

# ESTIMATION OF THE BLURRING KERNEL IN EXPERIMENTAL HR-pQCT IMAGES BASED ON MUTUAL INFORMATION

Y.Li<sup>\*</sup>, B.Sixou<sup>\*</sup>, F.Peyrin<sup>\*†</sup>

<sup>\*</sup>Univ.Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS  
UMR 5220, U1206, F-69621, LYON, France

<sup>†</sup> ESRF, 6 rue Jules Horowitz, F-38043, Grenoble Cedex France

## ABSTRACT

The analysis of trabecular bone micro structure from *in-vivo* CT images is still limited due to limited spatial resolution even with the new High Resolution peripheral Quantitative CT (HR-pQCT) scanners. In previous works, it has been proposed to exploit super resolution techniques to improve spatial resolution. However, the application of such methods requires to know the blurring kernel, which is challenging for experimental HR-pQCT images. The goal of this work is to determine the blurring kernel of these scanners in order to facilitate an increase of the resolution of the bone images and of the segmentation of the bone structures. To this aim, we propose a method based on mutual information and compare it with classical  $L_2$ -norm minimization methods.

**Index Terms**— Deconvolution, super-resolution, Total Variation, 3D CT images, bone micro-architecture

## 1. INTRODUCTION

X-ray CT techniques are well adapted to image bone structure at the organ scale. However, imaging bone micro architecture *in vivo* remains challenging due to the need of high spatial resolution (order of  $100\mu\text{m}$ ) while keeping the X-ray dose as low as possible. High Resolution peripheral Quantitative CT (HR-pQCT) has been developed to reach these requirements. HR-pQCT devices providing images with spatial resolution of about  $100\text{-}150\mu\text{m}$  have been installed for research purpose at different hospitals [1]. However, while the images enables 3D rendering of trabecular bone, the extraction of morphometric and topological quantitative parameters remains somehow limited. This is related to the segmentation which is an issue for structures that are of the same order of size as the spatial resolution. To improve spatial resolution, we proposed to investigate super resolution techniques based on a single image and prior. In a previous work, we have showed that Total Variation (TV) regularization allowed to improve image

quality based on artificially degraded images [2, 3]. The assumption is that the image degradation can be modelled by a linear spatial invariant system characterized by a point spread function or a blurring kernel. However, when the method has to be applied to experimental HR-pQCT images, the kernel is unknown and has to be determined. Since the accuracy of the kernel has a strong impact on the results, the blurring kernel involved in the image formation process must be estimated with care.

The aim of this work is to estimate the degradation kernel of HR-pQCT for given high resolution and low resolution images. The origins of blur are various and various methods have been developed to estimate blur kernels [4–6]. The blur determination can be regularized with a finite dimensional approximation [7] or patch-based priors [8]. Some specific methods for motion blurred images have been studied in [9, 10]. The more classical method to obtain the kernel is to minimize a  $L_2$  distance.

We propose to use a regularization term based on Mutual Information (MI) with a gradient descent. The kernels obtained with this approach and  $L_2$  regularization are then compared with TV regularization applied to real data. They are used to solve the super-resolution problem in order to improve the *in-vivo* HR-pQCT image resolution and evaluated with respect to quantitative parameters of the bone micro-structure.

This paper is organized as follows. In section 2, we present the image formation model and the regularization functional used for the kernel estimation. The optimization methods are detailed in section 3. In section 4, the numerical results are detailed. Concluding remarks are given at the end of the article.

## 2. THE INVERSE PROBLEM OF THE BLURRING KERNEL ESTIMATION

### 2.1. The image problem formulation

In order to estimate the degradation kernel of HR-pQCT images, we consider two experimental CT images of a given sample, a high resolution  $\mu\text{CT}$  image considered as the ground truth and a low resolution noisy HR-pQCT image.

