EFFICIENT IMAGE RECONSTRUCTION UNDER SPARSITY CONSTRAINTS WITH APPLICATION TO MRI AND BIOLUMINESCENCE TOMOGRAPHY

M. Guerquin-Kern, J.-C. Baritaux, and M. Unser

Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne

ABSTRACT

Most bioimaging modalities rely on indirect measurements of the quantity under investigation. The image is obtained as the result of an optimization problem involving a physical model of the measurement system. Due to the ill-posedness of the above problem, the impact of the noise on the reconstructed images must be controlled. The recent emphasis in biomedical image reconstruction is on regularization schemes that favor sparse solutions, which renders the optimization problem non-smooth. In this work, we show how step-size adaptation can be used to speed up the most recent multi-step algorithms (e.g. FISTA) employed in sparse image recovery. We present experiments in MRI and Fluorescence Molecular Tomography with specifically tailored step-adaptation strategies. Our results demonstrate the possibility of an order-of-magnitude speed enhancement over state-of-the-art algorithms.

Index Terms— Bioluminescence tomography, parallel MRI, inverse problem, FISTA, FWISTA.

1. INTRODUCTION

This paper is concerned with the general problem of image reconstruction from linear measurements, including MRI and various forms of X-ray and optical tomographies. The generic form of the image-formation model in vector notation is

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}_{\text{orig}} + \boldsymbol{b}, \tag{1}$$

with $\boldsymbol{y}$: the physical measurements acquired by the scanner, $\boldsymbol{x}$: the coefficients of the image in some suitable basis (e.g. pixels), $\boldsymbol{H}$: the "system matrix" which results from the discretization of the physical model, and $\boldsymbol{b}$: some additional measurement noise. The goal is to reconstruct the unknown image $\boldsymbol{x}$ given the noisy measurements $\boldsymbol{y}$.

This work was supported in part by the Swiss National Competence Center in Biomedical Imaging (NCCBI) and in part by the Swiss innovation promotion agency (CTI grant no. 9601.1 PFLS-LS).

This linear model is valid for most biomedical imaging modalities[1]. The reconstruction is especially difficult when the problem is underdetermined or when $\boldsymbol{H}$ is badly-conditioned, in which case a standard matrix inversion is not applicable because it dramatically amplifies noise. Under those circumstances, the common approach for reconstructing $\boldsymbol{x}$ is to introduce regularization constraints and to reformulate the problem as an optimization task. The reconstructed image is defined as the minimizer of a cost functional of the form

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 + \mathcal{R}(\boldsymbol{x}), \tag{2}$$

where $\mathcal{R}$ is a regularization term that penalizes undesirable solutions.

Choosing $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{R}\boldsymbol{x}\|_2^2$ leads the famous Tikhonov solution $\boldsymbol{x}^\star = \left(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} + \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y}$. The latter can be computed iteratively by efficient methods such as the conjugate gradient (CG) algorithm that exhibits a linear rate of convergence; i.e., $\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|_2 \leq \rho^n$ for $0 \leq \rho = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} < 1$ with κ the conditioning number of the matrix to be inverted.

With the recent advent of compressed sensing, the use of ℓ_1-norm regularization, which favors sparsity, has become extremely popular. This is motivated by the fact that most biomedical images are well represented by their K-term approximation (with K small) in an appropriate basis. ℓ_1 regularization has been found to lead to images of better quality than the classical linear reconstruction techniques (in particular, Tikhonov and Wiener-filter-type methods). The method owes much of its success to the iterative shrinkage/thresholding algorithm (ISTA) [2, 3], which has the ability to handle the large-scale problems encountered in bioimaging. Unfortunately, ISTA is relatively slow, which has prompted researchers into searching for ways to accelerate it [4, 5].

The current state of the art is the fast iterated shrinkage/thresholding algorithm (FISTA), which incorporates some of the latest technology in gradient-based convex optimization [4]. While FISTA is faster than ISTA, its convergence rate is still sublinear and often not competitive with the best linear methods such as CG. Since reconstruction duration remains a major concern in the

case of very large scale and ill-posed problems, our proposal here is to revisit and fine-tune FISTA by taking advantage of the specific structure of the problem in order to improve the bound constants of the algorithm.

This paper is structured as follows. First, we introduce our modified version of FISTA, FWISTA, that applies the principle of an optimized step size adaptation. We derive the associated convergence properties whose tighter form is indicative of a reduced reconstruction time. Second, we present experiments in bioluminescence tomography (BLT) and parallel magnetic resonance imaging (pMRI) with phantom data where our method brings a significant speedup over ISTA and FISTA.

