PHASE RETRIEVAL WITH SPARSITY PRIORS AND APPLICATION TO MICROSCOPY VIDEO RECONSTRUCTION

Yoann Le Montagner[1,2]*, Elsa D. Angelini*[2]*, Jean-Christophe Olivo-Marin*[1]

[1]Institut Pasteur, Unité d'Analyse d'Images Quantitative, CNRS URA 2582, F-75015 Paris
[2]Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, F-75013 Paris

ABSTRACT

The theory of compressed sensing (CS) predicts that structured images can be sampled in a compressive manner with very few non-adaptive linear measurements, made in a proper adjacent domain. However, is such a recovery still possible with non-linear measurements, such as optical-based Fourier modulus?

In this paper, we study the problem of Fourier phase retrieval required for optical Fourier CS imaging. We propose an algorithm to solve this problem, exploiting a specific TV-based regularization constraint. We demonstrate the performance of the proposed method on synthetic and real test sequences, in the context of microscopy video reconstructions.

***Index Terms*—** Phase retrieval, sparsity, video reconstruction, Fourier measurements, total variation.

1. INTRODUCTION

1.1. Compressed sensing background

The inverse problem tackled by CS can be formulated as follows: given a signal of interest $x \in \mathbb{R}^N$ measured through a *random* linear operator Φ that outputs a vector $y \in \mathbb{R}^M$ of observations with $M \ll N$, can x be recovered from y? The randomness of the measurement operator Φ should not be understood in strict meaning, but rather as the fact that Φ should spread the information contained in x over the whole vector y. Examples of such operators include random Gaussian or Bernouilli matrices [1], randomly subsampled Fourier or Hadamard transforms [2], or dedicated unitary matrices [3].

Previous results (see [4, 5]) establish that, up to some technical details, x can be recovered from y if it has a sparse decomposition on some known basis or dictionary. In that case, recovery is performed by solving a convex optimization program which typically involves a l_1 constraint on the coefficients representing x in this dedicated sparsifying domain.

1.2. Optical Fourier measurements

In this paper, we focus on the case where the sensing operator measures only the magnitude of the Fourier transform, with a uniform random selection of the sampled Fourier coefficients. The motivation to focus on this type of operator is that Fourier magnitude can be sensed using optical devices. This property can be used to design acquisition set ups performing an optical Fourier transform upstream from the actual photo-electric sensors, allowing a simplification of the whole device with a CS-like acquisition strategy.

Indeed, measuring the complex Fourier transform of a scene can be implemented optically, but remains challenging as photo-electric transducers such as a CCD or CMOS array return output signals that correspond to a measure of the energy of the incident photons, which is independent of the phase of the corresponding electro-magnetic wave. Measuring this phase relies on more complex optical set-ups which are not always compatible with the experimental arrangement.

As a workaround, the present work studies whether image reconstruction is still possible without measuring the phase of the Fourier samples, while relying on computational strategies to recover the missing phase information. Formally, this problem can be stated as follows: given a signal of interest x (here a 2D image), is it possible to recover this signal from a measurement vector y defined as:

$$y = |\Phi x| + \text{ noise} \tag{1}$$

Here, Φ represents a Fourier transform followed by a selection of a random subset of coefficients, and $|\cdot|$ stands for the pointwise modulus. The difference with standard CS lies in the lack of Fourier phase information, calling for dedicated phase retrieval strategy.

1.3. Phase retrieval background

The problem of recovering a signal from the modulus of its Fourier transform, known as the *phase retrieval* problem, has been studied for a long time: this reconstruction technique is used for instance for X-ray microscopy applications in crystallography (see [6, 7]). To recover a signal $x \in \mathbb{R}^N$ from a measurement vector y defined as (1), the algorithm proposed in [6] defines two subsets of $\mathbb{R}^N$:

- the *data set* $\mathcal{D}_{y,\epsilon}$, that contains all the signals x that correspond to the measured samples, with a certain tolerance ϵ that depends on the noise that affects these measurements:

$$\mathcal{D}_{y,\epsilon} = \left\{ x \in \mathbb{R}^N \text{ s.t. } \| y - |\Phi x| \|_2 \leq \epsilon \right\} \tag{2}$$

- a *regularization set* $\mathcal{R}$ that corresponds to all the signals that meet certain prior conditions which are known to be true for the actual solution. For crystallography applications, $\mathcal{R}$ typically consists in all the 2D images that are supported on a given subset of pixels.