This work is supported by China Scholarship Council (CSC) and LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

The reconstruction of a 3D image with an improved resolution from a single low-resolution image can be expressed as follows. We assume that the low and high resolution images are defined on a bounded region  $\Omega$ . The low-resolution image is obtained from a high-resolution image with blur, down sampling and noise, which can be written as:

$$g = U(h * (f + c)) + n \quad (1)$$

where  $g \in \mathbb{R}^N$  denotes the  $N$ -voxels 3D low-resolution noisy image, and  $f \in \mathbb{R}^{N'}$  denotes an  $N' = N \times p^3$ -voxels high-resolution image with super-resolution factor  $p = 2$  in each dimension,  $n$  is the additive noise component,  $U$  is the undersampling operator and  $h$  the convolution kernel accounting for the blurring of images. We also introduce a constant  $c$  corresponding to a shift between the grey levels of the images  $f$  and  $g$ . Due to the use of different experimental devices, recovering the  $\mu$ CT image  $f$  from the given HR-pQCT degraded image  $g$  is an ill-posed problem and must be regularized.

## 2.2. $L_2$ -norm minimization

The determination of the blurring kernel  $h$  from a couple of high and low resolution images can be considered as a linear inverse problem. The usual approach can be formulated as:

$$\hat{h} \in \arg \min_h J_1(h, c) = \|U(h * (f + c)) - g\|_2^2 + \alpha R(h) \quad (2)$$

where  $\alpha$  is a regularization parameter and  $R$  a regularization functional [11]. The regularization term  $R(h) = \|\nabla h\|_2^2$  is chosen to penalize the fast variations of the kernel.

The search for the minimizer of the regularization functionals are done using a gradient descent strategy. For both regularization functionals we use an alternate minimization method with respect to  $h$  and  $c$ . The convolution operator  $h \rightarrow (f + c) * h$  is a linear convolution operator. The gradient of the functional  $J_1$  with respect to  $h$  and  $c$  are:

$$\begin{aligned} \frac{\partial J_1(c, h)}{\partial h} &= (f + c) * U((f + c) * h) - (f + c) * g - \alpha \Delta h \\ \frac{\partial J_1(c, h)}{\partial c} &= 2c \|U(E * h)\|^2 + 2\langle U(h * f) - g, U(E * h) \rangle \end{aligned}$$

where  $\langle \cdot, \cdot \rangle$  is the scalar product associated to  $L_2(\Omega)$ , the space of the square integrable functions on  $\Omega$ .  $E$  is a matrix whose values are all equal to 1.

## 2.3. Mutual Information based approach

### 2.3.1. Why Mutual Information?

Other statistical dissimilarity measures [12] may be more efficient than the  $L_2$  norm for enhancing the similarity between the histogram distribution of  $U(h * (f + c))$  and  $g$ . To this aim, we propose to exploit the MI between  $U(h * (f + c))$

and  $g$ . The estimated kernel  $h$  will then be obtained with the minimization of the regularization functional  $J_2(h, c)$ :

$$J_2(h, c) = -MI(U(h * (f + c)), g) + \alpha \|\nabla h\|^2 \quad (3)$$

The MI between the two variables  $U(f * h)$  and  $g$  is based on their joint probability function. It gives an estimate of their statistical dependence and it has been used for multi-modal image matching [12]. It is defined by:

$$\begin{aligned} MI(U((f + c) * h), g) &= \sum_{i,j} P_{conj}(i, j) \cdot \\ &\log \frac{P_{conj}(i, j)}{P_{U(h*(f+c))}(i)P_g(j)} \end{aligned} \quad (4)$$

where  $P_{conj}(i, j)$  is the conjugate probability distribution of  $U(h * (f + c))$  and  $g$  for intensity values  $i$  and  $j$ , and  $P_{U(h*(f+c))}(i)$ ,  $P_g(j)$  are the marginal probability distribution of  $U(h * (f + c))$  and  $g$ . In order to evaluate the MI, we consider a Parzen estimator [13, 14] for the joint probability function with a Gaussian of variance  $\beta$ , and a normalization constant  $|\Omega|$ :

$$\begin{aligned} P_{conj}(i, j) &= \frac{1}{|\Omega|} \int_{\Omega} \exp\left(-\frac{(U(h * (f(x) + c)) - i)^2}{\beta^2}\right) \\ &\cdot \exp\left(-\frac{(g(x) - j)^2}{\beta^2}\right) dx \end{aligned}$$

The marginal probability distributions can be obtained as:

$$P_{U(h*(f+c))}(i) = \sum_j P_{conj}(i, j), \quad P_g(j) = \sum_i P_{conj}(i, j)$$

### 2.3.2. Minimization algorithm

Similarly to the  $L_2$ -norm minimization method, the minimization of  $J_2$  is implemented by a gradient descent strategy and an alternate minimization.

