Least Squares Image Estimation in the Presence of Drift and Pixel Noise

Lorenzo Piazzo[a,d], Maria Carmela Raguso[a], Javier Gracia Carpio[b] and Bruno Altieri[c]
a) DIET - Sapienza University - Rome, Italy b) MPE - Max Planck Institute - Garching, Germany
c) ESAC - ESA - Madrid, Spain d) Email: lorenzo.piazzo@uniroma1.it
This work has been partially funded by the ESA, Contract no. 4000115155/15/NL/IB (Unido).

***Abstract*— We discuss Least Squares (LS) image estimation for large data in the presence of electronic noise and drift. We introduce a data model where, in addition to the electronic noise and drift, also an additional type of noise, termed pixel noise, is considered. This noise arises when the sampling does not take place on a regular grid and may bias the estimate if not accounted for. Based on the model, we present an efficient Alternating Least Squares (ALS) algorithm, producing the LS image estimate. Finally, we apply the ALS to the data of the Photodetector Array Camera and Spectrometer (PACS), which is an infrared photometer onboard the European Space Agency (ESA) Herschel space telescope. In this context, we discuss the ALS implementation and complexity and present an example of the results.**

I. Introduction

We consider an acquisition instrument where the image to be acquired is scanned by one or more sensors. During the scan, each sensor's output is sampled at regular times, to produce a sequence of readouts which is called a time-series. Examples of such instruments include biomedical, e.g. [1], [2], [3], and astrophysical, e.g. [4], [5], imaging systems. The image has to be estimated from the time-series but this task is complicated by a variety of disturbances, depending on the specific system. Moreover, the data volume produced by modern instruments may be huge, which makes the implementation difficult and calls for efficient estimation algorithms.

In this paper, we discuss the estimation problem for large data in the presence of two different disturbances. The first one is the ubiquitous additive noise, due to the readout electronics. The second one is a drift, causing a slowly time varying deviation of the time-series from the baseline level, which affects several practical instruments, e.g. [3], [5]. In particular, we extend a previous work on the topic, namely [6]. In that paper, a data model accounting for the drift was presented and an efficient Alternating Least Squares (ALS) estimation method was described. However, as we show later, the model of [6] suffers from an accuracy loss which may bias the estimate. Therefore, we propose an improved model, where the accuracy loss is compensated by introducing an additional type of noise, termed pixel noise, which was not accounted for in [6]. Moreover, we extend the ALS algorithm to the new model. Finally, we consider the application to the data of the Photodetector Array Camera and Spectrometer (PACS) [5], which is an infrared photometer onboard the European Space Agency (ESA) Herschel space telescope. In this context, we discuss the ALS implementation and complexity and present an example of the results.

The paper is organized as follows. In section II we introduce some general image estimation concepts. In section III we summarize the data model of [6] and the ALS algorithm. In section IV we improve the data model by considering the pixel noise. In section V we extend the ALS algorithm. In section VI we describe the PACS data. In section VII we discuss the implementation of the ALS and its computational complexity. In section VIII we present an example of the results. In section IX we give the conclusions.

Notation. We use lowercase letters to denote vectors and uppercase letters to denote matrices. We use $\Re^N$ do denote the set of the column vectors of N elements and $\Re^{N\times M}$ to denote the set of the matrices with N rows and M columns. We use I to denote the identity matrix. We use a superscript T to denote matrix or vector transposition, e.g. A^T. We use $E\{.\}$ to denote expectation.

II. LS image estimation in the presence of noise

Consider the acquisition instrument sketched in figure 1. The image to be acquired is assumed to be band-limited[1] and is represented by N_m pixels. The image is scanned by a sensor, the output of which is sampled at frequency f_s. The dashed line represents the scan trajectory and the circles indicate the sampling points. The instrument's output is constituted by a sequence of N_d readouts, termed a time-series, and the corresponding pointing information, giving the sampling coordinates of each readout. By assuming that the instrument is linear, each readout is a linear combination of the pixels. Thus, by stacking the pixels into an *image* vector $m \in \Re^{N_m}$ and the readouts into a *data* vector $d \in \Re^{N_d}$, we can write

$$d = Pm + n = s + n \tag{1}$$

where $P \in \Re^{N_d\times N_m}$ is a full-rank matrix, termed the *pointing* matrix, which depends on the sensor and on the observation protocol and is better discussed in section IV, $n \in \Re^{N_d}$ is a random *electronic noise* vector, accounting for the noise injected by the electronics, and we introduced the *signal* vector $s = Pm$, representing the ideal, noise-less output. We assume that $N_d >> N_m$, i.e. that there is redundancy in the data,

[1]This is the image seen at the output of the instrument's analog front-end, after the convolution with a band-limited Point Spread Function (PSF).

