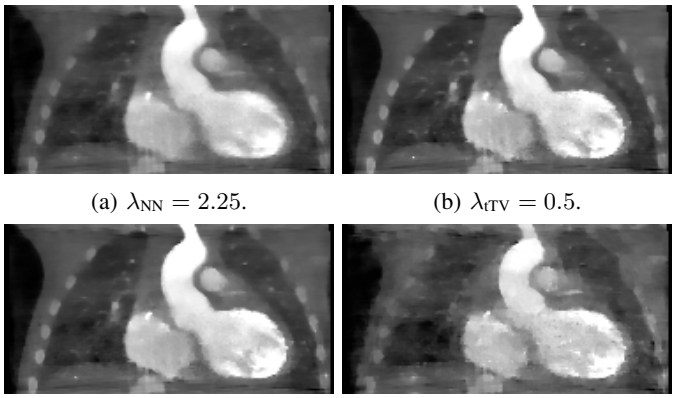


Fig. 2: RMSE for reconstructed images of the phantom data set over a range of parameter values around the optimum for each regularizer. While the solid line is the result after 30 iterations, the dotted lines correspond to previous iterations, visualizing the varying speeds of convergence.



(a) $\lambda_{\mathrm{NN}} = 2.25$. (b) $\lambda_{\mathrm{tTV}} = 0.5$.

(c) $\lambda_{\mathrm{TF}} = 0.00125$. (d) No temporal regularization.

Fig. 3: Reconstructed images of the clinical data set in end-diastole. The best parameterization determined in the phantom study was used for each regularizer (a, b, c). An image obtained without temporal regularization (d) is shown for comparison. All images are displayed with the same window at $[\text{center}, \text{width}] = [630, 3260]$ HU.

Similar results obtained in both the phantom and a clinical case using identical parameterization are reassuring w. r. t. generalizability and applicability, although further experiments are advised for validation. Nonetheless, we can conclude that exploiting redundancy in the temporal domain through regularization is a highly valuable tool for improving image quality of 4-D cardiac C-arm CT.

Disclaimer

The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[2] G. C. Mc Kinnon and R. Bates, "Towards imaging the beating heart usefully with a conventional CT scanner," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 2, pp. 123–127, 1981.

[3] K. Müller, A. Maier, C. Schwemmer, G. Lauritsch, S. D. Buck, J.-Y. Wielandts, J. Hornegger, and R. Fahrig, "Image artefact propagation in motion estimation and reconstruction in interventional cardiac C-arm CT," *Phys. Med. Biol.*, vol. 59, no. 12, pp. 3121–3138, 2014.

[4] O. Taubmann, A. Maier, J. Hornegger, G. Lauritsch, and R. Fahrig, "Coping with Real World Data: Artifact Reduction and Denoising for Motion-Compensated Cardiac C-arm CT," *Med. Phys.*, vol. 43, no. 2, pp. 883–893, 2016.

[5] O. Taubmann, G. Lauritsch, A. Maier, R. Fahrig, and J. Hornegger, "Estimate, Compensate, Iterate: Joint Motion Estimation and Compensation in 4-D Cardiac C-arm Computed Tomography," in *MICCAI*, 2015, pp. 579–586.

[6] H. Wu, A. Maier, R. Fahrig, and J. Hornegger, "Spatial-temporal Total Variation Regularization (STTVR) for 4D-CT Reconstruction," in *SPIE Med. Imag.*, 2012, p. 83133J.

[7] C. Mory, V. Auvray, B. Zhang, M. Grass, D. Schäfer, S. Chen, J. Carroll, S. Rit, F. Peyrin, P. Douek, and L. Boussel, "Cardiac C-arm computed tomography using a 3D + time ROI reconstruction method with spatial and temporal regularization," *Med. Phys.*, vol. 41, p. 021903, 2014.

[8] V. Haase, O. Taubmann, Y. Huang, G. Krings, G. Lauritsch, A. Maier, and A. Mertins, "Make the Most of Time: Temporal Extension of the iTV Algorithm for 4-D Cardiac C-Arm CT," in *Bildverarbeitung für die Medizin*, 2016, pp. 170–175.

[9] J.-F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao, "Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1581–1591, 2014.

[10] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restorations: total variation, wavelet frames and beyond," *J. Am. Math. Soc.*, vol. 25, pp. 1033–1089, 2012.

[11] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pac. J. Optim.*, vol. 6, no. 3, pp. 615–640, 2010.

[12] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, p. 1545, 2011.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[14] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Phys. Med. Biol.*, vol. 56, no. 13, p. 3787, 2011.

[15] M. Manhart, A. Fieselmann, and Y. Deuerling-Zheng, "Evaluation of a Tight Frame Reconstruction Algorithm for Perfusion C-arm CT Using a Realistic Dynamic Brain Phantom," in *CT Meeting*, 2012, pp. 123–126.

[16] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, pp. 4902–4915, 2010.

[17] A. Maier, H. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast Simulation of X-ray Projections of Spline-based Surfaces using an Append Buffer," *Phys. Med. Biol.*, vol. 57, no. 19, pp. 6193–6210, 2012.