The MI is a nonlinear functional of the kernel  $h$ . Given an initial estimate  $h_0$ , denote  $\text{NMI} = -MI$ , the minimization of the kernel with gradient descent of the MI is equivalent to the solution of the initial value problem:

$$\begin{cases} \frac{dh}{dt} = -\frac{\partial J_2(h, c)}{\partial h} = -\left(\frac{\partial \text{NMI}(c, h)}{\partial h} - \alpha \Delta h\right) \\ h(0) = h_0 \end{cases} \quad (5)$$

where  $\frac{\partial J_2(h, c)}{\partial h}$  is the gradient of  $J_2$  with respect to  $h$ . We will admit in this work that this first order differential equation is well-defined.

The first variation of the MI at  $h$  in the direction of  $k$  is defined by:

$$\delta_k \text{NMI}(h) = \left. \frac{\partial \text{NMI}(h + \epsilon k)}{\partial \epsilon} \right|_{\epsilon=0} \quad (6)$$

and the gradient is obtained with:

$$\delta_k \text{NMI}(h) = \left\langle \frac{\partial \text{NMI}(h, c)}{\partial h}, k \right\rangle \quad (7)$$

Considering the fact that  $P_{conj}$  is a probability distribution, we have [12]:

$$\delta_k \text{NMI}(k) = - \sum_{i,j} \delta_k P_{conj(i,j)} \ln \frac{P_{conj(i,j)}}{P_{U(h*(f+c))}(i)P_g(j)} \quad (8)$$

where

$$\delta_k P_{conj} = - \frac{2}{\beta^2} \int_{\Omega} \{U(k * (f(x) + c))(U(h * (f(x) + c)) - i) \exp(-\frac{(U(h * (f(x) + c)) - i)^2 + (g(x) - j)^2}{\beta^2})\} dx \quad (9)$$

Based on this formula, we can show that the gradient of the NMI (negative MI) with respect to  $h$  is given by:

$$\frac{\partial \text{NMI}(c, h)}{\partial h} = 2(f + c) * \sum_{i,j} \{(U(h * (f(x) + c)) - i) \cdot \exp(-\frac{(U(h * (f(x) + c)) - i)^2 + (g(x) - j)^2}{\beta^2}) \cdot \ln \frac{P_{conj(i,j)}}{P_{U(h*(f+c))}(i)P_g(j)}\} \quad (10)$$

On the other hand, the gradient of the negative MI with respect to the constant  $c$  is given by:

$$\frac{\partial \text{NMI}(c, h)}{\partial c} = \frac{\int \int \int M(i, j, x) \frac{2U(h * E) \cdot (U(h * (f + c)) - i)}{-\beta^2} didjdx}{\int \int \int M(i, j, x) \frac{(U(h * E))^2}{-\beta^2} didjdx} \quad (11)$$

with

$$M(i, j, x) = \exp(-\frac{(U(h * (f + c)) - i)^2 + (g - j)^2}{\beta^2}) \cdot \log \frac{P_{conj(i, j)}}{P_{U(h*(f+c))}(i)P_g(j)} \quad (12)$$

### 3. NUMERICAL EXPERIMENTS

#### 3.1. Simulation details

The tests for the determination of the blurring kernel were performed on 3D experimental images of trabecular bone obtained from HR-pQCT [1] and  $\mu$ CT [15]. The HR-pQCT image is considered as the low resolution image with a resolution  $82\mu m$ ; the image from  $\mu$ CT is the high resolution image,

with a resolution equal to  $41\mu m$  (registered from  $24\mu m$ ). The high and low resolution images were respectively cropped at  $N^3 = 200 \times 200 \times 200$  voxels and  $N = 100 \times 100 \times 100$  voxels. The initial kernel  $h_0$  is chosen as a Gaussian with a standard deviation of 4. Several support sizes for the kernel have been investigated with odd number ranging from 5 to 13. Different regularization parameters have also been tested. The kernels obtained with two regularization functionals based on the  $L_2$  norm or MI are used to solve a super-resolution problem with Total Variation regularization. The minimization method is based on the Alternated Direction Method of Multipliers (ADMM) [3]. Taking account of smoothing constraints, the augmented Lagrangian considered is:

$$\mathcal{L}_A = \frac{\mu}{2} \|U(h * (f + c)) - g\|^2 + \sum_i \{\|v_i\| + \frac{\beta}{2} \|v_i - D_i * f\|^2 - \lambda_i^t (v_i - D_i * f)\}$$

