TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray *in vivo* data

Xiaoli Yang,[1,*] **Ralf Hofmann,**[2,4] **Robin Dapp,**[1] **Thomas van de Kamp,**[2] **Tomy dos Santos Rolo,**[2] **Xianghui Xiao,**[3] **Julian Moosmann,**[2] **Jubin Kashef,**[2] **and Rainer Stotzka**[1]

[1]*Karlsruhe Institute of Technology, Institute for Data Processing and Electronics (IPE), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany*
[2]*Karlsruhe Institute of Technology, Institute for Photon Science and Synchrotron Radiation (IPS), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany*
[3]*Advanced Photon Source (APS), Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA*
[4]*ralf.hofmann2@kit.edu*
[*]*xiaoli.yang@partner.kit.edu*

Abstract: High-resolution, three-dimensional (3D) imaging of soft tissues requires the solution of two inverse problems: phase retrieval and the reconstruction of the 3D image from a tomographic stack of two-dimensional (2D) projections. The number of projections per stack should be small to accommodate fast tomography of rapid processes and to constrain X-ray radiation dose to optimal levels to either increase the duration of *in vivo* time-lapse series at a given goal for spatial resolution and/or the conservation of structure under X-ray irradiation. In pursuing the 3D reconstruction problem in the sense of compressive sampling theory, we propose to reduce the number of projections by applying an advanced algebraic technique subject to the minimisation of the total variation (TV) in the reconstructed slice. This problem is formulated in a Lagrangian multiplier fashion with the parameter value determined by appealing to a discrete L-curve in conjunction with a conjugate gradient method. The usefulness of this reconstruction modality is demonstrated for simulated and *in vivo* data, the latter acquired in parallel-beam imaging experiments using synchrotron radiation.

References and links

1. T. van de Kamp, P. Vagovič, T. Baumbach, and A. Riedel, "A biological screw in a beetle's leg," Science **333**, 52–52 (2011).
2. T. dos Santos Rolo, A. Ershov, T. van de Kamp, and T. Baumbach, "In vivo X-ray cine-tomography for tracking morphological dynamics," Proc. Natl. Acad. Sci. USA **111**, 3921–3926 (2014).
3. J. Moosmann, A. Ershov, V. Altapova, T. Baumbach, M. S. Prasad, C. LaBonne, X. Xiao, J. Kashef, and R. Hofmann, "X-ray phase-contrast in vivo microtomography probes new aspects of Xenopus gastrulation," Nature **497**, 374–377 (2013).
4. J. Moosmann, A. Ershov, V. Weinhardt, T. Baumbach, M. S. Prasad, C. LaBonne, X. Xiao, J. Kashef, and R. Hofmann, "Time-lapse X-ray phase-contrast microtomography for in vivo imaging and analysis of morphogenesis," Nat. Protoc. **9**, 294–304 (2014).

5. R. Hofmann, A. Schober, J. Moosmann, M. Hertel, S. Hahn, V. Weinhardt, D. Hänschke, L. Helfen, X. Xiao, and T. Baumbach, "Single-distance livecell imaging with propagation-based X-ray phase contrast," to be submitted .

6. A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging* (Society for Industrial and Applied Mathematics, 2001).

7. X. Yang, T. Jejkal, H. Pasic, R. Stotzka, A. Streit, J. van Wezel, and T. dos Santos Rolo, "Data intensive computing of X-ray computed tomography reconstruction at the LSDF," in "21st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)," (IEEE, 2013), pp. 86–93.

8. G. N. Hounsfield, "A method and apparatus for examination of a body by radiation such as X-ray or gamma radiation," (1972). Patent Specification 1283915.

9. R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," J. Theor. Biol. **29**, 471–481 (1970).

10. S. Kaczmarz, "Angenäherte Auflösung von Systemen Linearer Gleichungen," Bulletin International de l'Academie Polonaise des Sciences et des Lettres **35**, 355–357 (1937).

11. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE T. Inform. Theory **52**, 489–509 (2006).

12. T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," Mathematical Models of Computer Vision **17** (2005).

13. E. Y. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," J. X-ray Sci. Technol. **14**, 119–139 (2006).

14. A. Kostenko, K. J. Batenburg, H. Suhonen, S. E. Offerman, and L. J. van Vliet, "Phase retrieval in in-line x-ray phase contrast imaging based on total variation minimization," Opt. Express **21**, 710–723 (2013).

15. R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," Med. Phys. **12**, 252–255 (1985).

16. Y. Hu and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery," IEEE T. Image Process. **21**, 2559–2571 (2012).

17. E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).

18. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: two-step iterative shrinkage / thresholding algorithms for image restoration," IEEE T. Image Process. **16**, 2992–3004 (2007).

19. S. Becker, J. Bobin, and E. Candès, "NESTA: a fast and accurate first-order method for sparse recovery," SIAM J. Imaging Sci. **4**, 1–39 (2011).

20. T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen, "Implementation of an optimal first-order method for strongly convex total variation regularization," BIT **52**, 329–356 (2012).

21. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**, 183–202 (2009).

22. Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," SIAM J. Imaging Sci. **1**, 248–272 (2008).

23. J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data," IEEE J. Sel. Top. Signa. **4**, 288–297 (2010).

24. C. Li, W. Yin, and Y. Zhang, "User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms," CAAM Report (2009).

25. C. Li, *Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms* (PhD Thesis in Rice University, 2011).

26. M. Fukushima, "Application of the alternating direction method of multipliers to separable convex programming problems," Comput. Optim. Appl. **1**, 93–111 (1992).

27. T. Goldstein, B. O'Donoghue, and S. Setzer, "Fast alternating direction optimization methods," CAM report pp. 12–35 (2012).

28. T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," SIAM J. Imaging Sci. **2**, 323–343 (2009).

29. E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," CAM Report **9**, 31 (2009).

30. X. Tai and C. Wu, "Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model," in "Scale Space and Variational Methods in Computer Vision," (Springer, 2009), pp. 502–513.

31. P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," SIAM J. Sci. Comput. **14**, 1487–1503 (1993).

32. C. R. Vogel, "Non-convergence of the L-curve regularization parameter selection method," Inverse Probl. **12**, 535 (1996).

33. V. A. Morozov, "On the solution of functional equations by the method of regularization," in "Soviet Math. Dokl.," , vol. 7 (1966), vol. 7, pp. 414–417.

34. G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics **21**, 215–223 (1979).

35. L. Reichel and G. Rodriguez, "Old and new parameter choice rules for discrete ill-posed problems," Numer.

Algorithms **63**, 65–87 (2013).

36. J. Song, Q. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," Med. Phys. **34**, 4476–4483 (2007).
37. P. T. Lauzier, J. Tang, and G. Chen, "Prior image constrained compressed sensing: Implementation and performance evaluation," Med. Phys. **39**, 66–80 (2012).
38. W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," Pac. J. Optim. **2**, 35–58 (2006).
39. M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," Magn. Reson. Med. **58**, 1182–1195 (2007).
40. Y. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," SIAM J. Optimiz. **10**, 177–182 (1999).
41. D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, "Tikhonov regularization and the L-curve for large discrete ill-posed problems," J. Comput. Appl. Math. **123**, 423–446 (2000).
42. The energy dependence of the complex refractive index can be obtained for a variety of chemical compounds under http://henke.lbl.gov/optical_constants/getdb2.html .
43. Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," IEEE Signal Proc. Mag. **26**, 98–117 (2009).
44. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE T. Image Process. **13**, 600–612 (2004).
45. J. Moosmann, R. Hofmann, and T. Baumbach, "Single-distance phase retrieval at large phase shifts," Opt. Express **19**, 12066–12073 (2011).
46. R. Hofmann, J. Moosmann, and T. Baumbach, "Criticality in single-distance phase retrieval," Opt. Express **19**, 25881–25890 (2011).

1. Introduction

X-ray phase-contrast computed micro-tomography (XPCμT) is a weakly invasive three-dimensional (3D) imaging technique. Thanks to the high degree of spatio-temporal coherence and high flux density in state-of-the-art synchrotron radiation and due to impressive advances in detector technology, sub-micron spatial resolutions and exposure times well below one millisecond per projection are feasible. This enables four-dimensional (4D) imaging applications important for the elucidation of rapid dynamics. In particular, *in vivo* elucidations of internal processes on short time scales benefit from highest possible volume-acquisitions rates. Exemplarily, this was pointed out by investigation of a moving screw-and-nut-type hip joint [1] in the insect *Sitophilus granarius* (grain weevil) [2]. Moreover, dose reduction is an issue of key importance in all X-ray imaging applications. Namely, the usefulness of *in vivo* imaging strongly hinges on a good control of dose effects, thus enabling long time-lapse series of 3D images with weak perturbations induced by the imaging process itself. Recently, embryonic development in the vertebrate model organism *Xenopus laevis* (African clawed frog) was imaged *in vivo* via 3D time-lapse series [3, 4]. In cases like these a low number of tomographic projections is essential. The challenge for 3D reconstruction algorithms thus is to generate relevant information from limited input data. Note that propagation based phase contrast imaging dose reduction can also be achieved by exploiting spatial coherence. Namely, mean intensity contrast, measured at a single propagation z, rises with $\sqrt{z}$ whereas the signal to noise ratio (SNR) in the phase map increases with z [5]. Obviously, this implies that the same SNR can be obtained with a lower dose if z is increased accordingly. The degree of coherence in hard X-ray radiation provided by most 3rd-generation synchrotrons limits z to be about 1 m.

