

Denoising in Fluorescence Microscopy Using Compressed Sensing With Multiple Reconstructions and Non-Local Merging

Marcio Marim, Elsa Angelini and Jean-Christophe Olivo-Marin

Abstract—The cross-dependency of noise level and photobleaching in microscopy was discussed in a previous work and an efficient compressed sensing (CS) method was proposed to simultaneously reduce the noise level and the photobleaching. Here we present an improved CS denoising framework for fluorescence microscopy images, exploiting Non-Local means filtering to merge multiple reconstructions. This framework enables high-quality reconstruction of low exposed microscopy images based on random Fourier sampling schemes and multiple CS reconstructions. Practical experiments on fluorescence images demonstrate that even performing 10% of the measurements, the signal-to-noise ratio can be significantly improved while keeping reduced exposure time, preserving edges and the image sharpness.

Index Terms—Compressed sensing, denoising, fluorescence microscopy, NL-means filtering.

I. INTRODUCTION

In biological fluorescence microscopy, cellular components of interest in the specimen such as proteins are typically labeled with a fluorescent molecule called a fluorophore (green fluorescent protein (GFP), dyes, inorganic molecules) and can therefore be imaged with high specificity. Fluorophores lose their ability to fluoresce as they are illuminated through a process called photobleaching [1]. In microscopy, observation of fluorescent molecules is challenged by the photobleaching, as these molecules are slowly destroyed by the light exposure necessary to stimulate them into fluorescence. Loss of emission activity caused by photobleaching can be controlled by reducing the intensity or time-span of light exposure. Unfortunately, reducing the exposure time or intensity of the excitation also reduces the emission intensity but not the noise acquisition components, leading to a decrease of the signal-to-noise ratio SNR.

A host of denoising methods, well suited for piecewise smooth images, have been developed such as Non-Local Means (NL-means) [2], Total Variation Filtering (TV) [3], non-linear isotropic and anisotropic diffusion [4]. There are also methods which exploit the decomposition of the data onto wavelets, ridglets or curvelets basis functions and shrink the coefficients to eliminate noise components [5], [6]. Recently, efficient denoising methods were also developed based on sparsity and redundant representations over learned dictionaries [7], denoising image while simultaneously training a dictionary using the K-SVD algorithm [8], or based on sparse code shrinkage and maximum likelihood estimation of nongaussian variables [9]. However, most are post-processing techniques and require full data acquisition.

In this paper we propose an improved technique, based on compressed sensing (CS) that simultaneously enables reduction of exposure time or excitation light level and improvement of image SNR. Our CS-based method can simultaneously acquire and denoise data, based on random Fourier sampling, noise reconstruction

characteristics and signal modeling applied to microscopy images with low SNR. Here, we propose a non-local merging scheme to combine the multiple CS reconstructions.

Recent results from the field of compressed sensing (CS) [10], [11] have demonstrated that it is possible to reconstruct a compressible signal (redundant or sparse in some domain) from a small set of measurements. The main idea is that if the signal of interest $x \in \mathbb{R}^N$ has a S -sparse representation Ψx (i.e. S non-zero coefficients) for a particular spatial transform Ψ , then x can be reconstructed from $M \ll N$ indirect linear projections $y = \Phi x$ onto a more general basis, such as the Fourier domain. The almost exact solution \hat{x} to this problem is achieved by solving the convex problem:

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|\Psi x\|_{\ell_1} \quad \text{subject to} \quad Ax = y \quad (1)$$

where Ψx is a sparsifying transform $|\Psi x| \ll |x|$, with $|\cdot|$ corresponding to the cardinality. Indeed, the number of projections needed to solve this problem is approximately proportional to the sparsity $S = \|\Psi x\|_{\ell_0}$ and $\|\Psi x\|_{\ell_1} \approx \|\Psi x\|_{\ell_0}$. Considering noisy measurements $y = \Phi x + n$, a relaxed version of the problem (1) was also introduced in [12],

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|\Psi x\|_{\ell_1} \quad \text{s.t.} \quad \|y - \Phi x\|_{\ell_2} \leq \delta \quad (2)$$

for some δ tuned to match the noise level contaminating the observations x . In this case, the solution \hat{x} is an estimation of x and the noise discrimination will basically depend on the sampling scheme and the sparsity prior.

