SCALE INVARIANT DIVERGENCES FOR SIGNAL AND IMAGE RECONSTRUCTION

Henri Lantéri, Céline Theys and Claude Aime

Laboratoire Lagrange, UMR CNRS 7293,
Université de Nice Sophia-Antipolis, Observatoire de la Côte d'Azur

ABSTRACT

The subject of this paper is the reconstruction of a signal or an image under constraints of non negativity and of constant sum. The sum constraint is imposed by the use of scale invariant divergences, which allows the development of simple iterative reconstruction algorithms. Two families of divergences between two data fields p and q are considered, the α-divergence and the β-divergence. A procedure is applied to make them scale-invariant w.r.t. p and q. The resulting method is an interior point type algorithm useful in the context of ill-posed problems. Numerical illustrations are given for the deconvolution of a solar spectrum and an interferometric image.

Index Terms— Inverse problems, non-negativity and sum constrained minimization, scale invariant divergences.

1. INTRODUCTION

The problem we consider in this work appears frequently in various physical applications, such as deconvolution of spectra or images, for example. The task is to minimize w.r.t. the unknown x a divergence between the measured noisy data, here p, and the physical model $q = Hx$, with $h_{ij} \geq 0$ subject to the constraints $x_i > 0\ \forall i$, $\sum_i x_i = C,\ C > 0$. In the specific deconvolution problem considered here, the matrix H is highly ill-conditioned and moreover we have $\sum_i h_{ij} = 1$ so that $C = \sum_i p_i$.

This problem has been studied extensively in recent years and some solutions have been proposed mainly in the case of a quadratic cost function and KL divergence. Here, we propose a method applicable to any divergence constructed from a convex function. The non negativity constraint is taken into account by applying the KKT (Karush Kuhn Tucker) conditions, [1]. We have shown in previous papers, [2, 3, 4], that the sum constraint can be taken into account using a change of variables. In the present paper we show how to obtain a scale invariant divergence, so that the sum constraint can be fulfilled very simply in the iterative algorithms due to the scale invariance properties of such divergences.

The proposed iterative algorithms are detailed and illustrated on the deconvolution of two astrophysical examples. The first one is the deconvolution of solar spectrum showing pure emission lines in the extreme UV. The second one is the deconvolution of a possible Earth-like exoplanet observed with an interferometer in space.

2. SCALE INVARIANT DIVERGENCES

We consider divergences constructed from the convex function:

$$f(x) = \frac{1}{\gamma(\gamma-1)}\left[x^{\gamma} - \gamma x - (1-\gamma)\right]. \tag{1}$$

For this function, we have $f(1) = 0$ and $f'(1) = 0$. A Csiszár divergence, [5], between two data fields p and q built on a convex function f is defined by:

$$A(p\|q) = \sum_i q_i f\left(\frac{p_i}{q_i}\right) \tag{2}$$

and for the particular function given by (1), we obtain:

$$A_{\gamma}(p\|q) = \frac{1}{\gamma(\gamma-1)}\left[\sum_i \left(p_i^{\gamma} q_i^{1-\gamma} - \gamma p_i + (\gamma-1)\, q_i\right)\right]. \tag{3}$$

This is the so-called α-divergence widely analyzed by Amari et al., [6, 7]. In the same way, the Bregman divergence, [8], founded on the convex function f, is defined as:

$$B(p\|q) = \sum_i \left(f(p_i) - f(q_i) - (p_i - q_i)\, f'(q_i)\right), \tag{4}$$

and using the function f, we obtain:

$$B_{\gamma}(p\|q) = \frac{1}{\gamma(\gamma-1)}\left[\sum_i \left(p_i^{\gamma} + (\gamma-1)\, q_i^{\gamma} - \gamma p_i q_i^{\gamma-1}\right)\right]. \tag{5}$$

This is the β-divergence proposed by Basu et al., [9] and by Eguchi and Kano, [10].