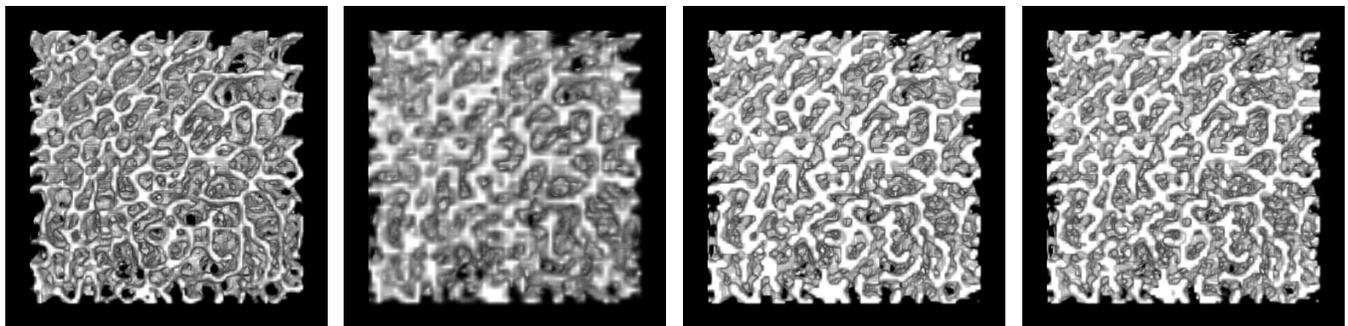
$\lambda_i$  is Lagrangian multipliers for TV,  $\beta$  a Lagrangian parameter and  $D_i$  is the gradient operator at the voxel  $i$ . The saddle point of the Lagrangian is estimated with successive updates.

In TV super resolution restoration, parameters are chosen to obtain the best decrease of the regularization functional,  $\mu = 10$ ,  $\beta = 7$  for both kernels. Iterations are stopped when  $\frac{\|f^{n+1} - f^n\|_2}{\|f^{n+1}\|_2} \leq 10^{-3}$ . In order to evaluate the reconstruction, we have calculated the Peak Signal-to-Noise Ratio (PSNR)  $-10 \log \frac{(f - f^n)^2}{(\max(f) - \min(f))^2}$ , where  $f$  is the high resolution image,  $f^n$  is a super-resolution image,  $(f - f^n)^2$  the mean square error. The Otsu segmentation method [16] is applied to calculate DICE ( $\frac{2A \cap B}{A + B}$ ,  $A$  and  $B$  are binary images so that the rate of overlaps of 2 images is evaluated). We have also estimated the BV/TV (bone volume/total volume) and bone connectivity density  $\frac{\beta_1}{TV}$ , where  $\beta_1$  is the first order Betti number evaluated from the Euler number and  $TV$  the total volume [17, 18].

#### 3.2. Numerical results

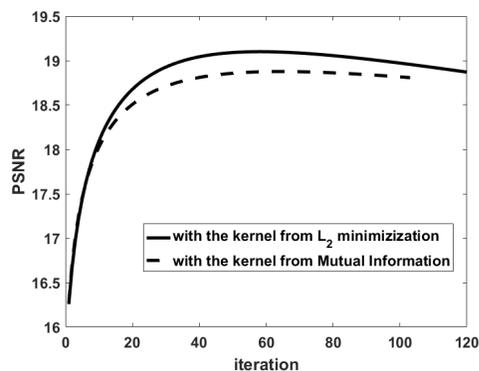
The images obtained by TV regularization with these two optimized kernels are displayed in Figure 1 together with the low and high resolution images. Figure 1(c)(d) display more structural details compared to the low resolution image (b), but the reconstructed structures are thick regarding to the high resolution image (a).