A multiple CS reconstruction was introduced in a previous work [13], allowing to simultaneously reduce the noise level and photobleaching. Averaging multiple CS reconstructions with uncorrelated noise-based patterns enables to generate a high-quality denoised image. In the previous work, the combination was performed via simple averaging. Here an improved CS-based denoising framework for microscopy images is proposed. The improvement results from the optimization of the combination process, we investigate the use of NL means filtering, better exploiting the high level of redundancy within the spatial domain and across the set of CS-reconstructed images. The paper is organized as follows: in Section II we recall the general CS reconstruction framework for noisy measurements. In Section III we recall our previous framework. In Section IV we introduce an improved CS-based denoising framework and in Section V we present experimental results and comparison.

II. COMPRESSED SENSING BACKGROUND

Compressed sensing offers a theoretical framework to remove non-sparse random noise components from a corrupted signal by optimally reconstructing a signal with explicit sparsity constraints [14]. Indeed, the performance on removing noise from $x + n$ will rely on the efficacy of Ψ at representing the signal x sparsely and the inefficacy at representing the noise n . Considering that the sparsity measure is the Total Variation (TV) $\|x\|_{TV} := \|\nabla x\|_{\ell_1}$, then this corresponds to assuming that ∇x is very sparse and

This work was funded by Institut Pasteur and DGA.

M. Marim and J.-C. Olivo-Marin are with Institut Pasteur, Unité d'Analyse d'Images Quantitative CNRS URA 2582, F-75015 Paris. marim@pasteur.fr, jcolivo@pasteur.fr, www.bioimageanalysis.org

E. Angelini is with Institut Télécom, Télécom ParisTech CNRS LTCl, F-75013 Paris. elsa.angelini@telecom-paristech.fr

that ∇n is non-sparse. In terms of measurements, we will have $\|\nabla(x+n)\|_{\ell_1} \gg \|\nabla x\|_{\ell_1}$.

As demonstrated in [15], if such a sparsifying transform exists in the spatial domain, it is possible to reconstruct an image $x \in \mathbb{R}^N$ from partial knowledge of its Fourier spectrum $y = F|_{\Gamma}$. Now suppose a noise corrupted subset of Fourier coefficients $F|_{\Gamma} + n$. The image reconstruction consists in solving a convex optimization problem that finds the image candidate \hat{x} of minimal complexity satisfying $\|\hat{F}|_{\Gamma} - F|_{\Gamma}\|_{\ell_2} \leq \delta$, where $F|_{\Gamma} \subseteq F$ is a partial subset of Fourier coefficients in the set Γ . Considering that the noise energy has a known bounded norm $\|n\|_{\ell_2} \leq \epsilon$ the convex optimization corresponds to solve the following problem

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \|\nabla x\|_{\ell_1} \text{ s.t. } \|\hat{F}|_{\Gamma} - F|_{\Gamma}\|_{\ell_2} \leq \delta \quad (3)$$

where δ is related to the norm of the noise.

We note here that the TV-based spatial sparsity constraint, will lead to sharp edges and removal of noise components, resulting in an error:

$$\|\hat{x} - x\|_{\ell_2} \leq \alpha + \beta \quad (4)$$

where α reflects the desired error (responsible for noise removal) from the relaxation of the constrain δ in (3) and β reflects the undesired error from Fourier undersampling of the signal. If TV represents x efficiently and n inefficiently, the term β vanishes and $\alpha \rightarrow C\delta$.

III. OVERVIEW OF OUR PREVIOUS WORK

A CS-based denoising scheme was proposed in a previous work [13] allowing to limit the scanning time and to avoid photobleaching effects. The proposed method consists in acquiring a single image with a short exposition time. Then multiple sampling matrices Φ_i are created and each sampling matrix chooses a different and unique set of random frequencies in the Fourier domain. Using different random sampling matrices exploit the fact that the estimation of the fluorescence signal x should be strongly correlated for all sampling matrices Φ_i , while the noise estimation n should not. Then, we perform multiple CS reconstructions of a single noisy image acquisition $x+n$, using different sampling matrices Φ_i , solving the following problem:

$$\hat{x}_i = \arg \min_{x \in \mathbb{R}^N} \|\nabla x\|_{\ell_1} \text{ s.t. } \|y_i - \Phi_i x\|_{\ell_2} \leq \delta \quad (5)$$

for $i = 1 \dots K$ and $\Phi_i x = F|_{\Gamma_i}$. The value of K is an empirical estimation of the number of reconstructions for which the image SNR becomes constant.

