An investigation into the Multiple Optimised Parameter Estimation and Data compression algorithm

Philip Graff,1⋆ Michael P. Hobson1 and Anthony Lasenby1,2

1Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE
2Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA

Accepted 2011 February 17. Received 2011 February 16; in original form 2010 October 28

ABSTRACT

We investigate the use of the Multiple Optimised Parameter Estimation and Data compression (MOPED) algorithm for data compression and faster evaluation of likelihood functions. Since MOPED only guarantees maintaining the Fisher matrix of the likelihood at a chosen point, multimodal and some degenerate distributions will present a problem. We present not only examples of scenarios in which MOPED does faithfully represent the true likelihood, but also cases in which it does not. Through these examples, we aim to define a set of criteria for which MOPED will accurately represent the likelihood and hence may be used to obtain a significant reduction in the time needed to calculate it. These criteria may involve the evaluation of the full likelihood function for comparison.

Key words: methods: data analysis – methods: statistical.

1 INTRODUCTION

Multiple Optimised Parameter Estimation and Data compression (MOPED; Heavens et al. 2000) is a patented algorithm for the compression of data and the speeding up of the evaluation of likelihood functions in astronomical data analysis and beyond. It becomes particularly useful when the noise covariance matrix is dependent upon the parameters of the model and so must be calculated and inverted at each likelihood evaluation. However, such benefits come with limitations. Since MOPED only guarantees maintaining the Fisher matrix of the likelihood at a chosen point, multimodal and some degenerate distributions will present a problem. In this Letter, we report on some of the limitations of the application of the MOPED algorithm. In the cases where MOPED does accurately represent the likelihood function, however, its compression of the data and consequent much faster likelihood evaluation does provide orders of magnitude improvement in runtime. In Heavens et al. (2000), the authors demonstrate the method by analysing the spectra of galaxies, and in Gupta & Heavens (2002) they illustrate the benefits of MOPED for estimation of the cosmic microwave background (CMB) power spectrum. The problem of ‘badly’ behaved likelihoods was found by Protopapas, Jimenez & Alcock (2005) for the problem of light transit analysis; none the less, the authors present a solution that still allows MOPED to provide a large speed increase.

We begin by introducing MOPED in Section 2 and define the original and MOPED likelihood functions, along with comments on the potential speed benefits of MOPED. In Section 3, we introduce an astrophysical scenario where we found that MOPED did not accurately portray the true likelihood function. In Section 4, we expand upon this scenario to another where MOPED is found to work and to two other scenarios where it does not. We present a discussion of the criteria under which we believe MOPED will accurately represent the likelihood in Section 5, as well as a discussion of an implementation of the solution provided by Protopapas et al. (2005).

2 DATA COMPRESSION WITH MOPED

Full details of the MOPED method are given in Heavens et al. (2000); here we will only present a limited introduction.

We begin by defining our data as a vector, x. Our model describes x by a signal plus random noise,

\[ x = u(\theta_T) + n(\theta_T). \]  

(1)

where the signal is given by a vector u(\theta) that is a function of the set of parameters \( \theta = \{ \theta_i \} \) defining our model, and the true parameters are given by \( \theta_T \). The noise is assumed to be Gaussian with zero mean and noise covariance matrix \( N_{jk} = \langle n_j n_k \rangle \), where the angle brackets indicate an ensemble average over noise realizations (in general, this matrix may also be a function of the parameters \( \theta \)).

The full likelihood for \( N \) data points in \( x \) is given by

\[
L^{\text{Original}}(\theta) = \frac{1}{(2\pi)^{N/2} \sqrt{|N(\theta)|}} \exp \left\{ -\frac{1}{2} [x - u(\theta)]^T N(\theta)^{-1} [x - u(\theta)] \right\}. 
\]

(2)

At each point, then, this requires the calculation of the determinant and inverse of an \( N \times N \) matrix. Both scale as \( N^3 \), so even for smaller data sets this can become cumbersome.
MOPED allows one to eliminate the need for this matrix inversion by compressing the $N$ data points in $x$ into $M$ data values, one for each parameter of the model. Additionally, MOPED creates the compressed data values such that they are independent and have unit variance, further simplifying the likelihood calculation on them to an $O(M)$ operation. Typically, $M \ll N$, so this gives us a significant increase in speed. A single compression is done on the data, $x$, and then again for each point in parameter space where we wish to compute the likelihood. The compression is done by generating a set of weighting vectors, $b_i(\theta_f)$ ($i = 1, \ldots, M$), from which we can generate a set of MOPED components from the theoretical model and data,

$$y_i(\theta_f) = b_i(\theta_f) \cdot x = b_i(\theta_f)^T x.$$  \hspace{1cm} (3)