Reconstruction is an inverse problem restoring the 3D distribution of the attenuation index throughout the object from the tomographic projections distributed in an angular range from 0° to 180° (see Fig. 1 for a schematic experimental setup). In this article we pursue a reconstruction algorithm that obtains a high accuracy of relevant features, using a low number of projections. A particular reconstruction technique, called filtered back-projection (FBP) [6], appeals to a combination of the Fourier-slice with the Nyquist-Shannon sampling theorem. It posits that the minimal number of equi-angularly spaced, tomographic projections, required
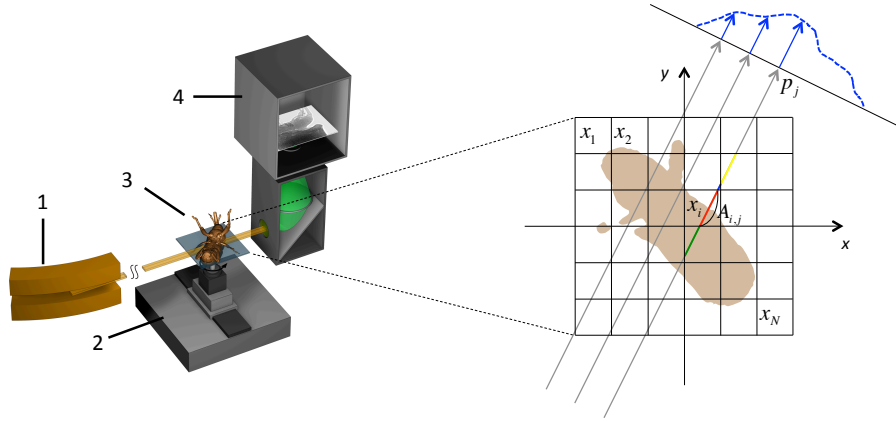
Fig. 1. *In vivo* X-ray tomography experiment with propagation-based contrast. Left panel: experimental setup showing (1) the X-ray source (here: bending magnet), (2) tomographic stage, (3) sample, and (4) detector system. Right panel: parallel-beam projection of attenuation index through 2D slice of object (forward model), for explanations of labels see text.

for reconstruction of a noiseless 3D image at a spatial resolution $2\Delta x$, is given as

$$P_{\text{FBP}} = \frac{\pi}{2}K\,. \tag{1}$$

Here $K \equiv L/\Delta x$, L representing the linear dimension of the (quadratic) field of view. FBP is a rapid reconstruction technique which does not require any parameter tuning. Sparsely sampled projection data result in a degraded quality of FBP reconstruction [7]. An alternative approach, the algebraic reconstruction technique (ART) [6, 8, 9], is not widely used due to long computing times and expert knowledge required in searching optimal solutions in the highly degenerate landscape of possible answers to the inverse problem. That is, there are many iterations for solving a highly dimensional, yet underdetermined system of linear equations [10]. A choice of appropriate solution demands some sort of a priori knowledge and/or the application of deep physical principles, usually in analogy to those of statistical thermodynamics. The process of constructing a unique solution to ART subject to such additional conditions is called regularisation.

Compressive sampling (CS) theory gives definite conditions on the number of required measurements for a sparsely represented object [11] under which it can be recovered within a prescribed error. In image reconstruction ART is able to accommodate such prior knowledge as the existence of a sparse representation, and thus one may hope that a limited set of noisy data suffices in reconstructing such an object with a good accuracy. To construct an optimal representation of the object, a sequence of base changes, transforming the original (spatial domain) into an increasingly sparse set of representations, is required. For example, pursuing this idea, a base change called discrete gradient transform (DGT) may lead to a sparse expansion of certain 2D or 3D images. The totality of associated expansion coefficients expresses the so-called total variation (TV) of the image [12]. Inspired by CS theory, ART with TV regularisation has been studied for both few-view and limited angle problems [13]. Also, phase retrieval in propagation based X-ray imaging was considered as a minimisation problem subject to TV regularisation in [14], based on a linear model for forward propagation.

Let us be more definite. ART searches a solution to a linear system of equations $\mathbf{Ax} = \mathbf{p}$, representing the forward imaging process. Here $\mathbf{A}$ is the system matrix which can be calculated,

e.g., by Siddon's algorithm [15] (weighting by length of line segments associated with the intersection of an X-ray with a given pixel as shown in the right panel of Fig. 1), $\mathbf{p}$ denotes a vector embodying the set of projections acquired during the tomographic imaging process, and $\mathbf{x}$ indicates the distribution of the attenuation index. Only parallel-beam geometry is considered in this paper, but the conclusions obtained are sufficiently general to be applicable to the fan or cone beam situation as well. ART now performs the inverse operation of recovering $\mathbf{x}$ in terms of $\mathbf{A}$ and $\mathbf{p}$. As mentioned above, this problem is ill-posed in the case of low number of projections which lies essentially below the bound of Eq. (1). ART augmented with TV regularization yields an optimal solution to $\mathbf{A}\mathbf{x} = \mathbf{p}$ in the sense of $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$ and $\|\mathbf{x}\|_{\mathrm{TV}}$ being minimal where the l_2 norm of vector $\mathbf{b}$ reads $\|\mathbf{b}\|_2 \equiv (\sum_i b_i^2)^{1/2}$ and $\|\mathbf{x}\|_{\mathrm{TV}}$ is the TV part to be introduced in detail in Sec. 2.1. One can impose the inverse reconstruction problem as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}}, \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{p}, \tag{2}$$

which aims to minimise the total variation function while enforcing the optimal solution $\mathbf{x}^*$ to be consistent with measurements $\mathbf{p}$. However, this problem may have no solution at all considering errors in experimental data. In this case a tolerance is required to obtain an approximate solution. In this spirit, one may formulate a variant of the problem by introducing a Lagrangian multiplier literally [16] as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 + \lambda \|\mathbf{x}\|_{\mathrm{TV}}. \tag{3}$$

λ is a parameter controlling the trade-off between data fidelity and sparse regularisation. The TV minimisation was introduced into clinical applications and developed into the fan-beam algorithm TV-POCS for CT reconstruction [13], which was latter improved towards better convergence and robustness against cone-beam artifacts [17]. Note that TV minimisation also is applied to image denoising and restoration. In the general context, many solvers have already been developed for the TV-based, large-scale minimisation problems (2) or (3) with fast convergence, such as TwIST [18], NESTA [19], UPN [20], and FISTA [21]. Moreover, another set of algorithms employs the splitting idea [22] in developing an alternating minimisation approach for image recovery with TV regularisation, such as RecRF [23], TVAL3 [24, 25], and ADMM [26, 27] utilizing techniques like the split Bregman algorithm [28], the augmented Lagrangian method or the alternating direction method and gaining also fast convergence. The interrelations of these techniques have been pointed out in [29] and [30].

In the present work, we pursue a simple implementation of the conjugate gradient method to solve the unconstrained minimisation problem (3) for a yet undetermined parameter λ. In solving problem (3) in the exact sense of problem (2) the minimum $\mathbf{x}^*(\lambda)$ would have to be inserted into the TV minimisation condition to yield the optimal value λ^* by algebraic inversion. As a consequence, $\mathbf{x}^*(\lambda^*)$ solves problem (2). To proceed like this, however, is impractical and could even be impossible since (i) the determination of λ^* requires an iterative approach, e.g., Newton's method, (ii) λ^* is usually not unique implying the need for manual selection, and (iii) a solution in this sense may not exist at all for inconsistent data. Even if it existed, the search for an exact solution to problem (2) may be too restrictive in the sense that reconstructions may ensue that are unrealistically flat. Therefore, we search a *reasonable* constraint in fixing the value of λ in (3) to a finite value in an automated way. This can be achieved by the celebrated discrete L-curve method [31] which we will employ in the present paper. For a comprehensive analysis of this method in the context of inverse problems, also pointing out limitations in view of its convergence properties, see [32].

When constrained minimisation problems are mapped into unconstrained ones more freedom can be introduced in terms of the so-called augmented Lagrangian method such that the number

of parameters exceeds the number of constraints. To determine these parameters by reasonable, additional conditions poses a multidimensional optimisation problem which is computationally expensive. Still, it is interesting to discuss a particular example called TVAL3 proposed in [25]. There total variation is generalised in terms of a variety of constraints (various generalised gradient transforms), formulated in terms of new variables ω_i in addition to x, to give rise to a new minimisation problem in ω_i and x. The formal substitution of these additional constraints into the minimisation problem gives back the higher dimensional equivalent to the old problem of Eq. (3). The important insight is to treat the additional constraints also in the sense of assigning Lagrangian multipliers $\nu_{i,1}$, the latter, however, never to be literally determined by resolutions of the associated constraint equations. In addition, higher powers of all constraints are included into the new cost function, subject to additional multipliers μ_1, μ_2, $\cdots$, $\nu_{i,1}$, $\nu_{i,2}$, $\cdots$ to allow for various 'speeds' of the approach to the solution of the constraint equations. The equivalent of the automated parameter fix in terms of the L-curve approach of Sec. 2.2 now is to find, say, the smallest distance of a surface to the origin, parametrised by the above set of multipliers, in a high-dimensional space spanned by all residuals. Practically, this is prohibitive in view of computational effort even though the 'augmented' Lagrangian method appears to exhibit superior convergence properties for minimising the cost function. Pragmatically, one may just set μ_1, μ_2, $\cdots$, $\nu_{i,1}$, $\nu_{i,2}$, $\cdots$ to values which visually lead to somewhat expected results (human judgment) which, however, prejudices the reconstruction and thus should be avoided in practical applications.