Then, an estimator $\hat{x}$ of the solution is obtained as a limit of alternated projections over the two sets $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$:

$$\hat{x} = \left(\Pi_{\mathcal{R}} \circ \Pi_{\mathcal{D}_{y,\epsilon}} \circ \Pi_{\mathcal{R}} \circ \cdots \circ \Pi_{\mathcal{D}_{y,\epsilon}} \right)(x_0) \tag{3}$$

where $\circ$ is the composition operator, x_0 is an initial guess of the solution, and $\Pi_{\mathcal{D}_{y,\epsilon}}$ and $\Pi_{\mathcal{R}}$ stand respectively for the projection operators over $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$:

$$\Pi_{\mathcal{D}_{y,\epsilon}}(x) = \underset{z \in \mathcal{D}_{y,\epsilon}}{\arg\min} \| z - x \|_2 \tag{4}$$

and similarly for $\Pi_{\mathcal{R}}$. It was shown in [6] that the sequence of estimators (3) converges toward the intersection of $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$.

It is important to note that, although the set $\mathcal{D}_{y,\epsilon}$ is not convex, (4) can be solved explicitly[1]. Consequently, the evaluation of $\Pi_{\mathcal{D}_{y,\epsilon}}(x)$ is exact and low-cost in terms of computation time[2].

Other works have refined the algorithm (3), mostly to improve the convergence rate (see e.g. [8] and references therein). However, to the best of our knowledge, all existing phase retrieval algorithms rely on the construction of a sequence of estimators that converges toward the intersection of two sets, one characterizing the valid signals with respect to the measurements, and the other one the prior information available on the solution.

In our case, the prior information is quite different from the one available in crystallography applications (i.e. support constraints). We therefore introduce in the next section a general reconstruction approach suited for microscopic video reconstruction which exploits a similar alternated projection iteration scheme as the one used in phase retrieval, but allows different *a priori* information to be used.

2. PHASE RETRIEVAL APPLIED TO VIDEO RECONSTRUCTION

2.1. Problem formulation

We formulate our video reconstruction problem as follows: how to recover a 2D frame $x \in \mathbb{R}^N$ from a video sequence, knowing both a vector $y \in \mathbb{R}_+^M$ of Fourier modulus measurements obtained from this frame (as in (1)) and the previous frame $a \in \mathbb{R}^N$ in the video?

Iteratively solving this problem on consecutive frames will permit to recover a full video sequence from a subset of Fourier modulus measurements and an initial key-frame, which is required to start the reconstruction process. In practice, propagation and amplification of reconstruction errors are likely to limit the number of frames that can be recovered starting from a single key-frame.

2.2. Hybrid total variation

To solve our problem using the phase retrieval framework presented in Sec. 1.3, we have to define the regularization set $\mathcal{R}$ that contains all the admissible solutions, based on an *a priori* regularity model. In the context of microscopy video reconstruction, similarly to what was proposed in [9], we assume that the following properties hold for the image frame x to be reconstructed:

- the 2D gradient map ∇x is sparse,
- the difference map $(x-a)$ between the reconstructed frame x and its predecessor a is sparse,
- the non-zeros coefficients in ∇x, which mainly correspond to edges in the frame x, are located close to the non-zero coefficients in $(x-a)$, i.e. the displacement fronts.

To enforce these properties, we introduce a weighted hybrid total variation energy over the space $\mathbb{R}^N$ of 2D images, defined as:

$$\|x\|_{\mathrm{TV},w,a} = \sum_P w(P) \sqrt{\|\nabla x(P)\|_2^2 + |x(P) - a(P)|^2} \quad (5)$$

where P visits every pixel, and $w \in \mathbb{R}^N$ is a weight map with $0 < w(P) < +\infty$ and tuned such that $w(P)$ is small on pixels P where we expect to reconstruct edges. Tuning of w is discussed in Sec. 3.3.