Note that the weighting vectors must be computed at some assumed fiducial set of parameter values, $\theta_f$. The only choice that will truly maintain the likelihood peak is when the fiducial parameters are the true parameters, but obviously we will not know these in advance for real analysis situations. Thus, we can choose our fiducial model to be anywhere and iterate the procedure, taking our likelihood peak in one iteration as the fiducial model for the next iteration. This process will converge very quickly and may not even be necessary in some instances. For our later examples, since we do know the true parameters we will use these as the fiducial ($\theta_f = \theta_t$) in order to remove this as a source of confusion (all equations, however, are written for the more general case). Note that the true parameters, $\theta_t$, will not necessarily coincide with the peak $\hat{\theta}_0$ of the original likelihood or the peak $\hat{\theta}_M$ of the MOPED likelihood (see below).

The weighting vectors must be generated in some order so that each subsequent vector (after the first) can be made orthogonal to all previous ones. We begin by writing the model equation with respect to the $i$th parameter as $\frac{\partial \theta_i}{\partial \theta_f} u_i(\theta_f) = b_i(\theta_f)$. This gives us a solution for the first weighting vector, properly normalized, of

$$b_1(\theta_f) = \frac{\mathcal{N}(\theta_f)^{-1} u_1(\theta_f)}{\sqrt{u_1(\theta_f)^T \mathcal{N}(\theta_f)^{-1} u_1(\theta_f)}},$$ \hspace{1cm} (4)

The first compressed value is $y_1(\theta_f) = b_1(\theta_f)^T x$ and will weight up the data combination most sensitive to the first parameter. The subsequent weighting vectors are made orthogonal by subtracting out parts that are parallel to previous vectors and are normalized. The resulting formula for the remaining weighting vectors is

$$b_m(\theta_f) = \frac{\mathcal{N}(\theta_f)^{-1} u_m(\theta_f) - \sum_{q=1}^{m-1} u_m(\theta_f)^T b_q(\theta_f) b_q(\theta_f)}{\sqrt{u_m(\theta_f)^T \mathcal{N}(\theta_f)^{-1} u_m(\theta_f) - \sum_{q=1}^{m-1} (u_m(\theta_f)^T b_q(\theta_f))^2}},$$ \hspace{1cm} (5)

where $m = 2, \ldots, M$. Weighting vectors generated with equations (4) and (5) form an orthonormal set with respect to the noise covariance matrix so that

$$b_i(\theta_f)^T \mathcal{N}(\theta_f) b_j(\theta_f) = \delta_{ij}.$$ \hspace{1cm} (6)

This means that the noise covariance matrix of the compressed values $y_i$ is the identity, which significantly simplifies the likelihood calculation. The new likelihood function is given by

$$\mathcal{L}_{\text{MOPED}}(\theta) = \frac{1}{(2\pi)^{M/2}} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^{M} (y_i(\theta_f) - \langle y_i(\theta_f) \rangle)^2 \right\},$$ \hspace{1cm} (7)

where $y_i(\theta_f) = b_i(\theta_f)^T x$ represents the compressed data and $\langle y_i(\theta_f) \rangle = b_i(\theta_f)^T u(\theta)$ represents the compressed signal. This is a much easier likelihood to calculate and is time limited by the generation of a new signal template instead of the inversion of the noise covariance matrix. The peak value of the MOPED likelihood function is not guaranteed to be unique as there may be other points in the original parameter space that map to the same point in the compressed parameter space; this is a characteristic that we will investigate.

MOPED implicitly assumes that the covariance matrix, $\mathcal{N}$, is independent of the parameters. With this assumption, a full likelihood calculation with $N$ data points would require only an $O(N^3)$ operation at each point in parameter space (or $O(N)$ if $\mathcal{N}$ is diagonal). In MOPED, however, the compression of the theoretical data is an $O(MN)$ linear operation followed by an $O(M)$ likelihood calculation. Thus, one loses on speed if $\mathcal{N}$ is diagonal but gains by a factor of $N/M$ otherwise. For the data sets we will analyse, $N/M > 100$. We begin, though, by assuming a diagonal $\mathcal{N}$ for simplicity, recognizing that this will cause a speed reduction but that it is a necessary step before addressing a more complex noise model. One can iterate the parameter estimation procedure if necessary, taking the maximum likelihood point found as the new fiducial and re-analysing (perhaps with tighter prior constraints); this procedure is recommended for MOPED in Heavens et al. (2000) but is not always found to be necessary. MOPED has the additional benefit that the weighting vectors, $b_i$, need only to be computed once provided the fiducial model parameters are kept constant over the analysis of different data sets. Computed compressed parameters, $(y_i)$, can also be stored for reuse and require less memory than storing the entire theoretical data set.

