HOW TO SEPARATE THE SIGNAL FROM THE BACKGROUND

W. VON DER LINDEN, V. DOSE AND R. FISCHER
Max-Planck-Institut für Plasmaphysik, EURATOM Association
D-85740 Garching b. München, Germany ¶

Abstract. In many applications like radio astronomy and electron spectroscopy three sources contribute to the measured signal: the desired image, the noise, and an unknown slowly varying background. The inferential problem is to separate the signal from the noise and the background. We accomplish this task employing Bayesian probability theory.

Key words: Maxent,Background

1. Introduction

Bayesian probability theory along with the entropic prior (quantified maxent) has been applied successfully in deconvolution, reconstruction, and inversion problems to separate the signal from the noise and to regularize incomplete data-information[1–4]. This is, however, very often not the full problem. Usually one has to get rid first of all of an unknown background, which can be a nontrivial task, e.g. in radio astronomy. In the commonly used quick and dirty approach the background is subtracted by hand. Our experience is that inappropriate subtraction can lead to artificial structures in the desired image. If too much is subtracted the data constraints cannot be fulfilled by a positive image and in the opposite case, standard maxent tends to introduce ringing and over-fitting [5]. This phenomenon is particularly present in the example depicted in fig.1. It is therefore important to subtract the background as accurately as possible. It is the purpose of this paper to illustrate how this can be achieved following the rules of probability theory.

2. Selfconsistent Background Analysis

We are interested in an image $\mathbf{f}$. For the present derivation it is not necessary to be more specific about the dimension of the image space or the detailed specification of $\mathbf{f}$. Three contributions add to the measured signal in channel l

$$g_l = g_l(\mathbf{f}) + b_l + \eta_l \qquad . \tag{1}$$

¶e-mail: wvl@ibmop5.ipp-garching.mpg.de

The sought-for image contributes to the signal via the transformation $g_l(\mathbf{f})$ which depends upon the application under consideration. In the following analysis the back ground b_l is eliminated from the measured data g_l. This leads to a modified likelihood term which can be used in a standard maxent analysis to separate the signal from the noise η_l and to invert the transformation $g_l(\mathbf{f})$ to extract the image $\mathbf{f}$. The following formalism is closely related to the analysis of Bretthorst used for spectrum analysis and parameter estimation [6]. The background is expanded in terms of suitable basis functions $\phi_\nu(x)$.

$$b_l = \sum_{\nu=1}^{E} \phi_\nu(x_l)\; c_\nu \qquad . \tag{2}$$

E stands for the expansion order which we assume to be known for the time being. We will call those basis functions which are included in Eq.2 "occupied". To separate the background from the signal it is necessary that the expansion of the background in this basis converges more rapidly than that of the signal. Anything else would fall into the realm of magic and not of probability theory. Typically we will use polynomials for the basis functions. The joint posterior for the image $\mathbf{f}$ and background coefficients $\mathbf{c}$ is

$$\begin{aligned} p(\mathbf{f},\mathbf{c}|\mathbf{g},E,\mathcal{I}) &= p(\mathbf{g}|\mathbf{f},\mathbf{c},E,\mathcal{I})\; p(\mathbf{c}|\mathbf{f},E,\mathcal{I})\; p(\mathbf{f}|E,\mathcal{I})\;/p(\mathbf{g}|E,\mathcal{I}) \\ &= p(\mathbf{g}|\mathbf{f},\mathbf{c},E,\mathcal{I})\; p(\mathbf{c}|E,\mathcal{I})\; p(\mathbf{f}|\mathcal{I})\;/p(\mathbf{g}|E,\mathcal{I}) \qquad . \end{aligned} \tag{3}$$

In the last line we discarded superfluous conditions. All implicit dependencies upon hyper-parameters, like noise-level and regularization parameter, are summaries in $\mathcal{I}$. They will only be mentioned explicitly when they are needed. As usual $p(\mathbf{g}|\mathbf{f},\mathbf{c},E,\mathcal{I})$ is the likelihood describing the error statistics of the experiment. Here we concentrate on the most important case of Gaussian errors

$$\begin{aligned} p(\mathbf{g}|\mathbf{f},\mathbf{c},E,\mathcal{I}) &= \frac{1}{Z_L} e^{-\frac{1}{2}\chi^2} \\ \chi^2 &= \sum_l (g_g - g_l(\mathbf{f}) - b_l)^2/\sigma^2 \qquad . \end{aligned} \tag{4}$$