The evolutions of PSNR, DICE, BV/TV, connectivity density (defined in 3.1) as well as data fitting term as a function of the number of iterations are illustrated in Fig 2-6. PSNR reflects the similarity of two images in grey level, while DICE, BV/TV and connectivity density compare binary images. As expected, we obtain a decrease of the reconstruction errors evaluated with PSNR or DICE with the two methods. The kernel obtained with  $L_2$  minimization performs slightly

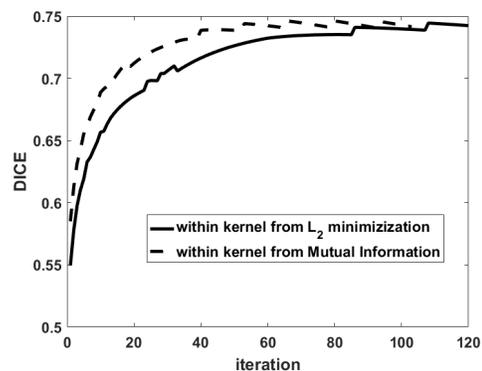


(a) binarized high resolution image (b) binarized low resolution image (c) binarized super resolution image with the kernel determined by  $L_2$ -norm (d) binarized super resolution image with the kernel determined by MI

**Fig. 1:** illustration of binarized high resolution, low resolution,  $L_2$  kernel restored image and MI kernel restored image.



**Fig. 2:** Evolution of PSNR as a function of iteration for the two kernels.



**Fig. 3:** Evolution of DICE as a function of iteration for the two kernels.

better in term of PSNR, whereas the one derived from MI is better regarding the DICE.

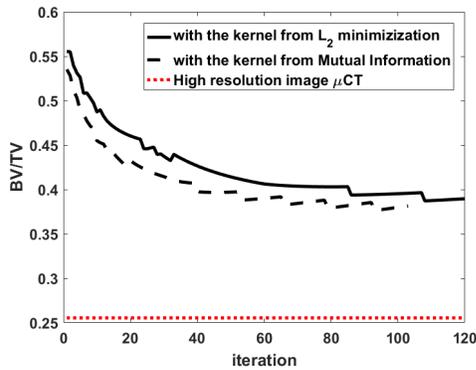
The obtained super-resolution images have thinner structures than the low resolution image, but the reconstructed structures are still too thick. This is also illustrated in Figure 4 where the red short dot line represents the BV/TV of the high resolution image. Both methods improves image quality with respect to BV/TV, but this ratio is still too high in the reconstructed images. Connectivity density is a topological criterion. Even though it has been improved as displayed in Figure 5, the final divergence tail shows that the determined kernels did not well recover the degradation of the topological structure. In addition, the kernel from  $L_2$  minimization changes the topological structure of the super-resolution image more slowly than the one from MI optimization. Figure 6 illustrates the data term evolution. This figure also shows the convergence of the TV reconstructions with the two optimized kernels. In general, the kernel deduced from  $L_2$  norm is more efficient than MI considering connectivity density and local artifacts, whereas the kernel from MI is slightly better regarding BV/TV and data fitting term.

#### 4. CONCLUSION

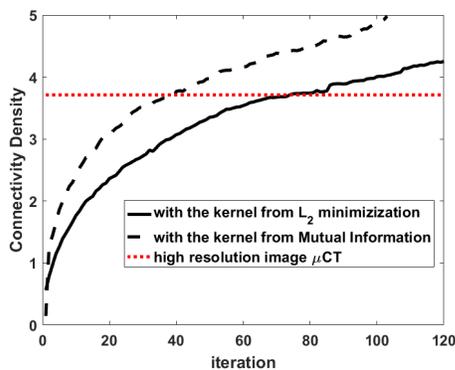
The determination of the kernel is a major issue in image restoration. In this paper, we have compared two methods of determination of the blurring kernel based on the optimization of MI or  $L_2$  norm. Similar results are obtained by both kernels derived from  $L_2$  norm or MI. The kernel from  $L_2$  norm preserves a better connectivity density, and the one from MI performs well in BV/TV. One possible improvement for our MI gradient descent method is to consider local information [12]. Hermosillo et al. proposed a local MI method, which will generate a spatially variant kernel that may improve the results. These results need to be further confirmed on other ROIs and other HR-pQCT images.

