Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

Skewness Maximization for Impulsive Sources in Blind Deconvolution

Patrik Pääjärvi and James P. LeBlanc

Luleå University of Technolgy, Div. of Signal Proc.
SE-971 87 Luleå, SWEDEN {patrik,leblanc}@csee.ltu.se

ABSTRACT

In blind deconvolution problems, a deconvolution filter is often determined in an iterative manner, where the filter taps are adjusted to maximize some objective function of the filter output signal. The kurtosis of the filter output is a popular choice of objective function. In this paper, we investigate some advantages of using skewness, instead of kurtosis, in situations where the source signal is impulsive, i.e. has a sparse and asymmetric distribution. The comparison is based on the error surface characteristics of skewness and kurtosis.

1. INTRODUCTION AND PROBLEM SETTING

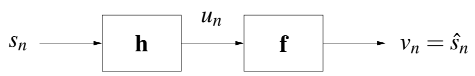

Figure 1: Block diagram of a deconvolution problem.

Fig. 1 shows a discrete-time deconvolution problem model. A source signal, s_n, whose characteristics are not completely known, is convolved with some unknown transfer function, **h**. The output signal, u_n, is then applied to a deconvolution filter, **f**, which, ideally, reconstructs s_n as $\hat{s}_n = v_n = u_n * \mathbf{f}$.

For geophysical applications in which the source signal has a sparse distribution (i.e. 'spiky' appearance), Wiggins [1] proposed a method called *minimum entropy deconvolution* (MED). The approach was to use the knowledge that the source signal had a sparse distribution, and try to find the deconvolution filter whose output distribution was as sparse as possible. As a measure of sparseness, Wiggins proposed the 'varimax norm' (similar to the more commonly known *kurtosis*) as a measure of the 'spikiness' of the deconvolution filter output. The varimax norm V for a filter output sequence v_n of M samples is defined as

$$V = \frac{\sum_{n=0}^{M-1} v_n^4}{\left(\sum_{n=0}^{M-1} v_n^2\right)^2}. \tag{1}$$

The MED method consisted of choosing an initial filter vector **f** and then iteratively adjust the filter towards maximizing V.

Most deconvolution methods are based on some knowledge about the distribution of s_n. By using some suitable objective function $O(v_n)$ of the deconvolution filter output, the filter can be adapted towards maximizing $O(v_n)$.

Donoho [2] generalized the theory behind minimum entropy deconvolution by considering a family of objective functions of a sequence v_n of length M,

$$O_s^r(v_n) = \frac{\frac{1}{M}\sum_{n=0}^{M-1} |v_n|^r}{\left(\frac{1}{M}\sum_{n=0}^{M-1} |v_n|^s\right)^{r/s}}, \tag{2}$$

of which the varimax norm is a scaled version of $O_2^4(v_n)$.

Donoho noted that, as a consequence of the central limit theorem, linear combinations of identically distributed random variables become 'more Gaussian' than the individual variables. Therefore, the transfer function output signal u_n will have a distribution that is more nearly Gaussian than the distribution of s_n. Any objective function should therefore be used to reduce 'the Gaussianity' of the deconvolution filter output.

One suitable measure of Gaussianity for an MED implementation would be the *kurtosis*, K_v, of v_n,

$$K_v = E\{v_n^4\}/(E\{v_n^2\})^2, \tag{3}$$

where $E\{\cdot\}$ denotes expectation. Wiggins varimax norm is a scaled approximation of K_v. Thus, its objective would be to find the filter whose output has a kurtosis value far from a Gaussian signal (the kurtosis value of all Gaussian distributed signals is 3).

However, for *impulsive* sources, the kurtosis may not perform well [3]. An alternative choice of objective function might be the *skewness*, S_v, of v_n, defined as

$$S_v = E\{v_n^3\}/(E\{v_n^2\})^{3/2}. \tag{4}$$

Note that skewness maximization clearly would not be suitable for deconvolution of symmetrically distributed source signals, since the skewness of any filtered version of such a signal is zero.