The first step in our analysis is to apply on the divergences, (3) and (5), a transformation that allows us to obtain new divergences that become insensitive to a positive multiplicative constant factor of the values of q. The aim is to obtain divergences whose values depend only of the shape of p and q, and not of their relative amplitudes.

The procedure suggested by Eguchi and Kato, [11] can be summarized as follows for a divergence $D(p\|q)$:

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

1. Considering the factor $T(p,q) > 0$, express $D(p\|Tq)$.
2. Solve, for $T > 0$ the equation $\frac{\partial D(p\|Tq)}{\partial T} = 0$.
3. Insert the expression of T so obtained in $D(p\|Tq)$.

The resulting factors $T(p,q)$ obtained for, respectively, the α-divergences and the β-divergences are :

$$TA\left(p,q\right) = \left[\frac{\sum_j p_j^\gamma q_j^{1-\gamma}}{\sum_j q_j}\right]^{\frac{1}{\gamma}}, \tag{6}$$

$$TB\left(p,q\right) = \frac{\sum_j p_j q_j^{\gamma-1}}{\sum_j q_j^\gamma}. \tag{7}$$

When $q \to p$, $TA\left(p,q\right)$ and $TB\left(p,q\right) \to 1$. The scale invariant divergences obtained from (3) and (5) are given by:

$$\begin{aligned} &AI_\gamma(p\|q) = \frac{1}{\gamma-1} \\ &((\sum_i q_i)^{\frac{\gamma-1}{\gamma}}(\sum_i p_i^\gamma q_i^{1-\gamma})^{\frac{1}{\gamma}} - \sum_i p_i) \end{aligned} \tag{8}$$

and

$$\begin{aligned} &BI_\gamma(p\|q) = \frac{1}{\gamma(\gamma-1)} \\ &(\sum_i p_i^\gamma - (\sum_i q_i^\gamma)^{1-\gamma}(\sum_i p_i q_i^{\gamma-1})^\gamma). \end{aligned} \tag{9}$$

These new divergences are insensitive to scale changes on q and appear as the difference of two terms. We can apply on each of these terms an increasing function without modifying the sign of the expression. We perform such operation using the *deformed or generalized logarithm* function defined for $u > 0$ by:

$$L_d(u) = \frac{u^d - 1}{d} \tag{10}$$

whose limits are $(u-1)$ if $d \to 1$ and $\log(u)$ if $d \to 0$. In this later case we obtain respectively:

$$\begin{aligned} LAI_\gamma\left(p\|q\right) = &\frac{1}{\gamma}\log\sum_i q_i - \frac{1}{(\gamma-1)}\log\sum_i p_i \\ &+ \frac{1}{\gamma\left(\gamma-1\right)}\log\sum_i p_i^\gamma q_i^{1-\gamma} \end{aligned} \tag{11}$$

which has been exhibited by Cichocki et al., [12, 13], and

$$\begin{aligned} LBI_\gamma\left(p\|q\right) = &\frac{1}{\gamma\left(\gamma-1\right)}\log\sum_i p_i^\gamma + \frac{1}{\gamma}\log\sum_i q_i^\gamma - \\ &\frac{1}{\gamma-1}\log\sum_i p_i q_i^{\gamma-1} \end{aligned} \tag{12}$$

which is the divergence of Fujisawa, [14] previously used in [4]. Note that the logarithmic forms of the divergences are also invariant to a scale change on p.

A similar analysis could be performed on the generalized $\alpha\beta$ divergence proposed by [15], from which the particular cases of (3) and (5) are recovered. For the sake of clarity, this generalized approach is not presented here.