The last step of the denoising algorithm involves the combination of the set of reconstructed images \hat{x}_i . In our initial work we proposed to average the set of x_i images to generate a single denoised image \hat{x} ,

$$\hat{x} = \frac{1}{K} \sum_{i=1}^K x_i \quad (6)$$

Each single reconstruction x_i is slightly different due to the differences in the sampling matrices Φ_i . Because of the random distribution of frequencies in Φ_i , the global frequency content of the x_i is similar and the images only differ in terms of details. Indeed, since $\|\nabla x\|_{\ell_1}$ is sparse and $\|\nabla n\|_{\ell_1}$ is non-sparse, noisy reconstructed patterns are uncorrelated for reconstructions using different set of measures. Nevertheless, averaging the set of images \hat{x}_i exploits the signal redundancy locally but it does not exploits the similarity of patches in the same image. In this work we propose a different combination of the set of reconstructed images, described

in Section IV, which exploits the non-local redundancy to achieve better denoising results.

IV. METHODS AND CONTRIBUTIONS

A. The Sampling Pattern for Denoising

In this section we deal with the Fourier sampling pattern adopted for the CS reconstruction. Since the goal is to perform an optimal denoising, the sampling pattern is of central importance. The design of the sampling matrix Φ can be based on different distributions, such as a independent and identically-distributed random distribution, or a line-based radial distribution, which are illustrated in Figure 1.

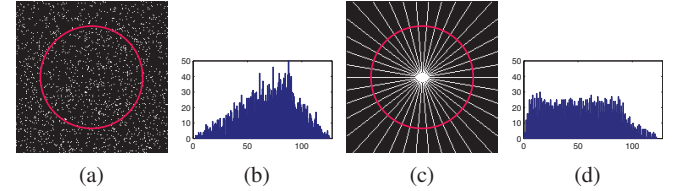


Fig. 1: Spatial distributions in the Fourier domain with $M = 0.1N$ (a) Random, (c) Radial and (b,d) corresponding frequency histogram.

The optimal sampling scheme would be the one which maximizes the incoherence between the sampling basis Φ and the sparsity basis Ψ . A fully random sampling often results in a high degree of incoherence and near-optimal solution. Indeed random distributions present some advantages such as simple mathematical proofs and satisfaction of the RIP conditions [10].

The Figure 1(b)-(d) displays the histograms of the spatial density of frequency measurements, characterized by the number of sampling points at a given distance from the $(0,0)$ center frequency in Figure 1(a)-(c). The histogram (b) shows that for random spatial distributions on a square, there are a majority of intermediate frequencies. On the other hand, the histogram (d) shows that for the radial distributions on a square, there are equal levels of low and intermediate frequencies while high frequencies have a minor density.

One interesting fact from Figure 1 is that when the sampling process is random, extreme high and low frequencies have the same low probabilities. These histograms can be modeled with the number of sample points contained on a circle of radius r ($0 \leq r \leq \sqrt{2}N/2$) and centered at $(0,0)$. We then distinguish two different behaviors for r below and above $N/2$, as observed in Figure 1(b)-(d). For a random distribution, we define a probability law of occurrences,

$$P(r) = \frac{(2\pi - 4\theta)r}{N^2} \quad (7)$$

where θ corresponds to the angle of one of the four external parts of the circle (illustrated in Figure 2). When the radius of the circle is smaller than $N/2$, the circle is completely inside the image, then $\theta = 0$ and $P(r) = \frac{2\pi r}{N^2}$, otherwise $\theta > 0$ and can be computed as:

$$\theta = 2 \arctan \left(\frac{\sqrt{r^2 - n^2}}{n} \right) \quad (8)$$

and r is the distance to the center of the image, N^2 is the total number of pixels and $n = N/2$. For $r \leq N/2$ the probability of all possible pixels increases linearly with r . For $r > N/2$ the probability starts decreasing since the perimeter is not completely inside the image, as illustrated in Figure 1 and 2. For the radial

