Minimum contrast estimation for super-resolution fluorescence microscopy using speckle patterns

Penghuan LIU[1], Jérôme IDIER[1], Sébastien BOURGUIGNON[1], Marc ALLAIN[2], Simon LABOUESSE[2], Anne SENTENAC[2]

[1]LS2N, École Centrale de Nantes/CNRS, 44321 Nantes, France

[2]Institut Fresnel, Aix-Marseille Université/CNRS/Centrale Marseille, 13013 Marseille, France

penghuan.liu@ls2n.fr, jerome.idier@ls2n.fr, sebastien.bourguignon@ls2n.fr
simon.labouesse@fresnel.fr, marc.allain@fresnel.fr, anne.sentenac@fresnel.fr

Résumé – La microscopie conventionnelle à large champ est un système limité par la diffraction, dont la résolution est limitée par la limite de Abbe. Il a été montré récemment qu'un facteur de super-résolution de deux peut être obtenu en illuminant l'objet avec des éclairements aléatoires (speckles), en exploitant les statistiques d'ordre deux des données dans un estimateur à minimum de contraste. Dans cet article, nous montrons que le cadre de l'estimation à minimum de contraste fournit un estimateur consistent lorsque le nombre d'illuminations augmente. Nous proposons finalement une approximation de cet estimateur permettant de réduire le coût de calcul.

Abstract – Conventional wide-field microscopy is a diffraction limited system with the resolution limit given by Abbe limit. It has been shown recently that a super-resolution factor of two can be achieved using unknown speckle patterns, by taking advantage of the second-order information of the acquired data associated with a minimum contrast estimator. In this paper, we show that such minimum contrast estimation provides a consistent estimator as the number of speckle pattern increases. An approximation of such estimator is finally proposed for efficient computations.

1 Introduction

The point spread function (PSF) of a classical fluorescence wide-field microscopy with perfect circular lens has a finite support in Fourier domain which is a disk of radius $\nu_{\text{psf}} = \text{NA}/\lambda$, where λ is the wavelength of the emitted light and NA is the numerical aperture of the system. Structured illumination microscopy (SIM) allow us to surpass this limit by illuminating the object ρ with several structured patterns I_m instead of uniform illumination [1]. For each pattern I_m, the recorded data y_m can be modeled as the convolution of the emitted density from the object ρ with PSF h :

$$y_m = h \star (\rho I_m) + \epsilon_m \tag{1}$$

with $\star$ represents the convolution operator and ϵ_m is the electrical noise. The high spatial frequencies in the object are down modulated by I_m to the support of the system and can be reconstructed by data post processing.

One disadvantage of SIM is that its reconstruction process rely on precise knowledge of illuminations. Normally controlling illumination is a difficult task and small errors will cause strong artifacts in the reconstructed image. Blind SIM is proposed to address this problem by using speckle illumination as a substitute for conventional harmonic illumination to obtain super-resolution (SR) [2, 3]. Moreover, the speckle pattern is easier to generate.

A joint reconstruction approach has been proposed in [2, 3] in which the object are reconstructed simultaneously with the speckle patterns. However the theoretical SR capacity one can expected form such a joint approach remains unclear. A marginal estimator based on the second-order statistics of the recorded data has been proposed in [4] with a known theoretical super-resolution capacity of factor of two. We demonstrate that such a estimator is statistical consistent in this paper. To reduce the computational complexity of the original method, a patch based approximation approach is proposed.

2 Marginal estimation principle

In a discrete form where each 2D image is displayed as a column vector, Eq. (1) can be written as :

$$\boldsymbol{y}_m = \mathbf{HRI}_m + \boldsymbol{\varepsilon}_m \tag{2}$$

with $\mathbf{H} \in \mathcal{R}^{N\times N}$ the convolution matrix and $\mathbf{R} \in \mathcal{R}^{N\times N}$ the diagonal matrix with diagonal values identify with $\boldsymbol{\rho}$. $\mathbf{I}_m \in \mathcal{R}^N$ where $m \in \{1, \cdots, M\}$ is the m-th realization of speckle with homogeneous intensity mean I_0 and covariance matrix $\mathbf{C}$. We consider that the speckle pattern and the collection of the emitted light of the sample are performed in the same optical device. In this case the covariance matrix $\mathbf{C}$ identifies with $\mathbf{H}$. The noise patterns ε_m are assumed jointly independent, centered, and spatially white, sharing a common covariance matrix $\boldsymbol{\Gamma}_\varepsilon$ proportional to the identity matrix. Now we can write the mean and the covariance of the measured image $\boldsymbol{y}_m$ as :