For simplicity we chose a constant absolute error σ. The background coefficients $\mathbf{c}$ are nuisance parameters which ought to be marginalized, since we are interested in posterior expectations $E(\mathcal{F}(\mathbf{f})|\mathbf{g},\mathcal{I})$ of functionals $\mathcal{F}(\mathbf{f})$ of the image. This brings us to the marginal posterior

$$\begin{aligned} p(\mathbf{f}|\mathbf{g},\{\mathbf{c}\},E,\mathcal{I}) &= \int d^E c\; p(\mathbf{g}|\mathbf{f},\mathbf{c},E,\mathcal{I})\; p(\mathbf{c}|E,\mathcal{I})\; p(\mathbf{f}|\mathcal{I})\;/p(\mathbf{g}|E,\mathcal{I}) \\ &= P(\mathbf{g}|\mathbf{f},\{\mathbf{c}\},E\mathcal{I})\; p(\mathbf{f}|\mathcal{I})\;/p(\mathbf{g}|E,\mathcal{I}) \qquad . \end{aligned} \tag{5}$$

We proceed in three steps. For given expansion order E we will first derive the marginal likelihood, then the marginal posterior for the image, and finally the expansion order will be determined according to the probability for E given the data.

2.1. MARGINAL LIKELIHOOD

The marginal likelihood $p(\mathbf{g}|\mathbf{f},\{\mathbf{c}\},E,\mathcal{I})$ in which the background has been integrated out, replaces the original likelihood term in the following maxent analysis. First we determine the prior for the background coefficients $p(\mathbf{c}|E,\mathcal{I})$ which should have the following properties:

1. the marginal likelihood should be invariant under linear transformations of the basis-set, since the latter can always be absorbed into the expansion coefficients $\mathbf{c}$ and should therefore vanish after marginalization.
2. following Bretthorst[6] we consider the background power as testable information: $\sum b_l^2 = \mu^2$. The details of the prior ought to be unimportant as they are overruled by the data constraints if the experimental data make any sense at all.
3. otherwise as ignorant (uninformative) as possible.

The constraint (2) can be cast into a more suitable form

$$\mu^2 \quad = \quad \sum_l b_l^2 = \sum_{\nu\nu'} c_\nu \left(\sum_l \phi_\nu(x_l)\phi_{\nu'}(x_l) \right) c_{\nu'} \; = c^T \; S \; c \qquad . \tag{6}$$

We can always assume that the basis functions are not linearly dependent and the modified overlap matrix S is therefore positive definite. We introduce the following abbreviations $\Delta\mathbf{g}_l = g_l - g_l(f)$ and $D_{l\nu} = \phi_\nu(x_l)$. In this notation the vector of the background signals reads $\mathbf{b} = D\mathbf{c}$. To fulfill the constraint (2) along with property (3) we employ Jaynes' MaxEnt leading to[7]

$$p(\mathbf{c}|E,\mu,\mathcal{I}) = \frac{\sqrt{\det(S)}}{(2\pi\mu^2)^{\frac{E}{2}}} e^{-\frac{1}{2\mu^2} c^T \; S \; c} \qquad . \tag{7}$$

Obviously Eq.7 satisfies all properties listed above: it has the desired mean background-power, the degree of ignorance is governed by μ^2, and, as we will see below, the marginal likelihood is invariant under linear transformations. The background-power is actually a nuisance parameter which we marginalize right from the start. The prior for the scale parameter μ is according to Jeffreys

$$p(\mu|\mathcal{I}) = \frac{z}{\mu} \qquad . \tag{8}$$

We do not have to worry about the improperness of Jeffreys' prior since the posterior probability will be proper due to the data-constraints. For a thorough discussion of improper priors see [8]. The principally unknown normalization constant z will drop out if we compare probabilities. The only limitation is that we cannot compare the hypotheses "*there is no background*" and "there is a background of expansion order E". But we can compare results to different expansion orders. The prior for $\mathbf{c}$ becomes