More complex parameter selection methods were proposed in the literature such as the discrepancy principle (DP) [33] and the generalized cross validation (GCV) [34]. For an elaborate comparison of the L-curve method with other ways of fixing model parameters in a priori under-determined inverse problems, see [35]. To the best of our knowledge the conjugate gradient minimisation of the cost function, associated with (3), together with the subsequent, automated determination of parameter λ using the discrete L-curve has not been used before for tomographic reconstruction.

This paper is organised as follows. In Sec. 2 we discuss peculiarities of the TV-based conjugate gradient method including possible definitions of TV. The criterion on how to determine the parameter value for λ is elucidated conceptually and exemplarily (Shepp-Logan phantom) in fixing an optimal point on the discrete L-curve. Also, we outline a strategy for automised reconstruction of entire 3D volumes given parallel-beam illumination. Sec. 3 first defines and discusses suitable error measures to judge reconstruction quality compared to a reference and subsequently applies our modality to successively increasing data complexity: noisefree simulated data (Barbara) and low-noise *in vivo* data obtained from X-ray propagation based microtomography, using intensity 'projections' in case of a weevil and phase projections (generated from intensity maps by high-resolution quasiparticle phase retrieval) representing a stage-17 frog embryo during neurulation. To point out certain advantages of the here-proposed modality we compare few- and many-view reconstruction results with those obtained by standard filtered back-projection. The concluding section of the paper summarises our results, speculates about an advantageous application to (low-dose) data with a high level of statistical noise, and discusses a future venue for an essential reduction of the computational effort.

2. Methods

In this section we apply the TV-based conjugate gradient method (CGTV) to solve problem (3). In terms of CGTV, for a specific sparse-object application see [36], for another implementation and comparison of performance with alternative minimisation procedures see [37]. For the parameter determination, the discrete L-curve method is used. Finally, we discuss implementations of this workflow for automated 3D reconstruction.

2.1. CGTV solver

For later use, it is convenient to represent the object in terms of a 2D matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ (n pixels in each of the two dimensions) where each entry is associated with the local value of the attenuation index. Thus, vector $\mathbf{x}$ in (3) can be composed out of the elements $\mathbf{X}_{i,j}$ ($i, j = 1, \cdots, n$) by column $\mathbf{X}_{i,1}$ being stacked on top of column $\mathbf{X}_{i,2}$ and so forth. Here the dimensionality of vector $\mathbf{x}$ is given as $N = n^2$ (total pixel number in the to-be-reconstructed slice).

The system matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ in (3) possesses M rows and $\mathbf{p} \in \mathbb{R}^{M \times 1}$. For exact reconstruction one would have to impose $M = N$ but usually one has $M < N$ such that a regularisation for a reasonable solution to the reconstruction problem in the sense of (3) is required. The function to be minimised in (3) consists of two terms: the data fidelity term, $F(x) = \|\mathbf{Ax} - \mathbf{p}\|_2^2$, and the total variation term, $T(\mathbf{x}) = \|\mathbf{x}\|_{\mathrm{TV}}$. To define the latter, we consider the l1-norm or the l2-norm of the gradient field, obtained by the so-called discrete gradient transform (DGT). For a given pixel (i, j) one has

$$\mathrm{DGT}_{\mathrm{l1},ij}(\mathbf{X}) \equiv |\triangle_{i,j}^h \mathbf{X}| + |\triangle_{i,j}^v \mathbf{X}| \tag{4}$$

or

$$\mathrm{DGT}_{\mathrm{l2},ij}(\mathbf{X}) \equiv [(\triangle_{i,j}^h \mathbf{X})^2 + (\triangle_{i,j}^v \mathbf{X})^2]^{1/2}, \tag{5}$$

where the difference operations $\triangle_{i,j}^h \mathbf{X}$ (h for horizontal) and $\triangle_{i,j}^v \mathbf{X}$ (v for vertical) on matrix $\mathbf{X}$ are defined as

$$\triangle_{i,j}^h \mathbf{X} \equiv \mathbf{X}_{i,j} - \mathbf{X}_{i-1,j} \tag{6}$$

and

$$\triangle_{i,j}^v \mathbf{X} \equiv \mathbf{X}_{i,j} - \mathbf{X}_{i,j-1}, \tag{7}$$

respectively. Thus, total variations, defined in terms of the l1-norm or l2-norm of the gradient field, see Eqs. (8) and (9), read

$$\|\mathbf{x}\|_{\mathrm{l1,TV}} \equiv \sum_{i,j} \mathrm{DGT}_{\mathrm{l1},ij}(\mathbf{X}) \tag{8}$$

or

$$\|\mathbf{x}\|_{\mathrm{l2,TV}} \equiv \sum_{i,j} \mathrm{DGT}_{\mathrm{l2},ij}(\mathbf{X}). \tag{9}$$

In practice, no essential differences arise when appealing to either of the definitions of total variation, Eq. (8) or Eq. (9). However, definition of (9) does not depend on the choice of Cartesian coordinates, and therefore we exclusively will use it throughout the remainder of this work. Thus, notationally, we let $\|\mathbf{x}\|_{\mathrm{l2,TV}} \to \|\mathbf{x}\|_{\mathrm{TV}}$.

Let us now turn to technical points in addressing the minimisation problem (3). To do this, we employ the nonlinear conjugate gradient method [38]. There are accelerated minimisation schemes such as TwIST [18], UPN [20], FISTA [21] and ADMM [27]. Their common feature is that at least one parameter controlling the trade-off between data fidelity and regularization is determined empirically. It is true that, compared to these fast algorithms, the conjugate gradient method possesses higher iteration complexity. However, as such it exhibits fairly general, stable convergence properties. This is why we simply choose to use the conjugate gradient method here.

Minimisation of the cost function $f(\mathbf{x})$, $f \equiv F(\mathbf{x}) + \lambda T(\mathbf{x})$ (see text before Eq. (4)), then requires the computation of the gradients of F and T in an n^2-dimensional vector space. The gradient of $F(\mathbf{x})$ is given as

$$\nabla F(\mathbf{x}) = \nabla(\|\mathbf{Ax} - \mathbf{p}\|_2^2) = 2\mathbf{A}^{\mathrm{T}}(\mathbf{Ax} - \mathbf{p}). \tag{10}$$

However, T is not differentiable when $\mathbf{X}_{i,j} = \mathbf{X}_{i-1,j}$ and $\mathbf{X}_{i,j} = \mathbf{X}_{i,j-1}$. This is overcome by introducing a mild modification of the modulus definition as [39]

$$|z| \approx (z^2 + \varepsilon)^{1/2}, \tag{11}$$

where ε is much smaller than typical values of z are. Therefore, we obtain for the partial derivative of $\|\mathbf{x}\|_{\text{TV}}$

$$
\begin{aligned}
\frac{\partial \|\mathbf{x}\|_{\text{TV}}}{\partial \mathbf{X}_{i,j}} = &\frac{2\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j} - \mathbf{X}_{i,j-1}}{[(\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j})^2 + (\mathbf{X}_{i,j} - \mathbf{X}_{i,j-1})^2 + \varepsilon]^{1/2}} \\
&+ \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}}{[(\mathbf{X}_{i+1,j} - \mathbf{X}_{i,j})^2 + (\mathbf{X}_{i+1,j} - \mathbf{X}_{i+1,j-1})^2 + \varepsilon]^{1/2}} \\
&+ \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}}{[(\mathbf{X}_{i,j+1} - \mathbf{X}_{i-1,j+1})^2 + (\mathbf{X}_{i,j+1} - \mathbf{X}_{i,j})^2 + \varepsilon]^{1/2}} .
\end{aligned}
\tag{12}
$$

The gradient of cost function $f(\mathbf{x})$ is now calculated using Eqs. (10) and (12). Algorithm 1 below shows how CGTV is implemented. In this algorithm, the parameter β^k controls the update of conjugate direction. In our implementation we apply the Dai and Yuan formula proposed in [40], showing robust and fast convergence property which is absent for other formulas such as Hestenes-Stiefel, Fletcher-Reeves, Polak-Ribière, etc., listed in [38].