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

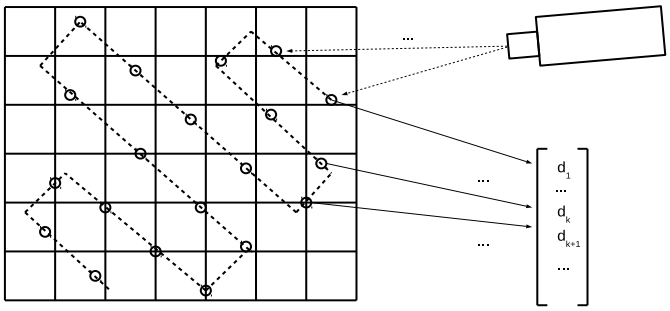

Fig. 1. Sketch of the acquisition system. The grid represents the pixels of the image to be acquired. The sensor is depicted as a stylised camera. The sensor pointing is moved along the scan trajectory, represented by the dashed line. At regular times, the sensor output is sampled: the first two samplings are indicated by the dotted arrows; the sampling points are represented by the circles. The readouts are orderly stored in the data vector, as indicated by the continuous arrows.

as needed to combat the noise. Moreover, we assume that the noise vector is obtained by sampling a zero mean noise process and denote the noise covariance matrix by $N = E\{nn^T\}$.

Given the data model of equation (1), we consider the problem of producing an estimate of m knowing d, P and the statistics of n. This is a classical problem having several established solutions. An important one is based on Generalised Least Squares (GLS) and is

$$\bar{m} = (P^T N^{-1} P)^{-1} P^T N^{-1} d. \tag{2}$$

The latter is the unbiased linear estimate with the minimum variance and, when the noise is Gaussian, it is the Maximum Likelihood (ML) estimate, e.g. [7].

When the electronic noise is stationary white noise, which we will assume from now on, the estimate is simpler. In this case, the covariance matrix is a scaled identity matrix, namely $N = \sigma^2 I$ where σ^2 is the noise variance, and the GLS estimate reduces to the Least Squares (LS) estimate, given by

$$\bar{m} = (P^T P)^{-1} P^T d, \tag{3}$$

where the matrix $(P^T P)^{-1} P^T$ is known as the pseudo-inverse of P.

Practical instruments are usually constituted by an array of sensors and not by a single one. Each sensor produces a time-series, with corresponding pointing matrix and noise vector. In this case, the model of equation (1) can still be used, together with the estimates of equation (2) and (3), assuming that d is obtained by stacking the time-series, n by stacking the noise vectors and P by stacking the pointing matrices.

III. LS IMAGE ESTIMATION IN THE PRESENCE OF DRIFT

An additional disturbance which may affect the time-series is a drift, causing the readouts to depart from the baseline level. The drift is slowly varying and can be modeled as a function depending on a few parameters. In order to include the drift, the data model of equation (1) can be extended as follows

$$d = Pm + Da + n \tag{4}$$

where $D \in \Re^{N_d \times N_a}$ is a full-rank matrix, termed the *drift* matrix, and $a \in \Re^{N_a}$ is a *coefficients* vector. As we see, in the latter model, the drift is represented as a vector Da which is a linear combination of the columns of the D matrix, with vector a giving the N_a coefficients. By properly selecting the matrix, several drift types can be modeled, for example polynomial [6] or low-pass [8]. In all cases of interest $N_m >> N_a$.

Given the data model of equation (4), we consider the problem of producing an estimate of m and a, knowing d, P, D and the statistics of n. This problem has been analysed in [6], where it has been shown that, in white and stationary noise, the LS estimates can be obtained efficiently, by using an algorithm based on Alternating Least Squares (ALS).