3. MINIMIZATION ALGORITHMS

In this section, we develop iterative algorithms based on the scale invariant divergences LAI_γ and LBI_γ and we denote by $\mathbf{p}$ the vector of noisy data and by $\mathbf{q}$ the vector describing the physical process: $\mathbf{q} = H\mathbf{x}$. We consider the problem:

$$\min_{\mathbf{x}} D\left(\mathbf{p}\|H\mathbf{x}\right) \quad s.t \quad x_i \geq 0, \quad \sum_i x_i = C. \tag{13}$$

where D is any divergence. The non-negativity constraint is taken into account using the KKT conditions. As developed in [1, 4], the KKT conditions for the non-negativity constraint expresses that at the convergence to the optimum $\mathbf{x}^*$, we must have:

$$x_i^*\left[\nabla_{\mathbf{x}} D\left(\mathbf{p}\|H\mathbf{x}^*\right)\right]_i = 0. \tag{14}$$

This condition can then be used as a descent direction in an iterative gradient descent algorithm as follows:

$$x_i^{k+1} = x_i^k + \alpha_i^k x_i^k\left[-\nabla_{\mathbf{x}} D\left(\mathbf{p}\|H\mathbf{x}^k\right)\right]_i \tag{15}$$

To fulfill the non-negativity of the estimate at each iteration, the maximum stepsize $\alpha_{\max}^k$ must be determinated first, then to ensure the convergence, the stepsize α^k is computed in the range $\left[0, \alpha_{\max}^k\right]$ using a line search method such as the Armijo method [16] for example.

We can deal with the constraint of constant sum by a change of variables as mentionned in [4]. For the scale invariant divergences, two more possibilities are offered due to the properties of these divergences:

i) either in the iterative minimization process, we proceed to a normalization of the solution after each iteration. This procedure does not modify the value of the objective function and can be associated to any algorithm.

ii) or we use a specific property of the gradient of these divergence ; in the following we focus on the corresponding algorithms.

The iterative algorithm, (15), developed for the divergence LAI_γ leads to:

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k\left[H^T\frac{\mathbf{p}^\gamma \circ \left(\mathbf{q}^k\right)^{-\gamma}}{\sum_j p_j^\gamma\left(q_j^k\right)^{1-\gamma}} - H^T\frac{\mathbf{1}}{\sum_j q_j^k}\right]_i \tag{16}$$

and for the divergence LBI_γ:

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k\left[H^T\frac{\mathbf{p}\circ\left(\mathbf{q}^k\right)^{\gamma-2}}{\sum_j p_j\left(q_j^k\right)^{\gamma-1}} - H^T\frac{\left(\mathbf{q}^k\right)^{\gamma-1}}{\sum_j\left(q_j^k\right)^\gamma}\right]_i \tag{17}$$

In these expressions, the symbol "$\circ$" is used for the Hadamard (pointwise) product.

It can be easily verified that for the algorithms (16) and (17), we have $\sum_i x_i^{k+1} = \sum_i x_i^k$, then, the sum constraint is fulfilled for all the successive estimates if the sum constraint is imposed on the initial estimate. This is the main property

connected to scale invariant divergences and consequently the proposed procedure is an interior point method, i.e all the successive estimates fulfill the constraints.

Obviously multiplicative algorithms could be proposed as well following [1], but the convergence of such algorithms is never demonstrated but for particular cases (KL divergence and least squares) due to the lack of tuning parameter (step-size). Moreover, the convergence speed of these algorithms, if they converge, is generally low and the property of spontaneous fixed sum appearing in procedure ii) is lost then the procedure i) must be used.

Let us note that the algorithms corresponding to the divergences AI_γ and BI_γ, (8) and (9), are very similar to those of (16) and (17). The difference appears as a multiplicative factor depending of the iteration k in the corrective part of the algorithms, beside α^k. This factor will be implicitly taken into account during the computations dealing with the step size.

4. SIMULATIONS RESULTS

4.1. Solar spectrum

The data used in this section is a solar spectrum extracted from the BASS 2000 Solar survey archive available on line at `http : //bass2000.obspm.fr/`. It corresponds to solar ultraviolet emission lines for wavelengths from 760 to 780 Å, (Figure 1 Top). It shows a few Ne VIII lines emitted from the transition region between the corona and the chromosphere.