Algorithm 1 TV-based conjugate gradient method (CGTV)

Input: set ε to a positive, small value (typically $\varepsilon = 10^{-6}$); take $\mathbf{x}^0$ as the initial guess (we assume $\mathbf{x}^0 \equiv 0$); prescribe parameter value for λ; maximum iteration number K.
Output: estimate $\mathbf{x}$
1: $k := 0$
2: **while** $k \leq K$ **do**
3:      calculate gradient direction $\nabla f(\mathbf{x}^k) = \nabla F(\mathbf{x}^k) + \lambda \nabla T(\mathbf{x}^k)$
4:      set Dai and Yuan formula [40] $\beta_{\text{DY}}^k = \|\nabla f(\mathbf{x}^k)\|_2^2 / [(\mathbf{d}^{k-1})^{\text{T}}(\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k-1}))]$ for $k \geq 1$, $\beta_{\text{DY}}^0 = 0$
5:      update the conjugate direction $\mathbf{d}^k = -\nabla f(\mathbf{x}^k) + \beta_{\text{DY}}^k \mathbf{d}^{k-1}$
6:      perform back-tracking line search algorithm to find step size α^k
7:      update the object estimate $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
8:      **if** $\|\alpha^k \mathbf{d}^k\| / \|\mathbf{x}^k\| < \varepsilon$ **then**
9:          **return** $(k+1)^{\text{th}}$ estimate $\mathbf{x}^{k+1}$
10:     **end if**
11:     $k = k + 1$
12: **end while**
13: **return** $(K+1)^{\text{th}}$ estimate $\mathbf{x}^{K+1}$

2.2. Discrete L-curve method to fix TV regularization

As already mentioned, it is advantageous to fix parameter λ in problem (3) with a reasonable criterion other than the strict imposition of $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ in the sense of an exact resolution of this constraint by the Lagrangian multiplier method. Rather we would like to maintain a definite balance between data fidelity and the sparsity condition.

Different values of λ change reconstruction quality. This can be seen from Fig. 2 which displays the results for the Shepp-Logan phantom of image size 256×256 reconstructed from 60 projections using the CGTV solver for various values of λ. If λ is small (Fig. 2(a) and 2(b))

then reconstruction retains a good data fidelity but streakline artifacts do appear in the reconstructed image. For λ approaching the value 2 (Fig. 2(c)) artifacts disappear. For larger values of λ (Fig. 2(d) and 2(e)) images becomes oversmoothed, thus losing details around edges. Therefore, we conclude visually that a good choice is $\lambda = 2$.
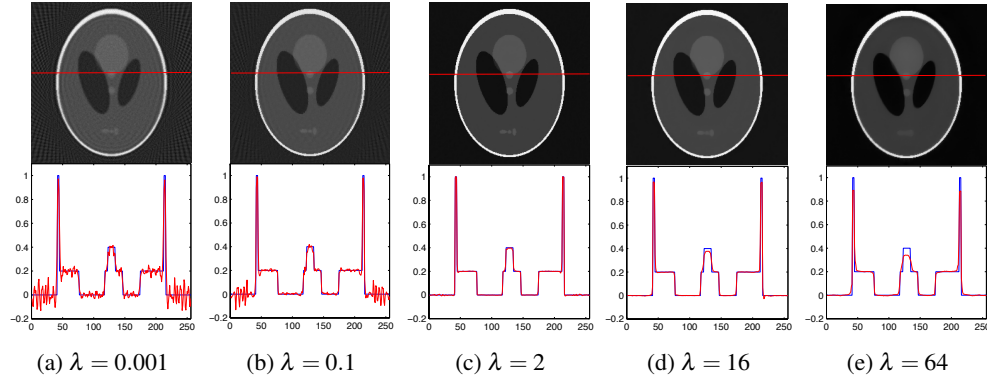

(a) $\lambda = 0.001$	(b) $\lambda = 0.1$	(c) $\lambda = 2$	(d) $\lambda = 16$	(e) $\lambda = 64$

Fig. 2. *Upper row:* reconstruction results for Shepp-Logan phantom (256×256) from 60 projections (402 projections required in FBP reconstruction to meet pixel resolution) using the CGTV solver subject to various values of λ. *Lower row:* Cuts along horizontal through respective images of upper row (marked by red lines). Corresponding cuts through the ground truth are blue lines.

Let us now investigate whether there is a way to automatically optimise λ. An established method to do this is the use of the L-curve criterion which does not require any prior information. Extensively used by Tikhonov in truncated singular value decomposition methods [41], the L-curve criterion surely is applicable to CGTV. This criterion evaluates the norm of the regularisation term versus the residual error in parametric dependence on λ. In our case this corresponds to plotting the parametric curve $\|\mathbf{x}\|_{\mathrm{TV}}$ versus $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$. When the regularisation term is overweighted, the algorithm reconstructs the object with a low regularisation and a high fidelity term. On the other hand, for small values of λ the fidelity term is small while the regularisation term is large. As a consequence, parametric fidelity- versus regularisation-term dependence is L-shaped, see Fig. 3. One may represent this L-curve on a log-log scale. However, the L-shape persists when representing the curve on a linear-linear scale. In our work residuals depend only power-like on λ, and thus we use a linear-linear representation of the L-curve.

Notice that the quantities $\|\mathbf{x}\|_{\mathrm{TV}}$ and $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$ plotted versus one another in the L-curve do in principle possess different dimensions. Setting the dimension of electron density equal to zero, the dimensionality of the "integrands" (integral: average over reconstruction plane) is 1/length (gradient, l1-norm, length given by pixel size $\Delta\mathbf{X}$) for $\|\mathbf{x}\|_{\mathrm{TV}}$ and length^2 (line integral, l2-norm, length again given by pixel size $\Delta\mathbf{X}$) for $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$. So, to obtain the same dimension, say, length^2, $\|\mathbf{x}\|_{\mathrm{TV}}$ would have to be multiplied by $\Delta\mathbf{X}^3$. Since we never change resolution in our reconstruction, we may, however, set $\Delta\mathbf{X} = 1$, and the above-mentioned multiplication needs not to be done. If a problem was to be addressed in which a derivation of the L-curve for a re-scaled pixel size $S\Delta\mathbf{X}$ from the L-curve of the original pixel size $\Delta\mathbf{X}$ was demanded, then this multiplication needs to be performed to exclude geometric rescaling effects. Since the to be reconstructed electron density *does* depend on the resolution, however, we do expect a change of the L-curve and hence its optimal value λ^* after this multiplication still obeying the same L-curve criterion (maximum curvature or minimum euclidean distance to the origin).
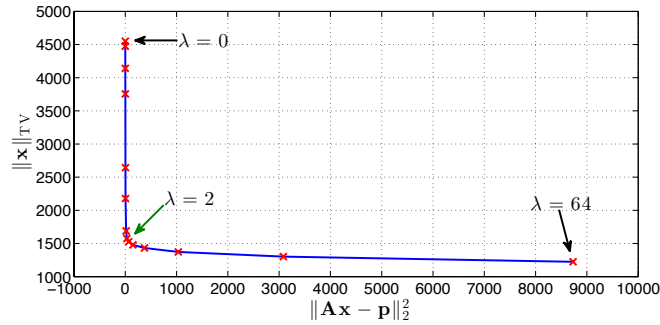
Fig. 3. Discrete L-curve for the reconstruction of the Shepp-Logan phantom (256×256) from 60 projections with respect to $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64$ (for information preserving FBP reconstruction 402 projections would be required). Green arrow indicates the optimal point on the L-curve, corresponding to $\lambda = 2$.

Numerically, it is appropriate to discuss the discrete L-curve method which is based on an interpolation of discrete points on the curve obtained from a finite set λ-values at a fixed number of iterations for the conjugate gradient minimisation where stagnation starts to set in. The L-curve criterion now states that both the regularisation and fidelity terms are comparably small. That is, the prejudice of TV minimisation is to be effective mildly only compared to what problem (2) in Sec. 1 demands. Intuitively, it is clear that the corner of the L-curve meets this requirement. But how can one identify the L-curve corner mathematically speaking? Two criteria can be considered [31]. First, one may seek the point of shortest distance to the origin. Alternatively, one may pick the point of maximum curvature. Numerically, a reliable calculation of curvature requires a higher-order-polynomial interpolation. We do not pursue this option this paper and resort to the first possibility, minimising the Euclidean distance $(F(\mathbf{x})^2 + T(\mathbf{x})^2)^{1/2}$.

For the L-curve in Fig. 3 reconstructions of the Shepp-Logan phantom in Fig. 2 were performed from 60 projections using the CGTV solver subject to 14 values of λ. Breaking the algorithm off after 200 iterations, $F(\mathbf{x})$ and $T(\mathbf{x})$ were calculated and parametrically plotted. From top right to down left points marked by red crosses correspond to the following values of λ: 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, and 64. In the sense of minimum distance to the origin the optimal value is $\lambda = 2$, highlighted by the green arrow. The reconstructed image corresponding to $\lambda = 2$ is displayed in Fig. 2(c). So, indeed, the L-curve criterion generates the same value of λ as identified above visually, indicating its reliability. For the remainder of this paper we refer to the CGTV solver subject to the discrete L-curve method as *optimised CGTV*.

2.3. *Strategy for automised reconstruction*

To reconstruct 3D objects from noisy intensity projections, various pre-processing steps need to be performed which include flat- and dark-field corrections to eliminate modulations e.g. introduced by beam- and detector-inhomogeneities as well as hot- and dark-pixel filtering. In the case of phase-contrast imaging phase retrieval is required in addition. Depending on the SNR, the application of various masks to Fourier-transformed (processed) intensity data may be required prior to phase retrieval [5].