In order to derive the ALS algorithm note that, by assuming that the coefficients vector is known and by introducing $\delta = d - Da$, which is the data vector depurated from the drift, equation (4) yields $\delta = Pm + n$, which is a standard LS problem, with solution

$$\bar{m} = (P^T P)^{-1} P^T \delta.$$

Similarly, if the image vector is given, by introducing $\eta = d - Pm$, which is the data vector depurated from the signal, equation (4) yields $\eta = Da + n$, which is a standard LS problem with solution

$$\bar{a} = (D^T D)^{-1} D^T \eta.$$

Based on the last equations, the ALS algorithm iteratively solves the two problems, using the current estimates of the image and of the coefficients to form the vectors δ and η. Specifically, the ALS algorithm amounts at the following steps

1. Initialise $\delta = d$. Repeat steps 2-5 till convergence
2. Estimate image: $\bar{m} = (P^T P)^{-1} P^T \delta$.
3. Remove signal: $\eta = \delta - P\bar{m}$.
4. Estimate coefficients: $\alpha = (D^T D)^{-1} D^T \eta$.
5. Remove drift: $\delta = \delta - D\alpha$.

In [6], it has been shown that the algorithm converges and that the vector $\bar{m}$ tends to the LS image estimate. Moreover, the LS estimate for the coefficients can be obtained by accumulating the vector α across the iterations.

IV. HARD POINTING AND PIXEL NOISE

We now discuss two ways of selecting the pointing matrix, which we refer to as soft- and hard-pointing. To describe soft-pointing note that the vector m hosts the values of the image at the pixels' center. Then, since the image is band-limited and assuming that the pixel size is small enough, the Sampling Theorem says that the value of the image at any point can be expressed as a linear combination of the elements of vector m. Using this fact, the soft-pointing matrix P can be formed by placing on its k-th row the coefficients of the linear combination for the sampling point of the k-th readout.

In general, the soft-pointing matrix P is a full matrix, with no zero elements. Handling this matrix, e.g. computing its pseudo-inverse, is heavy as soon the data size is not trivial. To

circumvent this problem, we make the simplifying assumption that value of each readout is simply equal to the value of the center of the pixel where the readout falls. Using this approximation, a hard-pointing matrix P can be formed, such that the element on the k-th row and i-th column is one if the k-th readout was taken in the i-th pixel and is zero otherwise.

The hard-pointing matrix is a sparse and binary matrix, all zero except for N_d elements which are one. Clearly, it is much simpler to handle than the soft one. For example, it can be efficiently stored. Moreover, the LS estimate of equation (3) is simply the image obtained by averaging the values of the readuts falling into each pixel, as is easy to check. Furthermore, the multiplication of a vector with the matrix, e.g. Pm, can be computed with just N_d memory access operations. For these reasons, hard-pointing is invariably used in practical systems.

The drawback of hard-pointing is that the data models become approximate. In practice, the accuracy loss is tolerable, at least as long as the pixel size is smaller than the instrument Point Spread Function (PSF). However, as we see in section VIII, in some cases the approximation is not adequate and can seriously bias the estimates. Fortunately, we can modify the data models in order to retain the simplicity of hard-pointing and restore accuracy. Specifically, we may assume that each readout is affected by an additional error, due to the hard-pointing approximation, given by the difference of the image value at the sampling point and the image value at the pixel center. This error is conveniently modeled as an additional form of random noise, which we term *pixel noise*. Then, by representing the pixel noise as a vector $x \in \Re^{N_d}$, we simply rewrite the data model of equation (1) as

$$d = Pm + n + x, \tag{5}$$

and the data model of (4) as

$$d = Pm + Da + n + x. \tag{6}$$

In order to use the latter models we need to investigate the statistics of the pixel noise. As a general comment, note that the pixel noise depends on the image morphology: for example, where the image is flat or slowly varying, the pixel noise is absent or negligible; on the contrary, a strong pixel noise is expected where the image has gradients and fast variations. Moreover, it is reasonable to assume that the pixel noise affecting one redaout is statistically independent of the pixel noise affecting the other readouts. Furthermore, we can assume that the pixel noise is zero mean. The latter two facts imply that the covariance matrix of the pixel noise, denoted by $X = E\{xx^T\}$, is diagonal. An additional plausible assumption is that all the readouts falling into the same pixel experience a pixel noise with the same variance. Therefore, the diagonal entries of X corresponding to the same pixel are equal. Finally, we can assume that the pixel noise is statistically independent of the electronic noise. Therefore, the covariance matrix of the total noise, denoted by $C = E\{(n+x)(n+x)^T\}$, is simply the sum of the two covariances, i.e. $C = N + X$. Thus, when the electronic noise is white and stationary, we have $C = \sigma^2 I + X$, which is a diagonal matrix. Based on these facts, it is possible to estimate the pixel noise covariance matrix directly from the data vector, as better discussed in section VI.