$$\boldsymbol{\mu}_y = I_0\mathbf{H}\boldsymbol{\rho}, \qquad \boldsymbol{\Gamma}_y = \mathbf{HRCRH}^T + \sigma^2\mathbf{I}_d, \tag{3}$$

where ^{T} denotes the transpose operator and $\mathbf{I}_d$ the identity matrix.

The principle of marginal estimation is to infer the sample $\boldsymbol{\rho}$ from the statistical characteristics of the collected data. A typical marginal estimation procedure would consist in maximizing the likelihood of the data as a function of ρ. However, because of the complexity of the speckle statistics, it is hard or even impossible to express the data likelihood in closed-form. A simpler alternative is to estimate $\boldsymbol{\rho}$ by minimizing the mismatch between the theoretical second-order data statistics (3) and their empirical moments :

$$\hat{\boldsymbol{\mu}}_y = \frac{1}{M}\sum_m \boldsymbol{y}_m, \qquad \hat{\boldsymbol{\Gamma}}_y = \frac{1}{M}\sum_m \boldsymbol{y}_m\boldsymbol{y}_m^{\mathrm{t}} - \hat{\boldsymbol{\mu}}_y\hat{\boldsymbol{\mu}}_y^T. \quad (4)$$

In [4], the Kullback-Leibler divergence is chosen to measure the difference between the theoretical and the empirical moments :

$$\begin{aligned} D_M(\boldsymbol{\rho}) &\propto D_{\mathcal{KL}}\Big(\mathcal{N}(\hat{\boldsymbol{\mu}}_y, \hat{\boldsymbol{\Gamma}}_y)||\mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y)\Big) \\ &= \log|\boldsymbol{\Gamma}_y| + \frac{1}{M}\mathrm{Tr}(\boldsymbol{\Gamma}_y^{-1}\mathbf{V}\mathbf{V}^T) + \mathrm{cst} \end{aligned} \quad (5)$$

where $\mathcal{N}$ denotes the normal distribution and $\mathrm{Tr}(\cdot)$ is the trace of a square matrix, and

$$\mathbf{V} = (\boldsymbol{v}_1|\cdots|\boldsymbol{v}_M), \qquad \text{with} \qquad \boldsymbol{v}_m = \boldsymbol{y}_m - \boldsymbol{\mu}_y \quad (6)$$

According to the law of large numbers, we have :

$$\lim_{M\to\infty}\hat{\boldsymbol{\mu}}_y \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_y^*, \qquad \lim_{M\to\infty}\hat{\boldsymbol{\Gamma}}_y \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_y^* \quad (7)$$

where $\boldsymbol{\mu}_y^*, \boldsymbol{\Gamma}_y^*$ indicate the true value of the corresponding variables and $\xrightarrow{\mathcal{P}}$ denotes the convergence in probability. Consequently, for a sufficiently large number of acquisitions, $D_M(\boldsymbol{\rho}^*)$ is expected to vanish.

The solution of (5) has no expression in closed form. We choose a gradient based iteration algorithm, L-BFGS [5] in our simulations to minimize the criterion (5) with the gradient given by [4] :

$$\nabla D_M(\boldsymbol{\rho}) = -2\Big([\boldsymbol{\Omega}^T(\frac{1}{M}\mathbf{V}\mathbf{V}^{\mathrm{t}} - \boldsymbol{\Gamma}_y)\boldsymbol{\Omega}]\circ\mathbf{C}\Big)\boldsymbol{\rho} - \frac{2}{M}I_0\boldsymbol{\Omega}^{\mathrm{t}}\mathbf{V}\mathbf{1} \quad (8)$$

in which $\boldsymbol{\Omega} = \boldsymbol{\Gamma}_y^{-1}\mathbf{H}$ and $\mathbf{1} = (1\cdots1)^T$.