Strictly speaking, CGTV reconstruction of a volume out of 2D projection data acquired in a parallel beam would necessitate the separate determination of the optimal value λ^* for each of the 2D slices comprising the reconstruction volume. Obviously, this task can be performed in parallel [7]. Lacking computing power, it often suffices to determine λ^* in a typical slice and

subsequently use it for near-to-optimal reconstruction of all other slices. Figure 4 displays the workflow of the CGTV reconstruction of a volume.
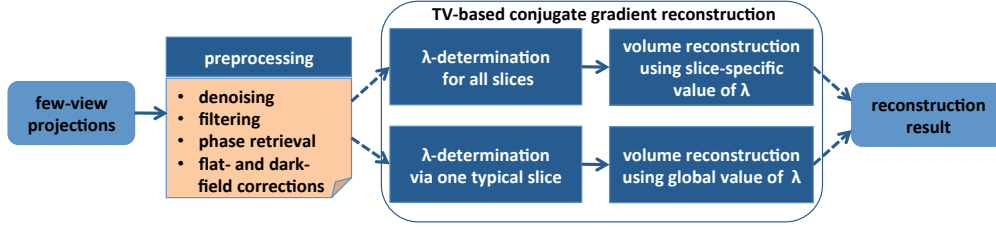

Fig. 4. Strategy of volume reconstruction using the optimised CGTV in the case of parallel-beam imaging.

3. Applications of optimised CGTV reconstruction

In this section we investigate the reconstruction quality of optimised CGTV when applied to a more realistic phantom (photograph of Barbara) and two sets of X-ray *in vivo* data, representing a weevil as well as a stage-17 frog embryo. While reconstructions of the weevil are obtained from a maximum of 400 intensity images, generated by a mix of phase and absorption contrast (no phase retrieval); reconstructions of the frog embryo rely on a maximum of 499 phase maps, retrieved from propagated intensity images. Practically, no absorptive contamination is present in the frog data. (The ratio between real decrement and imaginary part of the refractive index at an X-ray energy of ~ 30 keV is $\sim 10^3$ [42]). Thus, judging from the vantage point of image formation, reconstructions of the frog embryo use a lower dose than the reconstructions of the weevil do. We also compare the performance of optimised CGTV to that of conventional FBP.

3.1. Image quality assessment (IQA)

Subjective evaluation, which draws upon experience, is often used for CT reconstruction to assess its quality. In general, an analysis of research data should not rely on subjective evaluation. Reproducible results of low prejudice can only be generated by subjecting data assessment to sufficiently general scientific principles (a set of error metrics optimised to given noise types and the effects of missing input information). This is important for meaningful image analysis subsequent to the reconstruction procedure. For example, reconstructed CT data acquired in entomology or developmental biology serve as input to segmentation of tissues and cells, volumetry, cell-mass determination, estimates on the statistics of model parameters, feature extraction, inference of force exertion, flow-field analysis, etc., and each of these procedures propagates the reconstruction error.

3D reconstruction of tomographic data acquired in a parallel beam decomposes into 2D reconstruction of independent slices. Therefore, only a 2D assessment of reconstruction quality is considered here. The simplest and most widely used quality measure is the mean square error (MSE), defined as the average of the squared grey-value differences over all pixels representing reconstructed ($\mathbf{X}$) and reference ($\mathbf{Y}$) slices

$$\text{MSE}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} (\mathbf{X}_{i,j} - \mathbf{Y}_{i,j})^2. \tag{13}$$

An assumption behind using MSE as an error measure is that image pixels are statistically independent. Thus MSE exhibits a certain sensitivity to structural changes [43]. A low value

of $MSE(\mathbf{X}, \mathbf{Y})$ indicates good similarity between $\mathbf{X}$ and $\mathbf{Y}$. However, congruence between extended regions is detected more poorly. Therefore, we also consider another image quality index called structural similarity (SSIM) [44]. SSIM is based on the fact that images acquired of real objects are usually structured in a hierarchical way, and correlations between grey values in separated pixels occur. The SSIM index is a product of three factors based on a luminance measure $l(\mathbf{X}, \mathbf{Y})$, a contrast comparison $c(\mathbf{X}, \mathbf{Y})$, and association $s(\mathbf{X}, \mathbf{Y})$,

$$SSIM(\mathbf{X}, \mathbf{Y}) = l(\mathbf{X}, \mathbf{Y})^{\alpha} \times c(\mathbf{X}, \mathbf{Y})^{\beta} \times s(\mathbf{X}, \mathbf{Y})^{\gamma}, \tag{14}$$

where α, β, and γ are real-positive trade-off parameters which adjust their relative importance. For simplicity, here we set them equal to unity, $\alpha = \beta = \gamma = 1$. The reader is referred to [44] for explicit representations of l, c, and s. SSIM ranges within [0, 1]. Values close to unity (zero) indicate a great (small) structural similarity between the reconstructed and the reference image.

3.2. Reconstruction of simulated data

The discrete L-curve of the Shepp-Logan reconstruction of Sec. 2.2 was represented by fourteen values of λ. Figure 5 now shows the dependences of MSE and SSIM on the same λ values, demonstrating consistency with the discrete L-curve of Fig. 3 in the sense that the optimal value of $\lambda = 2$ is close to the minimum (maximum) of the MSE (SSIM) curve. Note that compared to the L-curve criterion the minimum (maximum) of the MSE (SSIM) curve yields a slight underestimation (overestimation) of this parameter value. This demonstrates consistency of the L-curve method in the sense of a compromise between MSE and SSIM based optimisation.
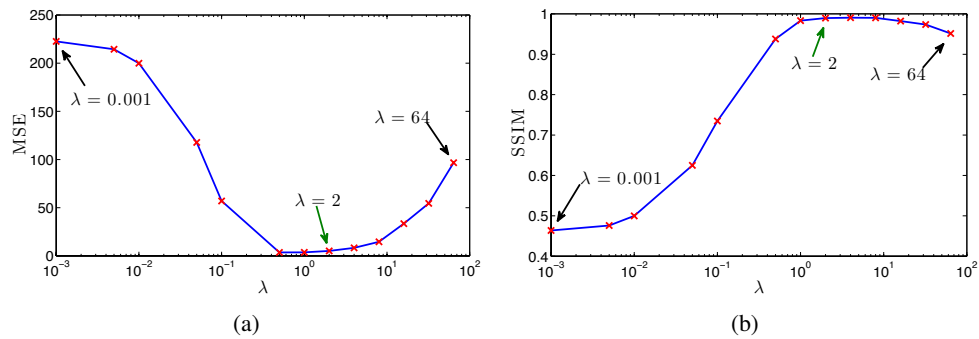

Fig. 5. MSE in (a) and SSIM in (b) upon evaluating the CGTV reconstruction of the Shepp-Logan phantom with fourteen different values of λ. Green arrows indicate the optimal value $\lambda = 2$ as determined from the discrete L-curve in Fig. 3.

Let us now apply this analysis to a more realistic phantom – the Barbara image (256×256) shown in Fig. 6(a). In many X-ray CT applications the gradient image is sparse only in an approximate sense. Namely, it is not guaranteed that the object, like the one in Fig. 6(a), enjoys the property of piece-wise constancy across its entire support as was the case for the Shepp-Logan phantom. Using the same forward model, see Sec. 1, 120 projections are sampled which are subsequently taken as input for the CGTV solver subject to the discrete L-curve criterion to optimally reconstruct the image.

The first step is to determine the optimal value of λ. Again, the same fourteen values of λ as used in case of the Shepp-Logan phantom are employed. Figures 6(b)–6(d) display reconstruction results for three different values of λ. Small values of λ, say $\lambda = 0.001$, generate noticeable streakline artifacts. In contrast, when λ goes large, details at edges are oversmoothed. The corresponding L-curve is shown in Fig. 7. The smallest distance to the origin occurs at

$\lambda = 4$, corresponding to the reconstruciton in Fig. 6(c). Theoretically, regions represented by high frequencies (texture) can only be smoothly interpolated from a few projections when using TV regularization due to lack of sparsity in the gradient-space representation. As shown in Fig. 6(a), areas marked by red ellipses lose texture information in the reconstruction of Fig. 6(c), but sparse areas, marked by green rectangles, preserve their edges well.
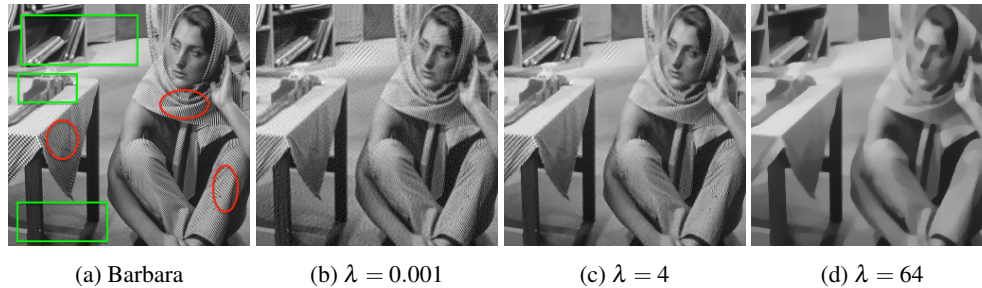

(a) Barbara   (b) $\lambda = 0.001$ (c) $\lambda = 4$ (d) $\lambda = 64$

Fig. 6. (a) original Barbara (256×256) and reconstruction results from 120 projections using the CGTV solver with (b) $\lambda = 0.001$, (c) $\lambda = 4$, and (d) $\lambda = 64$.
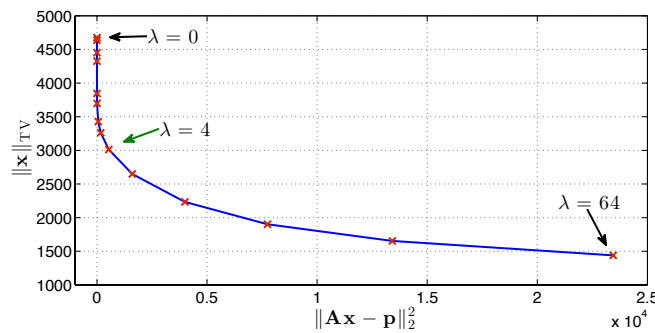
Fig. 7. Discrete L-curve for the reconstruction of the image Fig. 6(a) (256×256) from 120 projections (for information preserving FBP reconstruction 402 projections would be required). Values of λ, which are employed, read $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64$. According to the criterion of shortest distance to the origin, the optimal value is $\lambda = 4$, indicated by the green arrow.