This hybrid TV is similar to the mixed $l_{1,2}$ norms used in CS problems to favor block-sparse solutions (see e.g. [10, 11]): the spatial correspondence of non-zero coefficients in ∇x and $(x-a)$ is enforced by the fact that the coefficients corresponding to a given pixel P in these maps are grouped inside the same l_2 block in (5).

Finally, we define the regularization set $\mathcal{R}$ as a level set of the hybrid TV (5): $\mathcal{R}_{w,a,\tau} = \left\{x \in \mathbb{R}^N \text{ s.t. } \|x\|_{\mathrm{TV},w,a} \leq \tau\right\}$. The newly introduced parameter $\tau > 0$ becomes an input prior to the reconstruction problem or can be set adaptively during the reconstruction process, as proposed below.

2.3. Projection operator $\Pi_{\mathcal{R}_{w,a,\tau}}$

Using $\mathcal{R}_{w,a,\tau}$ as a regularization set for the reconstruction requires to be able to evaluate efficiently the projection operator $\Pi_{\mathcal{R}_{w,a,\tau}}$, which implies to solve several instances of the following problem:

$$\Pi_{\mathcal{R}_{w,a,\tau}}(x) = \arg\min_z \|z-x\|_2 \text{ s.t. } \|z\|_{\mathrm{TV},w,a} \leq \tau \quad (6)$$

To solve this convex optimization problem, we use an algorithm derived from the total variation projection method presented in [12]. The main idea behind this algorithm is to recast the constrained problem (6) into the following unconstrained one:

$$\arg\min_u \tau \|u\|_{\infty,\frac{1}{w}} + \left\langle u^{(1)}, a\right\rangle + \frac{1}{2}\left\|x - u^{(1)} + \mathrm{Div}\, u^{(2,3)}\right\|_2^2 \quad (7)$$

Here, u represents a 3D vector field (i.e. a 3-channel image) over the same domain than x, $u^{(1)}$ and $u^{(2,3)}$ denote respectively the first and the concatenation of the second and third components of u, Div is the adjoint operator of $-\nabla$, and $\|u\|_{\infty,\frac{1}{w}} = \max_P \frac{1}{w(P)} \|u(P)\|_2$ where P visits every pixel. Without going into details, it can be shown that (6) and (7) are equivalent, and that the corresponding optimal variables $z^\star$ and $u^\star$ are related through the equation $z^\star = x - u^{\star(1)} + \mathrm{Div}\, u^{\star(2,3)}$. Problem (7) is then solved using a Nesterov accelerated gradient descent scheme (see [12, 13, 14]).

This leads to an iterative scheme to evaluate the projection operator $\Pi_{\mathcal{R}_{w,a,\tau}}$: each iteration is performed in $\mathcal{O}(N \log N)$ operations, and the Nesterov acceleration ensures a quadratic convergence rate. This approach is much slower than what is needed to compute the other projection operator $\Pi_{\mathcal{D}_{y,\epsilon}}$, but we observed that a careful initialization of the gradient descent provides significant speed up of the convergence.

2.4. Reconstruction algorithm

The proposed reconstruction algorithm is based on alternated projections of the iterated reconstructions over the data set $\mathcal{D}_{y,\epsilon}$ and the regularization set $\mathcal{R}_{w,a,\tau}$, involving two parameters ϵ and τ. ϵ controls the size of the data set $\mathcal{D}_{y,\epsilon}$, and is set proportional to the noise level that affects the measurements. Setting τ is not straightforward, and we developed an adaptive heuristic to dynamically adjust this parameter during the iterative reconstruction process.

This dynamic adjustment process relies on the following observation. The alternated projection scheme (3) produces a sequence of estimators that converge to the intersection $\mathcal{R}_{w,a,\tau} \cap \mathcal{D}_{y,\epsilon}$, but this intersection is empty when τ is below a certain threshold $\tau^\star$ (if the image a is not constant, the set $\mathcal{R}_{w,a,\tau}$ itself is empty when $\tau = 0$): therefore, the algorithm becomes non-convergent if $\tau < \tau^\star$.