$$\begin{aligned} p(\mathbf{c}|E,\mathcal{I}) \quad &= \quad \int_0^\infty p(\mathbf{c}|E,\mu,\mathcal{I}) \; p(\mu|\mathcal{I}) d\mu \\ &= \quad \frac{z}{2}\sqrt{\det(S)}(2\pi)^{-\frac{E}{2}} (\frac{\mathbf{c}^T S \mathbf{c}}{2})^{-\frac{E}{2}} \Gamma(\frac{E}{2}) \qquad . \end{aligned} \tag{9}$$

We are now in the position to determine the marginal likelihood Eq.5

$$
\begin{aligned}
p(\mathbf{g}|\mathbf{f},\{\mathbf{c}\},\mathcal{I}) &= \frac{z\Gamma(\frac{E}{2})\sqrt{\det(S)}}{2Z_L(2\pi)^{\frac{E}{2}}}\int \left(\frac{\mathbf{c}^T S\mathbf{c}}{2}\right)^{-\frac{E}{2}} e^{-\frac{1}{2\sigma^2}(\Delta\mathbf{g}-D\mathbf{c})^2}\, d^E c \\
&= \frac{z\Gamma(\frac{E}{2})\sqrt{\det(S)}}{2Z_L(2\pi)^{\frac{E}{2}}}\int \left(\frac{\mathbf{c}^T S\mathbf{c}}{2}\right)^{-\frac{E}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{c}^T S\mathbf{c}-2\Delta\mathbf{g}^T D\mathbf{c}\,+\,\Delta\mathbf{g}^2)} d^E c \quad (10)
\end{aligned}
$$

We used $(D^T D)_{\nu\nu'} = \sum_l \phi_\nu(x_l)\phi_{\nu'}(x_l) = S_{\nu\nu'}$. The prior factor $(\mathbf{c}^T S\mathbf{c})^{-\frac{E}{2}}$ is structure-less and decreases monotonically with $\mathbf{c}$, whereas the data-factor is sharply peaked if the data are decent. We therefore replace the prior-factor under the integral by its value at the maximum $\mathbf{c} = \mathbf{c}^* = S^{-1}D^T\Delta\mathbf{g}$ of the likelihood term[6]. The remaining multi-normal integral can be performed analytically and the marginal likelihood reads

$$
p(\mathbf{g}|\mathbf{f},E,\{\mathbf{c}\},\mathcal{I}) = \frac{z\Gamma(\frac{E}{2})}{2Z_L}\left(\frac{\Delta\mathbf{g}^T\mathcal{P}\Delta\mathbf{g}}{2\sigma^2}\right)^{-\frac{E}{2}} e^{-\frac{1}{2\sigma^2}\Delta\mathbf{g}^T(1\!\!1-\mathcal{P})\Delta\mathbf{g}} \quad . \quad (11)
$$

We introduced the definition $\mathcal{P}_{ll'} = (DS^{-1}D^T)_{ll'} = \sum_{\nu\nu'}\tilde\phi_\nu(x_l)S^{-1}_{\nu\nu'}\tilde\phi_{\nu'}(x_{l'})$ of the matrix $\mathcal{P}$ which projects into the subspace spanned by the "occupied" basis functions $\{\tilde\phi_\nu(x_l)\}$. As demanded, the result is manifestly invariant under linear transformations of the basis set.

