

FORMATION, SIMULATION AND RESTORATION OF HYPERTELESCOPES IMAGES

D. Mary¹, C. Aime¹ and A. Carlotti²

Abstract. This article first provides a historical and detailed introduction to the image formation models for diluted pupils array and their densified versions called *hypertelescopes*. We propose in particular an original derivation showing that densification using a periscopic setting like in Michelson's 20-foot interferometer, or using inverted Galilean telescopes are fully equivalent. After a review based on previous reference studies (Tallon & Tallon-Bosc 1992; Labeyrie 1996; Aime 2008 and Aime *et al.* 2012), the introductory part ends with a tutorial section for simulating optical interferometric images produced by cophased arrays. We illustrate in details how the optical image formation model can be used to simulate hypertelescopes images, including sampling issues and their effects on the observed images.

In a second part of the article, we address the issue of restoring hypertelescope images and present numerical illustrations obtained for classical (constrained Maximum Likelihood) methods. We also provide a detailed survey of more recent deconvolution methods based on sparse representations and of their spread in interferometric image reconstruction.

The last part of the article is dedicated to two original and numerical studies. The first study shows by Monte Carlo simulations that the restoration quality achieved by constrained ML methods applied to photon limited images obtained from a diluted array on a square grid, or from a densified array (without spectral aliasing) on a grid, are essentially equivalent. The second study shows that it is possible to recover in hypertelescopes images quasi point sources that are not only far outside the clean field, but also superimposed on the replicas of other objects. This is true at least for the considered pupil array and in the limit of vanishing noise.

¹ Laboratoire Lagrange, UMR 7293, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, Campus Valrose, 06108 Nice Cedex 02, France

² Princeton University, Mechanical & Aerospace Engineering, Olden Street, Princeton, NJ 08544, USA

1 Introduction

The History of Science is that of a continuous quest for a better understanding of Nature. In particular, the history of Astronomy reflects the breakthroughs which have led to our modern conception of the Universe.

As researchers, the participants of the Fréjus School are probably all aware of the uncountable efforts that must be spent on various tasks – huge bibliographical studies, long and difficult theoretical calculations or ultra sensitive, and thus ultra irritating instrumental experiments – in order to obtain a single useful, correct and well understood result. These efforts constitute an invisible but necessary sand, which is paved by the bright success stories published in the official History of Sciences.

Success stories often exert on researchers the ambiguous attraction of perfect things: their essence is so bright, exemplary and rare that it appears often discouraging, not to say ridiculous, to imagine any comparison between such major achievements and research of its own. On the other hand, these stories tell us that research efforts need to sum up substantially before major works really happen to culminate. Besides, past successes constantly diffuse, as a constant background Moon for research groping in the darkness, a faint light of scientific glory that shines down into the deepest and most obscure offices of every research laboratory. We start with two such stories, which are connected to the topics of this article, Hypertelescopes.

One of the fundamental concern of Astronomy, of which we can find traces in Mesopotamian, Egyptian, Greek and Arabic Astronomy, is that of high angular resolution. A higher resolution of celestial objects means that the objects can be better understood because they are better seen. High resolution once allowed Mankind to discover that the celestial objects change and move. But Mankind had to wait a very long time before it was able to prove by observations that planets do not move in perfect circles, and that stars are not fixed on an hypothetic celestial sphere beyond which, as Aristotle once wrote, nothing shall exist, not even space or time. This brings us to our first story.

Probably because his own observations of the stars' positions relatively to each other did not match those of his predecessors, and because he had guessed that stars' positions and magnitudes could be variable³, the Greek astronomer and mathematician Hipparchus (2nd Century BC) collected the positions and apparent magnitudes of about a thousand stars. This catalogue, which constitutes one of the most audacious legacy to future research, was transmitted to Arabic and European astronomers via copies of the *Almagest* of Ptolemy.

This major book⁴ was written four centuries after Hipparchus' time, in the 2nd Century AD. After the prestigious, several-century-old Library of Alexandria was

³According for instance to Halley (1715).

⁴Ptolemy's geocentric model of the Universe, using spheres and epicycles, was based on previous models from Eudoxus of Cnidus (4th Century BC), Apollonius of Perga (3rd Century BC) and Hipparchus. This model constituted the standard model and was continuously made more complex until N. Copernic (1473-1543).

definitively destroyed, the Almagest was copied, studied and enriched in various places and epochs during one millenium, from the new capital Constantinopolis of the Orient Roman Empire, through Arabic observatories as Maragha or the active translation centers of Andalusia.

Quite a few years after Hipparchus' times – in the 1710 s – the British astronomers Halley and Flamsteed could show, by comparing their observations to those of Hipparchus, that the position of Sirius in the sky had moved (by proper motion) of an angle of about one lunar diameter since the Hellenistic period (as well as Arcturus and Aldebaran (Mignard & Martin 1997)). After almost 2000 years, Hipparchus had won his bet: his observations had finally served his successors in proving facts he could only suspect. Patient and repeated high angular resolution observations eventually forced the minds to open on a Universe totally different from what the most brilliant scientists of the Antiquity could imagine.

The history of Hipparchus' catalog is not a success, it is an absolute and multiple triumph. First, contributing to the observational evidence that stars could be moving and variable was bringing strong arguments against the old vivid idea that they were inherently immobile and ever lasting; this discovery had deep, cosmological and philosophical implications about the size and nature of the Universe. Second, the careful attention paid by each link of this long chain to the knowledge of his predecessors (Hipparchus to the Mesopotamian tables, Ptolemy to Hipparchus, Al-Tusi, many other Arabic astronomers, Halley and Flamsteed to Ptolemy) exemplifies how new science may succeed in building carefully on the experience of the past. Third, the project realized by Hipparchus with his catalogue comes as a striking remembrance that research thought in the long term (several centuries in this case) is not necessarily wasted research.

The second story is also about breakthroughs in experimental high angular resolution and about progressive accumulation of knowledge. It starts one day of 1868, at the Academy of Sciences in Paris, where H. Fizeau reports about a treatise on the “directions of ether vibrations in polarized light”.

At the end of his reading, Fizeau mentions that interference fringes can arise from two interfering apertures only if the source has a very small angular dimension. “Hence”, Fizeau pursues, “to mention this briefly, we might hope that by using this principle, and by forming, for instance by means of two large separated slits, interference fringes in the focal plane of instruments aimed at observing stars, it might become possible to obtain some new insights on the angular diameter of these stars”⁵ (Fizeau 1868). The name of the author of the treatise reported by Fizeau was covered, and it is not reported in the Comptes-Rendus of the Academy. The treatise was deposited under the title: *Sine experientia nihil sufficienter sciri potest* – without experience, nothing can be known sufficiently.

Five years later, in a letter to Fizeau communicated by the same Academy, É. Stephan (1874) textually recalls Fizeau words quoted above, in which he recognizes “a totally new path that might lead to results otherwise inaccessible to

⁵The english translation is ours.

the methods currently available in Astronomy". He formalizes the principle of the corresponding experiment and reports the first tentative measurements of the diameter of Sirius (undertaken at Marseille). The following year, in 1874, Stephan reports to the same assembly extensive measurements from which he concludes that all observed stars have diameter (much) less than 158 milliarcseconds (158 mas, Stephan 1874). These experiments were achieved by placing two slits on an 80 cm aperture telescope. Fizeau's 1868 comment had led to the first generation of stellar interferometers.

As early as in the 1880 s, A. Michelson had used an interferometer of Fizeau's type to measure the diameters of four satellites of Jupiter (≈ 100 mas, Michelson 1891). But according to Michelson later on, the method was not tested on stellar objects for the thirty following years, probably for two reasons. First, the success of such experience was supposed to require ideal seeing conditions. Second, diameters of the order of 10 mas would have required a distance between the apertures of 10 meters or more – a size entirely out of question at that time (Michelson 1920).

In 1919 however, A. Michelson discovers by tests that fringes can be obtained even with bad seeing conditions (Michelson 1920); and J. Anderson manages in 1920 to measure the separation of two components of Capella's system (54 mas) on a 100-inch (2.5 m) reflector (Anderson 1920). Stimulated by these results, Michelson proposes in the same article to use a setting that is similar to that of Fizeau, but where a periscopic mounting is introduced. In this Michelson stellar interferometer, the holes or slits of the Fizeau-Stephan mask are replaced by apertures that can be moved on the same mounting, allowing much larger separations than Fizeau's setting. The apertures' beams are redirected on a smaller telescope, in the focal plane of which the interferences fringes appear.

The experimental discovery that relatively steady fringes could be obtained, along with the successes of the experiments of 1920 led to the building of the famous 20-foot (about 6 m) Michelson stellar interferometer at Mount Wilson. This second generation interferometer allowed to determine the diameter of Betelgeuse as 47 mas, within 10% (Michelson 1921).

About 50 years later, at the Observatory of Nice, a third generation of interferometers appeared when A. Labeyrie obtained for the first time, using two independent telescopes separated by 12 m, optical interference fringes on Vega (Labeyrie 1975). This experimental success ignited the modern developments of optical high angular resolution interferometry, whose most exploited instruments are today the Very Large Telescope Interferometer (VLTI) in Chile and the Center for High Angular Resolution Astronomy (CHARA) array at Mount Wilson. These systems allow to create interference fringes from 4 to 6 independent telescopes, which are cophased pairwise.

The high angular resolution optical systems described in this paper inherit from the characteristics of the stellar interferometers mentioned above, of which they constitute a further evolution. These systems use a possibly very large number of apertures (an array), which are simultaneously cophased. The interference pattern of the whole array is recorded so as to allow *direct* imaging of the objects. They can be used either in Fizeau mode (in which case the array can be seen

as a huge masked aperture) or in Michelson mode (where the relative size of the apertures is changed with respect to their separation by the periscopic setting, a process called *subpupil densification*). Densification can also be obtained by using inverted Galilean telescopes. *Hypertelescopes* (Labeyrie 1996) refer generically to densified interferometers.

Our second story, which started in 1868, is thus not over yet. Actually, it even reaches a critical point because next generation high angular resolution interferometric arrays are currently subject to in depth comparative studies. These studies will allow to choose which observing technology should be pushed in the next decades.

Three main types of next generation optical interferometric arrays emerge, the apertures' number and configurations of which will tend to be similar to current radioastronomical arrays (see M. Bremer's article on radiointerferometry in these proceedings). First, a direct extension of the VLTI system: few (possibly extremely) large telescopes that remain relatively compactly disseminated (in the 10^2 m range). Second, few relatively large (8 m) telescopes separated by kilometeric distances. Third, a large number (in the hundreds) of small telescopes disseminated on kilometeric distances.

In the three cases, such optical systems are expected to reduce our uncertainty and maybe to solve questions whose cosmological and philosophical implications are comparable to those evoked at the beginning of this Introduction. The first of those is the existence of life in distant stellar systems, but many other fields can reveal important discoveries, like stellar physics (through spatio-spectral studies of their atmospheres), or Active Galactic Nuclei (see on these issues the articles of A. Labeyrie, D. Mourard, M. Hadjara and J. Kluska in these proceedings).

The paper continues with issues and topics that are also echoed in several other articles of this volume. Section 2 describes more precisely the differences in the optical models of Fizeau and hypertelescopes configurations. We address in this section Michelson's periscopic interferometer. The Appendix derives the densification operated by inverted Galilean telescopes, and uses for that purposes results from the theory of light propagation that are detailed in the article of Aime in this Volume. Section 3 can be seen as a tutorial to numerically simulate these systems. Section 4 turns to methods aimed at improving the images recorded in the focal plane of such instruments, and proposes a survey of restoration methods based on sparse representations. Articles of these proceedings connected to optimization issues arising in image restoration are those of M. Bertero, C. Theys, and É. Thiébaud. A substantial part of A. Bijaoui's article is in addition dedicated to methods based on sparsity. Section 5 proposes a comparison of Fizeau versus hypertelescopes configurations in a specific case. Section 6 presents original simulations aimed at detecting an object that is small, faint, and far from a central object. The last section summarizes and concludes the paper.

2 The *Fizeau* and *hypertelescopes* configurations

In a Fizeau configuration, the ratio between the distance between any subpupil and their diameter is the same for the input and the output pupil (input and

output pupils are homothetic). In hypertelescopes this ratio is allowed to change. This is illustrated in Figure 1. In the hypertelescope configuration using inverted Galilean telescopes, the diameter of the output pupils is magnified relatively to their separations, which remain unchanged. In the “periscopic Michelson” hypertelescope configuration, the diameters are unchanged but the relative separation is smaller. Indeed, all three configurations are unchanged by applying a global arbitrary scaling factor.

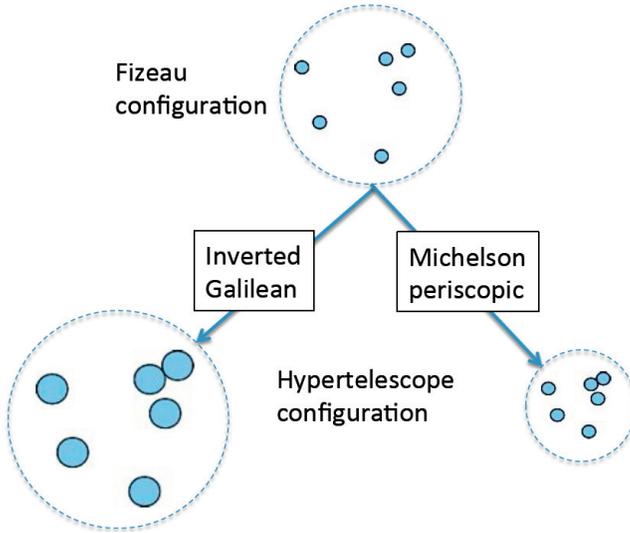


Fig. 1. Illustration of Fizeau (*top*) and hypertelescopes (*bottom*) configurations.

We will consider in this paper that the subpupils are all the same: circular, with diameter D . We will also consider that the atmospheric turbulence is negligible, and that the plane wave is monochromatic with wavelength λ .

We shall now see for both systems how we model the intensity distribution in the focal plane (the Image) that is obtained from a given celestial scene. The perfect (geometric) image of the celestial scene is called the Object. This section considers the continuous model (spatial and frequency variables are continuous). The discrete setting comes into play when images are sampled, and in numerical simulations of optical systems. Discretization will be addressed in Section 3.

The synthesis proposed below is a summary of Tallon & Tallon-Bosc (1992), Aime (2008) and Aime *et al.* (2012), papers to which we refer for a detailed treatment of the periscopic Michelson mode. The hypertelescope mode using inverted Galilean telescopes is detailed in the Appendix and leads to the same results.

2.1 Fizeau configuration

The image formation mechanism in Fizeau mode is described by the general and standard equations of convolutional optics. The Point Spread Function

(PSF, denoted by R) in its normalized form (*i.e.*, summing to one) can be written as a function of the angular coordinates $\beta(\beta_x, \beta_y)$ on the sky as:

$$R(\beta) = \frac{1}{S\lambda^2} \left| \widehat{P} \left(\frac{\beta}{\lambda} \right) \right|^2, \tag{2.1}$$

where superscript $\widehat{}$ denotes the Fourier Transform (FT), $\widehat{P}(\frac{\beta}{\lambda})$ is the scaled FT of the telescope aperture transmission $P(\mathbf{r})$ with $\mathbf{r}(r_x, r_y)$ the vector of position, λ is the light wavelength, S is the total surface of the telescope aperture. As for classical (monolithic) telescope, the cophased optical system acts as a linear (bandpass) filter in the Fourier space. The transfer function T of this filter is the FT of the PSF. By the Wiener-Kintchine theorem, T corresponds to the spatial autocorrelation function of the input diluted pupil. If we denote by $\mathbf{u}(u, v)$ the angular frequency vector, the normalized optical transfer function (OTF) $T(\mathbf{u})$ is defined by

$$T(\mathbf{u}) = \frac{1}{S} \iint P(\mathbf{r})P^*(\mathbf{r} - \lambda\mathbf{u})d\mathbf{r}. \tag{2.2}$$

The Object-Image relation, relating the object O to the image I^F in the Fizeau mode, is a convolution in the direct space

$$\underbrace{I^F(\beta) = O(\beta) \star R(\beta)}_{\text{Angular Convolution}}, \tag{2.3}$$

and a multiplication in the Fourier space

$$\underbrace{\widehat{I^F}(\mathbf{u}) = \widehat{O}(\mathbf{u}) T(\mathbf{u})}_{\text{Frequency Filtering}}. \tag{2.4}$$

Let us have a close look at the structure of the transfer function in the case of a diluted pupil composed of K subpupils of diameter D . In this case, the normalized OTF of each subpupil $T_0(\mathbf{u})$ (see Eq. (7.8)) corresponds to the autocorrelation function of a disk, which is sometimes called a “chinese hat” function. The support of $T_0(\mathbf{u})$ is a disk of diameter $2D/\lambda$.

Let the centers of the K subpupils be at spatial positions $\mathbf{r}_k, k = 1, \dots, K$. The autocorrelation function of the centers defines a set of central frequencies $\mathbf{u}_{kl} = (\mathbf{r}_k - \mathbf{r}_l)/\lambda$. The optical system composed of the diluted pupil samples frequencies located within a disk of radius $2D/\lambda$ around the central frequencies \mathbf{u}_{kl} . The transfer function of the diluted pupil can thus be written as

$$T(\mathbf{u}) = T_0(\mathbf{u}) + \frac{1}{K} \sum_{l=1}^K \sum_{k \neq l}^K T_0(\mathbf{u} - \mathbf{u}_{kl}), \tag{2.5}$$

where the double sum collects the contributions around the frequencies $\mathbf{u}_{kl}, k \neq l$.

The Fizeau image has frequency content

$$\widehat{I^F}(\mathbf{u}) = \widehat{O}(\mathbf{u})T(\mathbf{u}) = \widehat{O}(\mathbf{u}) T_0(\mathbf{u}) + \frac{1}{K} \sum_{l=1}^K \sum_{k \neq l}^K \widehat{O}(\mathbf{u}) T_0(\mathbf{u} - \mathbf{u}_{kl}). \tag{2.6}$$

The image I^F formed in the focal plane of a Fizeau interferometer is the inverse FT of (2.6). Clearly, a lot of information about O is missing in I^F because a lot of frequencies are zero in $\widehat{I^F}(\mathbf{u})$. While for monolithic telescopes the missing frequencies are the high frequencies (at low frequencies, the transfer function has no “hole” or zero value), for a diluted pupil the Fourier coverage may present voids in any frequency region. In the Fourier space, the transfer function may be seen as an “archipelago of emerged islands” in the middle of a black sea where no measurement is available (see Fig. 2, bottom images).

Here appears an aspect that is crucial for the comprehension of this problem: we see precisely what I^F is missing to be O . This calls for five remarks.