The spectrum sampled over 2048 data points is taken as the object $\mathbf{x}$. We assume that the spectrum is analyzed with a spectrometer producing a convolution effect. The blurring window H depends on diffraction effects and geometrical construction of the spectrometer. It is not represented here for the sake of conciseness.

The blurred data suffer a Poisson transformation (corresponding to photodetection) followed by the addition of an independent Gaussian noise. These effects can be mathematically described by:

$$\mathbf{p} = \mathcal{P}(H\mathbf{x}) + \mathbf{n} \tag{18}$$

with $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$, resulting p data are shown in the bottom curves of Figure 1 for a total number of 3×10^6 photoelectrons and a standard deviation of $\sigma = 100$ photons. The LAI_γ and LBI_γ, (16) and (17) are run for $\gamma = 3$. The step size has been computed by the Armijo rule. Top curves of Figure 2 represent a plot of the divergences LAI_3 (red line) and LBI_3 (blue line) of (11) and (12) as a function of the number of iterations. It is then verified that the criterion to be minimized is a decreasing function of k. Bottom curves of Figure 2 is a plot of the error of reconstruction between the reconstructed estimate $\mathbf{x}^k$ and the exact value $\mathbf{x}^*$ (known in this numerical simulation). Note that to be independent of the divergence, the reconstruction error computed is the normalized quadratic one,

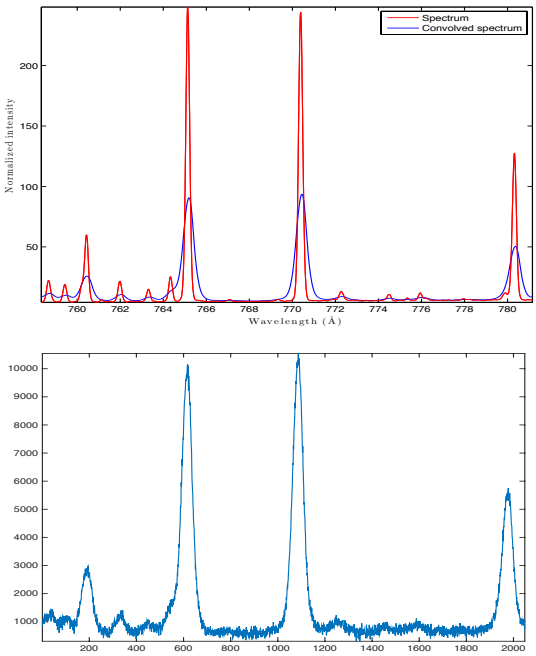

Fig. 1. Top: Emission lines $\mathbf{x}$ of the solar spectrum and convolved spectrum $\mathbf{q} = H\mathbf{x}$. Bottom: Noisy spectrum $\mathbf{p}$ with 3.10^6 photons in the whole spectrum and $\sigma = 100$.

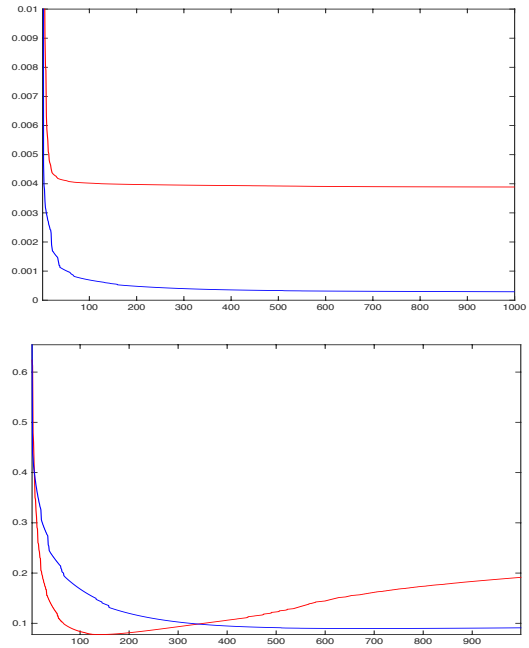

Fig. 2. Top: LAI_3 (red line) and LBI_3 (blue line) divergences as a function of k. Bottom : LAI_3 (red line $k = 139$) and LBI_3 (blue line $k = 678$) reconstruction error as a function of k.