Figure 8 depicts reconstructions of the Shepp-Logan phantom and Barbara with a much small number of projections (60 as opposed to 402 in former, 120 as opposed to 402 in latter case) than required for FBP reconstruction without loss of resolution relative to that imposed by pixel size. Note the streaklines in FPB reconstructions subject to these lower number of projections while optimal CGTV is void of such artifacts. Table 1, in terms of IQA indexes, shows that optimised CGTV performs much better than FBP for a limited number of projections. Note that due to its piecewise constancy optimised CGTV reconstruction of the Shepp-Logan phantom is essentially perfect. This is less so for the optimised CGTV reconstruction of Barbara which, however, still performs much better than FBP.

3.3. Reconstruction of experimental data

In this section, two different *in vivo* data sets are examined using FBP and optimised CGTV reconstruction, one representing a weevil imaged by a mixture of phase and absorption

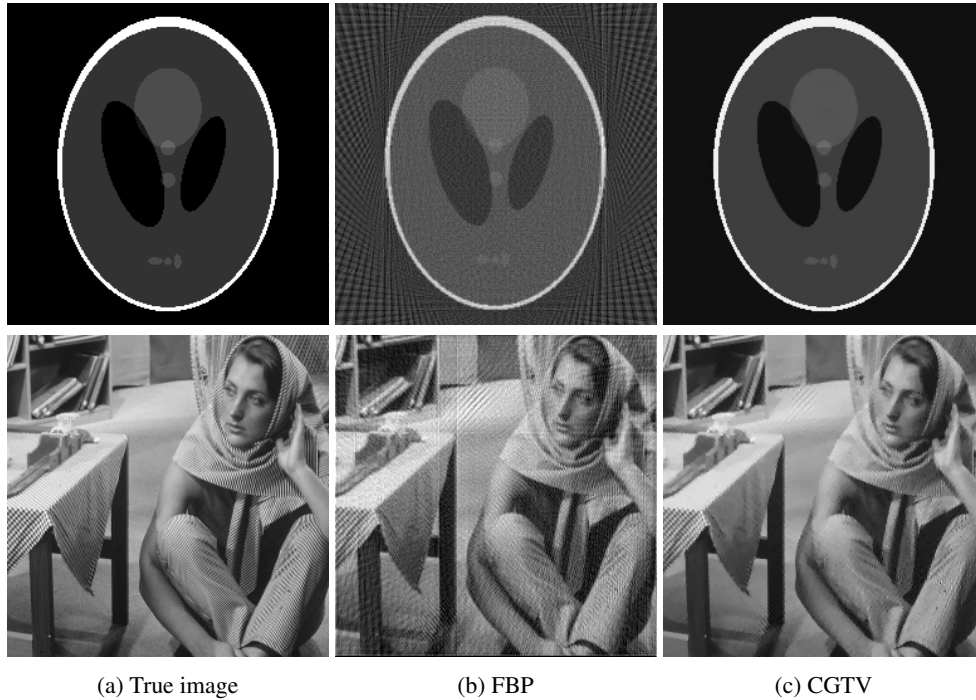(a) True image   (b) FBP   (c) CGTV

Fig. 8. True image and image reconstructions using FBP and CGTV for the Shepp-Logan phantom (first row, 60 projections) and Barbara (second row, 120 projections).

Table 1. IQA measures for reconstructions using FBP and optimised CGTV subject to low numbers of projections through the Shepp-Logan and Barbara phantoms.

IQA Measures	Shepp-Logan		Barbara	
	FBP	CGTV	FBP	CGTV
MSE	843.35	4.54	867.11	299.05
SSIM	0.21	0.99	0.40	0.75

propagation-based X-ray intensity contrast subject to direct reconstruction (no retrieval of the exit wave function), the other one embodying a stage-17 frog embryo during neurulation with a stack of exit-phase maps (projection of real decrement of refractive index) serving as input to the reconstruction. Phase retrieval is performed by a variant of the quasiparticle algorithm proposed in [45] and [46] with the high-frequency part of the intensity contrast being cut off at a radial point where noise starts to exceed the signal. Both data sets are subject to statistical noise, in the former case directly present in the intensity data while the stack of phase maps carries the noise contamination in an implicit way. Projections are constrained to the angular range $[0°, 180°]$.

For imaging the weevil, white-beam illumination with a critical energy of $E_c \sim 15\,\text{keV}$, a propagation distance of $z = 50\,\text{cm}$, a photon flux density of $\sim 10^{13}$ phs/mm^2/s, an effective pixel size of $\Delta x = 3.7\,\mu\text{m}$, and a field of view of $(7.4\,\text{mm})^2$, corresponding to 2000×2000 pixels, was employed at ANKA's TOPO-TOMO bending-magnet imaging beamline. The detection system uses a freestanding 50 μm-thick Cerium-doped lutetium aluminum garnet (LuAG:Ce)

scintillator orthosilicate scintillator, generating visible light by the absorption of X-rays. This latent visible-light image is relayed by an optical system with a three-fold magnification onto the pco.dimax camera that performs the actual signal detection. An acquisition of 400 intensity projections was performed per tomogram with an exposure time of 0.3 ms per projection.

The frog embryo was imaged at the undulator imaging beamline 32-ID of APS subject to monochromatic ($\Delta E/E \sim 10^{-4}$), highly coherent X-ray illumination of energy $E = 30\,\text{keV}$, a photon flux density of $\sim 10^{12}$ phs/mm^2/s, a propagation distance of $z = 70\,\text{cm}$, an effective pixel size of $\Delta x = 1.3$ µm, and a field of view of $(3.328 \times 2.808\,\text{mm})^2$, corresponding to 2560×2160 pixels, were employed. The detection system used in this experiment was a $100\,\mu\text{m}$ LuAG:Ce scintillator supplemented by a five-fold magnifying optics and a pco.edge camera. The number of projections acquired per tomogram was 499 with 60 ms of exposure time per projection. A rough estimate yields that the dose per projection, deposited into the object, is about 20 times smaller for weevil compared to frog-embryo imaging.

In order to perform few-view reconstructions in case of weevil imaging, one fifth of angular, equidistantly-spaced projections (80) are extracted out of 400 projections, while in the frog case a third (167) equidistant projections are taken as input for the reconstruction. This choice of data thinning is motivated by Figs. 15 and 16 to be discussed below. Sparsity holds fairly well for the weevil data (chitin skeleton, large cavities represent piecewise constant regions), thus few-view reconstruction should lead to good results. In contrast, the frog embryo exhibits, at a comparable imaging resolution, finer structures (cells of variable size, cell nuclei, yolk platelets, boundaries between densely and loosely packed cells) leading to appreciable variations within regions. In principle, more projections thus are necessary for faithful reconstruction.

Thanks to parallel-beam illumination, we focus on the reconstruction of one 2D slice only. Recall that the reconstruction of the 3D volume is accomplished by simple vertical stacking of such 2D slices, allowing parallel computing for fast reconstruction even though this produces an asymmetric treatment of sparsity in x, y versus z gradients. Reference slices are FBP-reconstructed using the full number of projections.

Figures. 9(a), 9(b), 9(d), 9(f) display weevil-reconstruction results in one and the same slice using CGTV subject to various values of λ. The level of Poisson noise in the intensity projections here is about 0.5 %. Such a low noise-level is expected to propagate to the reconstructed object, maintaining the same order of magnitude. Statistical noise thus is negligible for weevil reconstruction based on the original data. To check exemplarily how a substantial increase of Poisson noise affects the CGTV reconstruction we artificially impose a 3 % noise level (6 times as the original) on the average intensity of weevil data. The corresponding discrete L-curves are shown in Fig. 10 for both cases of 0.5 % and 3 % Poisson noise. Minimum distance to the origin occurs for both at the same value $\lambda = 0.5$ (green arrows in Fig. 10). Notice that for $\lambda = 0.5$ the values of $\|\mathbf{x}\|_{\text{TV}}$ are comparable while $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ is considerably larger for the case of 3 % Poisson noise. This indicates that optimal CGTV reconstruction universally annihilates small-scale fluctuations albeit subject to a mild alteration of the large-scale structure, comparing Fig. 9(d) with Fig. 9(e). However, we do expect a shift of the optimal value of λ in the right direction if further higher level of Poisson noise is added. Obviously, when λ is small there are streakline artifacts (Fig. 9(b)) and pronounced decrease in quality for the high noise data (Fig. 9(c)). Streakline artifacts are absent in Figs. 9(d) and 9(e) (optimal reconstruction) and 9(f), the latter, however, exhibiting poor resolution because of higher Poisson noise.