2.2. MARGINAL POSTERIOR FOR THE IMAGE

Now that we have the marginal likelihood we can determine the posterior probability for the image from Eq.5. We still assume that the expansion order is known. For the image we use the entropic prior[1,2]. The expansion order will not greatly exceed 1 and therefore the image dependence of $(\frac{\Delta\mathbf{g}^T\mathcal{P}\Delta\mathbf{g}}{2})$ in the prefactor in Eq.11 is negligible as the same expression appears exponentially in the last term of Eq.11. As far as the posterior probability for the image $\mathbf{f}$ is concerned the prefactor can therefore be taken as constant and we obtain for the marginal posterior

$$
p(\mathbf{f}|\mathbf{g},E,\{\mathbf{c}\},\alpha,\sigma,\mathcal{I}) \propto e^{\frac{1}{2}\tilde\chi^2}p(\mathbf{f}|\alpha,\mathcal{I})) \quad . \quad (12)
$$

We have now introduced explicitly the dependence upon the regularization parameter α of the entropic prior and the noise level σ as they will become important in the sequel. The χ^2-measure has been changed due to the marginalization of the background to $\tilde\chi^2 = \Delta\mathbf{g}^T(1\!\!1 - \mathcal{P})\Delta\mathbf{g}/\sigma^2$ as compared to its original form $\tilde\chi^2 = \Delta\mathbf{g}^T\ 1\!\!1\ \Delta\mathbf{g}/\sigma^2$ in Eq.4. The effect of the matrix $(1\!\!1 - \mathcal{P})$ is obvious: it projects into subspace of the unoccupied basis functions. The modified χ^2-measure is therefore invariant under transformations of the image $\mathbf{f}$ which lead to changes in $g(\mathbf{f})$ that can be described by the "occupied" basis functions. In other words, those components of the image $\mathbf{f}$ that yield structures in $\mathbf{g}(\mathbf{f})$ that can be described by the "occupied" basis functions have no influence on the modified misfit and can be chosen according to the prior requirements. It is expedient to introduce the following transformation. The matrix $\mathcal{Q} = (1\!\!1 - \mathcal{P})$ has eigenvalues $\lambda_i \geq 0$. It has at least E zero eigenvalues as the occupied basis function belong to the null space of $\mathcal{Q}$. Using the spectral representation $\mathcal{Q} = U\mathrm{diag}(\lambda_i)U^T$, where the columns of the unitary matrix U contain the eigenvectors of $\mathcal{Q}$ we have

$$\tilde{\chi}^2 = (U\Delta\mathbf{g})^T \mathrm{diag}(\lambda_l)(U\Delta\mathbf{g})/\sigma^2 = \sum_l (U\Delta\mathbf{g})_l^2/(\sigma^2/\lambda_l) \quad . \tag{13}$$

Therefore, if we use the transformations

$$\begin{aligned} g_l &\rightarrow \tilde{g}_l = \sum_{l'} U_{ll'} g_{l'} \\ g_l(\mathbf{f}) &\rightarrow \tilde{g}_l(\mathbf{f}) = \sum_{l'} U_{ll'} g_{l'}(\mathbf{f}) \\ \sigma^2 &\rightarrow \tilde{\sigma}^2 = \min(\sigma^2/\lambda_l, 10^{30}) \end{aligned} \tag{14}$$

we can retain the original form of the likelihood with the modified input data. The zero eigenvalues of $\mathcal{Q}$ lead to modified data constraints with infinite noise $\tilde{\sigma}_l$, which is another way of expressing the invariance features discussed above. Using the transformation (Eq.14) of the input data the exponential factor $\exp(-\frac{1}{2}\tilde{\chi}^2)$ is proportional to the original Gaussian likelihood in the transformed input data

$$e^{-\frac{1}{2}\tilde{\chi}^2} \propto p(\tilde{\mathbf{g}}|\mathbf{f},\tilde{\sigma},\mathcal{I}) \quad . \tag{15}$$

So the ultimate form of the posterior for $\mathbf{f}$ is simply

$$p(\mathbf{f}|\mathbf{g},E,\{\mathbf{c}\},\alpha,\sigma,\mathcal{I}) \propto p(\tilde{\mathbf{g}}|\mathbf{f},\tilde{\sigma},\mathcal{I})p(\mathbf{f}|\alpha,\mathcal{I})) \quad . \tag{16}$$

The old quantified maxent code can now be used without any changes.