2. THEORY

In this paper we denote $\boldsymbol{\Lambda_w}$ the diagonal matrix with entries given by the vector $\boldsymbol{w}$. We consider the variational problem

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \underbrace{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \|\boldsymbol{\Lambda_\lambda} \boldsymbol{x}\|_1}_{\mathcal{C}(\boldsymbol{x})}, \tag{3}$$

where $\boldsymbol{\lambda}$ is a vector of arbitrary regularization weights. In particular, $\boldsymbol{\lambda}$ accounts for the regularization parameter. We also denote $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{W}$, where $\boldsymbol{W}$ is the reconstruction matrix of the underlying basis (or frame) and $\boldsymbol{x}$ the corresponding signal coefficients.

We now slightly extend FISTA by adapting it to the form of the above problem. Our main point will be to show how we can theoretically improve the convergence by exploiting the structure of $\boldsymbol{A}$. We then substantiate the feasibility of such a convergence enhancement with a specific example of magnetic resonance imaging (MRI).

2.1. Fast Weighted ISTA (FWISTA)

The derivation of FWISTA follows the one of FISTA. Our key concept is to consider some weighted norms of the form $\|\boldsymbol{x}\|_{\boldsymbol{S}}^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x}$ where $\boldsymbol{S}$ is symmetric positive-definite matrix.

We define the thresholding operation with pixel-dependent threshold values as

$$\mathcal{T}_{\boldsymbol{\tau}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{v}} \|\boldsymbol{x} - \boldsymbol{v}\|_2^2 + \|\boldsymbol{\Lambda_\tau} \boldsymbol{v}\|_1 . \tag{4}$$

where $\boldsymbol{\tau}$ is an arbitrary weight vector. The important point is that this operation can be achieved in one step with a low computational cost.

We then formulate FWISTA in Alg. 1, as an extension of FISTA with adapted thresholds.

It is not difficult to show that the condition for Algorithm 1 to converge to $\boldsymbol{x}^\star$ is that $\boldsymbol{\Lambda_\tau}^{-1} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ must be positive-definite.

Algorithm 1: FWISTA solving (3)

input : $\boldsymbol{M} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, $\boldsymbol{a} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$, $\boldsymbol{x}_0$, $\boldsymbol{\Lambda_\tau}$.
Initialization: $n = 0$, $\boldsymbol{u}_0 = \boldsymbol{x}_0$, $t_0 = 1$;
repeat
  $\boldsymbol{x}_{n+1} \leftarrow \mathcal{T}_{\boldsymbol{\Lambda_\tau \lambda}}(\boldsymbol{u}_n + \boldsymbol{\Lambda_\tau}(\boldsymbol{a} - \boldsymbol{M}\boldsymbol{u}_n))$;
  $t_{n+1} \leftarrow \left(1 + \sqrt{1 + 4t_n^2}\right)/2$;
  $\boldsymbol{u}_{n+1} \leftarrow \boldsymbol{x}_{n+1} + \frac{t_n - 1}{t_{n+1}}(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n)$;
  $n \leftarrow n + 1$;
until *stopping criterion is met*;
return $\boldsymbol{x}_{n+1}$.

Standard FISTA corresponds to the case $\boldsymbol{\Lambda_\tau} = \frac{2}{L}\boldsymbol{I}$ and $\boldsymbol{\Lambda_\lambda} = \lambda \boldsymbol{I}$. Note that the key difference between FISTA and FWISTA is that the thresholding step in FWISTA can apply a different threshold value for each component of $\boldsymbol{x}$ depending on the entries of $\boldsymbol{\Lambda_\tau}$ and $\boldsymbol{\Lambda_\lambda}$.

Considering the weighted norm formalism, we revisit the convergence results of FISTA [4, Thm. 4.4].