Based on this remark, we propose a reconstruction algorithm where τ is initialized at an arbitrary high value τ_0, and then reduced until the algorithm becomes non-convergent (see the pseudo-code

[1] For $\epsilon = 0$, $\Pi_{\mathcal{D}_{y,\epsilon}}(x)$ is computed as follows: 1st) take the Fourier transform X of x; 2nd) for all the spatial frequencies k for which a measure $y(k)$ is available, replace the obtained modulus $|X(k)|$ with $y(k)$; 3rd) finally, inverse the Fourier transform. For $\epsilon > 0$, the second step is slightly more complex, but can still be run in $\mathcal{O}(N)$ operations.

[2] $\mathcal{O}(N \log N)$ operations.

function FRAME RECONSTRUCTION($x_0, \tau_0, \Delta_{\text{tolvar}}, \alpha$)
  $k \leftarrow 0$
  $\tau \leftarrow \tau_0$
  loop
    $k \leftarrow k + 1$
    $x_k \leftarrow \Pi_{w,a,\tau} \circ \Pi_{y,\epsilon}(x_{k-1})$
    $\delta_k \leftarrow \|x_k - x_{k-1}\|_2 \,/\, \|x_{k-1}\|_2$
    if $\delta_k < \Delta_{\text{tolvar}}$ **then**
      $x_{\text{candidate}} \leftarrow x_k$
      $\tau \leftarrow \alpha\tau$
    else if detect non-convergence **then**
      return $x_{\text{candidate}}$
    end if
  end loop
end function

Fig. 1. Pseudo-code of the iterative reconstruction algorithm. The algorithm takes 4 arguments as input: x_0 and τ_0, which are the initial values for the reconstructed frame and the partial 3D TV bound, and Δ_{tolvar} and α, that controls the general behavior of the algorithm. When the relative variation between successive iterates goes below Δ_{tolvar}, the bound τ is reduced by a factor α ($0 < \alpha < 1$), until the algorithm becomes non-convergent.

in Fig. 1). The algorithm returns the result (denoted as $x_{\text{candidate}}$) obtained with the smallest value of τ that leads to convergence.

However, detecting that the sequence of estimators does not converge for a given value of τ is a challenging issue, as we do not have any result on the theoretical convergence rate of this sequence of estimators. We developed an empirical approach based on the properties of the sequence (δ_k), which measures the relative variations between two successive iterates. More precisely, to detect whether the algorithm should be stopped at a given iteration $k^\star$, we perform the following test:

1. linear regression over the truncated sequence of values of $(\log \delta_k)$ for $k^\star - \Delta k + 1 \leq k \leq k^\star$, where Δk is a fixed parameter, returning a slope of evolution s;
2. stop if s is above a certain threshold $s_{\max}$.

The proposed non-convergence test evaluates the mean variation of the sequence $(\log \delta_k)$ over a window of Δk samples: if this sequence increases at a rate higher than $s_{\max}$, then we assume that the algorithm is diverging. Typical parameter values for this test are $\Delta k = 100$ iterations and $s_{\max} = -10^{-4}$ per iteration. Finally, to improve computation speed, we typically perform this test only every 25 iterations: as the linear regression is performed over a sliding window, the value of the resulting slope is not likely to change much from one iteration to the next, which justifies this approach.

3. RESULTS

3.1. Test sequences

We present here some preliminary results obtained with two test video sequences, both of size 256×256 with 80 frames:

- *Disks*: synthetic sequence representing disk shapes of random intensity levels and sizes (5 to 25 pixels diameter), and moving with random directions and speeds. The typical distance travelled by disks between two frames is about 1 to 3 pixels.
- *Amiba*: real microscopy video of moving and stretching cells having similar sizes and speeds than in *Disks*.

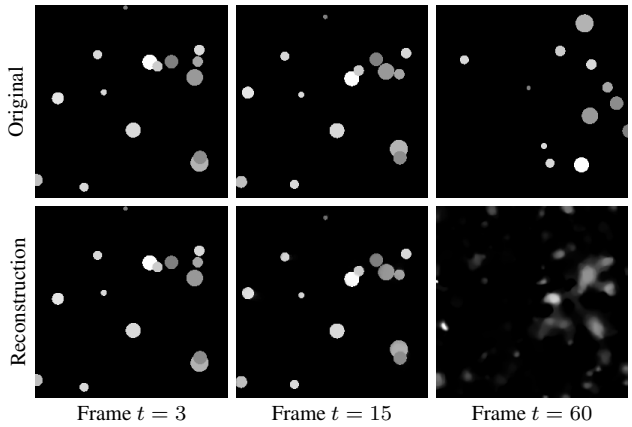

Fig. 2. Reconstruction result for the sequence *Disks* with a uniform weight map ($w = 1$).