2.3. EXPANSION ORDER

So far we have assumed that we know the expansion order. This is not really the case and we have to determine it following the rules of probability. For the probability $p(E|\mathbf{g},\mathcal{I})$ for expansion order E, the E-dependence of the prefactor in Eq.11 plays an important role as it is part of the Ockham factor, while the $\mathbf{f}$-dependence is weak and can be replaced by its value at the maxent solution $\mathbf{f}^*$. The misfit entering the prefactor is accordingly $\Delta\mathbf{g}^*$. The probability for the expansion order is then

$$\begin{aligned} p(E|\mathbf{g},\alpha,\sigma,\mathcal{I}) &= \int \mathcal{D}f \int d^E c\, p(E,\mathbf{f},\mathbf{c}|\mathbf{g},\alpha,\sigma,\mathcal{I}) \\ &= \int \mathcal{D}f\; p(\mathbf{g}|\mathbf{f},\{\mathbf{c}\},E,\sigma\mathcal{I}))\; p(\mathbf{f}|\alpha,\mathcal{I})\; p(E|\mathcal{I})/p(\mathbf{g}|\mathcal{I}) \qquad (17) \\ &\propto p(E|\mathcal{I})\;\Gamma(\frac{E}{2})(\frac{(\Delta\mathbf{g}^*)^T P\Delta\mathbf{g}^*}{2\sigma^2})^{-\frac{E}{2}} \int \mathcal{D}f\; p(\tilde{\mathbf{g}}|\mathbf{f},\tilde{\sigma}\mathcal{I})\; p(\mathbf{f}|\alpha,\mathcal{I}) \qquad (18) \end{aligned}$$

The integral in the last equation is proportional to the α-evidence in the transformed input quantities $p(\alpha|\tilde{\mathbf{g}},\tilde{\sigma},\mathcal{I})$[1] and the probability for the expansion order simplifies considerably

$$p(E|\mathbf{g},\alpha,\sigma,\mathcal{I}) \propto p(E|\mathcal{I})\,\Gamma(\frac{E}{2})(\frac{(\Delta\mathbf{g}^*)^T P \Delta\mathbf{g}^*}{2\sigma^2})^{-\frac{E}{2}}\; p(\alpha|\tilde{\mathbf{g}},\tilde{\sigma},\mathcal{I}) \qquad . \tag{19}$$

Again, we can use our old maxent program to determine the probability for the expansion order as it routinely determines the α-evidence. We use an uninformative flat prior for E is $p(E|\mathcal{I}) = const$. Before we apply our formalism to an example, we discuss its qualitative features. According to the exponential in Eq.11 the marginal likelihood term increases with increasing $(\Delta\mathbf{g}^{*T}\mathcal{P}\Delta\mathbf{g}^*)/\sigma^2$ as long as the total misfit χ^2 does not change. $(\Delta\mathbf{g}^{*T}\mathcal{P}\Delta\mathbf{g}^*)/\sigma^2$ measures the projected misfit between experimental signal and image in the subspace of occupied basis functions. In essence it is the background power in units of the noise level. For small E the probability, therefore, increases with E in order to eliminate all parts from the image which can be considered as background and which can be described by the first basis functions. The probability for E increases until the background is completely described within the noise-limits. Afterwards nothing is gained by increasing E as the image can only be described by the high order basis functions and the term $(\Delta\mathbf{g}^{*T}\mathcal{P}\Delta\mathbf{g}^*)/\sigma^2$ is roughly independent of E. Hence, for moderate E the probability starts to decrease due to the Ockham factor $(\Delta\mathbf{g}^{*T}\mathcal{P}\Delta\mathbf{g}^*)/\sigma^2)^{-E/2}$.

3. Example

To assess the selfconsistent background we apply it to mock data discribing the situation encountered in electron spectroskopy. The sought-for spectral function consists of two peaks as depicted by dashed lines in fig.1. The spectral function is convoluted by a Gaussian apparatus function which has twice the width of the spectral peaks. The broadened peak sits on top a quadratic background. The signal is corrupted by a noise with constant absolute value of 1 percent of the maximum of the spectral peaks. If take only a constant background into account, i.e. $E = 0$, then the main part of the background can still be found in the image. But the worrying part is the pronounced noise-fitting. If this were a real-world example one would be tempted to identify the various peaks with elementary excitations in the system. The reason for the strong noise-fitting is due to the broad apparatus function and the correspondingly small eigenvalues. A slightly worse result is obtained if the background is not subtracted at all. Fortunately, we are now in the position to eliminate the background properly and we see in fig.1 how the image improves nicely up to expansion order $E = 2$. High order basis fucntions have no influence on the image anymore.