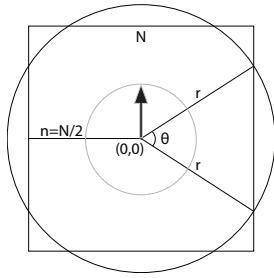


Fig. 2: Circle of probability.

distribution, the number of occurrences in a circle is constant for $r \leq N/2$ and depends on the number of lines n_l ,

$$P(r) = \frac{(2\pi - 4\theta) n_l}{2\pi N^2} \quad (9)$$

To perform our denoising we have the option to work with a set of random distributions or a set of rotated radial distributions. From (9) we can conclude that for a set of rotated radial distributions, low frequency samples are the same and the measurements are therefore redundant. Moreover, the small number of high frequency samples in the radial case constrains very loosely the CS reconstruction of high frequency details and noise. This does not mean that we want to reconstruct the noise but only that uncorrelated noise reconstruction is necessary to achieve a better denoising performance. Indeed, if sampling too few high-frequency components, there is a risk with CS to create pattern on the reconstructed image. Conversely, fully random sampling guarantees a better coverage of the high frequencies as illustrated in Figure 1(b). Therefore, random sampling combined with the known property of Non-Local means (NL-means) to better preserve high-frequency structures of an image [2], offers a very efficient framework for denoising.

B. The Improved Denoising Approach

In this subsection we describe the method based on the NL-means introduced above (section IV-A) to combine the set of images \hat{x}_i obtained in Equation (5). The main difference between this method and the original NL-means is that the noisy image is not a single image but a vector of similar images. Here we exploit the idea that images present a high degree of spatial redundancy in terms of objects appearance and shapes, and consequently, the set of images \hat{x}_i should present even more redundancy. As discussed before, the reconstructed images \hat{x}_i are slightly different, and most difference are high-frequency components.

The method does not make regularity assumptions on the original image, allowing noise reduction and preserving fine structures, details, and texture. The basic assumption is that small patches on objects from an image have many similar patches in the same image and averaging similar patches should correct the noisy component. The algorithm estimates the value of an image x as an average of the values of all the pixels whose Gaussian neighborhood looks like the neighborhood of x .

Consider a noisy image $v = \{v(i)|i \in \Omega\}$ defined on a discrete bounded domain $\Omega \subset \mathbb{R}^2$. Consider also a patch domain of fixed size $\mathcal{P}_i \subset \Omega$ where $v(\mathcal{P}_i)$ denotes a square patch centered at a pixel i . The estimated denoised image v_{NL} corresponds to a weighted average of all the pixels in the image:

$$v_{NL}(i) = \sum_{j \in \Omega} w(i, j) \cdot v(j) \quad (10)$$

where the weights $w(i, j)$ depend on the similarity between pixels

i and j and are defined as:

$$w(i, j) = \frac{1}{C(i)} e^{-\frac{\|v(\mathcal{P}_i) - v(\mathcal{P}_j)\|_2^2}{h^2}}, \quad (11)$$

where $\|v(\mathcal{P}_i) - v(\mathcal{P}_j)\|_2^2$ is the Euclidean distance between pixels from the patches $v(\mathcal{P}_i)$ and $v(\mathcal{P}_j)$ and $C(i)$ is a normalizing constant

$$C(i) = \sum_j e^{-\frac{\|v(\mathcal{P}_i) - v(\mathcal{P}_j)\|_2^2}{h^2}},$$

with two conditions (i) $0 \leq w(i, j) \leq 1$, (ii) $\sum_j w(i, j) = 1$.

However, the search of similar patches in Equation (11) is performed only for the first image \hat{x}_1 and used for the entire vector. Then, if the noisy image v is not a single image but a vector of similar images the final image is computed as a weighted average of all the pixels in the vector of images

$$\hat{x}_{NL}(i) = \frac{1}{K} \sum_{k=1}^K \sum_{j \in \Omega} w(i, j) \cdot \hat{x}_k(j), \quad (12)$$

with the condition $\frac{1}{K} \sum_{jk} w(i, j) = 1$. This method provides a much more discriminative combination of the set of images x_i than simple averaging, as initially proposed.

The simple averaging of \hat{x}_i in (6) exploit only the local redundancy from the set of images but not from the different patches of one single image. Solving our proposed method in (12) allows to exploit the redundancy locally beyond all images \hat{x}_i and also non-locally using similar patches from the same image.