3 SIMPLE EXAMPLE WITH ONE PARAMETER

In order to demonstrate some of the limitations of the applicability of the MOPED algorithm, we will consider a few test cases. These originate in the context of gravitational wave data analysis for the Laser Interferometer Space Antenna (LISA) since it is in this scenario that we first discovered such cases of failure. The full problem is seven-dimensional parameter estimation, but we have fixed most of these variables to their known true values in the simulated data set in order to create a lower dimensional problem that is simpler to analyse.

We consider the case of a sine-Gaussian burst signal present in the LISA detector. The short duration of the burst with respect to the motion of LISA allows us to use the static approximation to the response. In frequency space, the waveform is described by (Feroz et al. 2010)

$$h(f) = A \tilde{Q} \exp \left\{ -\frac{1}{2} Q^2 \left( f - f_c \right)^2 \right\} \exp(2\pi i \tilde{t}_0 f).$$ \hspace{1cm} (8)

Here, $A$ is the dimensionless amplitude factor, $Q$ is the width of the Gaussian envelope of the burst measured in cycles, $f_c$ is the central frequency of the oscillation being modulated by the Gaussian envelope and $\tilde{t}_0$ is the central time of arrival of the burst. This waveform is further modulated by the sky position of the burst source, $\theta$ and $\phi$, and the burst polarization, $\psi$, as they project on to the detector. The one-sided noise power spectral density of the LISA detector is given by (Feroz et al. 2010)

$$S_n(f) = 16 \pi^2 (2\pi f_{\text{th}})$$ \hspace{1cm} (9)

$$\times \left[ 2 \left( 1 + \cos(2\pi f_{\text{th}}) + \cos^2(2\pi f_{\text{th}}) \right) S_{\text{on}}(f) + (1 + \cos(2\pi f_{\text{th}})/2) S_{\text{off}} f^2 \right].$$
$S_{\text{med}}(f) = \left( 1 + \left( \frac{10^{-4} \text{Hz}}{f} \right)^2 \right) S_{\text{acc}} / f^2,$

where $t_c = 16.678 \text{s}$ is the light traveltime along one arm of the LISA constellation, $S_{\text{acc}} = 2.5 \times 10^{-8} \text{Hz}^{-1}$ is the proof mass acceleration noise and $S_{\text{sn}} = 1.8 \times 10^{-48} \text{Hz}^{-1}$ is the shot noise. This is independent of the signal parameters and all frequencies are independent of each other, so the noise covariance matrix is constant and diagonal. This less computationally expensive example allows us to show some interesting examples.

We begin by taking the one-dimensional case where the only unknown parameter of the model is the central frequency of the oscillation, $f_c$. We set $Q = 5$ and $t_0 = 10^4 \text{s}$; we then analyse a $2048 \text{s}$ segment of the LISA data, beginning at $t = 9.9 \times 10^4 \text{s}$, sampled at a $1 \text{s}$ cadence. For this example, the data was generated with random noise (following the LISA noise power spectrum) at an signal-to-noise ratio (S/N) of $\sim 34$ with $f_{c,T} = 0.1 \text{Hz}$ (thus $f_{c_B} = 0.1 \text{Hz}$ for MOPED). The prior range on the central frequency goes from $10^{-3} \text{Hz}$ to $0.5 \text{Hz}$. 10000 samples uniformly spaced in $f_c$ were taken and their likelihoods calculated using both the original and MOPED likelihood functions. The log-likelihoods are shown in Fig. 1. Note that the absolute magnitudes are not important, but the relative values within each plot are meaningful. Both the original and MOPED likelihoods have a peak close to the input value $f_{c,T}$.

We see, however, that in going from the original to MOPED log-likelihoods, the latter also has a second peak of equal height at an incorrect $f_c$. To see where this peak comes from, we look at the values of the compressed parameter $(y_1(f_c; f_{c, B}))$ as it varies with respect to $f_c$ as shown in Fig. 2. The true compressed value peak occurs at $f_c \simeq 0.1 \text{Hz}$, where $y_1(f_{c, B}) = (y_1(f_c; f_{c, B})$. However, we see that there is another frequency that yields this exact same value of $(y_1(f_c; f_{c, B}))$; it is at this frequency that the second, incorrect peak occurs. By creating a mapping from $f_c$ to $(y_1(f_c; f_{c, B})$ that is not one to one, MOPED has created the possibility for a second solution that is indistinguishable in likelihood from the correct one. This is a very serious problem for parameter estimation.