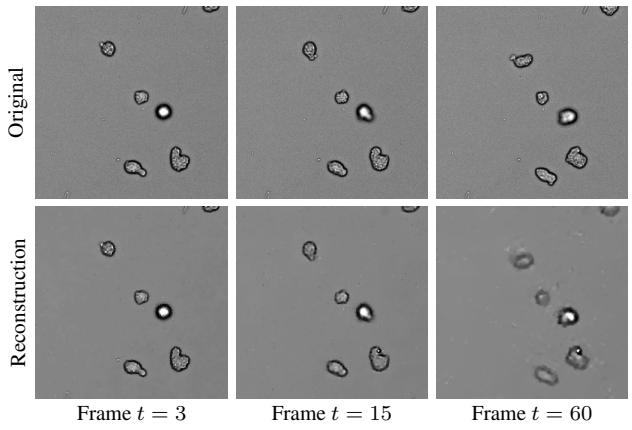

Fig. 3. Reconstruction result for the sequence *Amiba* with a uniform weight map ($w = 1$).

Simulations were conducted on the *Disks* sequence using 15% of magnitude Fourier measurements; on *Amiba*, we increased the sampling rate to 25% of Fourier samples to handle the more complex nature of the signal. In both cases, we used the first frame of the sequence as an input to initialize the process, and then we progressively recovered all the following frames (as described in Sec. 2.1).

3.2. Qualitative and quantitative results

Reconstructions obtained for the sequences *Disks* and *Amiba* are presented on Fig. 2 and Fig. 3. These results were obtained with a uniform weight map w for the regularization function $\|\cdot\|_{\text{TV},w,a}$. Fig. 4 presents the evolution of the frame-wise reconstruction error (measured as the mean squared error (MSE) between the original and reconstructed frames) as a function of frame index t (i.e. time).

These results show that the distortions introduced by the reconstruction method increase with time, i.e. distance to the initial keyframe: while in both sequences the reconstructed frames for $t < 10$ are quite similar to the original ones, errors become significant close to the end of the sequences, but exhibit different characteristics:

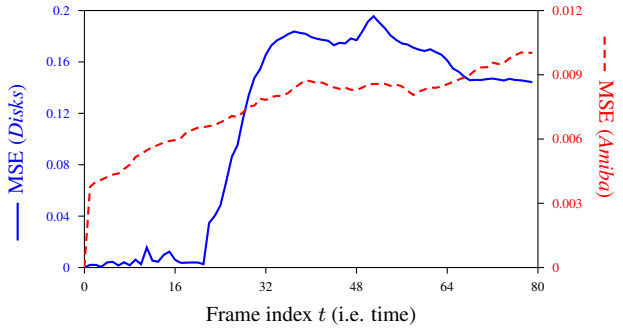

Fig. 4. Frame-wise MSE of the reconstructed video sequences *Disks* and *Amiba* versus time.

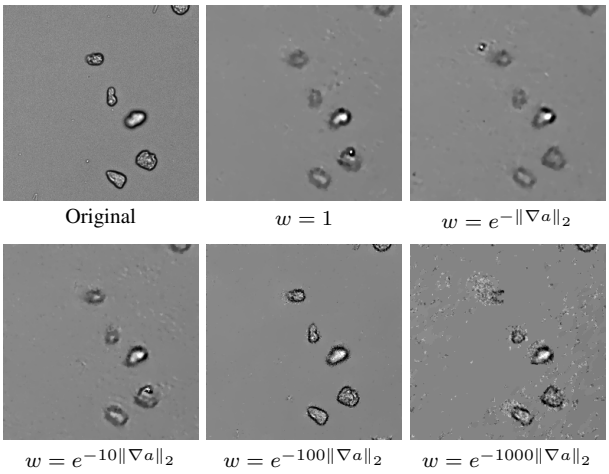

Fig. 5. Frame $t = 70$ of the sequence *Amiba* reconstructed with different spatially varying weight maps w.