V. Estimation in the presence of pixel noise

In this section, we discuss the GLS estimates and the ALS algorithm for the new models. Consider first the model of equation (5). The GLS estimate can be obtained by using (2) with N replaced by C, the total noise covariance matrix, and is

$$\bar{m} = (P^T C^{-1} P)^{-1} P^T C^{-1} d. \tag{7}$$

In general, the estimate is difficult to compute, as discussed in [9]. However, when the electronic noise is white and stationary, it is not difficult to show that the latter equation reduces to the standard LS estimate of equation (3). This is due to the fact that all the readouts falling in the same pixel experience the same pixel noise variance and are therefore given the same weight in the estimation.

Consider now the model of equation (6). The GLS estimate can be obtained by extending the ALS algorithm. In particular, we have to update the image and coefficients estimate of steps 2 and 4. However, assuming that the electronic noise is white, we have just seen that the image estimate coincides with the LS one. Therefore, step 2 does not need to be changed. Concerning the coefficients, step 4 has to be updated as follows

4a. Estimate coefficients: $\alpha = (D^T C^{-1} D)^{-1} D^T C^{-1} \eta$.

With the latter modification, we obtain a new algorithm, which will be denoted by ALS-X. The ALS-X converges and produces the GLS image estimate: the proof can be obtained by repeating the derivation of [6]. In the next sections, we apply the ALS-X to the data of the PACS instrument.

VI. PACS data

The Photodetector Array Camera and Spectrometer (PACS) [5] is an infrared instrument onboard the ESA's Herschel satellite. The PACS photometer consists of two arrays of bolometers. The first one operates in the 70 μm (blue) or in the 100 μm (green) band, the second one in the 160 μm (red) band. The field of view is approximately 45 squared arcsec, but a typical photometric observation covers a much larger sky area, up to some square degrees wide. To observe the area, the telescope is moved along a set of parallel scan lines, covering the area. During the scan, the bolometers are sampled at regular times, producing a sequence of readouts for each bolometer, i.e. a time-series, which is also called a *timeline* in the PACS jargon. Redundancy is guaranteed by scanning the area twice, along two orthogonal directions. The timelines, together with the corresponding pointing information, constitute the observation raw output and are collectively referred as the Time Ordered Data (TOD). A typical 70 μm TOD is constituted by $N_t \approx 4000$ timelines, with a total of $N_d \approx 10^9$ readouts; the image has $N_m \approx 10^7$ pixels.

The PACS timelines are affected by noise and drift[2]. The noise is injected by the readout electronics and is typically modeled as a zero-mean, stationary Gaussian process [5]. The noise covariance matrix is dictated by the noise spectrum, which is the sum of two terms, namely a white noise with flat spectrum $N_w(f) = N_0$ plus a correlated noise, also called the $1/f$ noise, with spectrum $N_c(f) = (f_0/f)^\epsilon N_0$ where f_0 is called the knee frequency and ϵ the frequency exponent. The drift is due to thermal variations on the focal plane and is well modeled as a polynomial curve of low order.

Concerning the data model, soft-pointing is not an option, due to the huge size of the data. Therefore we use hard-pointing and exploit the model of equation (6). The pointing matrix P is easy to derive, from the pointing information. For the drift, we assume a third order polynomial, depending on $N_p = 4$ coefficients. The corresponding drift matrix, D, is a block Vandermonde matrix [6] with $N_a = N_p N_t$ columns. For the electronic noise, we neglect the $1/f$ part and assume pure white noise: this is an approximation which makes the ALS-X algorithm directly applicable; the impact on the estimation quality is better discussed in the next section. Finally, the drift covariance matrix, X, can be estimated as follows: first, for all the pixels, we compute the variance of the corresponding readouts; next, if the k-th readout was taken in a pixel having variance σ_p^2, we set the k-th diagonal element of X to σ_p^2. This procedure is simple and normally produces satisfactory results. However, the estimate is biased, because the variance does not depend on the pixel noise only, but is affected by the electronic noise and the drift too. A better estimation procedure, capable of disentangling the electronic and pixel noise, is described in [9].

VII. Application of ALS-X to PACS data

We discuss the ALS-X implementation and complexity. Thanks to the hard-pointing matrix, steps 2 and 3 are simple to implement. In particular, as we mentioned, the image estimate of step 2 is just the average of the readouts of vector δ in each pixel and requires N_d additions and N_m divisions to be computed. Similarly, the computation of the η vector requires N_d memory access and subtractions.