4 RECOVERY IN A TWO-PARAMETER CASE

Interestingly, we find that even when MOPED fails in a one-parameter case, adding a second parameter may actually rectify the problem, although not necessarily. If we now allow the width of the burst, $Q$, to be a variable parameter, there are now two orthogonal MOPED weighting vectors that need to be calculated. This gives us two compressed parameters for each pair of $f_c$ and $Q$. Each of these may have its own unphysical degeneracies, but in order to give an unphysical mode of equal likelihood to the true peak, these degeneracies will need to coincide. In Fig. 3, we plot the contours in $(f_c, Q)$ space, where $(y_1)_{\theta, \theta_B} = (y_1)_{\theta_B}$ as $\theta$ ranges over $f_c$ and $Q$ values. We can clearly see the degeneracies present in either variable, but since these only overlap at the one location, near to where the true peak is, there is no unphysical second mode in the

![Figure 1](https://academic.oup.com/mnrasl/article-abstract/413/1/L66/1747595/1747595)

**Figure 1.** The original and MOPED log-likelihoods as a function of $f_c$ for the chosen template.

![Figure 2](https://academic.oup.com/mnrasl/article-abstract/413/1/L66/1747595/1747595)

**Figure 2.** The value of the MOPED compressed parameter as a function of the original frequency parameter.

![Figure 3](https://academic.oup.com/mnrasl/article-abstract/413/1/L66/1747595/1747595)

**Figure 3.** Contours of $(y_1)_{\theta, \theta_B} = (y_1)_{\theta_B}$ and $(y_2)_{\theta, \theta_B} = (y_2)_{\theta_B}$ as they vary over $f_c$ and $Q$. The one intersection is the true maximum likelihood peak.
Investigation into MOPED

Figure 4. Contours of the original and MOPED log-likelihoods (left-hand and right-hand panels, respectively). The MOPED likelihood has been multiplied by a constant factor so that its peak value is equal to the peak of the original likelihood. Contours are at 1, 2, 5, 10, 20, 30, 40, 50, 75 and 100 log-units below the peak going from the inside to the outside.

MOPED likelihood. Hence, when we plot the original and MOPED log-likelihoods in Fig. 4, although the behaviour away from the peak has changed, the peak itself remains in the same location and there is no second mode.

Adding more parameters, however, does not always improve the situation. We now consider the case where $Q$ is once again fixed to its true value and we instead make the polarization of the burst, $\psi$, a variable parameter. There are degeneracies in both of these parameters, and in Fig. 5 we plot the contours in $(f_c, \psi)$ space where the compressed values are each equal to the value at the maximum MOPED likelihood point. These two will necessarily intersect at the maximum likelihood solution, near the true value ($f_c = 0.1$ Hz and $\psi = 1.3$ rad), but a second intersection is also apparent. This second intersection will have the same likelihood as the maximum and will be another mode of the solution. However, as we can see in Fig. 6 in the left-hand panel, this is not a mode of the original likelihood function. MOPED has, in this case, created a second mode of equal likelihood to the true peak.

For an even more extreme scenario, we now fix to the true $\psi$ and allow the time of arrival of the burst $t_0$ to vary (we also define $\Delta t_0 = t_0 - t_{0,T}$). In this scenario, the contours in $(f_c, \Delta t_0)$ space where $\langle y_i \rangle(\theta; \theta_P) = \langle y_i \rangle(\hat{\theta}_M; \theta_F)$ are much more complicated. Thus, we have many more intersections of the two contours than just at the likelihood peak near the true values, and MOPED creates many alternative modes of likelihood equal to that of the original peak. This is very problematic for parameter estimation. In Fig. 7 we plot these contours so the multiple intersections are apparent. Fig. 8 shows the original and MOPED log-likelihoods, where we can see the single peak becoming a set of peaks.

5 DISCUSSION AND CONCLUSIONS

What we can determine from the previous two sections is a general rule for when MOPED will generate additional peaks in the likelihood function equal in magnitude to the true one. For an $M$-dimensional model, if we consider the $(M-1)$-dimensional hypersurfaces where $\langle y_i \rangle(\theta; \theta_P) = \langle y_i \rangle(\hat{\theta}_M; \theta_F)$, then any point where these $M$ hypersurfaces intersect will yield a set of $\theta$-parameter values with likelihood equal to that at the peak near the true values. We expect that there will be at least one intersection at the likelihood peak corresponding to approximately the true solution. However, we have shown that other peaks can exist as well. The set of intersections of contour surfaces will determine where these additional peaks are located. This degeneracy will interact with the model’s intrinsic degeneracy, as any degenerate parameters will yield the same compressed values for different original parameter values.