- 1. For a given frequency sampling (*i.e.* for a given pupil array) the problem of recovering O is both an interpolation and an extrapolation problem in the Fourier space. This suggests that the image restoration methods can be designed, or least interpreted, as methods controlling the way the voids of the Fourier space are filled in, while preserving the observed frequencies.
- 2. The restoration quality will be object- and sampling- dependent. To see this, imagine two objects observed via the same pupil array (fixed sampling). Assume the first object has the most energetic part of its frequency content in the support of $T(\mathbf{u})$, while the most energetic part of the second object falls in the “sea” (that is, outside the support of $T(\mathbf{u})$). Clearly, this array is good for the first object but bad for the second, or the first object is good for this sampling but not the second object. This dependance is true in general but fortunately, it cannot be arbitrarily uncontrollable. The reason is that most natural objects have their frequency contents mostly located at low frequencies. Hence, ensuring a fair coverage at low frequencies guarantees that at least some useful information will be sampled for most objects.
- 3. To best sample the object we would like to maximize the Fourier coverage of the pupil array (or, to refer back to our image, to minimize the area of the black sea). Hence, for a fixed number K , we would like to maximize the frequency support of $T(\mathbf{u})$ (ignoring the effects of noise). This leads to configurations that are called *non redundant* (Kopilovich & Sodin 2001): no spatial frequency is sampled more than once (except from the $\mathbf{0}$ frequency).
- 4. How many objects have the same frequency contents as that of I^F ? Remark that any object that has, outside the zero frequency, frequency content only “in the sea”, that is, outside the support of $T(\mathbf{u})$ create a totally flat (constant) image. Stated differently, this means that not only O leads to the observed image I^F , but also scaled versions of O plus *any* object that has frequency content outside the support of $T(\mathbf{u})$. This shows another very important point regarding the restoration we can hope to make from I^F : if we make no additional assumption on the geometrical properties of O , we can construct infinitely many different instances of objects that create the image I^F . Using the sole knowledge of I^F , the object remains thus unknown.
- 5. The informative frequency content that has been collected by $T(\mathbf{u})$ is (2.6), and the image that we obtain is the inverse FT of it. How many

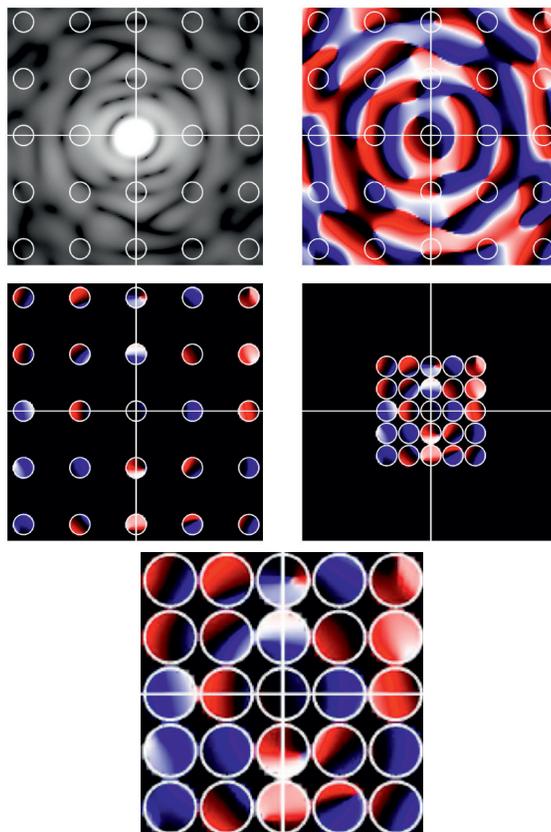


Fig. 2. Illustration of the Fizeau and hypertelescopes imaging properties in Fourier space. *Top left* (resp., *right*): absolute value (resp. phases) of the Fourier spectrum of the object. The white circles represent the boundaries of the elementary transfer function T_0 , *i.e.*, the zones inside which the spectral information is sampled by both systems. *Middle left*: the Fourier content of the Fizeau image. In the black region, no frequency is available. The centers of the circles are located at spatial frequencies \mathbf{u}_{kl} (see text). *Middle right*: the Fourier content of a Michelson image (*i.e.*, densified by a factor γ , full spectral densification is shown). The frequency content of each sampled disk of the Fizeau image has been translated by block, phases and moduli untouched. The centers of the circles are now located at spatial frequencies $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$. *Bottom*: the frequency sampling obtained with densification using inverted Galilean telescopes is a dilated (and thus fully equivalent) version of the Michelson mode with periscopic densification.

images contain the same contents of information about O as I^F ? To answer this, imagine that the frequency contents of I^F is modified in a reversible way, which preserves hermitian symmetry (thus ensuring that the corresponding

image is real) and which is further such that the resulting image is positive. For instance, appropriately permute the contents of some “islands”, or attribute the content of such island to an empty zone, and set the corresponding frequency content of the island to 0. There is a large number of such transformations: all transformations resulting in an OTF which is an autocorrelation function will work. From all the corresponding images (which will be very different from each other), we can take the FT (optically or numerically) and then numerically undo the transformation. We recover by doing so the originally sampled spectrum, and we can reproduce the image I^F . The Michelson densification by a factor γ in periscopic mode is one among such transformations: it leads to an image I_γ^P that is different from I^F , but the informational contents is the same, and the image I^F can be recovered from I_γ^P (at least for values of γ that are not too large). The Appendix shows that the densified image I_γ^G obtained using inverted Galilean telescopes has the same property.

2.2 Michelson configuration: Hypertelescopes

The *Michelson* configuration corresponds to a “densification” of the pupil because the diameters of the subpupils can be increased relatively to their separation. The degree of densification can be quantified by a densification factor called γ (Labeyrie 1996). We consider here the periscopic mode, in which the diameter of the subpupils D is fixed (see the Appendix for the case where the densification is obtained by dilation of the subpupils by a factor γ). The densification factor can in this case be defined as the ratio of the minimal distance between the subapertures before and after the densification: $\gamma = d/d'$. In the extreme case of maximal densification, some subpupils touch each other and are thus separated by a distance of $d' = D$, in which case $\gamma = d/D$.

The image formation model for the *Michelson* stellar interferometer in periscopic mode has been analyzed by Tallon & Tallon-Bosc (1992). The most important difference with the Fizeau configuration is that the image formation model is not a convolution anymore. In the Fourier space, the *Michelson* configuration involves a filtering corresponding to the diluted aperture before densification, followed by a translation of the frequency contents corresponding to the densification. During this translation, the spatial frequencies that are sampled by the input diluted aperture in a disk of width $2D/\lambda$ around frequency \mathbf{u}_{kl} are subsequently carried away, phase and modulus untouched, into a disk of same diameter but centered around the lower center frequencies $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$ (see Fig. 2).

The Fourier spectrum of the densified image I_γ^P is now (compare to (2.5) and (2.6))

$$\widehat{I}_\gamma^P(\mathbf{u}) = \widehat{O}(\mathbf{u}) T_0(\mathbf{u}) + \frac{1}{K} \sum_{l=1}^K \sum_{k \neq l}^K \widehat{O}(\mathbf{u} - \mathbf{u}'_{kl} + \mathbf{u}_{kl}) T_0(\mathbf{u} - \mathbf{u}'_{kl}), \quad (2.7)$$

where the term $T_0(\mathbf{u})$ is, as in Section (2.1) the elementary transfer function corresponding to one subaperture and K is the total number of subapertures. Each term of the double sum corresponds to the filtering, by the elementary transfer function centered at frequency $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$, of the object's spectrum translated by $\mathbf{u}'_{kl} - \mathbf{u}_{kl}$.

Again, the description in the Fourier space makes some important issues very clear:

- 1. We see that this frequency translation is a perfectly revertible transform⁶ as long as the zones around the new frequencies \mathbf{u}'_{kl} do not intersect. Hence, the Fizeau and Michelson images are actually equivalent. One difference arises however in presence of sampling: as visible in Figure 2, the densified image has a lower cut-off frequency than the Fizeau image. Hence, by the Shannon theorem, it may be sampled with less pixels than the Fizeau image.
- 2. When the translated frequency zones intersect, several frequencies melt into a single one at each point of the intersection zone. In this case, the transformation is not invertible, since there not a one-to-one mapping from the initial to the final frequency content. Because the disks corresponding to T_0 around \mathbf{u}_{kl} have width $2D/\lambda$, the lower center frequencies \mathbf{u}'_{kl} cannot be separated less than $2D/\lambda$ to avoid frequency overlap (aliasing). This means that the minimum separation d' between two subapertures in the densified pupil must not be less than $2D$ ($d'/\lambda \geq 2D/\lambda \Leftrightarrow d' \geq 2D$) to avoid information loss. The limiting case $d' = 2D$ is called FSP for Full Spectrum Densification in the literature. The case $d' = D$ (subpupils touching each other) is called FAD for Full Aperture Densification. This case indeed leads to frequency aliasing.

3 Numerical simulations of Fizeau and hypertelescopes interferometric images

3.1 Discretization and periodicity

Sampling and numerical simulations involve discrete approximations of continuous phenomena. We provide here a description of sampling issues which mostly relies on handy notions of Fourier analysis. A rigorous description of sampling theory requires to use distributions and Lebesgue integration, see *e.g.* Chap. 2, 3 and 5 of Mallat (2008).

The physical reference object is considered as constant (or sufficiently slowly varying) in time and as a continuous function of its space (or angle) variables. This object is of infinite resolution (the size of the smallest details in the object is vanishingly small).

⁶Hypertelescopes are an instance of transformations conserving the positivity while changing the image. Studying general properties of such transformations is an interesting point which is left out of the scope of this article.

In simulation, the discrete reference object we will consider is an approximation of this ideal reference. We shall assume that the discrete reference object has been obtained by a fine regular sampling from the reference. Let τ_r be the spatial sampling step. The discrete object is the multiplication of the continuous object by a Dirac comb of period τ_r . The numerical representation of the object assigns one number to each sampling cell $\tau_r \times \tau_r$, which is the pixel size. The discrete object is thus often represented as a “staircase” version of the ideal reference object, although a continuous version of the discrete object is indeed possible using other standard interpolation functions. This is illustrated in Figure 3.



Fig. 3. The discrete reference object (*middle*) is represented as a staircase (pixel) version of the ideal (continuous) reference (*left*). The approximation is visible when zooming (*right*): no detail smaller than the pixel size can be distinguished.

The discretization has indeed very deep implications on the image representations and processing. Let us think of the Fourier spectrum of the continuous reference object. This spectrum possess arbitrarily high frequencies because the smallest spatial structures can be arbitrarily small. Now let us consider the discrete reference object, obtained by multiplication of the continuous object with a Dirac comb. What is the Fourier spectrum of this object? By the convolution theorem, multiplication in one space translates to convolution in the dual space. The spectrum of the sampled object is the convolution of the true (infinite resolution) spectrum by the FT of the spatial sampling comb, that is, a Dirac comb of period $T_\nu = 1/\tau_r$. The Fourier spectrum of the discrete object is periodic, its frequency period is $T_\nu = 1/\tau_r$.

At this point we see that a discrete object has a continuous Fourier spectrum that is periodic. But, of course, this continuous spectrum cannot be stored as such in a numerical environment: it must be sampled. Well, the same reasoning as above can be applied to the periodic continuous spectrum. This spectrum is sampled with a step τ_ν in frequency, so it undergoes a multiplication by a Dirac comb with period τ_ν . What is the image corresponding to the resulting discrete, periodic spectrum? The multiplication by a Dirac comb in frequency results in a convolution by a Dirac comb in space, with period $T_r = 1/\tau_\nu$. Hence, if we consider discretization obtained by regular sampling, we end up with images and spectra that are periodic. Each period is sometimes called the “principal interval” (see Fig. 5). The Discrete FT (DFT) and its inverse compute the representations of

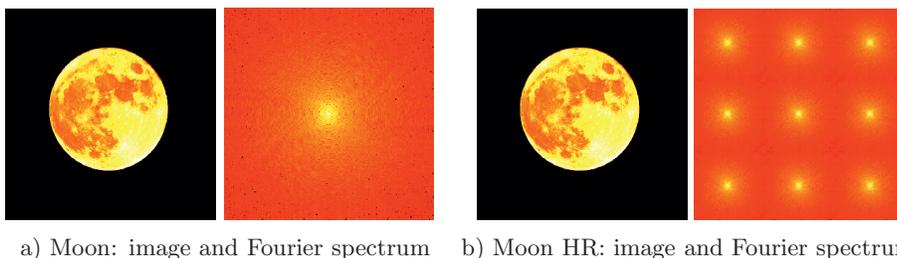


Fig. 4. a) Continuous-space reference object and corresponding continuous-frequency spectrum. The spectrum of the reference object is represented only for frequencies $\mathbf{u}(u, v)$: $|u| < T_\nu/2, |v| < T_\nu/2$ although of course the spectrum spreads at much higher frequencies. The discretized object and the corresponding (continuous) spectrum are shown in b). The spectrum is periodic: the Fourier space is paved with squares of size $T_\nu \times T_\nu$ that replicate the same continuous spectrum. Note that the continuous spectra in each such period are *not* equal to the continuous reference spectrum of the left figure, because the replicas on the right are obtained by superimposition of the spectrum of the left. Since the reference spectrum has no reason to be band-limited in a square of size T_ν (*i.e.* to be zero outside this square), higher frequencies contaminate the replica in the spectrum of the right (this is another instance of aliasing). If T_ν is high enough however, the continuous object will have little frequency content beyond T_ν , so that the replica of the periodic spectrum will be a good approximation of the reference spectrum at frequencies lower than T_ν . In all figures, the origin $(0, 0)$ in space and frequency variables are at the center of the image.

the object in both principal intervals. The DFT assumes that the discrete object and its spectrum are periodic, with a period of N points along each axis. The Fast FT (FFT) allows to compute the DFT in $\mathcal{O}(N \log_2 N)$ instead of N^2 additions and multiplications for monodimensional signals.

3.2 Numerical simulations of Fizeau and hypertelescopes images

The distinction made above between the continuous and discrete cases requires to differentiate the corresponding notation. In the following, a function that is continuous in its variable will be denoted by $f(t)$, and a discrete function by $f[n]$.

The first step is to choose a reference object. As we have seen above, this object must be discrete to be numerically manipulated, and is consequently an approximation of the corresponding continuous object. Anyhow, this discrete object (say, O) will become our reference. This object is composed of $N \times N$ pixels, with $N = 1024$ in all experiments below.

We now turn to the simulation of Equations (2.5) and (2.6) in the Fizeau mode, and of (2.7) in the Michelson (hypertelescope) mode⁷. To do this, we need to choose the subpupils of the array. They will be circular, with some diameter D .

⁷The numerical setting detailed here is essentially the same as that of Aime *et al.* (2012).

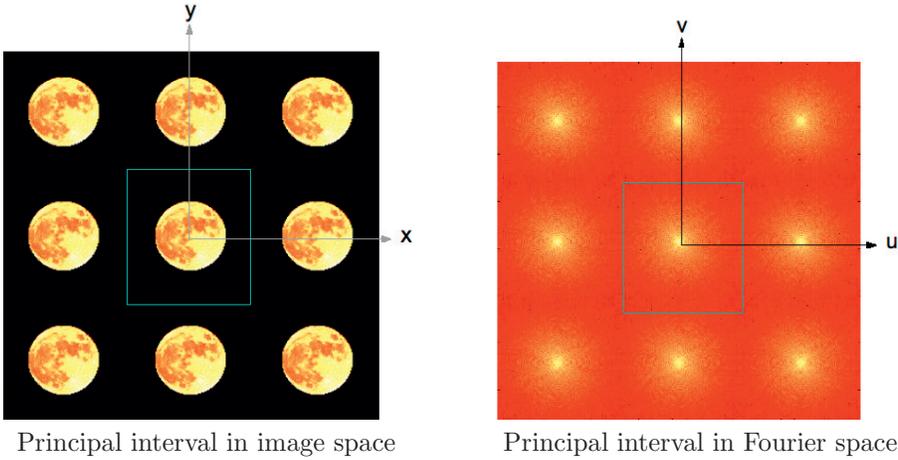


Fig. 5. The principal interval of a discrete reference object in the direct (*left*) and Fourier (*right*) space. Discrete objects of finite spatial and spectral extensions are made periodic when both spaces are related by the DFT. The size of the square cell in the direct space \mathbb{T}_r is related to the numerical resolution in frequency (sampling step): $\mathbb{T}_r = 1/\tau_\nu$. The size of the square cell in the Fourier space \mathbb{T}_ν is related to the numerical resolution in space (sampling step): $\mathbb{T}_\nu = 1/\tau_r$. The number of points N along each axis is $N = \frac{\mathbb{T}_r}{\tau_r} = \frac{\mathbb{T}_\nu}{\tau_\nu}$.

The center positions of the subpupils in the array are shown in Figure 7, along with their autocorrelation function which gives the central frequencies \mathbf{u}_{kl} defined in Section (2.1). In the Fourier domain, we have seen that the spectra are sampled on a square grid of step-size τ_ν . The transfer function of an elementary circular pupil is the continuous-frequency function $T_0(\mathbf{u})$. When sampled at step-size τ_ν , this function becomes a set of weighted discrete Dirac. In the present case, the support of T_0 spreads essentially over 9 samples (Fig. 7, bottom).

In the considered space-continuous array, the centers of the subpupils are located on the nodes of an integer grid, so the spatial frequencies \mathbf{u}_{kl} defined in Section (2.1) remain, once sampled, on a regular grid. While this may generally not be the case in practice, this setting makes the modeling of the densification easy.

Figure 8, top row, shows the locations where the Fourier samples of images are nonzero in Fizeau mode (left), in Michelson mode with full spectral densification (subpupils' centers are at least separated by $2D$), and in almost full aperture densification (some subpupils are almost touching each other). All samples are located on a 2-dimensional square grid with step-size τ_ν . A white sample means a nonzero sample. A fully white map would indicate that the N^2 samples of the DFT are available, so that the discrete reference object would be perfectly recovered by direct Fourier inversion.