$||\mathbf{x}^k - \mathbf{x}^*||/||\mathbf{x}^*||$. These curves exhibit the classical behavior: after a decrease of the reconstruction error, we observe an increase of this error. Obviously this phenomena known as "semi-convergence" does not depend on the form of the error function. It corresponds to the fact that when the iterations progress, high frequencies components appear in the solution x^k while simultaneously Hx^k does not changes.

It is found, in this example, that the LAI procedure converges more rapidly to the best estimate than the LBI one (iteration 139 vs 678). Of course, these numerical values, stopping iteration and corresponding error of the reconstruction, strongly depend on the parameters of the simulation.

As usual in the context of an ill-posed problem, the iterations number required to reach the minimum of the reconstruction error increases with the quality of the data, the iterations must be rapidly stopped for noisy data to prevent the amplification of noise. Results obtained with the two algorithms are very satisfactory in a qualitative sense and close one another, as it can be seen in Figure 3. Reconstructed spectra can be used effectively for a physical interpretation. It was, of course, checked that the integral of the reconstructed spectra has been preserved throughout the iterations and the representation at the scale of the object in Figure 3 is made for practical considerations.

Influence of γ. We have chosen $\gamma = 3$ in the previous experiments and it would be interesting to study the influence of gamma on the quality of reconstruction. A first study was done by computing the value of the reconstruction error as a function of k and γ in the case of the LBI_γ algorithm. and the result is plotted as a contour plot in the Figure 4. We can see that the minimum is obtained for $\gamma = 2$ at iteration 662.

4.2. Interferometric image

The data used in this section correspond to future observations in space with sparse array of interferometers as in the Luciola project of Labeyrie et al. [17]. A particular point of these observations is that spatial frequencies are transmitted in non-continuous regions of the Fourier plane. The reconstruction must fill the empty zones, as discussed in Aime et al. [18]. A thorough review of array configurations can be found in the book of Kopilovich and Sodin [19]. Here we used 35 identical apertures. As a result of our choice of a particular non-redundant configuration, the PSF (Fig. 5, top left) appears as the replication on a grid of the Airy pattern of the giant meta-telescope we seek to synthesize, with an added speckle-like structure due to the sparsity of the configuration (35 apertures instead of 2500 for a regular grid).

The object $\mathbf{x}$ is a possible exoplanet corresponding to the Earth-Moon system shown in Fig. 6. Due to the structure of the PSF, its pattern appears as replicas in the noiseless focal plane image $H\mathbf{x}$ of Fig. 5. The distance between the two bodies is chosen on purpose to introduce a visual ambiguity in $H\mathbf{x}$ whether the Moon is at the right or at the left of the Earth.

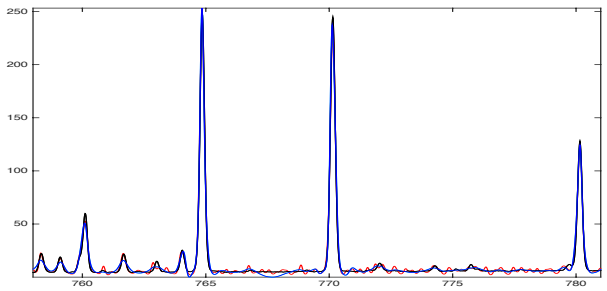

Fig. 3. LAI_3 (red line $k = 139$) and LBI_3 (blue line $k = 678$) reconstruction spectra at the minimum of the reconstruction divergence.