Proposition 1. *Let $\{\boldsymbol{x}_n\}$ be the sequence of estimates generated by Algorithm 1 with $\boldsymbol{\Lambda_\tau}$ such that $\boldsymbol{\Lambda_\tau}^{-1} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is positive-definite. Then, for any $n \geq 1$,*

$$\mathcal{C}(\boldsymbol{x}_n) - \mathcal{C}(\boldsymbol{x}^\star) \leq \left(\frac{2}{n+1}\right)^2 \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_{\boldsymbol{\Lambda_\tau}^{-1}}^2 . \tag{5}$$

If $\mathcal{C}$ is ϵ-strongly convex, i.e. $\epsilon = 2\lambda_{min}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) > 0$, then

$$\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|_2 \leq 2\sqrt{\frac{2}{\epsilon}} \frac{\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_{\boldsymbol{\Lambda_\tau}^{-1}}}{n+1}. \tag{6}$$

Proof. Consider the change of variable $\boldsymbol{z} = \sqrt{\boldsymbol{\Lambda_\tau}^{-1}}\boldsymbol{x}$ and apply FISTA with $L = 2$ to the new problem. It results in steps equivalent to the ones of FWISTA. Hence the convergence result of FISTA transposes directly $\mathcal{C}\left(\sqrt{\boldsymbol{\Lambda_\tau}}\boldsymbol{z}_n\right) - \mathcal{C}\left(\sqrt{\boldsymbol{\Lambda_\tau}}\boldsymbol{z}^\star\right) = \mathcal{C}(\boldsymbol{x}_n) - \mathcal{C}(\boldsymbol{x}^\star) \leq \left(\frac{2}{n+1}\right)^2 \|\boldsymbol{z}_0 - \boldsymbol{z}^\star\|_2^2 = \left(\frac{2}{n+1}\right)^2 \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_{\boldsymbol{\Lambda_\tau}^{-1}}^2$. In the strongly convex case, we have $\frac{\epsilon}{2}\|\boldsymbol{x}_n - \boldsymbol{x}^\star\|_2^2 \leq \mathcal{C}(\boldsymbol{x}_n) - \mathcal{C}(\boldsymbol{x}^\star)$. □

The convergence improvement over FISTA resides in the constant. Practically, we expect a significant speedup in the first iterations when the matrix $\boldsymbol{\Lambda_\tau}$ can be chosen such that $\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_{\boldsymbol{\Lambda_\tau}^{-1}}^2 \ll \frac{L}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$.

We note that FWISTA can be simply adapted in order to impose a monotonic decrease of the cost functional value, in the same fashion as MFISTA proposed by Beck and Teboulle, while preserving an identical convergence property.

2.2. Example of speedup analysis for MRI

For illustration, let us consider a simple MRI problem with one receiving channel and a full k-space Cartesian

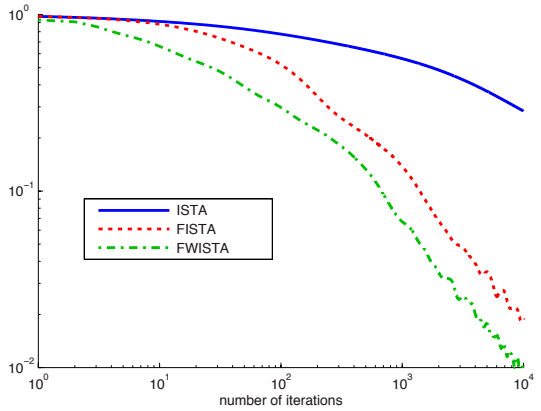

Fig. 1. PMRI simulation. $\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2$.

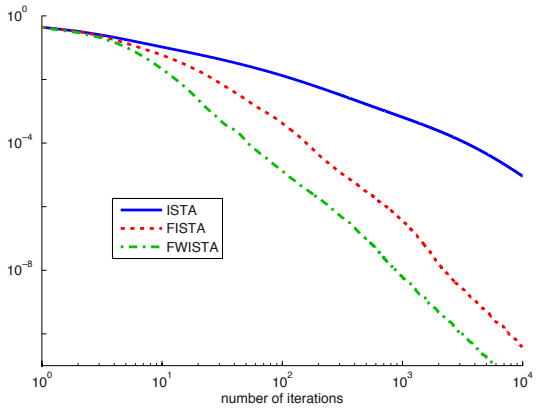

Fig. 2. PMRI simulation. $\mathcal{C}(\boldsymbol{x}_k) - \mathcal{C}(\boldsymbol{x}^\star)$.

sampling. In such a case, we get $\boldsymbol{H} = \boldsymbol{F}\boldsymbol{\Lambda}_{\boldsymbol{s}}$ where $\boldsymbol{F}$ is the unitary Fourier matrix and where $\boldsymbol{\Lambda}_{\boldsymbol{s}}$ is the diagonal matrix of coil sensitivities.