V. RESULTS

We apply our approach to a fluorescence microscopy image of hair follicle. The specimen is labeled with a green and a red marker. The total acquisition time is 15 seconds per image and the 3-dimensional stack has 65 frames (each channel). The results illustrated here concern the frame 31 of each channel taken from the 3D volume. Then each channel (red and green) is reconstructed independently such as grayscale image.

In Figure 3 we show some reconstruction results of the noisy images from the green and red channels in Figures 3(a) and 3(f), with size 400×400 pixels. The Figures 3(b) and 3(g) correspond to one single CS reconstruction with 10% of measurements in the Fourier domain, such as described in (3). In Figures 3(c) and 3(h) we display the denoised images by our previous work, solving the problem described in (6) and in 3(e) and 3(j) the denoised images with the improved method described in (12). For both methods, described in (6) and (12) with $K = 10$, we used the same sampling matrices with 10% of measurements in the Fourier domain. For comparison, Figures 3(d) and 3(i) display the noisy images denoised by the NL-means algorithm [2], with variance $\sigma^2 = 20.4$ (2^8 gray levels). We note that the results for NL-means present many artifacts due to the noise characteristics.

Since there is no ground-truth such as a perfect noise-free image, the denoising performances were evaluated via signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) measurements. However, the noise statistics corresponds to a mixture of *Poisson* and *Gaussian* noise, then the SNR of the images can be estimated from the noise model as:

$$\text{SNR} = \frac{\zeta A}{\sqrt{\zeta^2(A + \lambda_B) + \sigma^2}}$$

where ζ is the overall gain of the detector, A is the object intensity, λ_B is the mean intensity of the background, λ_i models the photon counting of a Poisson variable $\mathcal{P}(\lambda_i)$ and μ is the mean intensity

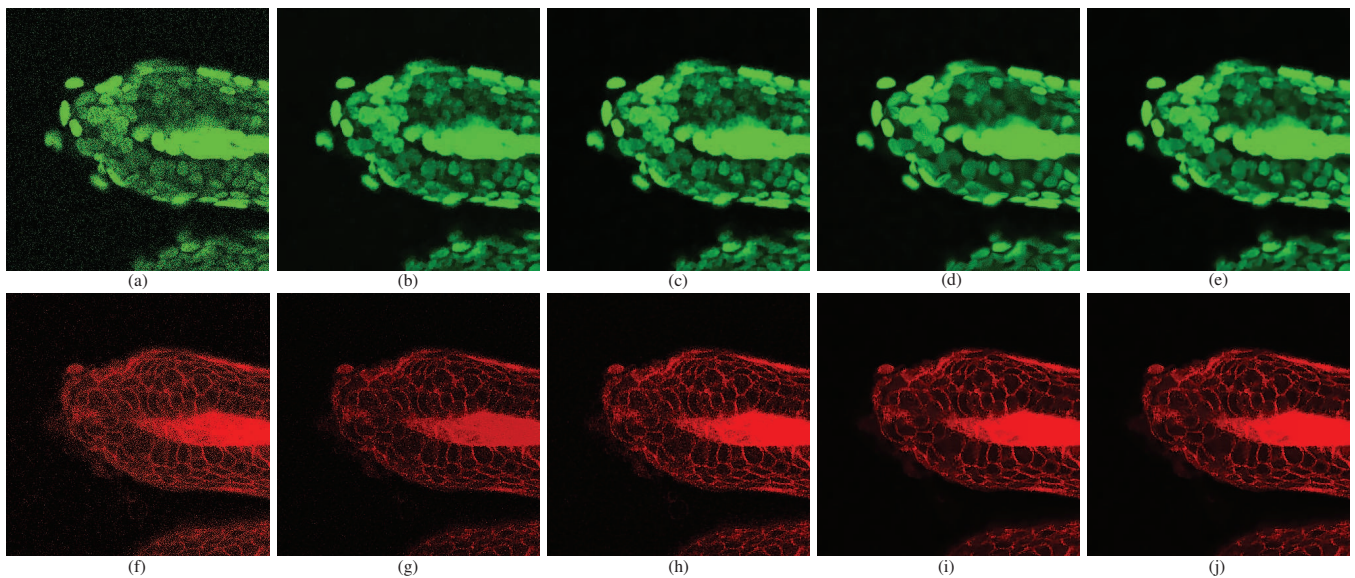


Fig. 3: Green and red channels; fluorescence microscopy image of hair follicle. (a)(f) Original noisy image. (b)(g) One single CS reconstruction. (c)(h) Multiple CS with averaging (6). (d)(i) NL-means filtering of the noisy image. (e)(j) Multiple CS with NL-means 12.