3 Asymptotic analysis

The statistical principle behind such an inferential principle is called *minimum contrast estimation* [6], or alternatively, *M-estimation* [7, Chap. 5]. Let $\boldsymbol{y}_m$, $m = 1, \ldots,$ be independent, identically distributed data vectors, each taking its values in $\mathcal{Y}$, with a common probability distribution depending on a parameter vector $\boldsymbol{\theta}^*$ in $\boldsymbol{\Theta}$. Let $C : \mathcal{Y}\times\boldsymbol{\Theta} \to \mathbb{R}$ be a real-valued function. The theory of minimum contrast estimation relies on the following definitions.

Definition 1. *The statistical expectation $J(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \mathrm{E}\,[C(\boldsymbol{y}, \boldsymbol{\theta})]$ is said to be a contrast function if it has a strict minimum at true value $\boldsymbol{\theta}^*$.*

Definition 2. *For arbitrary large M, if the empirical mean*

$$J_M(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} C(\boldsymbol{y}_m, \boldsymbol{\theta}) \quad (9)$$

converges towards $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ in probability for each value of $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$, $J_M(\boldsymbol{\theta})$ is called a contrast process.

Associated with a contrast process, a minimum contrast estimator $\hat{\boldsymbol{\theta}}_M$ is defined as the minimum of (9). The estimator $\hat{\boldsymbol{\theta}}_M$ has consistent property (i.e. $\hat{\boldsymbol{\theta}}_M$ converges to $\boldsymbol{\theta}^*$ in probability as $M\to\infty$) if the following conditions hold [6, 7] :

Theorem 1. *Let $\boldsymbol{\Theta}$ be a bounded open set in $\mathcal{R}^N$ and $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ be a continuous function on the closure of $\boldsymbol{\Theta}$. If*

— *$J_M(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$*

— *$J_M(\boldsymbol{\theta})$ converges uniformly to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, i.e.*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |J_M(\boldsymbol{\theta}) - J(\boldsymbol{\theta}^*, \boldsymbol{\theta})| \longrightarrow 0$$

then $\hat{\boldsymbol{\theta}}_M$ is a consistent estimator of $\hat{\boldsymbol{\theta}}$.

In our application, we define $\boldsymbol{\theta} = \mathcal{S}\boldsymbol{\rho}$ with $\mathcal{S}$ denoting the ideal low-pass filter with frequency support given by $2\nu_{\mathrm{PSF}}$. Then a contrast function can be defined as :

$$J(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = D^*(\mathcal{S}^+\boldsymbol{\theta}) \quad (10)$$

with $D^*(\boldsymbol{\rho}) = D_{\mathcal{KL}}(\mathcal{N}(\boldsymbol{\mu}_y^*, \boldsymbol{\Gamma}_y^*)||\mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y))$ and $\mathcal{S}^+$ the pseudo-inverse of $\mathcal{S}$. $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ coincide with our definition of contrast function since the true value $\boldsymbol{\theta}^*$ is identifiable from $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ [4]. Similarly we can define the contrast process by $J_M(\boldsymbol{\theta}) = D_M(\mathcal{S}^+\boldsymbol{\theta})$.

The continuous demonstration of $J_M(\boldsymbol{\theta})$ is simple. We see in formula (5) that every component in $D_m(\boldsymbol{\rho})$ is continuous. According to the rules for constructing continuous functions [8], it is obvious that $D_M(\boldsymbol{\rho})$ is a continuous function, so is $J_M(\boldsymbol{\theta})$.

For the uniformly convergence of $J_M(\boldsymbol{\theta})$ to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, we note that :

$$D_M(\boldsymbol{\rho}) - D^*(\boldsymbol{\rho}) = \mathrm{Tr}\Big(\boldsymbol{\Gamma}_y^{-1}(\hat{\boldsymbol{\Gamma}}_y - \boldsymbol{\Gamma}_y^*)\Big) + \log\frac{|\boldsymbol{\Gamma}_y^*|}{|\hat{\boldsymbol{\Gamma}}_y|} \quad (11)$$