Figure 8 is an illustration of what Equations (2.5)-(2.6) and (2.7) become in our setting. This Figure shows the discrete equivalent of Figure 2. The top left

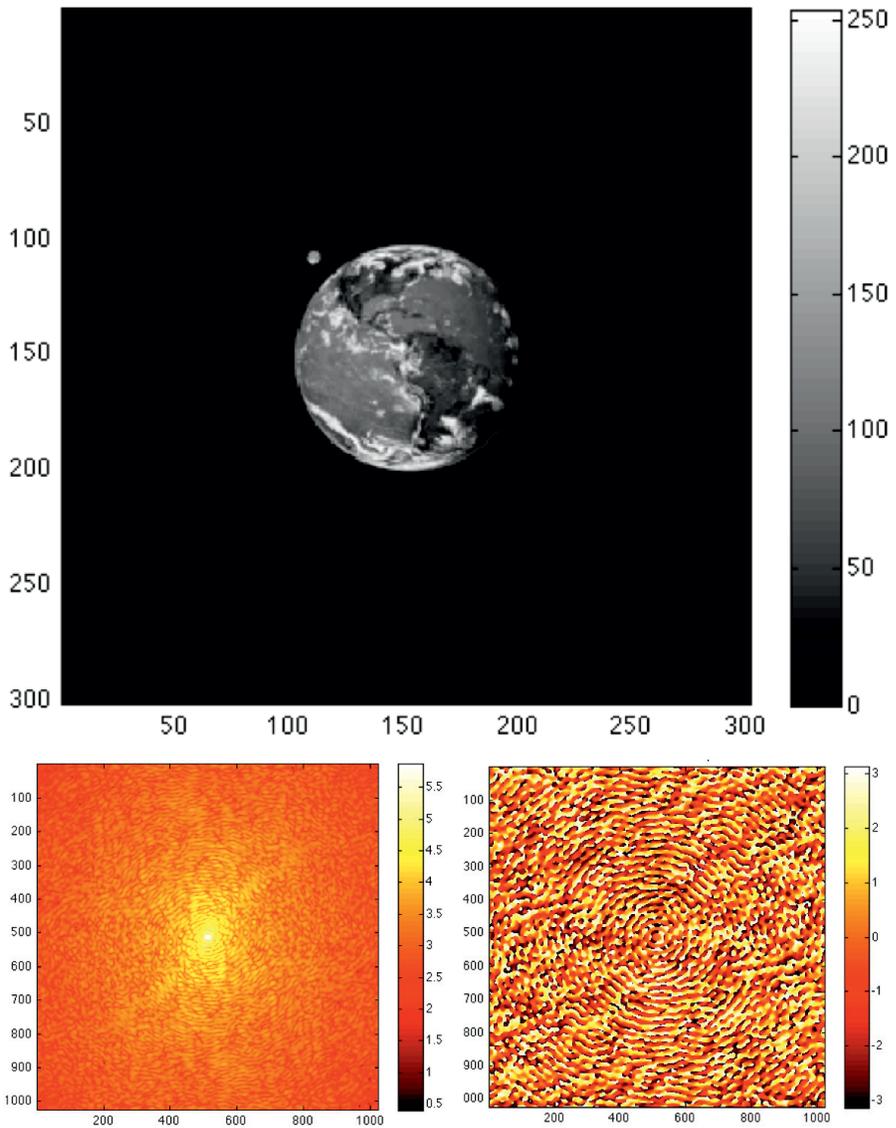


Fig. 6. *Top:* example of a (zoomed) reference object $O[\alpha]$. *Bottom, left:* moduli of $\widehat{O}[\mathbf{u}]$. *Bottom, right:* phases of $\widehat{O}[\mathbf{u}]$.

figure shows the support of $T(\mathbf{u})$ in Equation (2.5), once sampled. The sampling “islands” discussed in Section 2.2 for the Fizeau mode are visible in the figures of the second row, as little light squares in the middle of a dark sea of zero samples. These islands correspond to the frequencies sampled around the frequencies \mathbf{u}_{kl} and \mathbf{u}'_{kl} by the elementary OTFs (disks, which once sampled give rise to sets of 9 samples).

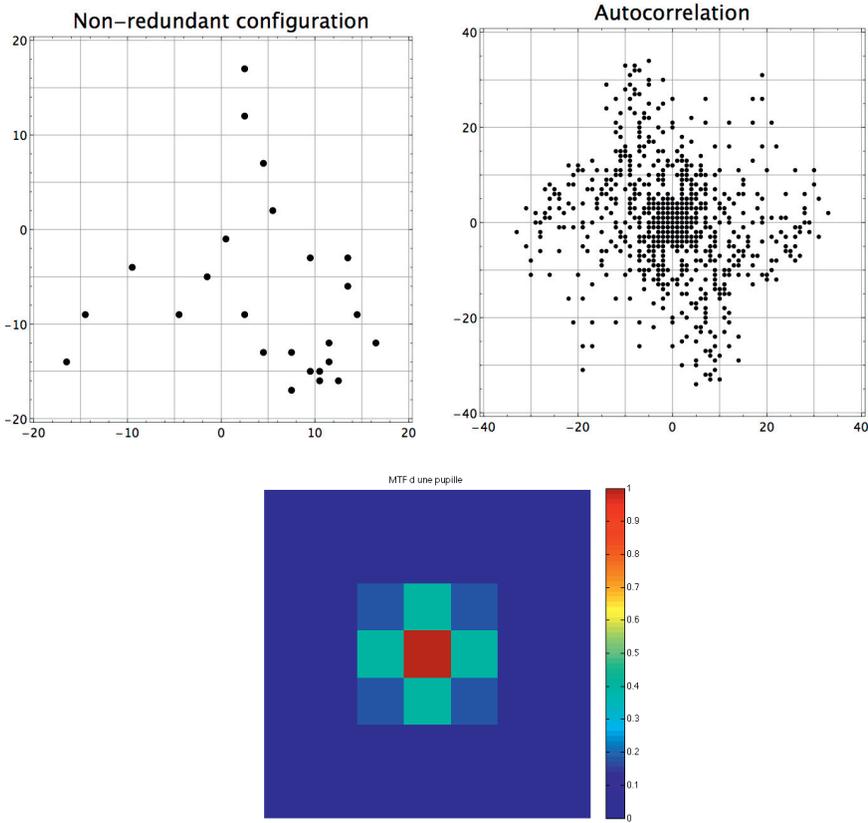


Fig. 7. *Top, left:* position of the centers of the subpupils in the array. These centers are all located at nodes of an integer grid. *Top right:* autocorrelation function of the centers of the subpupils. This function gives the central frequencies \mathbf{u}_{kl} defined in Section 2.1. *Bottom:* sampled OTF $T_0[\mathbf{u}]$ of an elementary aperture.

In the Michelson mode (hypertelescopes), the nonzero frequency contents are moved block-wise towards the frequency origin, and the center frequencies \mathbf{u}_{kl} become $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$. In the considered Fizeau mode the minimum separation between two center frequencies \mathbf{u}_{kl} is $7\tau_\nu$. In the FSD mode, for which the translated contributions of the elementary transfer function touch each other, the minimum separation in frequency between the \mathbf{u}'_{kl} is $3\tau_\nu$, so $\gamma = 7/3$. If we increase the densification, the contributions of the elementary transfer function overlap. In the considered (quasi)FAD case, $\gamma = 7/2$. The densification factors can easily be translated in terms of subpupilar distance. In the FSD mode, $d' = 2D$ and in the (quasi)FAD mode $d' = 4/3D$ (some subpupils almost touch each other). Figure 9 summarizes the principle for simulating the formation of hypertelescope images.

We see that starting on integer grid for the Fizeau configuration allows easily to obtain the frequency content for a set of Michelson configurations by simply

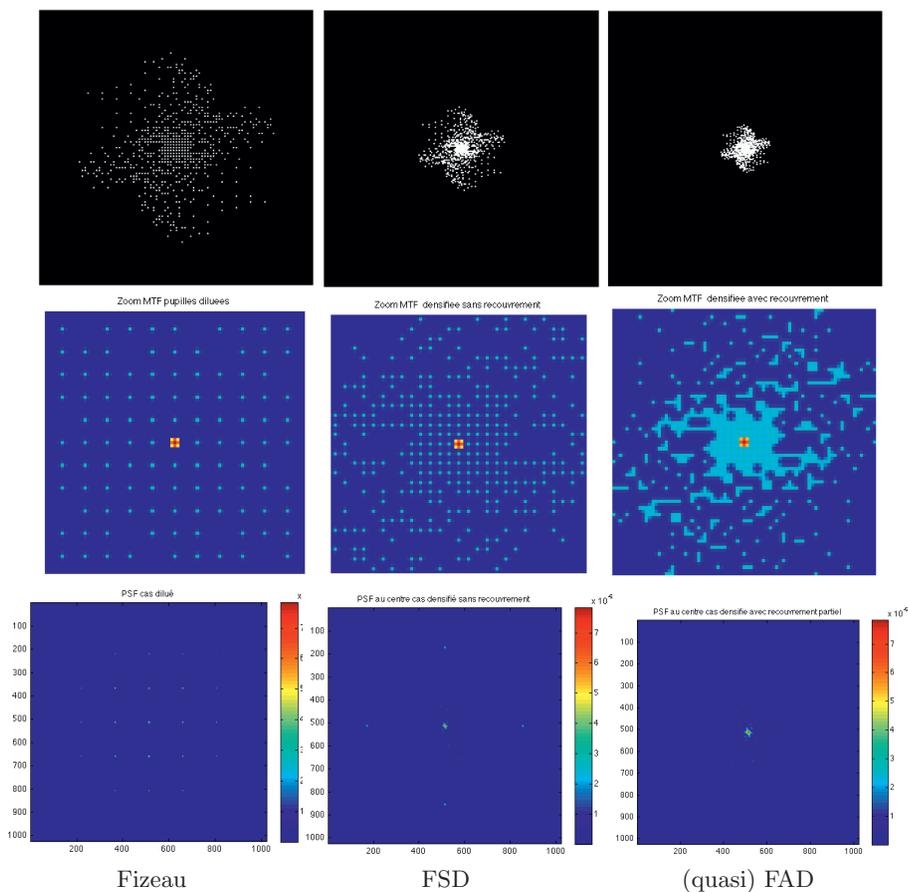


Fig. 8. *Top row:* compared locations of the nonzero Fourier samples that create the observed images in the three settings (Fizeau: *left column*; Michelson FSD: *center column*; Michelson FAD: *right column*). *Middle row:* zooms of the central parts of the top images. *Bottom row:* response to a point source on the optical axis (PSF).

removing the appropriate lines and columns of zeros. In the FSD example, the frequency content in each interval of the FSD image is obtained by removing 3 consecutive lines out of 7, and this process is repeated periodically in the rows $[u]$ and the columns $[v]$ with a period of 7. The outer region of each interval is then zero-padded so that the total number of N^2 samples is conserved. The samples of the densified image are thus again on an integer frequency grid (of same frequency resolution τ_ν).

3.3 Worked-out examples

Examples of images corresponding to the three considered sampling schemes are shown in Figure 10.

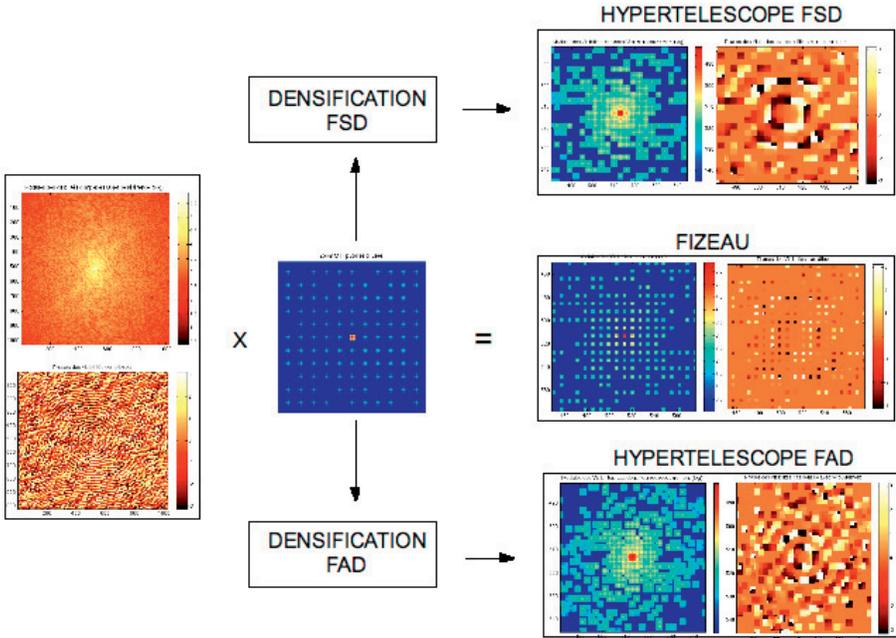


Fig. 9. Simulation of hypertelescope images. The discrete FT of a discrete reference object (*left*: moduli and phases, principal intervals are shown) is multiplied by the transfer function of the pupil array. The densification leads to the frequency contents (moduli and phases) shown on the right. The corresponding images are obtained by inverse DFT.

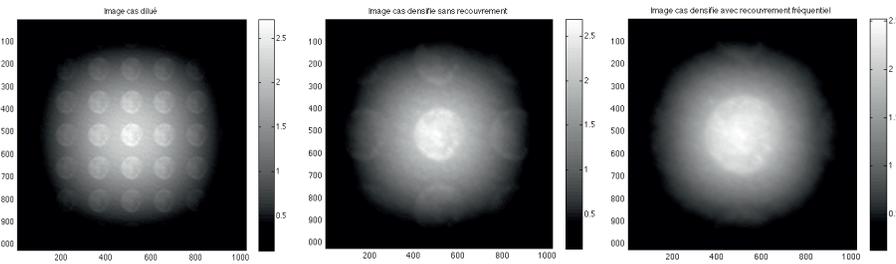


Fig. 10. Examples of images formed in the focal plane for 25 pupils on a grid. *Left*: Fizeau configuration. *Middle*: Michelson FSD. *Right*: Michelson quasi FAD.

There are several important aspects to be noted here.

- 1. First, all images are darker in the outer region. This is caused by the elementary contribution of each subaperture (which are all the same here). Let us consider the Fizeau image. In the Fourier space, the transfer function T is obtained by convolving the autocorrelation function of the pattern

created by Dirac impulsions (Fig. 7, top right) with the elementary transfer function T_0 . This convolution acts in the image space as a multiplication by the FT of T_0 , that is, by the PSF of the subapertures. What we see here is the diffraction envelope corresponding to the subapertures: imaging is done *inside* the pupils' PSF.

The width of this diffraction envelope defines the total field, which is a disk with diameter about $2.4\lambda/D$. In the Michelson mode this diffraction envelope is obviously visible as well.

In the numerical modeling/sampling, the principal interval should thus have extension $\mathbb{T}_r = 1/\tau_\nu$ that is about this size, $2.4\lambda/D$, since this is the zone we are interested in.

- 2. A striking particularity of these images is the presence of replicated patterns that resemble the reference object. The reason is that the pupils centers are located on a grid. The elementary OTF are in turn located on nodes of the frequency grid, which are here separated by 7 units, that is, by $7\tau_\nu$. If there were not 25 but many more pupils, so that all these nodes would correspond to the center of an elementary OTF, the sampling function would consist in a Dirac comb (of period $7\tau_\nu$) convolved by the elementary OTF T_0 . We see that the effect of this system would be, in the image space, to convolve the object intensity distribution with a Dirac comb of period $1/(7\tau_\nu) = \mathbb{T}_r/7$ (and the diffraction envelope would further tapper the result). The effect of this convolution would be to create, in a zone of extension $\mathbb{T}_r \times \mathbb{T}_r$, 7 replicas in both horizontal and vertical dimensions. This is essentially what we see in the Fizeau image of Figure 10. The difference is that for the Fizeau mode there are only 25 pupils. In this case, the Fizeau sampling function can be seen as the previous one multiplied by a function of ones and zeros which kills the frequencies where Fizeau's transfer function has no contribution, and leaves the others unchanged. This function is shown of Figure 8, top left, where the 1 are in white and the 0 in black. It has no regular shape. The Fizeau image is the convolution of the 7 replicas (the image obtained with the full grid) by the inverse FT of this function. The convolution by this "halo function" leads to an irregular and diffuse halo which explains the fuzziness of the Fizeau image. This halo blurs the image and removes some frequency contents (essentially the high frequencies, but also low frequencies in the voids of the sampling function).
- 3. Let us now turn to the Michelson configuration images. Densification translates frequencies block-wise, and thus performs a frequency modulation. This operation is not equivalent to downsampling. Downsampling by a factor k (*i.e.*, keep every k^{th} other sample) contracts the frequency axis uniformly. Each spatial frequency \mathbf{u} is moved to the frequency \mathbf{u}/k , so that the object whose spectrum is subsampled appears zoomed (and possibly aliased) by a factor k . In densification, the frequency axis is not uniformly contracted. Only the center frequencies $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$ are contracted by a factor γ . For all other frequencies, the result of the translation is only approximately a

contraction by γ . Densification is equivalent to downsampling (followed by lowpass filtering) only in the limit of point pupils: in this case T_0 tends to a Dirac impulsion and only central frequencies \mathbf{u}'_{kl} are sampled. The densified image is not a zoomed version of the Fizeau image. It may be considered as a zoom only in a first approximation (by neglecting the spatial extension of the pupils). This effect nevertheless explains why the replicas appear larger in densified images.

The magnification performed by densification is visible in the FSD image: in the Fourier space, 3 samples out of 7 are kept in the $[u]$ and $[v]$ frequency directions (see Fig. 8), resulting in a magnification of approximately $7/3$ (≈ 2.3) of the size of the replicated pattern. Needless to say, this magnification zoom comes with no gain in resolution at all (the frequency information about the reference objet is the same in both cases, as visible in Figure 2, and in Figure 8, middle left and middle center figures).

The number of replicas is decreased by the same amount $\gamma = 7/3$, because the nodes of the grid on which the frequencies $\mathbf{u}'_{kl} = \mathbf{u}_{kl}/\gamma$ fall are now $\gamma = 7/3$ closer than in the Fizeau case (they are separated by $3\tau_\nu$ instead of $7\tau_\nu$). Thus, the corresponding periodicity in the image space is now $1/(3\tau_\nu) = \mathbb{T}_r/3$: 3 replicas are visible in the vertical and horizontal directions. As in the Fizeau case, the sampling does not yield elementary OTF centered at all the nodes of the frequency grid of step $3\tau_\nu$. Hence, the replicas are convolved by the inverse FT of the 0/1 function shown Figure 8, top middle. This function is not a scaled (contracted) version of the corresponding Fizeau sampling function, unless the subpupils have negligible diameter. So, we see that the densification by a factor γ yields an image which, only in the limit of very small subpupil diameters, corresponds in the diffraction envelope to a zoomed (magnified) version of the Fizeau image.