Applying FISTA ($\boldsymbol{\Lambda}_{\boldsymbol{\lambda}} = \lambda \boldsymbol{I}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\tau}} = \|\boldsymbol{s}\|_\infty^{-2} \boldsymbol{I}$) results in a gradient step of the form $\boldsymbol{x}_n + \|\boldsymbol{s}\|_\infty^{-2} \boldsymbol{\Lambda}_{\boldsymbol{s}}^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{s}} (\boldsymbol{x}_{\mathrm{orig}} - \boldsymbol{x}_n) + \|\boldsymbol{s}\|_\infty^{-2} \boldsymbol{\Lambda}_{\boldsymbol{s}}^{\mathrm{T}} \boldsymbol{F}^{\mathrm{T}} \boldsymbol{b}$. The latter tends to be ineffective for the pixels whose locations correspond to a relatively small sensitivities, this fact being measured by the condition number of $\boldsymbol{\Lambda}_{\boldsymbol{s}}$; i.e., the product $\|\boldsymbol{s}\|_\infty \|1/\boldsymbol{s}\|_\infty$. In regard to the convergence result in Proposition 1, we see that the constant term of the upper bound for FISTA is proportional to $\|\boldsymbol{s}\|_\infty^2$ as it amounts to $\|\boldsymbol{s}\|_\infty^2 \sum_i |(x_0)_i - x_i^\star|^2$.

FWISTA, by contrast, can take advantage of step sizes (gradient and threshold) that are tailored to the individual pixel sensitivities. In our simple example, we apply FWISTA with $\boldsymbol{\Lambda}_{\boldsymbol{\lambda}} = \lambda \boldsymbol{I}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}^{-1} = \tau \boldsymbol{\Lambda}_{\boldsymbol{s}}^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{s}}$. This results into the simplified gradient step $\boldsymbol{x}_n + \tau(\boldsymbol{x}_{\mathrm{orig}} - \boldsymbol{x}_n) + \boldsymbol{F}^{\mathrm{T}} \boldsymbol{b}$. For the convergence bound, we achieve a constant that is $\sum_i |s_i|^2 |(x_0)_i - x_i^\star|^2$ and tends be significantly smaller than the one obtained with FISTA, especially if the range of s_i is large.

3. EXPERIMENTS

To assess the performance of FWISTA in practical situations, we ran simulations for two biomedical imaging modalities. The two related inverse problems are both notorious for their ill-conditioning.

3.1. Parallel MRI

We considered a 2-D brain imaging setup. The data are recorded by four head coils with known sensitivity maps distributed around the sample. Meanwhile, a radial k-space trajectory with 90 lines that supports a 176×176 reconstruction is imposed. Our simulation is achieved with a 704×704 rasterized version of the Shepp-Logan phantom. We are also using realistic coil sensitivities computed using Biot-Savart's law. The 2-levels Haar wavelet basis is chosen to impose the sparsity constraints. The sum-of-squares sensitivity is considered for the weights of FWISTA. We exploit the localization properties of the wavelets to impose the weights on the wavelet coefficients. The wavelet regularization matrix $\boldsymbol{\Lambda}_{\boldsymbol{\lambda}}$ is chosen with a constant value λ on its diagonal except for the entries concerning coarse coefficients where the value 0 was imposed. The value λ is optimized for MSE performance. The initial estimate is $\boldsymbol{x}_0 = \boldsymbol{0}$.

We compare the convergence speed of ISTA, FISTA, and FWISTA in the reconstruction task described above. The reference minimizer was determined by running FISTA for 100 000 iterations. Results are shown on Fig. 1 and Fig. 2. We observe that FWISTA yields nearly three-fold acceleration over FISTA in terms of cost functional. The asymptotic rates of FISTA and FWISTA both on the cost functional value and the distance to the minimizer are similar and the speedup of FWISTA seems primarily due to a better constant, which is consistent with our theoretical prediction.

3.2. Bioluminescence Tomography

In bioluminescence tomography (BLT), the spatial distribution of a luminescing dye buried inside tissue is reconstructed from boundary light measurements [6, 7]. Due to the high scattering and absorption characterizing living tissue, the propagation of light in tissue is well described by a diffusion process. The BLT imaging kernel therefore contains a smoothing component, which leads to a severely ill-conditioned inverse problem. In 3-D, the free-space Green's function of the propagation equation is:

$$G(\boldsymbol{r}, \boldsymbol{r}_0) = \frac{\exp\left(-\frac{\mu_a \|\boldsymbol{r} - \boldsymbol{r}_0\|}{D}\right)}{\|\boldsymbol{r} - \boldsymbol{r}_0\|},$$