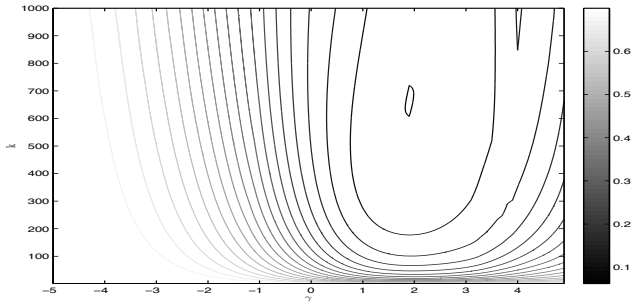

Fig. 4. LBI_γ normalized quadratic reconstruction error as a function of k and γ, minimum for $\gamma = 2$ and $k = 662$.

The image $\mathbf{p}$ (bottom figure) result of the transformation described by Eq. 18 for 3×10^7 photons and $\mathbf{n} \sim \mathcal{N}(0, 29)$. Numerical data are of 1024×1024 points.

During iterations, both LAI and LBI algorithms gradually suppress image replicas and improve the final image. Best results obtained for LAI_2 and LBI_2 are given in Fig. 5. Comparison of Fig. 5 and Fig. 6, either grayscale or cuts, shows the excellent results obtained by the application of these algorithms. Much better results are obtained with less iteration steps with LBI_2 ($k = 37$) than with LAI_2 ($k = 429$), the minimum of the reconstruction error being 0.2328 and 0.2691, respectively.

5. CONCLUSION

Iterative algorithms of minimization subject to constraints of non negativity and constant sum have been proposed and have given excellent results for 1D and 2D data. The constant sum constraint is taken into account by the use, as objective functions, of scale invariant divergences. This leads to interior point type algorithms that allow the stopping of iterations before convergence, useful in the context of ill-posed problems. More experiments need to be conducted, for example the difference of behavior of LAI and LBI algorithms observed in

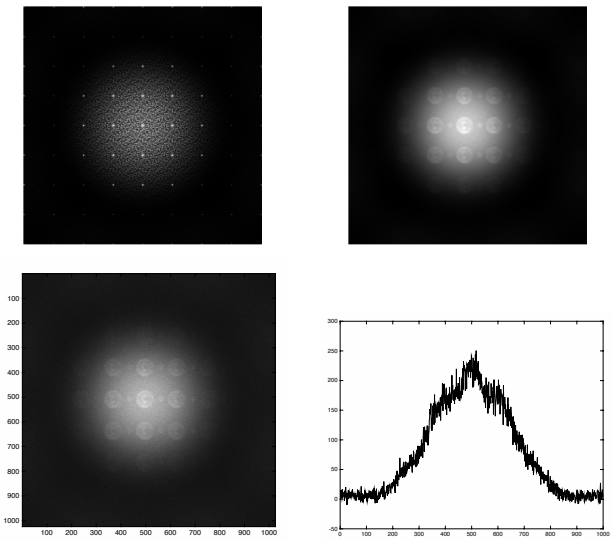

Fig. 5. Top: PSF (left) and noiseless focal plane image $H\mathbf{x}$ (right). Bottom: noisy image $\mathbf{p}$ for 30 M photons and $\sigma^2 = 29$ (left) and a radial cut (right).

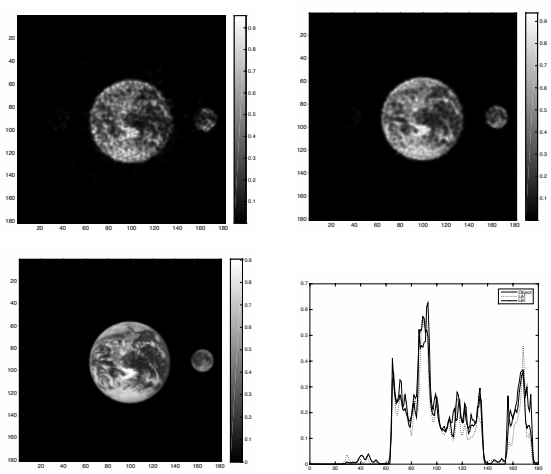

Fig. 6. Top: best reconstructed images for LAI_2 ($k = 429$) and LBI_2 ($k = 37$). Only the central 180×180 central points of the images are shown there. Bottom: the object $\mathbf{x}$ (the exo-Earth and its Moon), and radial cuts of the reconstructed objects.