- 4. In the FAD image finally, mainly one replica is visible in the center (actually two halves exist in the borders). This case seems to be a straightforward limiting case of the middle image: increased densification, larger magnification of the center replica, almost total disappearance of off-axis replicas. This is however not the case. The frequency contents in FAD mode has been modified (reduced) because of the too strong spectral densification (Fig. 8, right). During the FAD operation, some frequency cells have collapsed. Overlapping spatial frequencies have been melted, and the overall support size is smaller in the FAD than in the Fizeau/FSD cases. This operation is not invertible. Because of these reasons, it seems better for both purposes of direct imaging and of image restoration to stop the densification at the FSD limit. For direct imaging, FAD and quasi FAD produce images of reduced fidelity with respect to the object. For restoration purpose, FAD increases the difficulty of the inversion, because it adds un-invertibility to the imaging system.
- 5. The left figure in Figure 10 is the convolution of the reference object (Fig. 6) by the PSF shown in Figure 8, bottom left. However, as discussed in Section 2.2, the FSD and quasi FAD images are not the convolution of the

reference object by the PSF shown in Figure 8, bottom center and bottom right. A convolution may be retrieved in the FSD mode only in the limit of vanishingly small diameters (infinite fields of view). This suggests two regimes for densification (hypertelescopes), depending on whether the diameter D is much smaller than the smallest subpupil separation d or not. For hypertelescopes made of very large bases (in the kilometer range) and of many small telescopes (centimeters), $D \ll d$, and a convolution model may be a good approximation, at least close to the optical axis. For VLTI-like hypertelescopes, made of moderately large bases (in the hundreds of meter) and of a few large telescopes (in the tens of meters), $D \approx d$, and the image formation models strongly depart from convolution.

3.4 Noise

Real images will be affected by several perturbations. Effects caused by perturbations on the phase and by chromaticity are not addressed here. We consider two types of noise: Poisson noise (the number of detected photons in a pixel receiving a constant light flux is Poisson distributed), and Gaussian noise (which usually models the detector read-out noise, or approximates the Poisson distributed in the limit of large fluxes). The Figure below shows examples of simulated noisy data images that would be obtained for a hypertelescope. The image in FSD mode (left) is sampled on a 1024×1024 pixels CCD detector. The average flux falling on each pixel corresponds to 0.8 photons. The middle image is what the detector sees with Poisson noise (the recorded number of photons in a pixel is the realization of a Poisson process having for mean and variance the noiseless flux on this pixel). In this case the detector noise is negligible with respect to the photon noise. On the other hand, the right image is an instance of what the detector records with a zero mean Gaussian noise having a standard deviation of 1 photon.

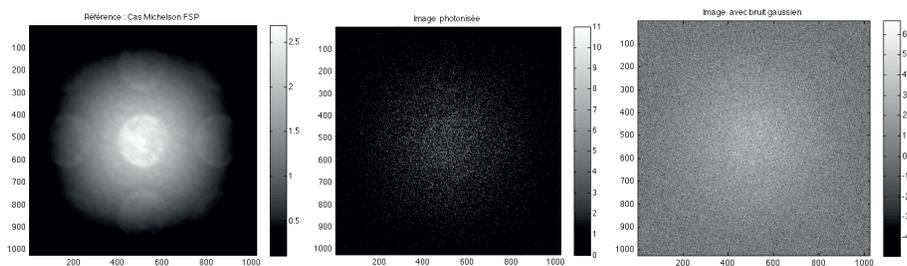


Fig. 11. *Left:* noiseless FSD hypertelescope image. *Middle:* corresponding photonized image. *Right:* noisy image with Gaussian noise.

Clearly, the data images in these cases are quite degraded versions of the noiseless image. However, this visual quality loss is partly illusory, because the noise component in the middle and right images have frequencies in the whole Fourier

space – while we know that our imaging system has measured only the frequencies that belong to the support of the transfer function T . Hence, we can safely (and numerically) filter out all the frequencies outside this support without any degradation of the astrophysical information. This is simply achieved by a DFT (or FFT) of the noisy image, multiplication the result by the corresponding indicator function of Figure 8, top row, and inverse FFT. We obtain the images of Figure 12.

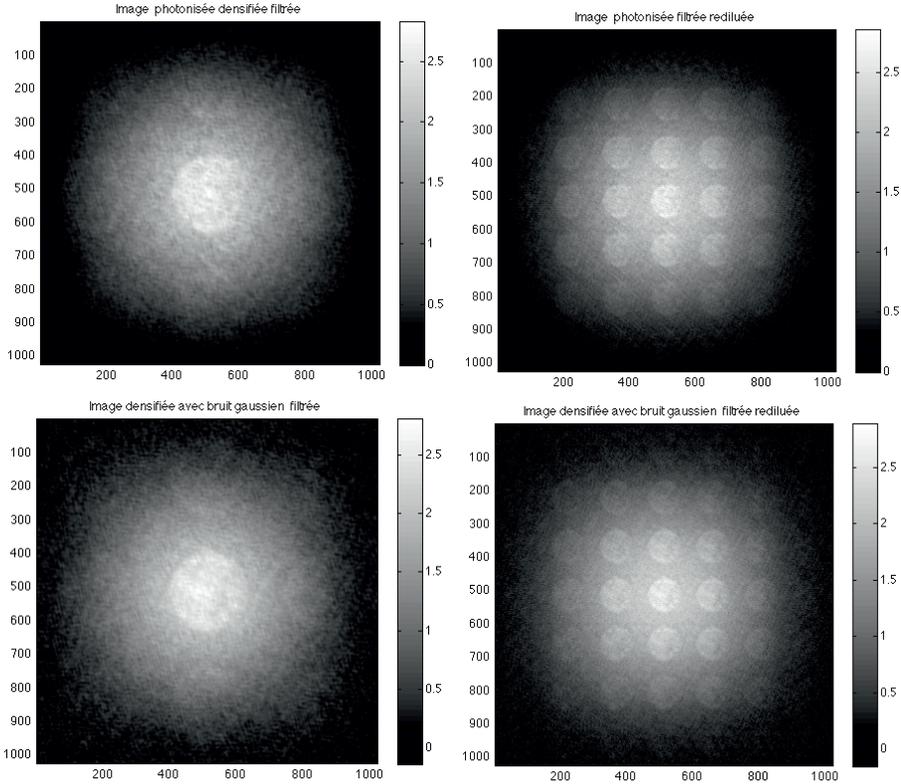


Fig. 12. *Top row, left:* filtered version of the photonized FSD hypertelescope image of Figure 11, middle. *Top row, right:* corresponding numerically rediluted image. *Bottom row, left:* filtered version of the FSD hypertelescope image with Gaussian noise of Figure 11, *right.* *Bottom row, right:* corresponding numerically rediluted image.

Clearly, the noise is much less adversarial than suggested by the data image. Note also that a numerical post-processing allows to create, from densified image, images that would have been obtained (up to the noise realization) with the corresponding Fizeau imaging setting. This is simply achieved by computing a FFT, translating back the frequency samples to the location they occupy in the Fizeau sampling, and computing an inverse FFT. This is interesting because now

the model relating the object to the rediluted data is again a convolution, and this allows to use classical deconvolution algorithm even in the densified case.

This example shows that the image quality (*i.e.*, its closeness to the object) can be improved by post-processing. This is the topics of the next Section.

4 Restoration algorithms

This section focuses on methods aimed at improving the estimation of the object from the image. What we want to do here is to infer from the data image, and from the mathematical model of the image-object relation, which object has generated the data. The process of finding back O from I^F , I_γ^P or I_γ^G is sometimes given the catchy name of “inverse crime” in the literature.

We first describe the inversion problem, and then illustrate three classical deconvolution algorithms. The last part proposes a review of a particular class of restoration methods that have received an increasing amount of attention in the last decades. These methods use the notion of sparse representations.

4.1 Inversion

Because restoration algorithms involve digital data and filtering techniques, we will consider from now on a fully discrete model. In this model, let us write the data image and the unknown reference object as vectors \mathbf{y} and \mathbf{o} respectively. These vectors are simply vectorized versions of the discrete arrays containing the intensity values in each pixel of the image and reference object. Considering principal intervals, the reference object and the image have N pixels, and their discrete Fourier spectra have N frequencies. In the case of subpupils placed on an integer grid, the sampled spatial frequencies are also on a grid. The noiseless image formation model in the Fizeau case becomes

$$\mathbf{y}_f = \mathbf{F}^\dagger \mathbf{T} \mathbf{F} \mathbf{o} = \mathbf{H}_f \mathbf{o}, \quad (4.1)$$

where \mathbf{F} is the matrix form of the DFT, superscript[†] denotes conjugate transpose, $\mathbf{T} = \text{diag}\{T[1, 1] \dots T[N, N]\}$ is the diagonal matrix representing the transfer function of the diluted array, and \mathbf{H}_f is a discrete circular convolution operator (a circulant matrix, because the considered discrete images and spectra are N -periodic). As we have now understood, the inversion is impossible because the solution is not unique: an infinity of objects can lead to the data, because of the zeros of the transfer function. In addition, the transfer function may be nonzero but very small at some frequencies, and the frequency content of the object at some sampled frequencies may be very small relatively to noise. In such cases, the data samples contain essentially noise. Inverting the transfer function at such frequencies leads in the image to fake oscillatory components that can have large amplitudes, a phenomenon called noise amplification. For these reasons, our inversion problem is said to be *ill-posed*, and the illness indeed comes from the instrument, not from the mathematical formulation.

The principal problem of the restoration can be seen as filling “cleverly” the zeros of the transfer function. Cleverly means that the recovered frequencies should, in some sense that remains to be defined, be close to the original frequencies of the object.

The noiseless image formation model in the case of hypertelescopes without spectral overlap (*i.e.*, for densification up to FSD or less) becomes

$$\mathbf{y}_{FSD} = \mathbf{F}^\dagger \mathbf{M}_{FSD} \mathbf{T} \mathbf{F} \mathbf{o}, \quad (4.2)$$

where \mathbf{T} is the Fizeau transfer function, and \mathbf{M}_{FSD} is the operator implementing the spectral densification (frequency modulation). This operator is a permutation matrix whose 1 specify the positions to which the frequency samples of the input (Fizeau) image are assigned in the Michelson image. This operator is linear, non diagonal (each column and each row of \mathbf{M}_{FSD} have exactly one 1 and $N - 1$ zeros) and obviously invertible ($\mathbf{M}_{FSD}^{-1} = \mathbf{M}_{FSD}^t$: transposing yields the inverse). The operator $\mathbf{F}^\dagger \mathbf{M}_{FSD} \mathbf{T} \mathbf{F}$ does not correspond to a convolution because \mathbf{M}_{FSD} is not diagonal (convolution is diagonalized by the Fourier transform).

If, however, we redilute the densified image

$$\mathbf{y}_{redil} = \mathbf{F}^\dagger \mathbf{M}_{FSD}^{-1} \mathbf{F} \mathbf{y}_{FSD} = \mathbf{F}^\dagger \mathbf{M}_{FSD}^{-1} \underbrace{\mathbf{F} \mathbf{F}^\dagger}_{\mathbf{I}} \mathbf{M}_{FSD} \mathbf{T} \mathbf{F} \mathbf{o} = \mathbf{H}_f \mathbf{o}, \quad (4.3)$$

from which we see that rediluting allows to retrieve the convolution model (4.1).

In the case of a hypertelescope with spectral overlap (or aliasing), that is, in the range of densification from FSD to FAD, the model is

$$\mathbf{y}_{FAD} = \mathbf{F}^\dagger \mathbf{M}_{FAD} \mathbf{T} \mathbf{F} \mathbf{o} = \mathbf{H}_{FAD} \mathbf{o}, \quad (4.4)$$

where \mathbf{M}_{FAD} is not a permutation matrix and is not invertible anymore. \mathbf{H}_{FAD} is not a convolution operator, and we cannot redilute this image because \mathbf{M}_{FAD} is not invertible. In the following, we consider only cases that can be described by a convolution: the Fizeau configuration, and spectrally densified images, without aliasing, and further rediluted. This model will be generically denoted by $\mathbf{y}_0 = \mathbf{H} \mathbf{o}$, and we will consider perturbations on \mathbf{y}_0 caused by Gaussian and Poisson noises.

We now turn to some standard methods aimed at estimating \mathbf{o} from noisy data \mathbf{y} . In estimation theory, the Maximum Likelihood (ML) method is a systematic method aimed at building estimators of parameters considered deterministic. The likelihood of the data is assessed using the model (image-object relationship in our case), and is defined as the probability of observing the data conditioned to the parameters. The (unconstrained) ML method looks for the value of the parameters (the object, in our case) that is the most likely given the data. Unconstrained ML is extremely popular in the context of multiple measurements of the same parameter, because it is often asymptotically (in the number of measurements n) unbiased⁸

⁸This means that if θ denotes the parameter of interest and $\hat{\theta}_{MV}[n]$ its ML estimate using n measurements, $\lim_{n \rightarrow \infty} \mathbb{E} \hat{\theta}_{MV}[n] = \theta$.

and consistent⁹. In addition, if an estimator that achieves the Cramer-Rao lower bound¹⁰ exists for a finite number of measurements, the ML finds it.

In the framework of images however, the problem is not posed in terms of multiple measurements (if the data image size increases, the number of parameters increases as well). Moreover, the unconstrained ML leads generally, for non invertible operators \mathbf{H} , to dramatic noise amplification. Consequently, the sought solution must somehow be constrained for the inversion to be possible. The most obvious constraint is to impose that the object \mathbf{o} is non-negative ($\forall i, \mathbf{o}_i \geq 0$). The two following methods use this constraint, for Poisson and Gaussian data likelihoods respectively. They are relatively popular in the astronomical and optical communities where they have been published, and are synthetically exposed below. A detailed and unified treatment of regularized maximum likelihood methods with non-negativity constraint can be found in Lanteri *et al.* (2002a,b).

4.2 Richardson-Lucy algorithm (1972, 1974)

In the case of Poisson noise, the statistical model is $\mathbf{y} = \mathcal{P}(\mathbf{H}\mathbf{o})$. The likelihood of the data \mathbf{y} is

$$\mathcal{L}(\mathbf{y}; \mathbf{o}) = \prod_{i=1}^N \frac{([\mathbf{H}\mathbf{o}]_i)^{\mathbf{y}_i}}{\mathbf{y}_i!} e^{-[\mathbf{H}\mathbf{o}]_i}, \quad (4.5)$$

where the product comes from the independence of the components, guaranteed by that of the noise realization from one pixel to another. By using Stirling's formula, maximizing the likelihood above leads to minimizing

$$J_{Poisson}(\mathbf{o}) = \sum_{i=1}^N [\mathbf{H}\mathbf{o}]_i - \mathbf{y}_i \ln[\mathbf{H}\mathbf{o}]_i, \quad \text{subject to } \forall i, \mathbf{o}_i \geq 0, \quad (4.6)$$

which leads to the iterations

$$RL: \quad \mathbf{o}^{(k+1)} = \mathbf{o}^{(k)} \cdot \mathbf{H}^t \frac{\mathbf{y}}{\mathbf{H}\mathbf{o}^k}. \quad (4.7)$$

In these iterations, the division of \mathbf{y} by $\mathbf{H}\mathbf{o}^k$ is made element-wise. The results is left multiplied by \mathbf{H}^t (which in practice is implemented in the Fourier space using the structure of \mathbf{H} in (4.1)), and the multiplication by the previous estimate $\mathbf{o}^{(k)}$ is again element-wise. The question of deciding when to stop the iterations is a difficult one, as RL (and ISRA) do not possess a natural stopping criterion. Stopping the iterations to some number performs a kind of regularization, although non explicit. Pseudo-codes for RL and ISRA below can be found in Thiébaud (2005).

⁹*i.e.*, $\lim_{n \rightarrow \infty} \mathbb{E}\{\widehat{\theta}_{MV}[n] - \theta\}^2 = 0$.

¹⁰*i.e.*, the smallest Mean Square Error that is achievable by any unbiased estimator, and that is caused the uncertainty inherent to the stochastic perturbations.

4.3 Image space reconstruction algorithm (1986)

When the data are spoiled by a Gaussian noise that is pixel-wise independent but possibly non identically distributed, the model is $\mathbf{y} = \mathbf{H}\mathbf{o} + \mathbf{b}$, $\mathbf{b} \sim \mathcal{N}(0, \mathbf{\Sigma})$, with $\mathbf{\Sigma} = \text{diag}[\sigma_1^2 \dots \sigma_N^2]$. The ISRA algorithm (Daube-Witherspoon & Muehlehner 1986) produces the *non-negative* solution that is the most likely according to the noise model. The likelihood of the data \mathbf{y} is

$$\mathcal{L}(\mathbf{y}; \mathbf{o}) = P([\mathbf{y}_1 \dots \mathbf{y}_N]^t; \mathbf{o}) = \prod_{i=1}^N (2\pi\sigma_i^2)^{-\frac{1}{2}} e^{-\frac{(\mathbf{y}_i - [\mathbf{H}\mathbf{o}]_i)^2}{2\sigma_i^2}}. \quad (4.8)$$

Maximising this function on non-negative \mathbf{o} is equivalent to minimize

$$J_{Gauss}(\mathbf{o}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{o}\|_{\mathbf{\Sigma}^{-1}}^2 = \frac{1}{2} \sum_i \frac{(\mathbf{y}_i - \mathbf{H}\mathbf{o}_i)^2}{\sigma_i^2}, \quad \text{subject to } \forall i, \mathbf{o}_i \geq 0. \quad (4.9)$$

This leads to the iterations

$$ISRA: \quad \mathbf{o}^{(k+1)} = \frac{\mathbf{o}^{(k)}}{\mathbf{H}^t \mathbf{\Sigma}^{-1} \mathbf{H}\mathbf{o}^{(k)}} \cdot \mathbf{H}^t \mathbf{\Sigma}^{-1} \mathbf{y}, \quad (4.10)$$

where the division and multiplications are elemen-twise. In practice however, some data may be negative, in which case the iterations above do not guarantee the non-negativity of $\mathbf{o}^{(k)}$ over the iterations. To overcome this problem, one should work on a data image \mathbf{y}' that is shifted by its minimum value $m = \min_i(\mathbf{y}_i)$ (Lanteri *et al.* 2002b). If we denote by \mathbf{d} the vector with entries $\mathbf{d}_i = -m, \forall i$, then $\mathbf{y}' = \mathbf{y} + \mathbf{d}$ is non-negative and the ISRA iterations become

$$\mathbf{o}^{(k+1)} = \mathbf{d} + \frac{\mathbf{o}^{(k)} - \mathbf{d}}{\mathbf{H}^t \mathbf{\Sigma}^{-1} \mathbf{H}\mathbf{o}^{(k)}} \cdot \mathbf{H}^t \mathbf{\Sigma}^{-1} \mathbf{y}'. \quad (4.11)$$

When the iterations are stopped at some iteration number k_{stop} , the estimated object is $\hat{\mathbf{o}} = \mathbf{o}^{(k_{stop})} - \mathbf{d}$.

4.4 Numerical illustrations

The following example (Fig. 13) illustrates what can typically be achieved by such algorithms. This example shows some results of the RL deconvolution algorithm as a function of the iterations. The considered reference object is that of Figure 6, and the data image is a photonized version of the FSD hypertelescope (Fig. 11, middle). Before deconvolution, we first redilute (Fig. 12, top right). We can also, before starting the restoration, filter out the noise. This is not necessary, because the first iteration of the RL algorithm will do it; but seeing the filtered image indicates that the situation is not as bad as visually suggested by the data image (Fig. 11, middle). This helps realizing that some merit should indeed be attributed to the restoration algorithm, but not too much.