Concerning the coefficients estimation of step 4a, we note that, since the drift matrix D is block Vandermonde and the noise covariance matrix C is diagonal, the matrix $(D^T C^{-1} D)^{-1}$ is block diagonal with N_t blocks of size $N_p \times N_p$ and can be directly computed and stored before the iterations begin. During the iterations, in step 4a, the data vector η is multiplied by C^{-1}, which is simple to do, since C is diagonal. The result is next multiplied by $(D^T C^{1-} D)^{-1} D^T$, which is simple to do, thanks to the structure of D: in practice this operation amounts at separately fitting each timeline to a polynomial and can be implemented with $O(N_p N_d)$ multiply and add operations. Finally, step 5 amounts at evaluating a polynomial for each timeline and can be realised with $O(N_p N_d)$ multiply and add operations. In summary, the complexity of one iteration is dominated by steps 4a and 5 and is $O(N_p N_d)$. In all our experiments, convergence was achieved in less than 10 iterations.

[2]The timelines are affected by other impairments too, like offsets, saturations, glitches and jumps. We assume that these impairments have been removed in a prior stage or are negligible.

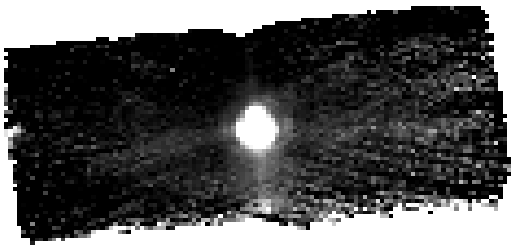

Fig. 2. The LS estimate for Ceres. Note the background noise. Note that the upper and lower edges are at different levels, because of the drift.

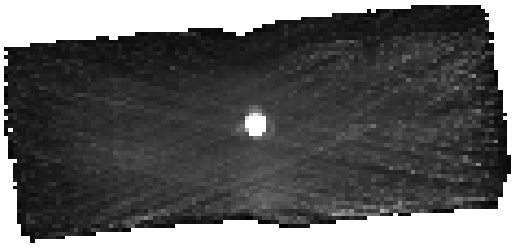

Fig. 3. The variance image, used to estimate the pixel noise covariance matrix. Note the strong pixel noise affecting the Ceres body.

As a final comment, recall that we have neglected the $1/f$ noise affecting the PACS timelines. As a result, the ALS-X estimates are sub-optimal and better estimates could be obtained. In particular, we should modify the ALS in order to account for the $1/f$ noise, which, in principle, is not difficult[3]. However, the computational complexity would become prohibitive. Luckily, in practice, the problem is serious for the image estimate only, while the drift estimate is usually accurate. In fact, the drift depends on a few parameters and the redundancy is normally so high that the estimate is accurate even when the noise is not white. Based on this observation, we can use the ALS-X to estimate the drift and subtract it from the data vector, to produce an updated data vector which is, ideally, free of drift. Next, in a separate step, we estimate the image from the updated data vector, properly accounting for the $1/f$ noise, using, for example, the GLS procedure described in [10]. In other words, for PACS data, the ALS-X is used as a pre-processing step, in order to estimate and remove the drift. It was verified that this approach yields nearly optimal results [6].

[3]Just use the GLS image estimate of equation (7) in step 2.

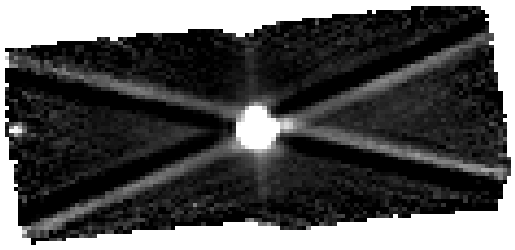

Fig. 4. The ALS estimate for Ceres. Note the strong distortion centered on the Ceres body, due to the pixel noise.

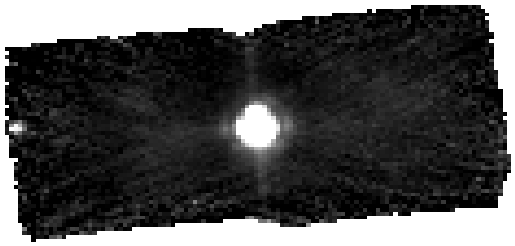

Fig. 5. The ALS-X estimate. Note the absence of distortion. With respect to the LS estimate, the background noise is reduced and the edges level is equalised.