Figures 11(b)–11(d) display according CGTV reconstructions of the stage-17 frog embryo. The level of Poisson noise in intensity projections is comparable to the original weevil data, roughly of the order of 0.3 %, which is low. Note, however, that towards the bottom part of the slice there is motion induced blur (systematic error) due to an onset of developmental dynamics during the tomographic scan. The L-curve is shown in Fig. 12 with the minimum distance to the

(a) reference, 0.5 % noise   (b) $\lambda = 0.001$, 0.5 % noise   (c) $\lambda = 0.001$, 3 % noise

(d) $\lambda = 0.5$, 0.5 % noise   (e) $\lambda = 0.5$, 3 % noise   (f) $\lambda = 4$, 0.5 % noise
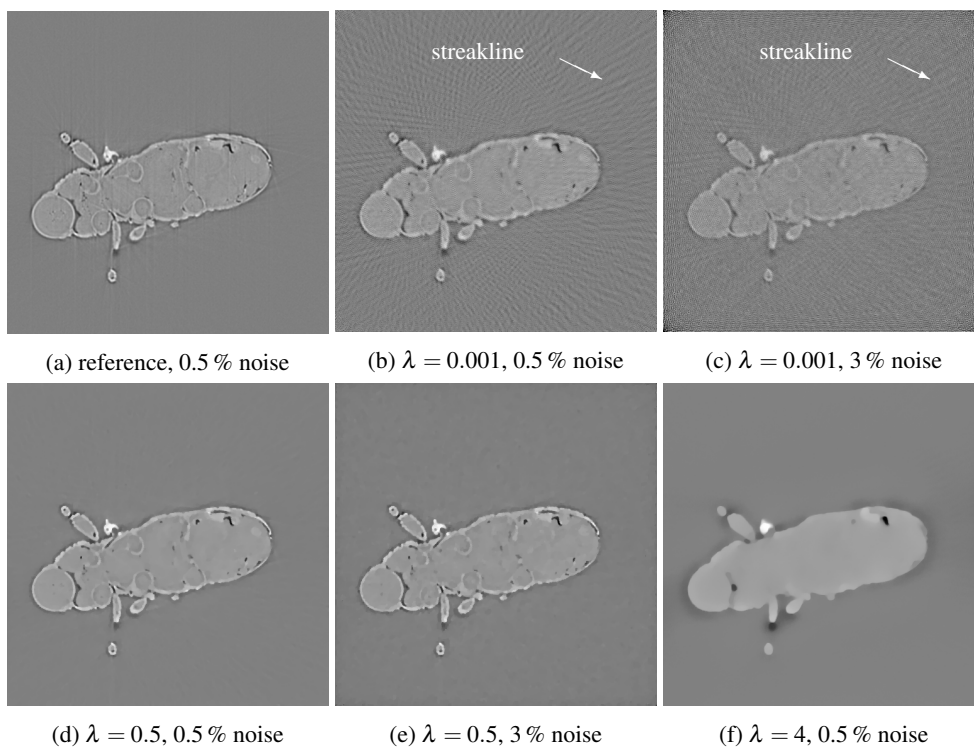
Fig. 9. Reference image and λ dependent reconstruction results from intensity data of weevil subject to different levels of Poisson noise and 80 projections: (a) FBP reconstruction using full set of low-noise projections; (b) and (c) CGTV reconstructions for a small value of $\lambda = 0.001$ for 0.5 % and 3 % noise data, respectively; (d) and (e) respective CGTV reconstructions for $\lambda = 0.5$ which is optimal for both, 0.5 % and 3 % noise data; (f) CGTV reconstruction using $\lambda = 4$ for 0.5 % noise data.
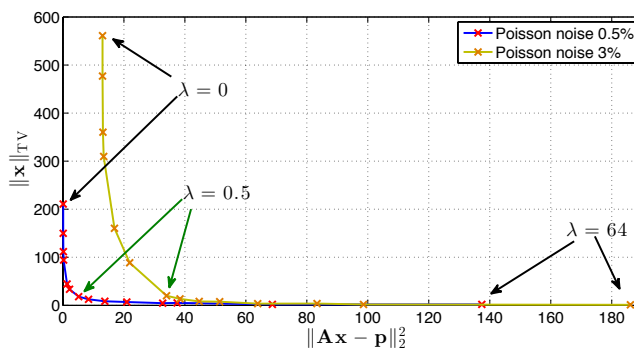
Fig. 10. Discrete L-curves for CGTV reconstruction from 80 projections of the weevil data with 0.5 % (blue curve) and 3 % (lime green curve) Poisson noise using $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64$. Both curves exhibit minimum distance to the origin at the same value $\lambda = 0.5$.

| (a) reference | (b) $\lambda = 0.001$ | (c) $\lambda = 0.1$ | (d) $\lambda = 4$ |
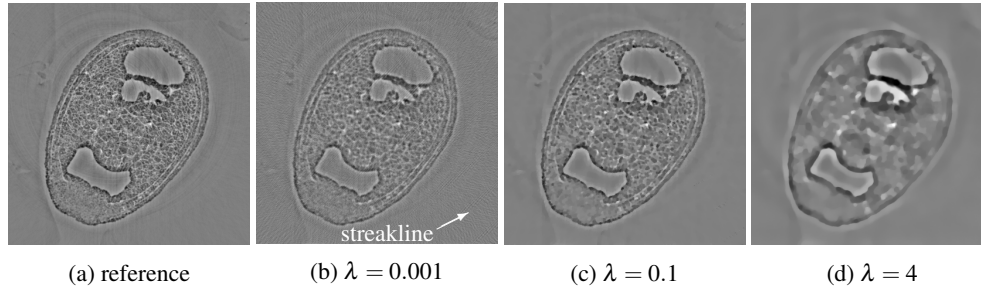
Fig. 11. Reference image and λ dependent reconstruction results from phase maps retrieved from intensity projections of stage-17 frog embryo: (a) FBP reconstruction using full set of projections; (b)–(d) CGTV reconstructions using 167 projections for different values of λ.
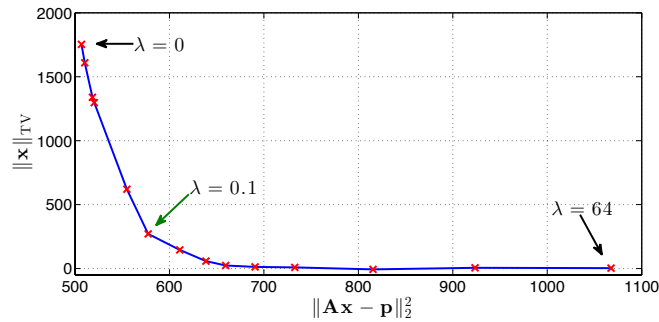
Fig. 12. Discrete L-curve for CGTV reconstruction from 167 phase maps of stage-17 frog embryo using $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64$.

origin occurring for $\lambda = 0.1$ (green arrow in Fig. 12). The associated reconstruction is depicted in Fig. 11(c). Again, streakline artifacts occur in Fig. 11(b), which are absent in Figs. 11(c) (optimal reconstruction) and 11(d).

Interestingly, the presence of a lightly weighted TV minimisation constraint for few-view CGTV reconstruction appears to mimick FBP reconstruction as far as the occurrence of streakline artifacts is concerned. This is shown in Fig. 13. Thus, by a small value of λ the degeneracy in finding the (global) minimum of $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ (underdetermined linear system of equations) is lifted in the sense of unique FBP reconstruction.

To point out the loss of resolution even for optimised CGTV we present in Fig. 14(b) a line cut through the frog-embryo slice shown in Fig. 14(a) with reconstructions performed by FBP subject to the full number of projections ($P = 499$, reference, red curve) and optimised CGTV using only $P = 167$ (blue curve). Clearly, the blue line misses high-frequency components (noisy structure in flat regions) which are present in the red line. There is high-frequency information in the optimised CGTV reconstruction with $P = 499$ (green curve), occasionally albeit not always representing that of the red line. We conclude that low-dose (few-view) reconstruction using optimised CGTV is free of streakline artifacts which necessarily are introduced by FBP. Optimised CGTV thus operates cleanly with an acceptable loss of resolution towards long *in vivo* time-lapse series in the following sense: it depicts, say, cellular shapes by precise reconstruction of cell boundaries, it excludes the appearance of FBP few-view reconstruction artifacts, but it does not resolve small intracellular details like yolk platelets and nuclei. For dose-intense (many-view) reconstruction of sub-cellular resolution FBP is to be preferred due
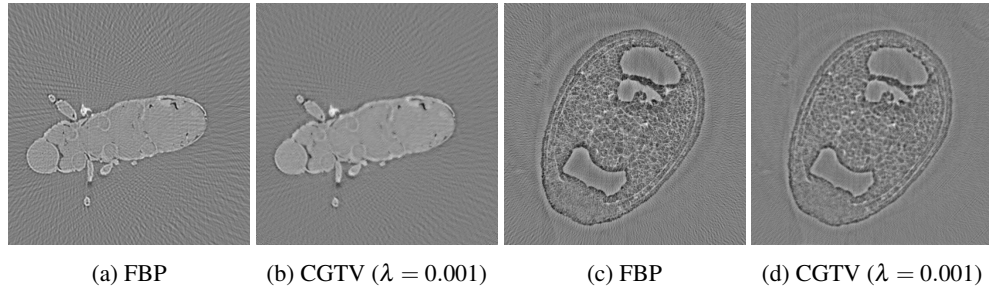
(a) FBP   (b) CGTV ($\lambda = 0.001$)   (c) FBP   (d) CGTV ($\lambda = 0.001$)