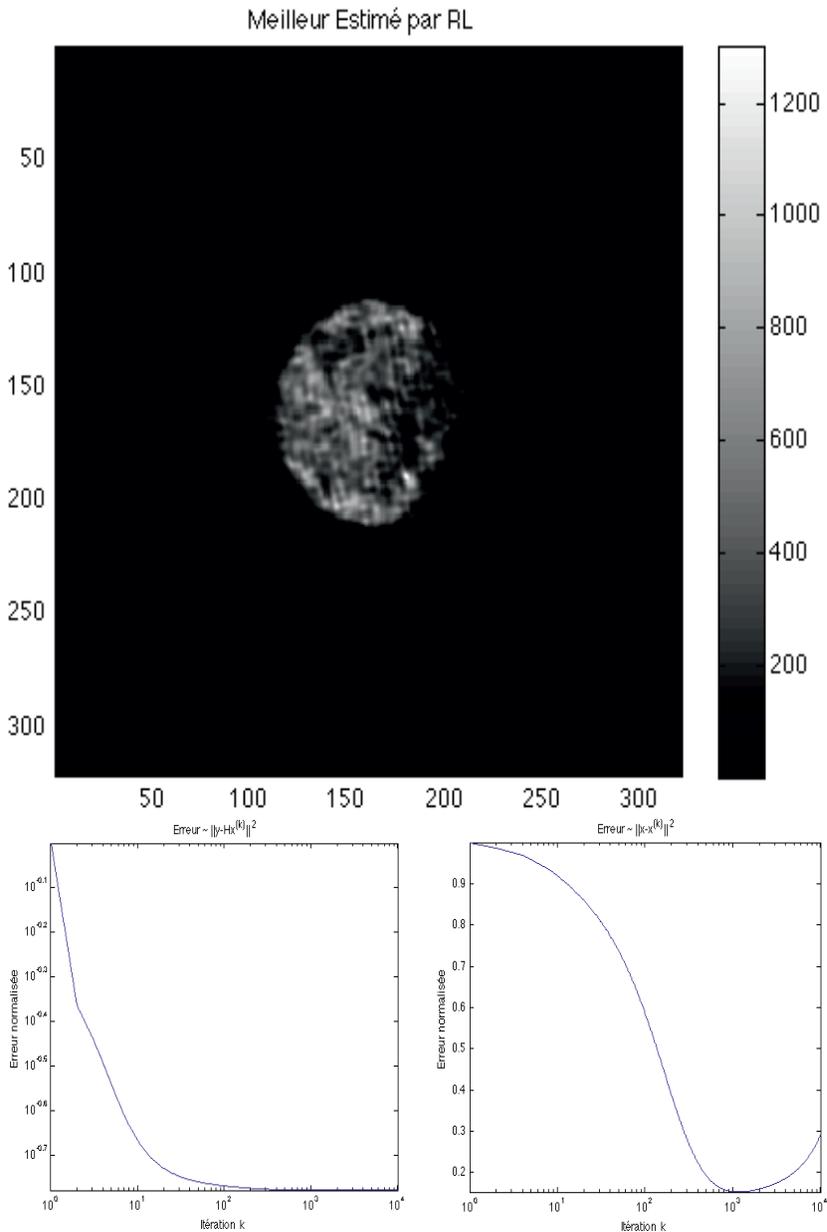


Fig. 13. *Top:* best (in the euclidean distance sense) recovered object for the RL algorithm. The companion is not reproduced. *Bottom, left:* data approximation error *vs.* iteration number. *Bottom, right:* error with respect to the reference object *vs.* iteration number.

The restored object in Figure 13 is, with respect to the true object, quantitatively and qualitatively inaccurate in the details, but rather fair in the overall shape. The companion is not recovered, and the surface intensity on the main planet has extremely high values. The overall flux of the solution is the same as the flux in the data, because RL iterations structurally preserve the flux. This flux is very close the flux of the reference object, because the 0 frequency is measured by the optical system: the flux is thus known up to the measurement error caused by the noise.

The two lower figures represent the data approximation error ($\|\mathbf{y} - \mathbf{H}\mathbf{o}^{(k)}\|^2/\|\mathbf{y}\|^2$) and the distance with respect to the object (measured here as $\|\mathbf{o} - \mathbf{o}^{(k)}\|^2/\|\mathbf{o}\|^2$). The data approximation error continuously decreases. This behavior always happens for ISRA (since ISRA iteratively reduces the convex criterion $J_{Gauss}(\mathbf{o}^{(k)})$), and usually for RL as well. On the other hand, the distance of $\mathbf{o}^{(k)}$ to \mathbf{o} decreases and then increases: the best solution in the euclidean distance sense, which is shown in the top figure, is here obtained for an iteration number of $k \approx 1000$. The best iteration number is always impossible to know in practice, because the object is not known. The behavior exhibited by these curves is very general.

This example illustrates important characteristics that share RL and ISRA algorithms. The positivity constraint is easily imposed (and any support constraint could be possibly imposed as well, owing to the multiplicative form of this algorithms), and their implementation is easy. On the other hand, the number of iterations to reach an acceptable solution can be large (slow convergence), and it is difficult to know when to stop the iterations in practice. Finally, the interpolation and extrapolation that are performed in the Fourier space are very difficult to formalize.

An uncountable number of inversion methods can be found in the literature. A large class of those injects an explicit regularization term in the criterion coming from the likelihood. This term reflects *a priori* information about the solution. Explicitly regularized methods allow to put well defined constraints on how the missing frequency content should be added to the data during the iterations, or at least, to impose that the solution exhibits specifically desired properties (*e.g.* smoothness, piecewise constant aspect, etc.). Another set of approaches uses the concept of sparse representations, and a survey is proposed below.

4.5 Deconvolution based on sparse representations

4.5.1 CLEAN and sparsity

Long before the optical interferometry era, radio astronomers had devised various techniques to recover estimates of \mathbf{o} from \mathbf{y} in convolution models. The ancestor of sparsity-based techniques in the radioastronomical community is the CLEAN algorithm, which is used routinely in radioastronomy since almost 40 years now. Numerous variants have been developed (see Cornwell 2008), and in practice CLEAN remains a benchmark method in radioastronomy. It is worth detailing a bit on

this algorithm, as it can directly be applied to the deconvolution of hypertelescope images, and because of its link with sparsity.

CLEAN was proposed by Högbom (1974), and was related to earlier methods in Scharz (1978). See also Cornwell (2009) for a description of the wide impact of CLEAN on Astronomy and beyond. CLEAN can be seen as a method which essentially models \mathbf{o} as a collection of a few (relatively to the number of pixels) point sources. The image \mathbf{y} is thus in turn modeled as a collection of shifted and scaled PSFs. The algorithm subtracts iteratively scaled and shifted PSFs from the image residual, until a stopping criterion is reached. Högbom recommended in its 1974's paper a threshold on the maximum value of the residual. This threshold forces a limited number of iterations, and thus a limited number of detected point sources – this is where sparsity comes in.

This iterative process is very similar to the Matching Pursuit algorithm, which is widely used in the signal processing community (Mallat & Zhang 1993). With a difference, yet: at the end of the iterative deconvolution process of CLEAN, the residual is added back to the synthesized detection map. As a matter of fact, CLEAN works very well when only a few points sources constitute \mathbf{o} , and thanks to the residual trick, it remains also relatively efficient even for extended sources, despite the apparent irrelevance of its sparse model for such sources.

The excessive simplicity of CLEAN's "point source" model for general astrophysical sources was nevertheless worked out in the 80 s by several researchers, who looked for more elaborated models of extended sources. Several methods based on multiresolution approaches followed the works by Wakker & Schwarz (1988). An overview of CLEAN's evolutions is given in Rau *et al.* (2009).

4.5.2 Analysis and synthesis sparsity

Sparse representations in dictionaries are ubiquitous in a large number of modern signal processing methods. The reason for this is perhaps double: first, they rely on simple (linear) statistical models; second, they naturally operate a dimension reduction, by focusing on subspaces of reduced dimension where the information of interest actually lies.

Sparse representations can be seen as a generalization of CLEAN's model. In the synthesis model, which is the most intuitive and has historically benefited from more efforts, the object \mathbf{o} is assumed to be well modeled by a linear combination of a few elementary shapes (not just point sources), called *atoms*.

Promoting sparsity relatively to appropriate dictionaries offers a straightforward way to fill the missing frequency contents caused by the zeros of the transfer function. Indeed, sparse methods essentially detect which atoms are present in the data using the measured frequencies: some missing frequencies are then automatically filled with the frequencies of the detected atoms.

Let us now describe in more detail the differences between analysis and synthesis sparsity (Elad *et al.* 2007). We consider the object-image model $\mathbf{y} = \mathbf{H}\mathbf{o} + \mathbf{n}$, with $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathcal{I})$ for simplicity. In the synthesis approach, the unknown intensity distribution \mathbf{o} (of size $N \times 1$, say) is assumed to be sparsely synthesizable by a

few atoms of a given full rank dictionary \mathbf{S} of size $N \times L$. Hence, we write \mathbf{o} as $\mathbf{o} = \mathbf{S}\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ (the synthesis coefficients vector) is sparse. This assumption is widely used in data modeling for denoising, compression, pattern recognition or inpainting applications for instance, because natural signals and images are approximately sparse in appropriate spaces (Mallat 2008).

A synthesis-sparse solution to the deconvolution problem posed by $\mathbf{y} = \mathbf{H}\mathbf{o} + \mathbf{n}$ can be obtained by solving:

$$\mathbf{o}_S^* = \mathbf{S} \left\{ \arg \min_{\boldsymbol{\eta}} \frac{1}{2} \|\mathbf{HS}\boldsymbol{\eta} - \mathbf{y}\|_2^2 + \mu_p \|\boldsymbol{\eta}\|_p^p \right\}, \quad (4.12)$$

where μ_p is a hyper parameter that tunes the desired sparsity degree, and $\|\boldsymbol{\eta}\|_p^p = \sum_i |\eta_i|^p$, $0 \leq p \leq 1$, is a function favoring zero values.

The solution \mathbf{o}_S^* is also interpretable in the Bayesian framework as a Maximum A Posteriori (MAP) solution, in which case μ_p is related to the parameters of a Generalized Gaussian prior on $\boldsymbol{\eta}$. The ℓ_0 quasi-norm (which counts the number of nonzero coefficients in $\boldsymbol{\eta}$), obtained for $p = 0$, is the most natural sparsity measure. To ensure the convexity of the resulting cost function, it is often replaced by the ℓ_1 norm $\|\cdot\|_1$, which still promotes sparsity (and correspond to a Laplacian prior on $\boldsymbol{\eta}$).

In contrast, the analysis approach consists in finding the solution \mathbf{o} that is not correlated with some atoms of a dictionary \mathbf{A} of size $N \times L$: $\mathbf{A}^T\mathbf{o}$ is sparse. An analysis-sparse solution can be obtained by solving:

$$\mathbf{o}_A^* = \arg \min_{\mathbf{o}} \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \mu_p \|\mathbf{A}^T\mathbf{o}\|_p^p. \quad (4.13)$$

Note that the synthesis prior is on the synthesis coefficients $\boldsymbol{\eta}$, while the analysis one is on the projection $\mathbf{a} = \mathbf{A}^T\mathbf{o}$ of the signal on the analysis dictionary \mathbf{A} .

While both approaches are equivalent when \mathbf{A} and \mathbf{S} are square and invertible, with $\mathbf{A} = \mathbf{S}^{-1}$, they yield in general different solutions for overcomplete dictionaries ($N < L$). Such dictionaries are required for efficient image modeling (see next Subsection). Since natural images can often be approximated by few atomic elements in such dictionaries, the synthesis approach is considered as more intuitive. Its design simplicity (in greedy approaches like CLEAN¹¹) has also made it more popular in image processing applications. However, the synthesis solution is restricted to a column subspace of the synthesis dictionary, and the significance of each selected atom is important. On the other hand, the analysis approach may be more robust to “false detections” since the signal is not built from a few number of atoms. Besides, note in Equation (4.12) and (4.13) that the number of unknown in the synthesis case (the number of atoms in the dictionary) can be much larger than in the analysis case (where it remains in the number of pixels).

¹¹The greediness of CLEAN is visible in the atom’s (shifted PSF) selection rule: select the atom that is the most correlated with the data, that is, the one that most decreases the norm of the residual.

Thus, analysis-based optimization strategies can be computationally much more efficient for large dictionaries. The comparison of both models is a very active field of research in fundamental signal processing (see the references in Gribonval *et al.* 2009, 2012).

Once a sparsity promoting model is chosen, we still need to choose an appropriate dictionary to perform the deconvolution.

4.5.3 Dictionaries

The sparsity is expressed via dictionaries, which correspond to representation spaces. Dictionaries express geometrical features that are likely to describe the unknown object. In synthesis, the columns of the dictionary are simply the atoms. In analysis, the rows may be atoms as well, or operators (gradient for instance, leading with ℓ_1 norm to the total variation regularization). These dictionaries can be orthonormal transforms (corresponding to orthonormal bases), or more generally redundant (overcomplete) dictionaries. A large variety of representations has been elaborated in the image processing literature, *e.g.*, canonical basis indeed (corresponding to point-like structures), Discrete Cosine Transform (DCT, 2-D plane waves), wavelets (localized patterns in time and frequency), isotropic undecimated wavelets, curvelets, ridgelets, shapelets and many others, see Mallat (2008) and Starck *et al.* (2010) for detailed reviews.

The choice of a dictionary is made with respect to a class of images. In Astronomy, wavelets dictionaries are widely used, but they are known to fail representing well anisotropic structures. In such cases other transforms can be used, that have been designed to capture main features of specific classes of objects. Among them, curvelets sparsify well curved, elongated patterns such as planetary rings or thin galaxy arms for instance; shapelets sparsify well various galaxy morphologies, etc. All these dictionaries have shown empirical efficiency for some specific types of images.

In order to model efficiently complex images with various and different features, several authors have proposed to concatenate dictionaries into a larger dictionary (Chen *et al.* 1998; Gribonval *et al.* 2003). However, the efficiency of a dictionary also critically depends on its size and on the existence of fast operators, without which iterative algorithms cannot run in reasonable time. This is especially true in radiointerferometry and in (possibly polychromatic) optical interferometry where the number of Fourier samples and of pixels can be of the order of hundreds of thousands.

4.6 CS and sparsity in astronomical deconvolution

Since the Compressed Sensing (CS) theory has emerged, providing exciting and beautiful mathematical results about sparse recovery in various cases, the sparsity ideas have benefited from considerable new strengths in the fields of signal and image processing (Donoho 2006; Candès *et al.* 2006).

In the context of the image restoration problem posed by interferometric measurements, the CS theory has provided theoretical proofs that, in idealized

situations, exploiting sparsity is indeed useful. For instance, the CS theory explains why a few point sources may be recovered from random Fourier measurements that are in number far less than specified by the sampling theorem. Of course, this possibility is exploited and implemented in long-existing restoration methods like CLEAN for instance; sparse methods have grown and evolved on their own before CS. They have lead to several types of elaborated sparsity-based algorithms, whose use evidences decades of empirical success.

As in many other applicative fields, references to CS have started hatching in quite a few recent publications about astronomical deconvolution. Yet, one key ingredient in that matter – sparsity – is exploited since at least Högbom’s time. Besides, the theoretical CS proofs invoked in the introduction of many such publications turn not to help much in the subsequently proposed restoration methods. This poses the question of the real benefits brought by CS to astronomical deconvolution.

Let us consider the question from an operational point of view, that is, with the concern of better estimating \mathbf{o} from the data \mathbf{y} . The benefits from CS with this respect are real but indirect, and they appear to be the following. First, CS clearly drained an increased research effort in fundamental models for sparse representations, like those of Equations (4.12) and (4.13). This in turn lead to improved reconstruction methods, through more elaborated statistical data models. Second, a lot of efficient optimization strategies have been designed to solve problems of the type (4.12) and (4.13), thanks to the new strengths in this field brought by the appealing theoretical results of CS.

In the recent years, sparsity promoting methods were used in interferometry by Wiaux *et al.* (2009a,b), using Basis Pursuit DeNoising (Chen *et al.* 1998) with wavelets dictionaries, and by Vannier *et al.* (2010), with Matching Pursuit algorithms in unions of bases (wavelets/Dirac). Li *et al.* (2011) adopted a synthesis approach with an IUWT (Isotropic Undecimated Wavelet Transform) synthesis dictionary, and solved a Basis Pursuit synthesis criterion through the ISTA minimization algorithm (Iterative Soft-Thresholding Algorithm) and its fast version, FISTA (Fast Iterative Shrinkage-Thresholding Algorithm, Beck & Teboulle 2009). Carillo *et al.* (2012) applied a reweighted ℓ_1 analysis algorithm promoting average signal sparsity over multiple redundant dictionaries, and relying on convex optimization techniques. Dabbech *et al.* (2012) have proposed an hybrid analysis-by-synthesis approach: \mathbf{o} is modeled using sparse synthesis priors as a sum of few objects which, as opposed to classical synthesis-based priors, are unknown. These atoms are iteratively estimated through analysis-based priors, the analysis being based on an IUWT dictionary.

Note that in optical interferometry, the 2012 international Beauty Contest also witnessed an increasing number of sparsity based methods (Baron *et al.* 2012). Finally, in polychromatic optical interferometry, Thiébaud *et al.* (2012) proposed to favor spatial sparsity and spectral grouping of the sources through an alternating direction method of multipliers, a method also issued from the convex optimization literature (see the Article of É. Thiébaud in these proceedings).

We will not include numerical results about restoration algorithms exploiting sparsity in this paper. Examples of results obtained with sparsity promoting methods (comparison of CLEAN, MP and BP) in the case of diluted apertures (Fizeau configuration) can for instance be found in Section 4 of Vannier *et al.* (2010).

To conclude this part, we see that numerous techniques are emerging. They offer sophisticated alternatives to the more traditional and robust constrained ML methods. Indeed, most of these methods comply with the non-negativity constraint. The algorithms described above rely on recent progresses in sparse representations and convex optimization techniques. They allow to solve large scales optimization problems involving complex image models, and they are becoming increasingly popular in interferometry.

We now turn to simulations results aimed at illustrating the effect of densification in the presence of photon noise. We then investigate the possibility of recovering small objects that are far from the optical center.

5 To densify, or not to densify

Of course, the ambition of this section is not to provide a general answer to this question, as many factors should be accounted for (for instance the number of detector pixels, the noise level and its statistical nature, the subpupil configurations, etc.). As already emphasized, the FSD densified and the Fizeau images contain the same frequency information. Since the densified image has a lower frequency, it needs less detector pixels to be properly sampled than its Fizeau counterpart. This should be an advantage of hypertelescopes, which will not be illustrated here as this is a straightforward consequence of the sampling theorem. See for instance Lardière *et al.* (2007) for useful insights on these issues.