VIII. Results

The ALS-X was integrated into the Unimap software[4] [10] and in this section we present the results obtained by processing two observations of the Ceres asteroid, in the red band. Initially, we neglect the drift and apply the model of equation (1), with hard-pointing and stationary white noise. The corresponding image estimate is the LS one, of equation (3), and is shown in figure 2. We see that there is noise in the background, which is due to the drift and to the electronic noise. We also see that the upper and lower edges of the image are at different levels, which is due to the drift.

In figure 3 we show the image of the pixels variance, which is used to estimate the pixel noise covariance matrix. The sharp peak in the center corresponds to the Ceres body and indicates a strong pixel noise. In the rest of the image, since the signal is negligible, the pixel noise is absent and the variance is due to the drift and the electronic noise.

We now apply model (4) with hard-pointing, thereby accounting for the drift but not for the pixel noise. The corresponding image estimate is obtained by exploiting the ALS algorithm and is shown in figure 4. Observing the figure we immediately see the detrimental effect of the pixel noise, which inflicts a strong distortion, taking the form of a cross centered on the Ceres body. In this case, the ALS severely distorts the image. However, it should be noted that we selected a limiting case, where the pixel noise is particularly strong. Usually the distortion is less pronounced.

Finally, we apply model (6) with hard-pointing, thereby accounting for both the drift and the pixel noise. The corresponding image estimate is obtained by exploiting the ALS-X algorithm and is shown in figure 5. Observing the figure we see that no distortion is introduced. Moreover, the background is less noisy and the level of the image edges is equalised, indicating that the drift has been successfully removed.

To conclude, we mention that we compared the ALS and ALS-X on a huge database, including both true and simulated data. The test campaign confirmed the facts just highlighted. Specifically, whenever the image has a significative signal component, the ALS is affected by a distortion which is absent in the ALS-X image.

IX. Conclusion

We discussed LS image estimation in the presence of electronic noise and drift. We introduced a data model accounting for the drift and discussed two ways of handling the pointing information, namely soft- and hard-pointing. Hard-pointing greatly simplifies the image estimation but is approximate and may distort the image. Therefore, we extended the model by considering and additional form of noise, the pixel noise, which arise when hard-pointing is used. We discussed the LS estimation problem for the new model and presented an efficient way to solve it, namely the ALS-X algorithm. Finally, we applied the ALS-X to the data of the PACS instrument, discussing its implementation and complexity and presenting an example of the results.

[4]Since release 6.2. Previous releases used the ALS.

References

[1] C. Matteau-Pelletier, M. Dehaes, F. Lesage, and J. Lina "1/f Noise in Diffuse Optical Imaging and Wavelet-Based Response Estimation", IEEE Trans. on Medical Imaging, Vol. 28, No. 3, pp. 415-422, March 2009.

[2] F. G. Meyer, "Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series", IEEE Transactions on Medical Imaging, Vol. 22, pp 315-322, March 2003.

[3] H. Luo and S. Puthusserypady, "Analysis of fMRI Data With Drift: Modified General Linear Model and Bayesian Estimator", IEEE Trans. on Biom. Eng., vol. 55, no. 5, pp. 1504-1511, May 2008.

[4] D. Lemke et al., "ISOPHOT - capabilities and performance", Astronomy and Astrophysics, vol.315, pp. L64-L70, Nov. 1996.

[5] A. Poglitsch et al., "The photodetector array camera and Spectrometer (PACS) on the herschel space observatory", Astronomy and Astrophysics, Volume 518, Issue 4, 2010.

[6] L. Piazzo. P. Panuzzo, M. Pestalozzi, "Drift removal by means of Alternating Least Squares with application to Herschel data", Signal Processing, vol. 108, pp. 430-439, 2015

[7] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice-Hall, New Jersey, 1993.

[8] L. Piazzo, "An Alternating Least Squares Algorithm with Application to Image Processing", Int.Conf on Applied Mathematics, Modelling and Simulation, ASM, pp. 379-384, Florence, Italy, 2014.

[9] L. Piazzo 'Generalised Least Squares Image Estimation in the Presence of 1/f and Pixel Noise', submitted to the IEEE Trans. on Image Processing, 2016.

[10] L. Piazzo, L. Calzoletti, F. Faustini, M. Pestalozzi, S. Pezzuto, D. Elia, A. di Giorgio and S. Molinari: 'Unimap: a Generalised Least Squares Map Maker for Herschel Data', Monthly Notices of the Royal Society, vol. 447, pp. 1471-1483, 2015.