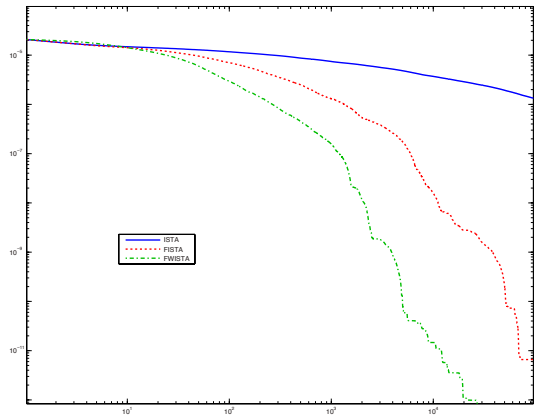

Fig. 3. BLT simulation. $\mathcal{C}(\boldsymbol{x}_k) - \mathcal{C}(\boldsymbol{x}^\star)$.

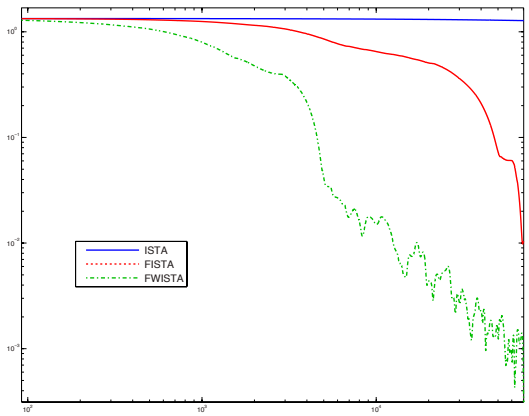

Fig. 4. BLT simulation. $\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2$.

where μ_a is the absorption coefficient, and D the diffusion coefficient of the considered medium. The ill-posedness results from the rapid decay of the Green's function which causes the detectors to have highly inhomogeneous sensitivities. In this example, we employ the sensitivities of the detectors to build a problem-specific diagonal approximation $\boldsymbol{\Lambda}_{\boldsymbol{\tau}}$ of $\mathbf{H}^T\mathbf{H}$, where $\mathbf{H}$ denotes the BLT system matrix. The entry $(\boldsymbol{\Lambda}_{\boldsymbol{\tau}})_{ii}$ corresponding to pixel i, is the sum of the sensitivities for that pixel squared.

We simulate a 2-D BLT system composed of a 50mm $\times$ 20mm slab, with 28 detectors disposed on the periphery. The optical coefficients are $\mu_s = 1.5\text{mm}^{-1}$ and $\mu_a = 0.02\text{mm}^{-1}$, which corresponds to realistic tissue values. The reconstruction grid is 50×20, and therefore we have a highly underdetermined system with a 28×1000 system matrix. Two point sources are placed inside the medium. The sensitivities of the detectors are again used in the regularization matrix $\boldsymbol{\Lambda}_{\boldsymbol{\lambda}}$ by posing $(\boldsymbol{\Lambda}_{\boldsymbol{\lambda}})_{ii}$ equal to the sum of the sensitivities. Note the double role of the sensitivities in this example. They are employed to determine the FWISTA steps and thresholds, as well as penalty in the regularization.

Convergence speed of ISTA, FISTA and FWISTA on the BLT problem is illustrated on Fig. 3 and Fig. 4. Again, we observe that the asymptotic convergence rate of FISTA and FWISTA are comparable while FWISTA appears to exhibit a much better constant. On this case, we observe a ten-fold acceleration.

4. CONCLUSIONS

Modern sparsity-based image reconstructions tend to yield better image quality compared to linear solvers but they are typically much slower. Our study suggests that there is still potential for speed improvement by taking into account the problem's specific structure. The method proposed here adapts FISTA and tailors it to two biomedical imaging modalities. The simulations of pMRI and BLT setups resulted in promising speedups in reconstruction. Further works should validate the acceleration capabilities of the method with experimental data.

5. REFERENCES

[1] A. Ribés and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 84–99, July 2008.

[2] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[3] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Signal Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[4] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[5] C. Vonesch and M. Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 509–523, 2009.

[6] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: the evolution of whole-body photonic imaging," *Nat Biotech*, vol. 23, pp. 313–320, Mar 2005.

[7] W. Cong, G. Wang, D. Kumar, Y. Liu, M. Jiang, L. Wang, E. Hoffman, G. McLennan, P. McCray, J. Zabner, and A. Cong, "Practical reconstruction method for bioluminescence tomography," *Opt. Express*, vol. 13, no. 18, pp. 6756–6771, 2005.