The question on which we focus here is the following. We are given a reference object \mathbf{o} (the one of Fig. 6) and two sampling schemes (the previously described Fizeau and Michelson FSD configurations, Fig. 8), leading to one-million-pixel images that are contaminated by photon noise (Fig. 11). In these conditions, which scheme leads statistically to the best restored images using a RL algorithm?

We propose to answer this question empirically, by running Monte Carlo simulations. We generated 50 photonized Fizeau images and the same amount of photonized FSD images. The images of the latter set were numerically rediluted so that the object-image model is a convolution. The difference between the two sets of images is in the noise statistics. While it is Poissonian for the first set, this is not the case for the set of rediluted images. These images can (and actually do) exhibit negative values. Thus the RL algorithm, seen as a ML method, is not justified any more because both the image positivity and the Poisson statistics are lost. However, RL can be (as ISRA) simply taken as a deconvolution method which minimizes some loss function between the data and convolved model (the Kullback-Leibler divergence for RL, and the quadratic error loss for ISRA). If non-negative restored objects are expected, care must be taken in this case that the data images are non-negative. This is achieved by setting to 0 the negative values of the rediluted images. We observed for the considered noise level that negative

data values in the rediluted images are very few (typically 5 out of 10^6) and close to 0 (typically less than 0.01 in absolute value). So this non-negativity precaution causes a negligible information loss.

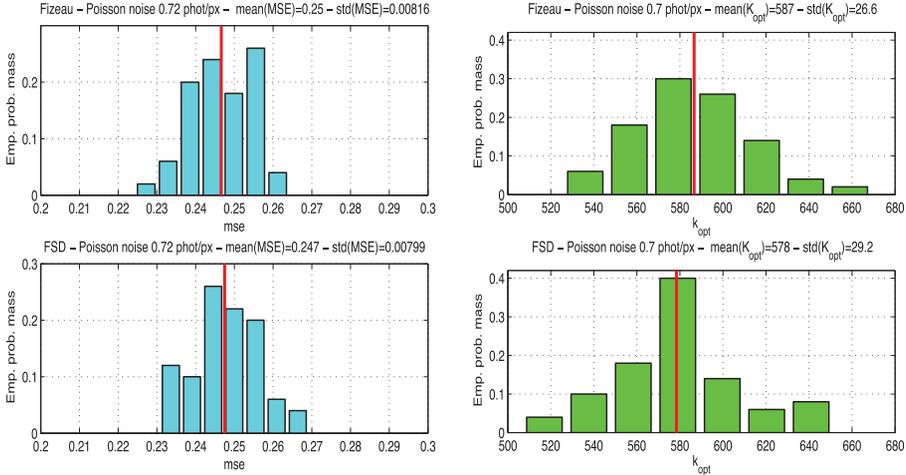


Fig. 14. *Left:* empirical distributions of the best reconstruction errors obtained by RL for photonized Fizeau images (*top*) and rediluted photonized FSD images (*bottom*), for 50 realizations. The average photon noise in the data images corresponds to 0.72 photon/pixel, and the images have 1024×1024 pixels. The vertical line shows the empirical mean of the distribution (which is also indicated in the titles of the figures along with the observed dispersion). *Right:* empirical distributions of the iteration number leading to the best reconstruction (*vertical line*: empirical means; values for means and standard deviations are indicated in the titles).

The results are presented in Figure 14. It is clear from this experiment that the results are essentially equivalent in terms of the quality of the reconstruction error and of the optimal number of iterations. The equivalence in terms of information that holds between Fizeau and FSD configurations appears conserved in images affected by photon noise.

6 Noiseless recovery of a small object outside the “clean” field

6.1 Introduction: Objectives and simulation parameters

We are interested here in the restoration of noiseless images obtained in the Fizeau configuration, or in any densified configuration without spectral aliasing. The objective of this study is to investigate whether a quasi-point source which is located outside the “clean” field (see below) can or not be restored by the RL algorithm. The noise is not considered in order to focus on the effects of the sampling.

The “clean” field (Lardière *et al.* 2007) is the central zone of the image of dimension λ/s , where s is the smallest spatial distance between two subpupils. The global field corresponds to the diffraction envelope of the elementary pupils.

The considered sampling is the same as described in Section 3: 25 non redundant circular apertures on an integer grid. These apertures have the same diameter D , and the principal lobe of the diffraction envelope defines the global field, which has diameter $2 \times 1.22\lambda/D$. For the considered array, the centers of the elementary OTF in the Fizeau sampling are separated by $7\tau_\nu$. Thus, the central part of the Fizeau image is essentially replicated 7 times in each direction, and λ/s corresponds to $1024/7 \approx 146$ pixels (*cf.* Fig. 15, bottom left).

The object we consider is presented in Figure 15. The flux ratio between the central planet and the satellite is $\approx 4.8 \times 10^{-3}$, which corresponds to a difference in magnitude of ≈ 5.8 .

6.2 Recovery without spatial aliasing

In this first simulation, the small object is not located on a replica of the main object, but it is quite far from the center (close to the limit set by the global field), and thus highly attenuated. This source is centered around the pixel coordinates $(x = 140, y = 513)$, and is 373 pixels away from the centre $(x = 513, y = 513)$ of the object. This angular distance represents $\approx 2.5 \times \lambda/s$, or $\approx 0.73 \times 1.22\lambda/D$.

Figure 16 illustrates the evolution of the deconvolution along the iterations.

Interestingly, we see that the algorithm first reconstructs a satellite close to the central object, and then transfers the flux from this position to the left by discrete jumps of λ/s (*i.e.*, the clean field), to finally reach the good position:

- Iteration 100: the central planet appears, the replicas and the halo have almost disappeared. No satellite yet.
- Iteration 3500: the central planet starts being fairly well estimated, and a quasi-point source is restored in the vicinity of the planet, at pixel $(x = 432, y = 513)$, *i.e.* at $2\lambda/s$ right of the real position of the satellite ($432 = 140 + 2 \times 146$).
- Iteration 4500: a second point source appears at pixel $(x = 286, y = 513)$, *i.e.* at λ/s right of the real position. The flux of the first (fake) quasi-point source has decreased with respect to iteration 3500.
- Iteration 5800: The first fake satellite at $(x = 432, y = 513)$ has disappeared.
- Iteration 7800: A third satellite at the right position $(x = 140, y = 513)$ appears. The flux of the second satellite at $(x = 286, y = 513)$ decreases.
- Iteration 30 000: The second fake satellite $(x = 286, y = 513)$ has almost disappeared. The algorithm has (almost) converged to a correct reconstruction of the central planet and of its satellite.

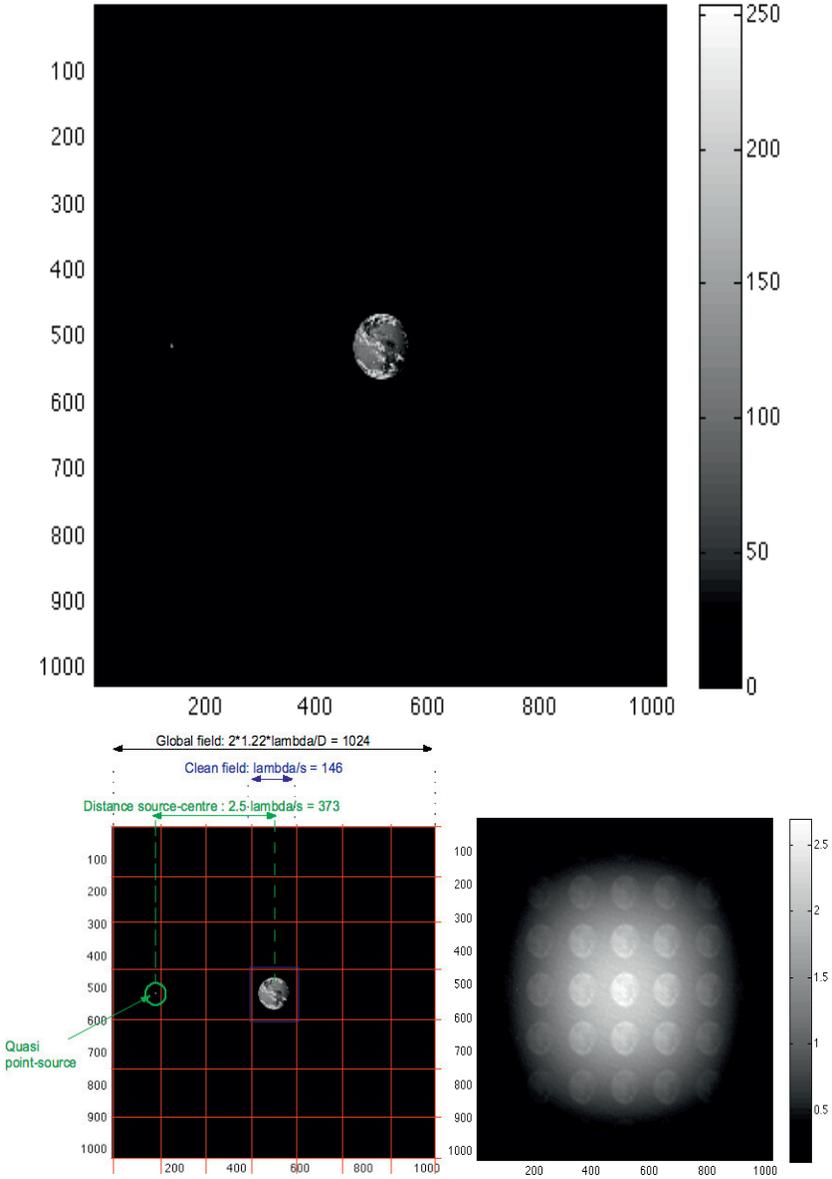


Fig. 15. *Top:* reference object (planet-satellite). *Bottom, left:* same object with the fields shown. The global field represents 1024 pixels in each direction, and the clean field 146 pixels. The quasi-punctual source contributes flux in 28 pixels. It is 373 pixels away from the centre, which corresponds to $\approx 2.5\lambda/s$. *Bottom right:* image obtained with the considered diluted pupil.

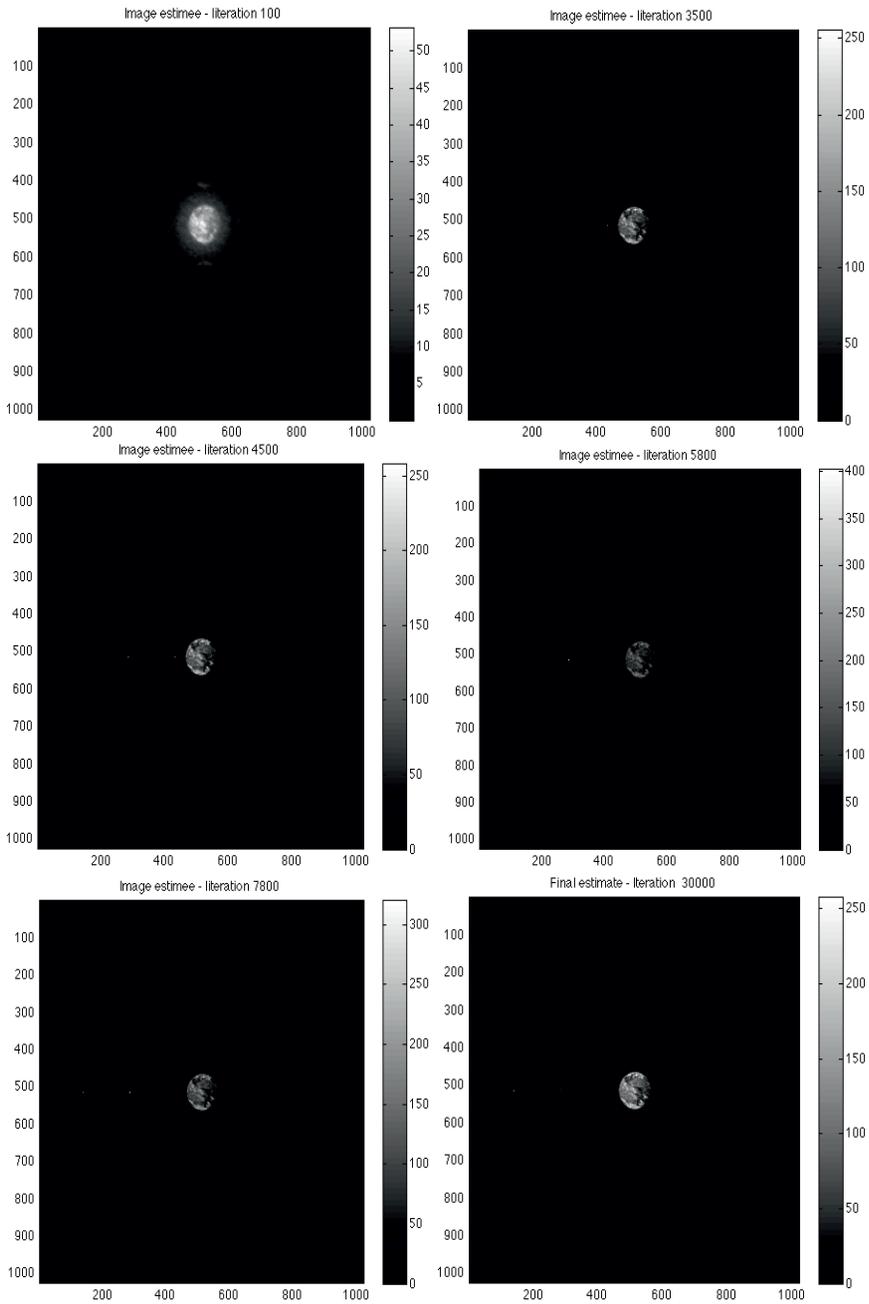


Fig. 16. Snapshots of restored object for some iterations of the RL algorithm. In lexicographic order $k = 100, 3500, 4500, 5800, 7800, 30\,000$.

The satellite reconstruction by jumps of extension λ/s can be followed in the Fourier space. As illustrated below, the RL algorithm fills the Fourier space by progressive interpolation of the spectrum around the available samples. The moduli of the frequency samples where information is available are represented in Figure 17, left, and the total spectral information to be recovered is in Figure 17, right. Figure 18 shows the same for the phases. The satellite information appears essentially as a modulation on the moduli and on the phases of the central planet's spectrum.

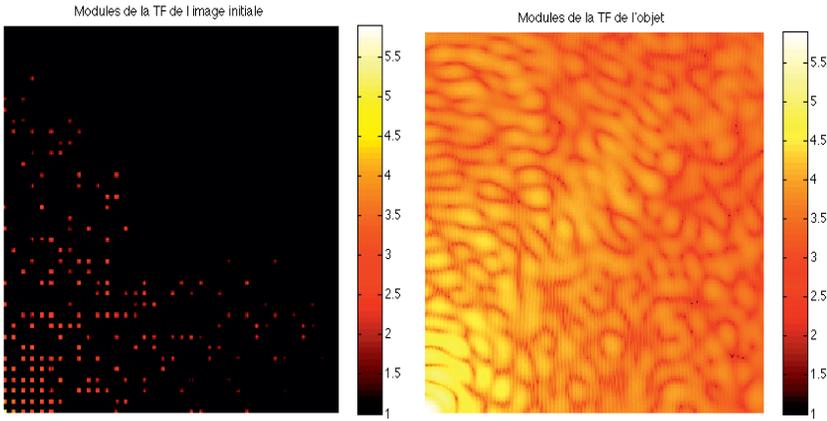


Fig. 17. Zoom on the moduli of the Fourier spectra. *Left:* available moduli (the missing samples are in black). *Right:* moduli of the spectrum of the considered planet-satellite system.

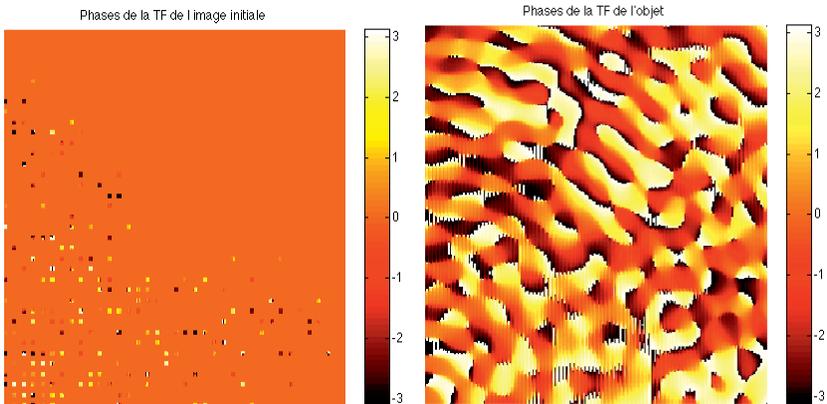


Fig. 18. Same as Figure 17, but for the phases.

As the iterations go, the “holes” at low frequencies are progressively filled, and the high frequencies are then estimated, as illustrated in Figure 19.

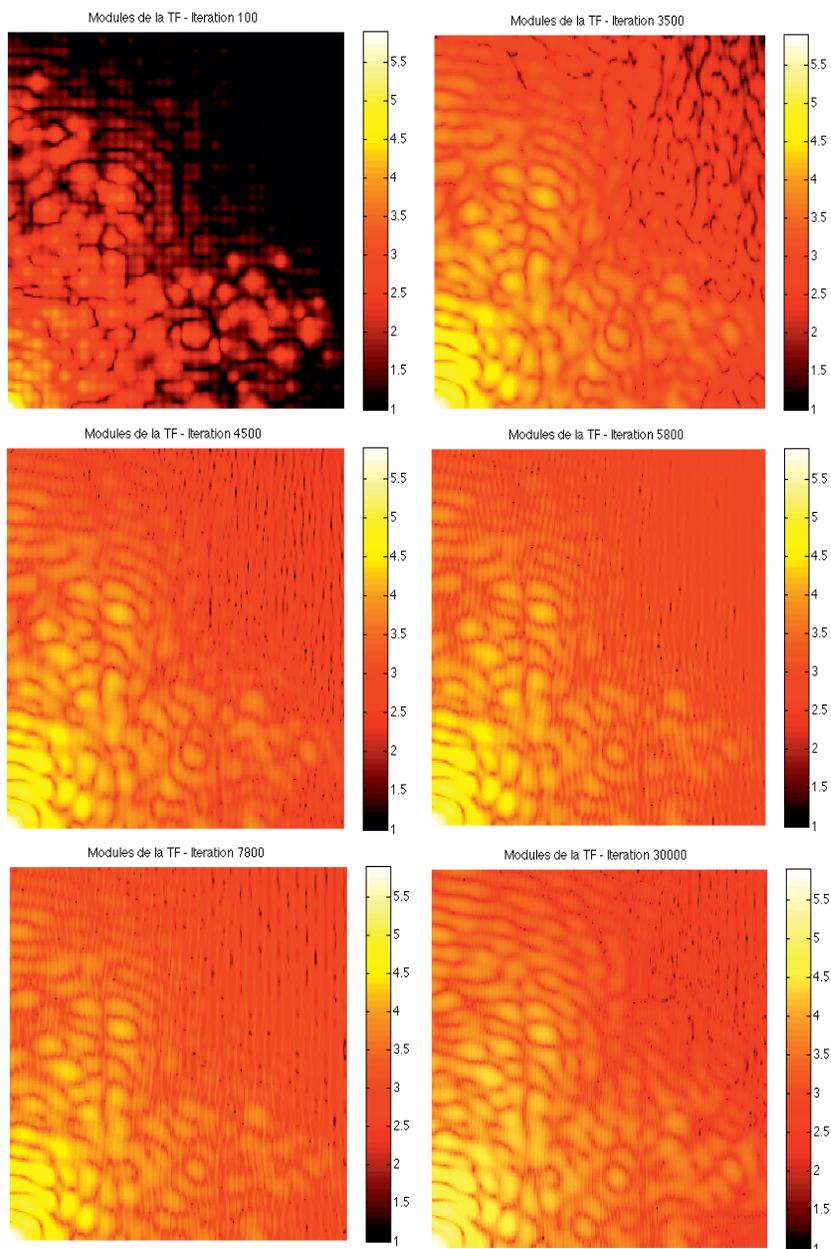


Fig. 19. Zoom on the moduli of the estimated objects for the iterations of the RL algorithm shown in Figure 16 (in lexicographic order $k = 100, 3500, 4500, 5800, 7800, 30000$).