Since the covariance matrix $\mathbf{HRCRH}^{\mathrm{t}}$ is a positive semi-definite matrix, applying Weyl's inequality [9] to formula (3), we have :

$$\lambda_{\min}(\boldsymbol{\Gamma}_y) \geqslant \lambda_{\min}(\boldsymbol{\Gamma}_\varepsilon) \quad (12)$$

where $\lambda_{\min}(\mathbf{A})$ means the smallest eigenvalue of the matrix $\mathbf{A}$. Since $\boldsymbol{\Gamma}_y^{-1}$ is a symmetric matrix, we can define its norm by its maximum eigenvalue :

$$\left\|\boldsymbol{\Gamma}_y^{-1}\right\| = \lambda_{\max}(\boldsymbol{\Gamma}_y^{-1}) = \frac{1}{\lambda_{\min}(\boldsymbol{\Gamma}_y)} \leqslant \frac{1}{\sigma^2} \quad (13)$$

Combining (7)(11)(13), we have :

$$\sup_{\boldsymbol{\rho}\in\mathcal{R}^N} |D_M(\boldsymbol{\rho}) - D^*(\boldsymbol{\rho})| \longrightarrow 0 \quad (14)$$

Restricting the domain of $D_m(\boldsymbol{\rho})$ to $\{\boldsymbol{\rho}|\boldsymbol{\rho} = \mathcal{S}^+\boldsymbol{\theta}, \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$, we obtain :

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |D_M(\mathcal{S}^+\boldsymbol{\theta}) - D^*(\mathcal{S}^+\boldsymbol{\theta})| \longrightarrow 0 \quad (15)$$

Now we have $J_M(\boldsymbol{\theta})$ converges uniformly to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$.

4 Patch based marginal estimator

To obtain the gradient (8) we need to inverse the covariance matrix Γ_y, the computational complexity for which is $\mathcal{O}(N^3)$. This computation burden is too high for realistic image sizes. One possible solution is to cut the image into a set of patches, and neglect the correlation between pixels from different patches. This corresponds to a new optimization problem :

$$\min_{\boldsymbol{\rho}} F'(\boldsymbol{\rho}) = \sum_p D_{\mathcal{KL}}(\mathcal{N}(\hat{\boldsymbol{\mu}}_{yp}, \hat{\boldsymbol{\Gamma}}_{yp})||\mathcal{N}(\boldsymbol{\mu}_{yp}, \boldsymbol{\Gamma}_{yp})) \quad (16)$$

where $\boldsymbol{\mu}_{yp}, \boldsymbol{\Gamma}_{yp}$ denote the mean and covariance for patch p, respectively. The gradient for (16) can be written similarly as in (8). We can prove that the computational complexity for (16) can be reduced to $\mathcal{O}(N^2(\log N + L))$ with L the number of pixels in each patch. What is more, the optimization for the patched version can be solved parallelized.

5 Simulation results

We use a simulated target whose fluorescence density in the polar coordinates given by : $\rho(r, \theta) \propto [1 + \cos(40\theta)]$ (Fig. 1a) as the true object. The point spread function is chosen as :

$$h(r, \theta) = \Big(\frac{J_1(NAk_0r)}{k_0r}\Big)^2 \frac{k_0^2}{\pi} \quad (17)$$

where J_1 is the first order Bessel function of the first kind, NA is the objective numerical aperture set to 1.49 and $k_0 = \frac{2\pi}{\lambda}$ is the free-space wavenumber with λ the emission and the excitation wavelengths. The covariance of speckle patterns are set to $\mathbf{H}$. To circumvent the boundary effects, we perform simulations with convolution matrix $\mathbf{H}$ with a block-Toeplitz with Toeplitz-block (BTTB) structure.

In our simulations a common pixel size of λ/20 for both raw images and the super-resolved reconstruction is adopted, which is finner than the Nyquist criterion $\lambda/8NA \approx \lambda/12$ for a super-resolution factor of two.

To evaluate the resolution enhancement of the reconstructed objects, we define the modulation contrast function $C(R)$ as a function of of radius R as :

$$C(R) = 2\tilde{f}_R(1/L(R))/\tilde{f}_R(0) \quad (18)$$

where $L(R) = 2\pi R/40$ denotes the period of the pattern taken on a circle and $\tilde{f}_R$ is the 1D Fourier transform of $f_R(s) \propto 1 + \cos(2\pi s/L(R))$ with s the arc length along the circle. For the true object, $C(R) = 1$ for all radius R.