Unfortunately, there is no simple way to find these contours other than by mapping out the $\langle y_i \rangle(\theta; \theta_P)$ values, which is equivalent in procedure to mapping the MOPED likelihood surface. The benefit
Figure 7. Contours of \( \langle y_1 \rangle (\theta; \theta_F) = \langle y_1 \rangle (\hat{\theta}; \theta_F) \) and \( \langle y_2 \rangle (\theta; \theta_F) = \langle y_2 \rangle (\hat{\theta}; \theta_F) \) values as they vary as functions of \( f_c \) and \( t_0 \). We can see many intersections here other than the true peak.

Figure 8. Contours of the original and MOPED log-likelihoods (left-hand and right-hand panels, respectively). The MOPED likelihood has been multiplied by a constant factor so that its peak value is equal to the peak of the original likelihood. Contours are at 1, 2, 5, 10, 20, 30, 40, 50, 75 and 100 log-units below the peak going from the inside to the outside.

comes when this procedure is significantly faster than mapping the original likelihood surface. The mapping of \( \langle y_1 \rangle (\theta; \theta_F) \) can even be performed before data are obtained or used, if the fiducial model is chosen in advance; this allows us to analyse properties of the MOPED compression before applying it to data analysis. If the MOPED likelihood is mapped and there is only one contour intersection, then we can rely on the MOPED algorithm and will have saved a considerable amount of time, since MOPED has been demonstrated to provide speed ups of a factor of up to \( 10^7 \) in Gupta & Heavens (2002). However, if there are multiple intersections, then it is necessary to map the original likelihood to know if they are due to degeneracy in the model or were created erroneously by MOPED. In this latter case, the time spent finding the MOPED likelihood surface can be much less than that which will be needed to map the original likelihood, so relatively little time will have been wasted. If any model degeneracies are known in advance, then we can expect to see them in the MOPED likelihood and will not need to find the original likelihood on their account.

One possible way of determining the validity of degenerate peaks in the MOPED likelihood function is to compare the original likelihoods of the peak parameter values with each other. By using the maximum MOPED likelihood point found in each mode and evaluating the original likelihood at this point, we can determine which one is correct. The correct peak and any degeneracy in the original likelihood function will yield similar values to one another, but a false peak in the MOPED likelihood will have a much lower value in the original likelihood and can be ruled out. This means that a Bayesian evidence calculation cannot be obtained from using the MOPED likelihood; however, the algorithm was not designed to be able to provide this.

The solution for this problem presented in Protopapas et al. (2005) is to use multiple fiducial models to create multiple sets of weighting vectors. The log-likelihood is then averaged across these choices. Each different fiducial will create a set of likelihood peaks that include the true peaks and any extraneous ones. However, the only peaks that will be consistent between fiducials are the correct ones. Therefore, the averaging maintains the true peaks but decreases the likelihood at incorrect values. This was tested with 20 random fiducials for the two-parameter models presented and was found to leave only the true peak at the maximum likelihood value. Other, incorrect, peaks are still present, but at log-likelihood values five or more units below the true peak. When applied to the full seven-parameter model, however, the S/N threshold for signal recovery is increased significantly, from \( \gtrsim 10 \) to \( \gtrsim 30 \).

The MOPED algorithm for reducing the computational expense of likelihood functions can, in some examples, be extremely useful and provide orders of magnitude of improvement. However, as we have shown, this is not always the case and MOPED can produce erroneous peaks in the likelihood that impede parameter estimation. It is important to identify whether or not MOPED has accurately portrayed the likelihood function before using the results it provides. Some solutions to this problem have been presented and tested.

ACKNOWLEDGMENTS

PG’s PhD is funded by the Gates Cambridge Trust.

REFERENCES

Feroz F., Gair J., Graff P., Hobson M. P., Lasenby A. N., 2010, Classical Quantum Gravity, 27, 075010
Gupta S., Heavens A. F., 2002, MNRAS, 334, 167
Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965
Protopapas P., Jimenez R., Alcock A., 2005, MNRAS, 362, 460

This paper has been typeset from a \TeX\!/\LaTeX\! file prepared by the author.