The probability for the expansion order E is given in fig.2. We see that it is sharply peaked at $E = 2$ in agreement with the background added to the mock data. As usual the probability distribution is asymmetric. The left flank of the peak is governed by the data-constraints while the right flank is governed by the Ockham factor.

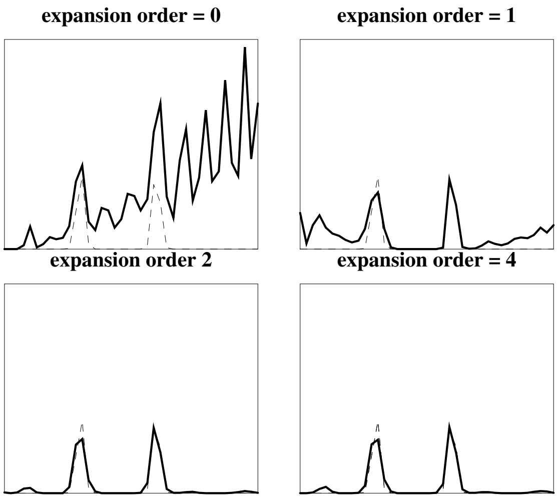

Figure 1. Sequence of reconstructed images (solid) using maxent for increasing expansion order E. The mock data are depicted by dashed lines.

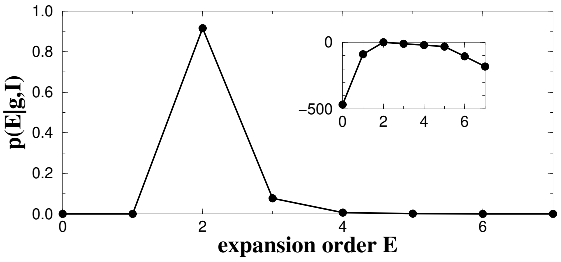

Figure 2. Probability for the expansion order E. The inset shows the result in dB.

4. Summary

We have considered the problem that the measured signal contains an additional part stemming from some unknown background. Following the rules of Bayesian probability theory the background is eliminated and the marginal likelihood function is derived. It appeared that the marginal likelihood is formally identical to the original likelihood with modified input such as experimental data and noise. Therefore the selfconsistent treatment of unknown background can be included in existing maxent codes with insignificant changes. We have illustrated that the correct treatment of the background can improve the image considerably.

References

1. S. Gull, "Developments in maximum entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 53–71, Kluwer Academic Publishers, Dordrecht, 1989.
2. J. Skilling, "Quantified maximum entropy," in *Maximum Entropy and Bayesian Methods*, P. Fougère, ed., pp. 341–350, Kluwer Academic Publishers, Dordrecht, 1990.
3. B. Buck and V. Macaulay, eds., *Maximum Entropy in Action*, Clarendon Press, Oxford, 1991.
4. W. von der Linden, "Maximum-entropy data analysis," *Appl. Phys. A*, **60**, p. 155, 1995.
5. R. Fischer, W. von der Linden, and V. Dose, "On the importance of α marginalization in maximum entropy," in *Maximum Entropy and Bayesian Methods*, R. Silver and K. Hanson, eds., Kluwer Academic Publishers, Dordrecht, 1996. to be published.
6. G. Bretthorst, "Excerpts from bayesian spectrum analysis and parameter estimation," in *Maximum Entropy and Bayesian Methods in Science and Engineering*, G. Erickson and C. Smith, eds., vol. 1, p. 75, Kluwer Academic Publishers, Dordrecht, 1988.
7. R. D. Rosenkrantz, ed., *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, 1983.
8. A. Ramaswami, W. von der Linden, and V. Dose, "Improper priors and evidence approximation," in *Maximum Entropy and Bayesian Methods (1996)*, J. S. S. Sibisi, ed., Kluwer Academic Publishers, Dordrecht, 1997. to be published.