Reconstructed objects under different number of speckle patterns are shown in Fig. 1, and their modulation contrast values as a function of period $L(R)$ are shown in Fig. 2. The estimation under infinite speckle patterns are addressed by considering $\hat{\boldsymbol{\mu}}_y = \boldsymbol{\mu}_y^*$ and $\hat{\boldsymbol{\Gamma}}_y = \boldsymbol{\Gamma}_y^*$, with $\boldsymbol{\mu}_y^*$ and $\boldsymbol{\Gamma}_y^*$ obtained by setting $\boldsymbol{\rho} = \boldsymbol{\rho}*$ in (3). In general the resolution are defined as the period whose contrast is above 0.1 [2]. As we can see, the deconvolution of the wide-field image bring no super-resolution (Fig. 1b) for patterns bigger than the Abbe limit $L_0 = \lambda$/2NA, while the marginal estimator has recover the resolution corresponding to $\frac{1}{2}L_0$ as expected. Some shading artifacts were viewed in the low frequency domain in the reconstructed objects (Fig.1c,d) when the number of speckle patterns is not large enough to give an accurate estimation of the theoretical covariance. This degradation disappears as the number of speckle patterns increases.

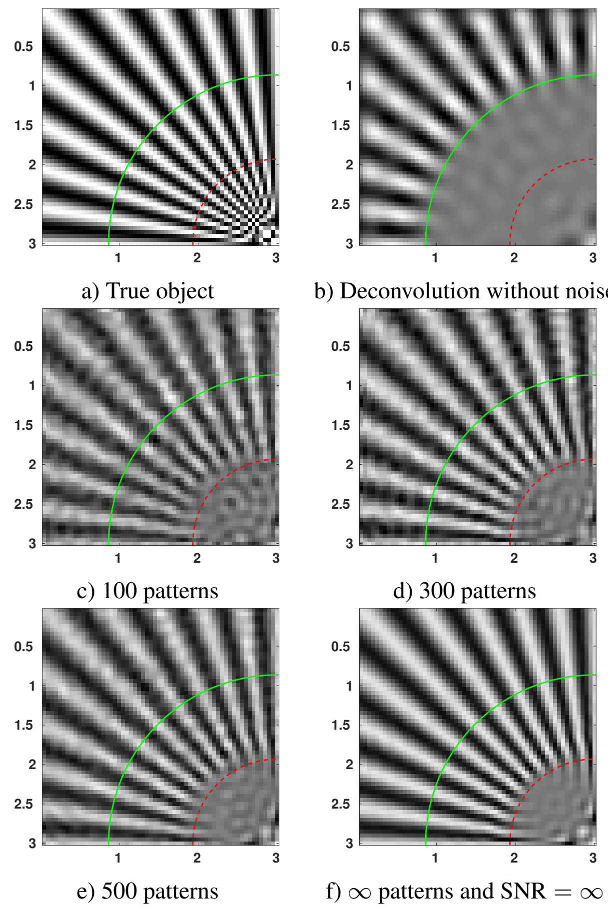


FIGURE 1 – **Reconstructed objects with L-BFGS algorithm with different number of speckle patterns.** (a) A quarter of the true object. (b) Deconvolution of wide-field image. (c,d,e) Marginal reconstruction with 100, 300, 500 speckle patterns respectively under SNR 40dB. (f) Marginal reconstruction under asymptotic condition with infinity number of speckle patterns and SNR. The green solid lines (resp. red dashed lines) correspond to spatial frequencies transmitted by OTF support (resp. 2 times OTF support) and the graduation in the images represents the wavelength λ.

To verify the SR capacity after introducing patches, we do simulations under asymptotic conditions (infinite speckle and SNR) with different patch sizes and the reconstructed objects are displayed in Fig. 3. Even we only consider the correlation of pixels with themselves (Fig. 3a), the resolution is better than the deconvolution case (Fig. 1b). And the super-resolution extends quickly as the patch size increases. The average time elapsed in one iteration after introducing patches using a standard Mat-