After 30 000 iterations, the object restored by the algorithm is relatively close to the original, at least as far as the satellite position is concerned. The Figure 20 presents a zoom on the central planet (left column) for the reference object (top, left) and on the deconvolved object (middle, left). Similarly, the right column shows a zoom on the satellite of the reference object (top, right) and deconvolved (middle, right). The total flux of the deconvolved satellite is about 60% of the total flux of the reference satellite. Of course the deconvolution is not perfect (and in cannot be, as too many frequency are lost by the sampling). But the result is comparable to the direct image that would be produced by a monolithic Extremely Large Telescope having the same diameter as the largest base of the hypertelescope¹² (Fig. 20, bottom row).

The important thing is that the point source is fairly recovered, a point which was not obvious considering the ambiguity posed by the sampling scheme. This result is encouraging, efforts for the quest of high angular resolution do not seem to be vain. We may pause here to think that some day in the future, the detection of such a faint little point in the dark corner of a real hypertelescope image might be the origin of a great discovery for the Human civilization.

Enough dreams for now, *sine experientia nihil sufficienter sciri potest*: let us come back to the prosaic reality of simulations and try a more difficult recovery.

6.3 Recovery with spatial aliasing

The considered object is still of the planet-satellite type but the satellite replicas are superimposed on the replicas of the central object in the data image, see Figure 21.

The point source is now at coordinates $(x = 213, y = 513)$, which is 300 pixels away from the centre $(x = 513, y = 513)$. This represents $\approx 2\lambda/s$. As visible in Figure 21 right, this is a clear case of spatial aliasing. The results of the deconvolution for some iterations are presented in Figure 22.

- Iteration 100: the planet starts being well restored. No satellite in the vicinity. Note that a bright spot, of about the satellite size, is created on the central planet. This spot comes from the spatial aliasing (replica of the satellite superimposed on the planet).
- Iteration 3000: the reconstruction seems to stabilize on an object without satellite, with a surface spot at the place of the satellite. It is unclear whether the algorithm will be able to distinguish between a satellite to be placed further away, and a bright surface spot.
- Iteration 5000: a satellite appears around pixel $(x = 359, y = 513)$, that is, at λ/s right of the real position. In the same time, the bright artifact at the center is less visible than in the previous iterations.

¹²This ideal telescope is called “Metatelescope” in Aime *et al.* (2012).

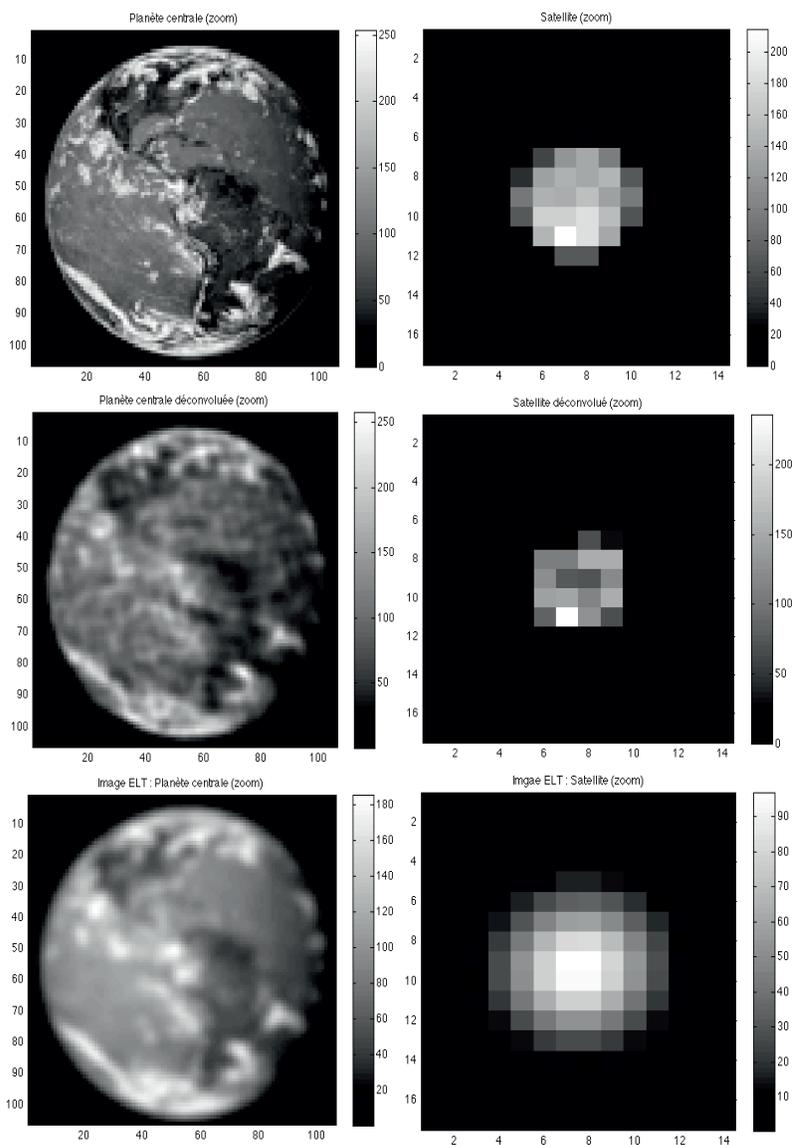


Fig. 20. Zoom on the central planet (*left column*) and on the satellite (*right column*). *Top row*: reference object. *Middle row*: deconvolved object after 300 000 RL iterations. *Bottom row*: Image of a monolithic ELT having the same high frequency cut-off as the diluted array.

- Iteration 40 000: a second satellite appears at the real position. The flux attributed to the first fake satellite decreases, and the artifact at the centre of the planet is almost not visible anymore.

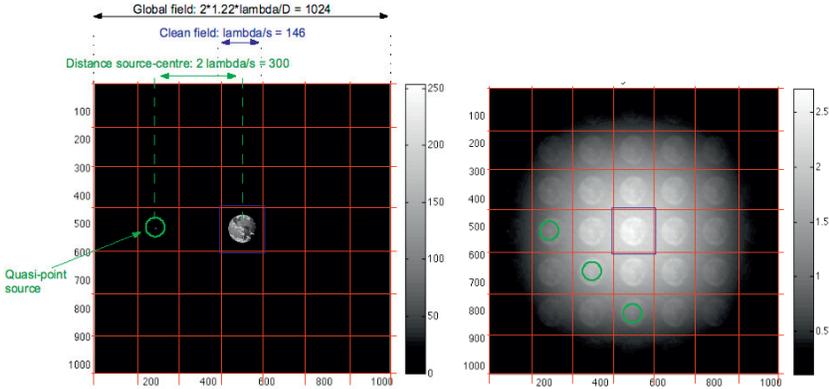


Fig. 21. Case where the satellite “falls” in a replica of the central planet. *Left:* considered system planet-satellite. *Right:* image produced by the diluted array. The light circles evidence the spatial aliasing on some replicas (it is the case for all of them).

- Iterations 80 000 et 200 000: the flux is progressively transferred from the first to the second satellite, but the convergence is very slow. Figure 23 zooms on the object deconvolved at iteration 200 000: zoom on the central planet (left) and on the satellite position (right). The flux estimated for the satellite at the real position is still insufficient (compare to Fig. 22).

We see that although the convergence is slow, the algorithm is on the way to find the right configuration. Figure 24 presents, for each estimated object $\mathbf{o}^{(k)}$ at iteration k , $k = 1, \dots, 200\,000$, the normalized error in approximating the data (left), and the normalized error with respect to the true object (right).

We see that the convergence is very slow. Note also that the error with respect to the object is not constantly decreasing. An intermediate solution corresponding to a local maximum ($k \approx 8000$) corresponds to an estimated object with one satellite that is too close to the planet. But this solution does not perfectly explain the data. Some flux then starts being injected at the right position, so that the error decreases again. These results suggest that the right configuration can be recovered, even if the flux is not perfectly estimated, at least with negligible noise.

How is the algorithm able to find out, from data where the satellite is everywhere superimposed on the surface of the central object, that there is satellite, and that the surface has no bright spot? The reason is that if the central planet had a bright surface spot, this spot should be less bright in replicas that are further away from the center (because of the diffraction envelope). But this is not the case for the replicas of the satellite: the brightest replicas are the ones that are close to the true position of the satellite. This discrepancy makes the algorithm to eventually inject the flux at the right position. In other words, it is the diffraction envelope which saves the reconstruction here.

A last remark. To see to which solution the algorithm would eventually converge, and how accurate the recovery would be in this case, there should be several

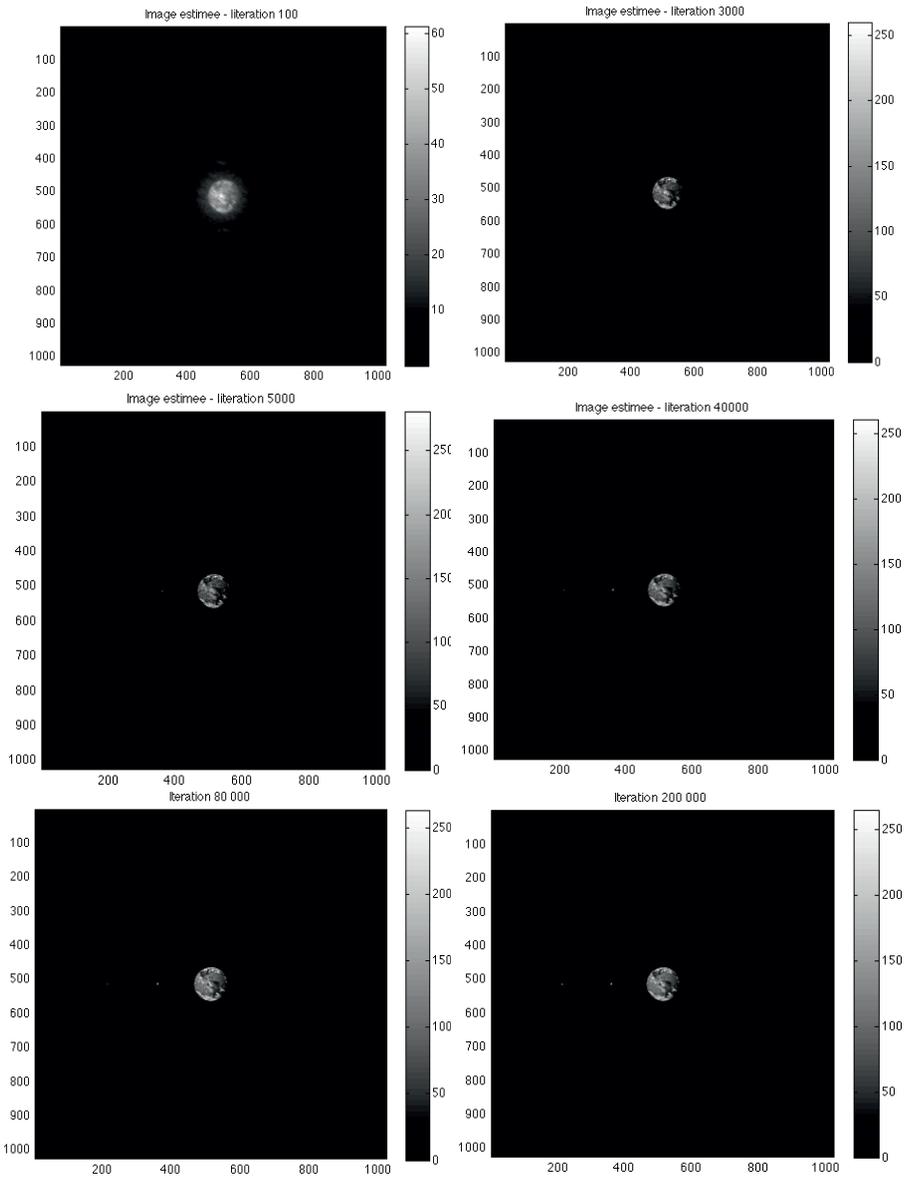


Fig. 22. Estimated objects for some iterations of the RL algorithm. By lexicographic order: $k = 100, 3000, 5000, 40\,000, 80\,000, 200\,000$. At iteration $k = 40\,000$ a source starts being visible at the right position.

hundred thousands iterations more. This is very time consuming: 200 000 RL iterations on 1024×1024 images represent ≈ 80 h on a standard laptop. This clearly illustrates the importance of designing fast algorithms for image restoration.

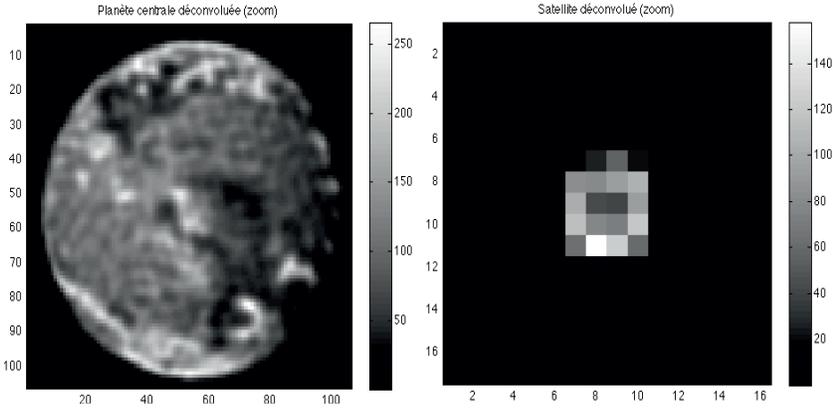


Fig. 23. Deconvolved object in the case of spatial aliasing by RL after 200 000 iterations: zoom on the central planet (*left*) and on the satellite position (*right*).

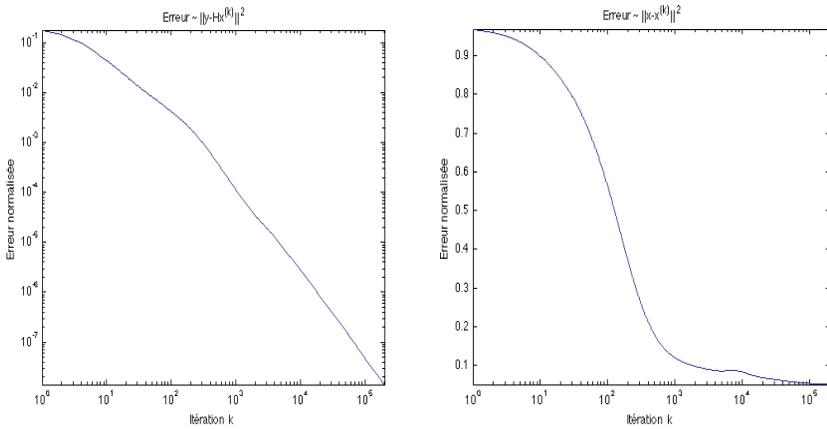


Fig. 24. Normalised error in approximating the data $\frac{\|y - H\mathbf{o}^{(k)}\|^2}{\|y\|^2}$ (*left*), and normalized error with respect to the true object $\frac{\|\mathbf{o} - \mathbf{o}^{(k)}\|^2}{\|\mathbf{o}\|^2}$ (*right*) as a function of the iteration number.

7 Summary and conclusions

This article tried to provide a detailed introduction to the description of the image formation models for diluted pupils array and their densified versions called hypertelescopes. These optical systems represent one of the main promises for the next generation of high angular resolution instruments.

The introduction underlined using historical elements how essential have been high angular resolution observations, transmission of knowledge, and reliance on long term research projects to our current representation of the Universe.

A substantial part of the paper was devoted to the explanations of sampling issues, of their effects on the observed images, and of possible settings that can be used to simulate hypertelescopes images.

In the Fizeau mode (no densification), the image model is a convolution. The densified mode corresponds to hypertelescopes and can be done using either a periscopic setting or inverted Galilean telescopes. We showed in the Appendix that both settings are fully equivalent. In the densified case, convolution generally disappears because frequencies are modulated (translated block-wise), and FAD yields information loss while FSD does not. A convolution may be retrieved in the FSD mode only in the limit of vanishingly small subapertures (infinite fields of view).

This suggests two modeling regimes for densification (hypertelescopes), depending on whether the diameter D is much smaller than the pupil separation d or not. For hypertelescopes made of very large bases (in the kilometer range) and of many small telescopes (centimeters), $D \ll d$, and a convolution model may be a good approximation, at least close to the optical axis. For VLTI-like hypertelescopes, made of moderately large bases (in the hundreds of meter) and of a few large telescopes (in the tens of meters), $D \approx d$ and the image formation models strongly departs from convolution.

We also addressed the issue of restoring such images, and presented classical methods of constrained ML for Gaussian and Poisson noises (RL and ISRA). Faster and regularized deconvolution algorithms should be preferred to RL and ISRA. We provided a detailed survey of such recent methods based on sparse representations.

The two last sections of the paper were dedicated to original studies.

The first study showed that the restoration quality achieved by constrained ML from photon limited images obtained from a diluted array on a grid, or from a densified (but free from spectral aliasing) array are essentially equivalent. We still expect a gain of densified w.r.t. Fizeau images because of the relatively lower cutoff frequency of the former, although we did not provide results supporting this assertion.