- For *Amiba*, the MSE increases progressively and quite regularly with time t, which is characteristic of error accumulation. Visually, this results in an increasing blurring effect.
- For *Disks*, the MSE increases sharply at time $t = 22$, and then continues to grow over the next 10 frames, leading to a reconstruction that is completely inconsistent with the original sequence for $t > 30$. This behavior is due to the fact that the algorithm outputs an erroneous reconstruction at $t = 22$ which is then propagated. On the contrary, frames corresponding to $t < 22$ were almost perfectly recovered.

3.3. Weight map w

Results presented in Sec. 3.2 were obtained with the weighting parameter w set to a uniform value ($w(P) = 1$ for all pixels P). However, as mentioned in Sec. 2.2, using a spatially varying weight map designed such that $w(P)$ is small on pixels where edges are expected in the reconstruction should improve the results.

Different maps were tested with $w(P) = e^{-\kappa\|\nabla a(P)\|_2}$ where $\kappa > 0$ is a parameter and a represents the frame that precedes the one being reconstructed. The underlying assumption guiding this choice is that edges in the reconstructed frames are expected to be located close to the edges in a.

Results presented on Fig. 5 for *Amiba* show that a careful choice of the weight map can indeed reduce the reconstruction artifacts, although this improvement is not observable in terms of MSE. In particular, in the sequence obtained with $\kappa = 100$, we were able to remove the blurring effect. However, automatic calibration of the parameter κ remains challenging: for instance, all our attempts to reconstruct *Disks* using a non-uniform weight map (i.e. $\kappa > 0$) resulted in a degradation of the reconstructed sequence, compared to what we obtained with $w = 1$.

Other forms of weight maps w were also tested, such as $w(P) = \left(\eta + \|\nabla a(P)\|_2\right)^{-1}$ with $\eta > 0$ a regularization parameter, but led to unsatisfactory results and numerical instability.

4. CONCLUSION

In this paper, we presented a new CS imaging framework for video reconstruction from partial Fourier modulus measurements, as required for Optical Fourier sensing. The proposed reconstruction scheme is inspired by previous variational phase retrieval techniques, and introduces a novel regularization set exploiting an original weighted hybrid TV regularization energy. Preliminary results demonstrate the feasibility of video reconstruction from partial Fourier measurements deprived of phase information. Some issues remain with the setting of the specialized parameters involved in the introduced regularization constraint, to ensure robustness of the iterative temporal estimation process. Their tuning seems to require more supervision for injection of prior image-content information, which might be an issue for practical implementation of such CS imaging paradigm into a real scanning device.

5. REFERENCES

[1] Emmanuel Candès and Terence Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[2] Emmanuel Candès and Justin Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, June 2007.

[3] Thong T. Do, Lu Gan, Nam H. Nguyen, and Trac D. Tran, "Fast and Efficient Compressive Sensing Using Structurally Random Matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, 2012.

[4] David L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] Emmanuel Candès, Yonina C. Eldar, Deanna Needell, and Paige Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2010.

[6] James R. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.

[7] Jianwei Miao, Pambos Charalambous, and Janos Kirz, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. July, pp. 342–344, 1999.

[8] S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," *Review of scientific instruments*, vol. 78, no. 1, Jan. 2007.

[9] Yoann Le Montagner, Elsa Angelini, and Jean-Christophe Olivo-Marin, "Video reconstruction using compressed sensing measurements and 3D total variation regularization for bio-imaging applications," in *International Conference on Image Processing*, Orlando, 2012, pp. 917–920, IEEE.

[10] Mihailo Stojnic, Farzad Parvaresh, and Babak Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3075–3085, 2009.

[11] Yonina C. Eldar and Moshe Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[12] Jalal M. Fadili and Gabriel Peyré, "Total variation projection with first order schemes," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 657–669, Mar. 2011.

[13] Yurii Nesterov, "Gradient methods for minimizing composite objective function," Tech. Rep., Université Catholique de Louvain, 2007.

[14] Pierre Weiss, *Fast convex optimization algorithms: application to image restoration and change detection*, Ph.D. thesis, Université de Nice - Sophia Antipolis, 2008, document in French.