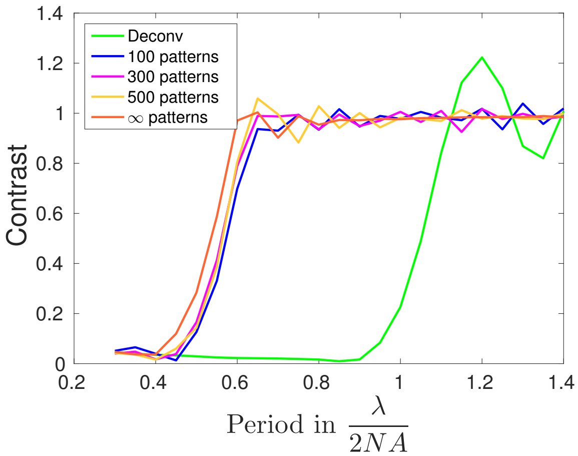
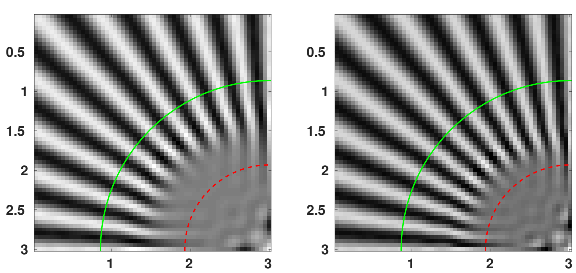


FIGURE 2 – **Modulation contrast of the reconstructed objects as a function the period extracted from images shown in Fig. 1.**

lab implementation on a normal computer is shown in Fig. 4. We see that except the bias when the patch size is small (due to the inefficiency for-loop processing in Matlab), the time elapsed for one iteration increases almost linearly as the patch size grows.



a) Pixel patch b) 2×2 patch

c) 3×3 patch d) 4×4 patch

FIGURE 3 – **Reconstructed objects with different patch size under ∞ speckle patterns and SNR $= \infty$.**

6 Conclusion and perspectives

For the super-resolution image reconstruction problem in SIM using speckle patterns, a marginal estimator was proposed based on the second-order moments of recorded data in [4]. We demonstrate in this paper that this marginal estimator is statistically consistent in a minimum contrast estimation approach. To reduce the computational complexity of the estimation, an approximation is introduced by cutting each recorded image into a set of patches and by neglecting the correlation between pixels from different patches. Simulation results show that the super-resolution is kept even under conditions when the patches are very small. We stress here that compared with classical SIM, the total photon budget does not increase even though a larger number of speckle patterns is used.

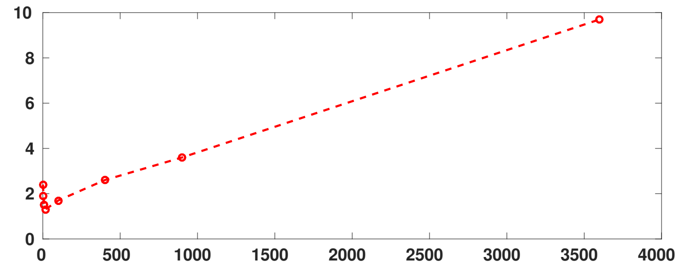


FIGURE 4 – **Time (seconds) elapsed per iteration (Y-axis) with respect to number of pixels in each patch (X-axis).**

Acknowledgements We would like to thank the China Scholarship Council (CSC) and the GDR ISIS for their partial funding to this work.

References

[1] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy", *J. Microscopy*, vol. 198, no. 2, pp. 82–87, 2000.

[2] E. Mudry, K. Belkebir, J. Girard, J. Savatier, E. Le Moal, C. Nicoletti, M. Allain, and A. Sentenac, "Structured illumination microscopy using unknown speckle patterns", *Nature Photonics*, vol. 6, no. 5, pp. 312–315, 2012.

[3] S. Labouesse, A. Negash, J. Idier, S. Bourguignon, T. Mangeat, P. Liu, A. Sentenac, and M. Allain, "Joint reconstruction strategy for structured illumination microscopy with unknown illuminations", *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 1–14, May 2017.

[4] J. Idier, S. Labouesse, P. Liu, M. Allain, S. Bourguignon, and A. Sentenac, "A theoretical analysis of the super-resolution capacity of imagers using speckle illuminations", *arXiv preprint arXiv :1512.06260*, 2015.

[5] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization", *SIAM J. Sci. Comp.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[6] P. J. Bickel and K. A. Doksum, *Mathematical Statistics : Basic Ideas and Selected Topics*, vol. 1, Prentice Hall, 2nd edition, 2001.

[7] A. W. van der Vaart, *Asymptotic Statistics*, vol. 3, Cambridge University Press, 1998.

[8] J. R. Munkres, *Topology*, Prentice Hall, 2000.

[9] D. Serre, "Matrices : Theory and applications", 2002.