of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with standard deviation σ . The mixed *Poisson-Gaussian* noise parameters $(\zeta, \lambda_B, \mu, \sigma)$ are estimated using the cumulant method, matching the first four cumulants of I with the k -statistics of the samples in a uniform image region [16]. This follows from the property that the k -statistics are the minimum variance unbiased estimators for cumulants. Finally, the CNR is estimated as:

$$\text{CNR} = \frac{|I_{\mathfrak{R}_1} - I_{\mathfrak{R}_2}|}{\sigma}$$

where $I_{\mathfrak{R}_1}$ and $I_{\mathfrak{R}_2}$ are the mean intensity on the region \mathfrak{R}_1 and \mathfrak{R}_2 , and σ is the standard deviation of the noise distribution. The SNR and CNR measures are given for all methods in Table I.

Image	SNR (dB)		CNR	
	green	red	green	red
Original	3.62	4.53	13.97	9.51
CS 1	10.45	5.14	89.42	10.89
CS-mean	18.32	7.74	106.71	20.72
NL-means	18.80	19.96	103.72	71.81
CS NL-means	20.29	22.37	121.78	71.09

TABLE I: SNR, CNR measures for the green and red channel.

VI. CONCLUSION

This paper proposes an improved approach of a CS-based denoising method exploiting a non-local filtering scheme to merge multiple reconstructions with random Fourier projections and TV spatial constraints. The approach presents several advantages over traditional denoising methods by combining a reduced number of measurements and spatial regularization constraints in a single framework. Experiments on fluorescence microscopy images of hair follicle demonstrate improvements of SNR and CNR values even with a very limited number of measurements. Our proposed method leads to a denoising performance equivalent or better than the NL-means method with the advantage of taking only 10% of measurements.

VII. ACKNOWLEDGMENTS

We thank Inès Sequeira and Jean-François Nicolas from Institut Pasteur for the microscopy fluorescence images.

REFERENCES

- [1] L. Song, E. Hennink, T. Young, and H Tanke, "Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy," *Biophysical Journal*, vol. 68, pp. 2588–2600, 1995.
- [2] A. Buades, B. Coll, and JM. Morel, "A review of image denoising algorithms, with a new one," *SIAM J. Multiscale Model. Simul.*, vol. 4 (2), pp. 490–530, 2005.
- [3] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
- [4] J. Weickert, B.M. Ter Haar Romeny, and M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. Image Processing*, vol. 7, pp. 398–410, 1998.
- [5] R.A. DeVore and B.J. Lucier, "Fast wavelet techniques for near-optimal processing," *IEEE Military Communications Conference*, pp. 48.3.1–48.3.7, 1992.
- [6] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Density estimation by wavelet thresholding," Tech. Rep. 426, Statistics, Stanford University, 1995.
- [7] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *Conference on Computer Vision and Pattern Recognition*, 2006.
- [8] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD algorithm," in *Proceedings of SPARS'05, Rennes, France*, 2005.
- [9] A. Hyvärinen, P. Hoyer, and E. Oja, "Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation," *Adv. Neural Inf. Process. Syst.*, vol. 11, pp. 473–479, 1999.
- [10] E. Candès, "Compressive sampling," *Int. Congress of Mathematics*, vol. 3, pp. 1433–1452, 2006.
- [11] D. L. Donoho, "Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 52(4), pp. 1289–1306, April 2006.
- [12] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52(2), pp. 489–509, 2006.
- [13] M. Marim, E. Angelini, and J.-C. Olivo-Marin, "Compressed sensing in biological microscopy," in *Proc. SPIE Wavelets XIII*, 2009, vol. 7446, p. 744605.
- [14] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. on Information Theory*, vol. 52, pp. 6–18, 2006.
- [15] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Statist.*, 2005.
- [16] C. Rose and M. D. Smith, *Mathematical Statistics with Mathematics, chapter 7.2C: k-Statistics: Unbiased Estimators of Cumulants*, Springer-Verlag, 2002.