Fig. 13. Comparison of FBP and small-λ CGTV reconstruction of weevil and frog-embryo slices ($\lambda = 0.001$ in both cases) using a small number of projections, $P = 80$ and $P = 167$ (as opposed to maximum numbers $P = 400$ and $P = 499$), respectively. Note that streakline artifacts appear for both FBP and small-λ CGTV reconstruction.
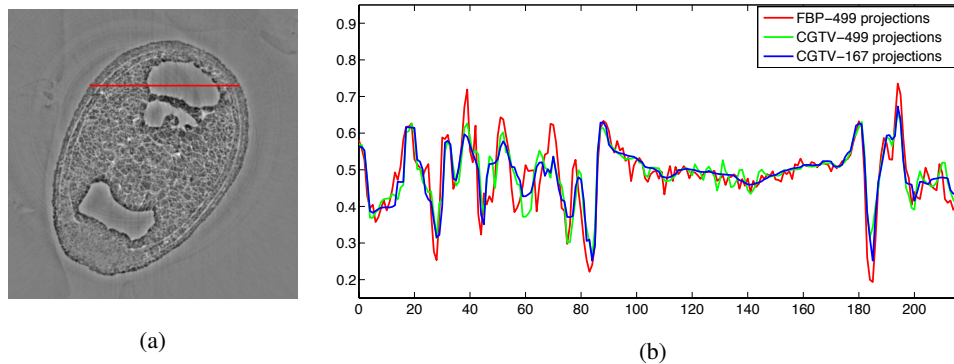

(a)   (b)

Fig. 14. (a) Definition of line through reconstructed slice of stage-17 frog embryo, (b) reconstruction results (normalised to [0, 1]) along this line for FBP subject to the full number of projections ($P = 499$, reference, red curve) and optimised CGTV using $P = 167$ (blue curve) and $P = 499$ (green curve).

to its computational efficiency, however.

Finally let us, using the two error metrics MSE and SSIM as defined in Sec. 3.1, ask the question how the reconstruction quality depends on the number of projections P, taking the full-projection FBP reconstruction as a reference. As Fig. 15 (weevil) and Fig. 16 (stage-17 frog embryo) indicate both error measures saliently saturate at $P \sim 80$ and $P \sim 167$ (green arrows), respectively. This justifies our use of these P values in the analysis above.

3.4. Convergence analysis and computational performance

To demonstrate relatively fast convergence for the regularisation of ART by TV-minimisation using the conjugate gradient solver, we depict in Fig. 17 the behaviour of the cost function for the data set of weevil, corresponding to (3), in dependence of the iteration number k for the weevil data set. Five curves corresponding to different values of λ are shown, in which red denotes $\lambda^* = 0.5$. For λ values in the vicinity of λ^* no appreciable decent occurs for $k > 30$. Regarding computational performance, forward and backward operations (such as $\mathbf{Ax}$, $\mathbf{A^Tx}$) are required and take a considerable time in each iteration. Typically, the computational time required for a CGTV volume reconstruction with a given value of λ takes about 25 min employing the parallel reconstruction strategy and hardware configuration in [7].
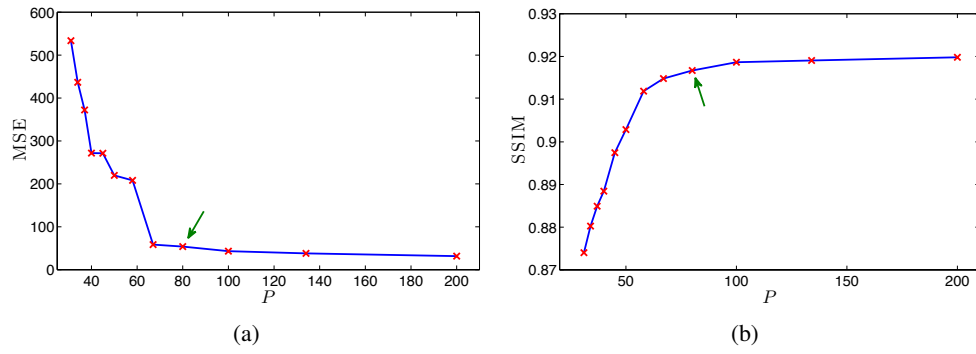
Fig. 15. Error of optimised CGTV with respect to full-view FBP reconstruction ($P = 400$) of weevil slice using error metrics MSE and SSIM, defined in Sec. 3.1, versus the number of projections P. Green arrows mark approximate onset of saturation.
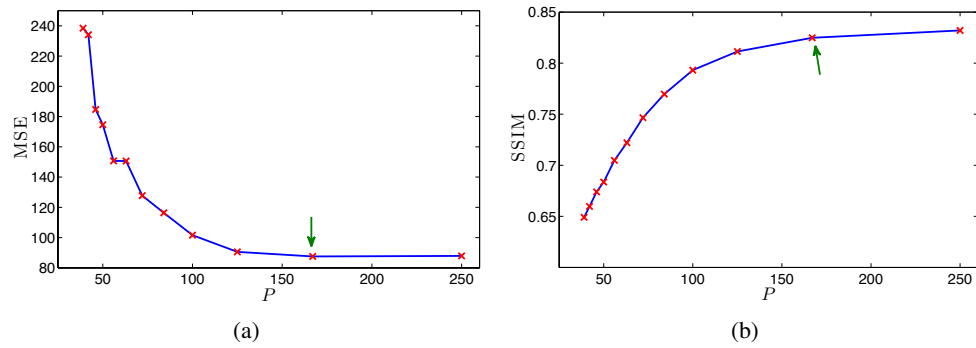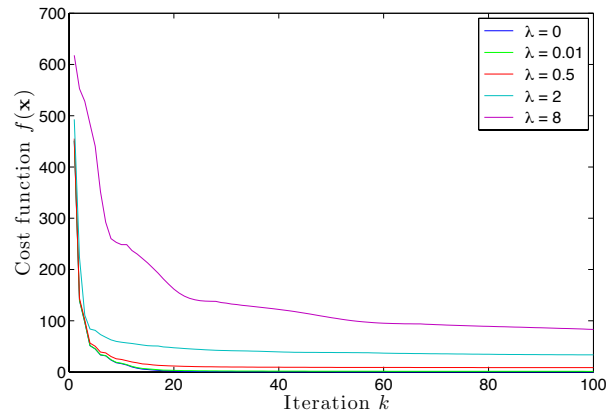

Fig. 16. Error of optimised CGTV with respect to full-view FBP reconstruction ($P = 499$) of stage-17 frog-embryo slice using error metrics MSE and SSIM, defined in Sec. 3.1, versus the number of projections P. Green arrows mark approximate onset of saturation.

Fig. 17. Convergence analysis of CGTV reconstruction of weevil data for five different values of λ. The optimal value is $\lambda^* = 0.5$ denoted in red, and in its vicinity minimisation saturates for a number of iterations $k > 30$.

4. Conclusions

In this work we propose and evaluate a particular algebraic reconstruction technique (optimised CGTV), minimising the total variation (TV) in a conjugate-gradient fashion in using a Lagrangian multiplier formulation. The optimal value of the multiplier is automatically obtained by imposing an independent L-curve criterion. By 'independent' we mean that the multiplier is not determined literally by resolving the constraint equation. Rather, a physically reasonable flexibility is introduced doing justice to the principle of least prejudice on the reconstruction.

Applying optimised CGTV to reconstruct phantom and *in vivo* data, acquired by propagation based X-ray imaging with and without phase retrieval, we conclude that few-view reconstruction represents a promising low-dose venue. Namely, it maintains resolution at an acceptable level, continues to realistically reproduce edges which bound sufficiently large structures (cells, tissue confines), quickly saturates the image quality as the number of projections increases, and is void of streakline artifacts as generated by filtered back-projection (FBP). This is important, e.g., for automatic image analysis such as structure segmentation, appealing to a priori set grey-value thresholds. That is, moderately sacrificing the spatial resolution is not a strong limitation if the subject of biological research is to understand coarse phenotypal dynamics: tissue separation, cell division and migration, and cavity formation. All these only require a clear segmentation of the associated boundaries. Because of this, optimised CGTV should be interesting for 4D livecell imaging, aspiring to the acquisition of long time-lapse series [3, 4]. The *in vivo* data, as reconstructed in the present work, is not contaminated by a high level of statistical photon noise. Exemplarily, we show for the weevil data that with an artificial increase of Poisson noise optimal CGTV performs well in annihilating noise-induced small-scale fluctuations at the same time well preserving objects features seen in CGTV reconstruction based on the low-noise data. A more systematic investigation of optimal CGTV reconstruction under various sources of noise is certainly in order. We plan to carry out such an analysis in future work based on *in vivo* tomographic data. In such data high-resolution reconstruction obtained by techniques of many-view reconstruction can still be accommodated, after the acquisition of a long time-lapse series few-view reconstructed by optimised CGTV, to spatiotemporally pin down and correlate interesting events.

It is perceivable that the L-curve method proposed in the present work can be applied to other task-specific reconstruction problems that appeal to some sort of sparsity, say, to focus on enhanced surface reconstruction or absorptive contamination in images that primarily use phase contrast.