our two examples should be further investigated. The proposed algorithms can be easily extended to the blind deconvolution problem due to commutativity of the convolution product. The method can be directly applied to the linear unmixing problem with the important advantage that generally the matrix H is not ill-conditioned in this case. The application to NMF is not immediate and concerns uniquely the abundance vectors, because the matrix product is not commutative.

6. REFERENCES

[1] H. Lantéri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints- multiplicative algorithms," *Inverse problems*, vol. 18, pp. 1397–1419, 2002.

[2] H. Lantéri, C. Theys, C. Richard, and C. Févotte, "Split gradient method for nonnegative matrix factorization," in *Proc. 18th European Signal Processing Conference (EUSIPCO'10), Aalborg, Denmark*, 2010.

[3] C. Theys, N. Dobigeon, J.Y Tourneret, and H. Lantéri, "Linear unmixing of hyperspectral images using a scaled gradient method," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Cardiff, Wales, UK, Aug. 2009, pp. 729–732.

[4] H. Lantéri and C. Theys, "La gamma divergence de Fujisawa-Eguchi, une alternative pour imposer la contrainte de somme. application au démélange linéaire.," in *GRETSI*, 2013.

[5] I. Csiszär, "Information measures: a critical survey," in *Transactions of the 7^{th} Prague conference*, 1974.

[6] S. I. Amari, "Alpha divergence is unique belonging to the f-divergence and bregman divergence classes," *IEEE Transactions on Information Theory*, 2009.

[7] S. I. Amari and H. Nagaoka, *Methods of information geometry*, Oxford University Press, New york, 2000.

[8] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.

[9] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones, "Robust and efficient estimation by minimizing a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[10] S. Eguchi and Y. Kano, "Robustifying maximum lakelihood estimation," Tech. Rep., Institute of statistical mathematics, 2001.

[11] S. Eguchi and S. Kato, "Entropy and divergences associated with power function and the statistical application," *Entropy*, 2010.

[12] A. Cichocki and S. Amari, "Families of alpha-beta and gamma divergences: flexible and robust measures of similarities," *Entropy*, vol. 12, 2010.

[13] A. H. Phan A. Cichocki, R. Zdunek and S.I. Amari, *Non negative matrix and tensor factorization*, J. Wiley, 2010.

[14] H. Fujisawa and S. Eguchi, "Robust parameter estimation with small bias against heavy contamination," *Multivariate Analysis*, vol. 99, no. 9, 2008.

[15] A. Cichocki, S. Cruces, and S. I. Amari, "Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization," *Entropy*, vol. 13, no. 1, pp. 134–170, 2011.

[16] L. Armijo, "Minimization of functions having continuous derivatives," *Pacific Journal of Mathematics*, , no. 16, pp. 1–3, 1966.

[17] A. Labeyrie, H. Le Coroller, J. Dejonghe, O. Lardière, C. Aime, K. Dohlen, D. Mourard, R. Lyon, and K. G. Carpenter, "Luciola hypertelescope space observatory: versatile, upgradable high-resolution imaging, from stars to deep-field cosmology," *Experimental Astronomy*, vol. 23, pp. 463–490, Mar. 2009.

[18] C. Aime, H. Lantéri, M. Diet, and A. Carlotti, "Strategies for the deconvolution of hypertelescope images," *Astronomy and Astrophysics*, vol. 543, pp. A42, July 2012.

[19] L. E. Kopilovich and L. G. Sodin, Eds., *Multielement system design in astronomy and radio science*, vol. 268 of *Astrophysics and Space Science Library*, 2001.