#### 5. REFERENCES

- [1] Stephanie Boutroy, Mary L Bouxsein, Françoise Munoz, and Pierre D Delmas, "In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography," *The Journal of Clinical Endocrinology & Metabolism*, vol. 90, no. 12, pp. 6508–6515, 2005.
- [2] Alina Toma, Loïc Denis, Bruno Sixou, Jean-Baptiste Pialat, and Françoise Peyrin, "Total variation super-resolution for 3D trabecular bone micro-structure segmentation," in *2014 22nd*



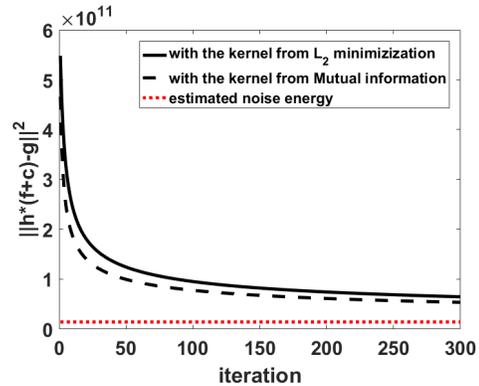
**Fig. 4:** Evolution of BV/TV as a function of iteration for the two kernels.



**Fig. 5:** Evolution of connectivity density as a function of iteration for the two kernels.

*European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 2220–2224.

- [3] Alina Toma, Bruno Sixou, and Françoise Peyrin, “Iterative choice of the optimal regularization parameter in TV image restoration,” *Inverse Problems & Imaging*, vol. 9, no. 4, 2015.
- [4] Neel Joshi, Richard Szeliski, and David J Kriegman, “PSF estimation using sharp edge prediction,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
- [5] Stanley J Reeves and Russell M Mersereau, “Blur identification by the method of generalized cross-validation,” *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 301–311, 1992.
- [6] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman, “Removing camera shake from a single photograph,” in *ACM Transactions on Graphics (TOG)*. ACM, 2006, vol. 25, pp. 787–794.
- [7] Loïc Denis, Eric Thiébaud, Ferréol Soulez, Jean-Marie Becker, and Rahul Mourya, “Fast approximations of shift-variant blur,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 253–278, 2015.
- [8] Libin Sun, Sunghyun Cho, Jue Wang, and James Hays, “Edge-based blur kernel estimation using patch priors,” in *Computational Photography (ICCP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–8.
- [9] Michael Cannon, “Blind deconvolution of spatially invariant image blurs with phase,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 1, pp. 58–63, 1976.
- [10] Michael M Chang, A Murat Tekalp, and A Tanju Erdem, “Blur identification using the bispectrum,” *IEEE transactions on signal processing*, vol. 39, no. 10, pp. 2323–2325, 1991.
- [11] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen, *Variational methods in imaging*, vol. 320, Springer, 2009.
- [12] Gerardo Hermosillo, Christophe Chefd’Hotel, and Olivier Faugeras, “Variational methods for multimodal image matching,” *International Journal of Computer Vision*, vol. 50, no. 3, pp. 329–343, 2002.
- [13] Emanuel Parzen, “On estimation of a probability density function and mode,” *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [14] Denis Bosq, *Nonparametric statistics for stochastic processes: estimation and prediction*, vol. 110, Springer Science & Business Media, 2012.
- [15] Andrew J Burghardt, Thomas M Link, and Sharmila Majumdar, “High-resolution computed tomography for clinical imaging of bone microarchitecture,” *Clinical Orthopaedics and Related Research®*, vol. 469, no. 8, pp. 2179–2193, 2011.
- [16] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285–296, pp. 23–27, 1975.
- [17] Anders Odgaard, “Three-dimensional methods for quantification of cancellous bone architecture,” *Bone*, vol. 20, no. 4, pp. 315–328, 1997.
- [18] Joachim Ohser, Werner Nagel, and Katja Schladitz, “Miles formulae for boolean models observed on lattices,” *Image Analysis & Stereology*, vol. 28, no. 2, pp. 77–92, 2011.



**Fig. 6:** Evolution of data fitting term as a function of iteration. In order to show the whole evolution of the data fitting term, this figure didn’t stop iteration by relative changes, which is  $\|f^{n+1} - f^n\|_2 / \|f^{n+1}\|_2 \leq 10^{-3}$ .