The second study (last section) showed that it is possible to recover or at least to “detect” in hypertelescopes (or more generally, interferometric) images quasi point sources that are not only far outside the clean field, but also superimposed on the replicas of other objects. This is true at least for the considered pupil array and in the limit of no noise. Further studies should investigate the effect of noise on the recovery, and of the magnitude difference for the satellite to be recoverable.

Appendix: Densification using Galilean inverted telescopes and recovery of former periscopic expressions

The densification of hypertelescopes can be operated in two ways: using a periscope as in Michelson’s stellar interferometer, or using Galilean inverted telescopes. In the first case, the distance between the subapertures is reduced in the output pupil with respect to the input pupil, while their diameter remains fixed. In

the second case, the relative distances between the subapertures is conserved but their diameter is magnified. In practice of course, the images are rescaled in both settings. We show here, by treating in detail the image formation model of Galilean inverted telescopes, that these two settings are equivalent.

Let us consider a monochromatic plane wave of amplitude $A(\boldsymbol{\beta})$ coming from an angular direction $\boldsymbol{\beta} (\beta_x, \beta_y)$ on the sky, and emitted by an object of intensity $O(\boldsymbol{\beta}) = |A(\boldsymbol{\beta})|^2$. This wave produces at position $\mathbf{r} (r_x, r_y)$ in the plane of the input pupil of a telescope an amplitude $A(\boldsymbol{\beta}) \exp(2i\pi\boldsymbol{\beta}\cdot\mathbf{r}/\lambda)$, where the phase factor accounts for the tilt of the wavefront and the bold dot means scalar product.

The wave $\Psi_1(\mathbf{r}, \boldsymbol{\beta})$ in the input pupil plane of an interferometer made of an array of K cophased identical apertures $P_0(\mathbf{r})$ centered at spatial positions \mathbf{r}_k can thus be written as

$$\begin{aligned} \Psi_1(\mathbf{r}, \boldsymbol{\beta}) &= A(\boldsymbol{\beta}) \sum_{k=1}^K P_0(\mathbf{r} - \mathbf{r}_k) \exp\left(2i\pi\frac{\boldsymbol{\beta}\cdot\mathbf{r}}{\lambda}\right) \\ &= A(\boldsymbol{\beta}) \sum_{k=1}^K P_0(\mathbf{r}) \exp\left(2i\pi\frac{\boldsymbol{\beta}\cdot\mathbf{r}}{\lambda}\right) \star \delta(\mathbf{r} - \mathbf{r}_k) \exp\left(2i\pi\frac{\boldsymbol{\beta}\cdot\mathbf{r}_k}{\lambda}\right), \end{aligned} \quad (7.1)$$

where the last form was first used by Tallon & Tallon-Bosc (1992) to treat the effect of the periscopic transformation in Michelson interferometry. This form explicits the separation between the positions and the geometry of the elementary apertures.

For a hypertelescope, the densification using the periscopic mode basically consists of translating the apertures images from the positions \mathbf{r}_k to the new positions $\mathbf{r}'_k = \mathbf{r}_k/\gamma$, where γ is called the densification factor (Labeyrie 1996). These aspects have been presented in several papers (Tallon & Tallon-Bosc 1992; Labeyrie 1996; Lardiere *et al.* 2007; Aime 2008; Aime *et al.* 2012) and will not be further detailed here.

In contrast to these papers, we present here the formalism for the densification using inverted Galilean telescopes, and show that it leads to results that are identical to the periscopic technique. From a physical point of view this is expected since the two images of the resulting apertures are identical, up to an irrelevant magnifying factor. Nevertheless, the presentation of the theory for the inverted Galilean telescope approach is of interest, at least from a pedagogic point of view.

Using inverted Galilean telescopes for densification amounts to applying a magnification by a real factor $\gamma > 1$ of the wave on each elementary aperture, leaving unchanged the center positions \mathbf{r}_k . In this operation the amplitude of the light is divided by the factor γ , to keep unchanged the energy. In Equation (7.1) this aperture reshaping consists in applying the dilation factor γ to the first term of the convolution. Let us denotes $\Psi_\gamma(\mathbf{r}, \boldsymbol{\beta})$ this amplitude:

$$\Psi_\gamma(\mathbf{r}, \boldsymbol{\beta}) = \frac{A(\boldsymbol{\beta})}{\gamma} \sum_{k=1}^K P_0\left(\frac{\mathbf{r}}{\gamma}\right) \exp\left(2i\pi\frac{\boldsymbol{\beta}\cdot\mathbf{r}}{\gamma\lambda}\right) \star \delta(\mathbf{r} - \mathbf{r}_k) \exp\left(2i\pi\frac{\boldsymbol{\beta}\cdot\mathbf{r}_k}{\lambda}\right). \quad (7.2)$$

For $\gamma = 1$ we obviously recover the original wavefront $\Psi_1(\mathbf{r}, \boldsymbol{\beta})$ of Equation (7.1).

Now let $A_\gamma(\boldsymbol{\alpha}, \boldsymbol{\beta})$ denote the complex amplitude of the wave in the focal plane of the telescope, at angular position $\boldsymbol{\alpha}(\alpha_x, \alpha_y)$ in this plane. This wave is obtained by a scaled Fourier transform of $\Psi_\gamma(\mathbf{r}, \boldsymbol{\beta})$ (see Aime *et al.* in these proceedings):

$$\begin{aligned}
 A_\gamma(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{i\lambda} \iint \Psi_\gamma(\mathbf{r}, \boldsymbol{\beta}) \exp\left(-2i\pi \frac{\mathbf{r} \cdot \boldsymbol{\alpha}}{\lambda}\right) d\mathbf{r} \\
 &= \frac{\gamma A(\boldsymbol{\beta})}{i\lambda} \sum_{k=1}^K \widehat{P}_0\left(\frac{\gamma\boldsymbol{\alpha} - \boldsymbol{\beta}}{\lambda}\right) \exp\left(-2i\pi \mathbf{r}_k \cdot \left(\frac{\boldsymbol{\alpha} - \boldsymbol{\beta}}{\lambda}\right)\right). \tag{7.3}
 \end{aligned}$$

The factor γ appears now at the numerator after a change of variable in the 2D integral. The elementary intensity in the case of inverted Galilean telescopes $I_\gamma^G(\boldsymbol{\alpha}, \boldsymbol{\beta})$ produced by the point source coming from the direction $\boldsymbol{\beta}$ at position $\boldsymbol{\alpha}$ in the focal plane can therefore be written as:

$$\begin{aligned}
 I_\gamma^G(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= |A_\gamma(\boldsymbol{\alpha}, \boldsymbol{\beta})|^2 = A_\gamma(\boldsymbol{\alpha}, \boldsymbol{\beta}) A_\gamma^*(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
 &= O(\boldsymbol{\beta}) \frac{\gamma^2}{\lambda^2} \sum_{k=1}^K \sum_{l=1}^K \left| \widehat{P}_0\left(\frac{\gamma\boldsymbol{\alpha} - \boldsymbol{\beta}}{\lambda}\right) \right|^2 \exp\left(2i\pi(\mathbf{r}_k - \mathbf{r}_l) \cdot \left(\frac{\boldsymbol{\alpha} - \boldsymbol{\beta}}{\lambda}\right)\right), \tag{7.4}
 \end{aligned}$$

where superscript * denotes complex conjugate. $I_\gamma^G(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is indeed real (the imaginary parts of the complex exponentials involving $\mathbf{r}_k - \mathbf{r}_l$ and $\mathbf{r}_l - \mathbf{r}_k$ cancel by pairs); the notation with complex exponentials will later evidence a Fourier transform that will be used in Equation (7.7).

The image in the focal plane $I_\gamma^G(\boldsymbol{\alpha})$ is obtained by summing all contributions coming from the object:

$$I_\gamma^G(\boldsymbol{\alpha}) = \iint I_\gamma^G(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\beta}. \tag{7.5}$$

Taking the Fourier transform of $I_\gamma^G(\boldsymbol{\alpha})$, we also have:

$$\widehat{I}_\gamma^G(\mathbf{u}) = \iint \widehat{I}_\gamma^G(\mathbf{u}, \boldsymbol{\beta}) d\boldsymbol{\beta}, \tag{7.6}$$

where \mathbf{u} is the angular frequency associated to $\boldsymbol{\alpha}$.

Using the notation $\mathbf{u}_{kl} = (\mathbf{r}_k - \mathbf{r}_l)/\lambda$ in Equation (7.4), the expression of $\widehat{I}_\gamma^G(\mathbf{u}, \boldsymbol{\beta})$ can be written as:

$$\begin{aligned}
 \widehat{I}_\gamma^G(\mathbf{u}, \boldsymbol{\beta}) &= \iint I_\gamma^G(\boldsymbol{\alpha}, \boldsymbol{\beta}) \exp(-2i\pi \mathbf{u} \cdot \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
 &= O(\boldsymbol{\beta}) \sum_{k=1}^K \sum_{l=1}^K \exp(-2i\pi \frac{\boldsymbol{\beta}}{\gamma} \cdot (\mathbf{u} + (\gamma - 1)\mathbf{u}_{kl})) \iint |\widehat{P}_0(\boldsymbol{\xi})|^2 \exp(-2i\pi \boldsymbol{\xi} \cdot \frac{1}{\gamma}(\mathbf{u} - \mathbf{u}_{kl})) d\boldsymbol{\xi} \tag{7.7}
 \end{aligned}$$

If we denote by S the telescope area and by $T_0(\mathbf{u})$ the normalized optical transfer function (OTF) defined by

$$T_0(\mathbf{u}) = \frac{1}{S} \iint P(\mathbf{r}) P^*(\mathbf{r} - \lambda \mathbf{u}) d\mathbf{r}, \tag{7.8}$$

Equation (7.7) becomes

$$\widehat{I}_\gamma^G(\mathbf{u}, \boldsymbol{\beta}) = O(\boldsymbol{\beta}) \sum_{k=1}^K \sum_{l=1}^K \exp\left(-2i\pi \frac{\boldsymbol{\beta}}{\gamma} \cdot (\mathbf{u} + (\gamma - 1)\mathbf{u}_{kl})\right) ST_0\left(\frac{1}{\gamma}(\mathbf{u} - \mathbf{u}_{kl})\right). \quad (7.9)$$

Substituting this expression into Equation (7.6), we obtain:

$$\widehat{I}_\gamma^G(\mathbf{u}) = \sum_{k=1}^K \sum_{l=1}^K \widehat{O}\left(\frac{1}{\gamma}(\mathbf{u} + (\gamma - 1)\mathbf{u}_{kl})\right) ST_0\left(\frac{1}{\gamma}(\mathbf{u} - \mathbf{u}_{kl})\right). \quad (7.10)$$

This expression depends on the collecting surface of the telescope. We can get rid of this surface by dividing by KS , which finally leads to:

$$\begin{aligned} \widehat{I}_\gamma^G(\mathbf{u}) &= \frac{1}{KS} \widehat{I}_\gamma^G(\mathbf{u}) \\ &= \frac{1}{K} \sum_{k=1}^K \sum_{l=1}^K \widehat{O}\left(\frac{1}{\gamma}(\mathbf{u} + (\gamma - 1)\mathbf{u}_{kl})\right) T_0\left(\frac{1}{\gamma}(\mathbf{u} - \mathbf{u}_{kl})\right) \\ \widehat{I}_\gamma^G(\mathbf{u}) &= \widehat{O}\left(\frac{\mathbf{u}}{\gamma}\right) T_0\left(\frac{\mathbf{u}}{\gamma}\right) + \frac{1}{K} \sum_{k=1}^K \sum_{l \neq k}^K \widehat{O}\left(\frac{1}{\gamma}(\mathbf{u} + (\gamma - 1)\mathbf{u}_{kl})\right) T_0\left(\frac{1}{\gamma}(\mathbf{u} - \mathbf{u}_{kl})\right). \end{aligned} \quad (7.11)$$

We see that the sampling in this case operates on a dilated version of the spectrum $\widehat{O}\left(\frac{\mathbf{u}}{\gamma}\right)$ using transfer functions that are dilated as well. Performing the change of variable $\boldsymbol{\nu} = \mathbf{u}/\gamma$, we recover the periscopic mode:

$$\widehat{I}_\gamma^G(\boldsymbol{\nu}) = \widehat{O}(\boldsymbol{\nu}) T_0(\boldsymbol{\nu}) + \frac{1}{K} \sum_{k=1}^K \sum_{l \neq k}^K \widehat{O}\left(\boldsymbol{\nu} + \mathbf{u}_{kl} - \frac{\mathbf{u}_{kl}}{\gamma}\right) T_0\left(\boldsymbol{\nu} - \frac{\mathbf{u}_{kl}}{\gamma}\right), \quad (7.12)$$

which shows that images obtained by densification in periscopic mode or using inverted Galilean telescopes are the same (compare to Eq. (2.7) and see Fig. 2).

In both cases, if we take $\gamma = 1$ we indeed recover the Fizeau mode of Equation (2.6):

$$\widehat{I}_1^G(\mathbf{u}) = \widehat{I}_1^P(\mathbf{u}) = \widehat{I}^F(\mathbf{u}) = \sum_{k=1}^K \sum_{l=1}^K \widehat{O}(\mathbf{u}) T_0(\mathbf{u} - \mathbf{u}_{kl}). \quad (7.13)$$

References

- Aime, C., 2008, *A&A*, 483, 361
 Aime, C., Lantéri, H., Diet, M., & Carlotti, A., 2012, *A&A*, 543, A42
 Anderson, J.A., 1920, *ApJ*, 51, 263

- Baron, F., *et al.*, 2012, The 2012 Interferometric Imaging Beauty Contest, Proc. SPIE: Astronomical Telescopes and Instrumentation Conference (Amsterdam)
- Beck, A., & Teboulle, M., 2009, Siam J. Imaging Sciences, 2, 183
- Candès, E. J., Romberg, J., & Tao, T., 2006, IEEE Trans. Inf. Theory, 52, 489
- Carrillo, R.E., McEwen, J.D., & Wiaux, Y., 2012 [[arXiv:1205.3123](https://arxiv.org/abs/1205.3123)]
- Chen, S.S., Donoho, D.L., & Saunders, M.A., 1998, SIAM J. Scientific Computing, 20, 33
- Cornwell, T.J., 2009, A&A, Special issue, 500, 65
- Cornwell, T.J., 2008, IEEE J. Selected Topics Signal Proc., 2, 793
- Dabbech, A., Mary, D., & Ferrari, C., 2012, Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 3665
- Daube-Witherspoon, M.E., & Muehllehner, G., 1986, IEEE Trans. Med. Imaging, 5, 61
- Donoho, D.L., 2006, IEEE Trans. Inf. Theory, 52, 1289
- Elad, M., Milanfar, P., & Rubinstein, R., 2007, Inverse Probl., 23, 947
- Fizeau, H., 1868, C. R. Hebd. Seanc. Acad. Sci. Paris, 66, 934
- Fornassier, M., 2010, Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter; 1 edition (2010)
- Giovannelli, J.-F., & Coulais, A., 2005, A&A, 439, 401
- Gribonval, R., 2009-2012, <http://small-project.eu/publications>
- Gribonval, R., & Nielsen, M., 2003, IEEE Trans. Inf. Theory, 49, 3320
- Edmond Halley, 1715, A short History of several New-Stars. Phil. Trans, XXIX, 354 (1715) Available online
- Högbom, J.A., 1974, A&AS, 15, 417
- Labeyrie, A., *et al.*, 2012, Optical and Infrared Interferometry III, Proceedings of the SPIE, Vol. 8445
- Labeyrie, A., 1996, A&AS, 118, 517
- Labeyrie, A., 1975, ApJ, 196, L71
- Lantéri, H., Roche, M., & Aime, C., 2002, Inverse Probl., 18, 1397
- Lantéri, H., *et al.*, 2002, Signal Proc., 82, 1481
- Lardièrre, O., Martinache, F., & Patru, F., 2007, MNRAS, 375, 977
- Li, F., Cornwell, T.J., & de Hoog, F., 2011, A&A, 528, 31
- Lucy, L.B., 1974, AJ, 79, 745
- Magain, P., Courbin, F., & Sohy, S., 1998, ApJ, 494, 472
- Mallat, S., 2008, A wavelet tour of signal processing: the sparse way, 3rd edition (Academic Press)
- Mallat, S., & Zhang, Z., 1993, IEEE Trans. Sig. Proc., 41, 3397
- McEwen, J.D., & Wiaux, Y., 2011, MNRAS, 413, 1318
- Michelson, A.A., 1891, Nature, 45, 160
- Michelson, A.A., 1920, ApJ, 51, 257
- Michelson, A.A., & Pease, F.G., 1921, ApJ, 53, 249
- Mignard, F., & Martin, C., 1997, Pour la Science, 235
- Pirzkal, N., Hook, R.N., & Lucy, L.B., 2000, In ASP Conference Series, Astronomical Data Analysis, Software and Systems IX, Paris, ed. N. Manset, C. Veillet & D. Crabtree, 216, 657

- Rau, U., Bhatnagar, S., Voronkov, M.A., & Cornwell, J.T., 2009, *Proc.*, 97, 1472
- Richardson, W.H., 1972, *J. Opt. Soc. Am.*, 62, 55
- Schwarz, U.J., 1978, *A&A*, 65, 417
- Starck, J.L, Pantin, E., & Murtagh, F., 2002, *PASP*, 114, 1051
- Starck, J.L, Murtagh, F., & Fadili, M.-J., 2010, *Sparse Image and Signal Processing - Wavelets, Curvelets, Morphological Diversity* (Cambridge University Press)
- Stephan, E., 1873, *C. R. Hebd. Seanc. Acad. Sci. Paris*, 76, 1008
- Stephan, E., 1873, *C. R. Hebd. Seanc. Acad. Sci. Paris*, 78, 1008
- Kopilovich, L.E., & Sodin, L.G., 2001, *Multielement System Design in Astronomy and Radio Science, Astrophysics and Space Science Library* (Kluwer, 2001)
- Tallon, M., & Tallon-Bosc, I., 1992, *A&A*, 253, 641
- Thiébaud, E., 2005, *NATO ASIB Proc. 198: Optics Astrophys.*, 397
- Thiébaud, É., Soulez, F., & Denis, L., 2012, *J. Opt. Soc. A*, submitted
- Vannier, M., *et al.*, 2010, *Spectral regularization and sparse representation bases for interferometric imaging*, *Proc. SPIE: Astronomical Telescopes and Instrumentation* (Conference, San Diego)
- Wakker, B.P., & Schwarz, U.J., 1988, *A&A*, 200, 312
- Wenger, S., Darabi, S., Sen, P., Glassmeier, K.H., & Magnor, M., 2010, *Proc. IEEE Int. Conf., Image Process., IEEE Signal Process. Soc.*, 1381
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A.M.M., & Vanderghenst, P., 2009a, *MNRAS*, 395, 1733
- Wiaux, Y., Puy, G., Boursier, Y., & Vanderghenst, P., 2009b, *MNRAS*, 400, 1029