Next, we compare kurtosis and skewness when used as objective functions for blind deconvolution of impulsive signals. This comparison considers the error surface topologies, i.e. K_v and S_v as functions of the filter coefficients. The error surface topology will affect the convergence of MED algorithms. In particular, the number of stationary points (i.e. points where the gradient of the error surface is zero) is an important characteristic, as an excessive number of saddle points (stationary points having a non-definite Hessian) 'stall' gradient-based filter adaptions.

1.1. Notation and Definitions

To enable a comparison between skewness maximization and kurtosis maximization for blind deconvolution, we introduce notation of such gradient-based methods.

The deconvolution filter $\mathbf{f}$ used is assumed to be an FIR filter of order $N+1$, represented by the column vector

$$\mathbf{f} = [f_0 \; f_1 \; \cdots \; f_N]^{\mathrm{T}}, \tag{5}$$

where f_m denotes the m^{th} filter coefficient. The filter output at time n is given by the convolution sum

$$v_n = \sum_{k=0}^{N} f_k u_{n-k}. \tag{6}$$

A simple strategy for maximizing any objective function, $O(v_n)$, is to use a gradient method wherein the filter coefficients are adapted iteratively towards increasing $O(v_n)$, regarding it ultimately as a function of $\mathbf{f}$, $O(\mathbf{f})$. Denote the filter vector after i iterations as $\mathbf{f}^{(i)}$, the next filter vector will be chosen as

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} + \mu \nabla_{O(\mathbf{f}^{(i)})}, \tag{7}$$

where

$$\nabla_{O(\mathbf{f})} = \left[\frac{\partial O}{\partial f_0} \; \frac{\partial O}{\partial f_1} \; \cdots \; \frac{\partial O}{\partial f_N} \right]^{\mathrm{T}} \tag{8}$$

is the gradient vector of $O(\mathbf{f})$, and μ is some fixed or variable stepsize.

The convergence of filter adaption algorithms based on gradient ascent, such as (7), depends mainly on two factors: the choice of stepsize, μ, and the topology of the error surface $O(\mathbf{f})$.

The stepsize choice is an implementation issue. It must be chosen small enough to allow convergence to a (possibly local) maximum (the stability issue), while choosing a too small stepsize incurs excessive iteration steps. The error surface topology, however, depends on the algebraic structure of the objective function used. The error surfaces of common blind deconvolution objective functions are well known to be multimodal (i.e. to have multiple local maxima). The number of stationary points for kurtosis has been explored [4], [5], but similar results for skewness has not been found.

2. COMPARISON OF ERROR SURFACE TOPOLOGIES

An important characteristic of an error surface $O(\mathbf{f})$ is the number of stationary points, i.e. the number of points where the gradient, $\nabla_{O(\mathbf{f})}$, is zero. More stationary points generally means slower convergence of the gradient algorithm.

By writing out (3) as a function of the filter coefficients, we obtain

$$K_v = \frac{E\left\{ \left(\sum_{k=0}^{N} f_k u_{n-k} \right)^4 \right\}}{\left(E\left\{ \left(\sum_{k=0}^{N} f_k u_{n-k} \right)^2 \right\} \right)^2}. \tag{9}$$

Taking the gradient of (9) with respect to the m^{th} filter coefficient, f_m, and equating to zero, we obtain the following; for $m, i = 0 \dots N$,

$$\begin{aligned} \sum_i f_i^3 \mathcal{R}_{\substack{0\\0\\m-i}} + 3\sum_{i \neq j} f_i^2 f_j \mathcal{R}_{\substack{0\\j-i\\m-i}} + \sum_{i \neq j \neq k} f_i f_j f_k \mathcal{R}_{\substack{j-i\\k-i\\m-i}} & \\ - \sigma_v^2 K_v \sum_i f_i \mathcal{R}_{m-i} = 0, & \end{aligned} \tag{10}$$

where

$$\sigma_v^2 = E\{v_n^2\} = E\left\{ \left(\sum_{k=0}^{N} f_k u_{n-k} \right)^2 \right\}, \tag{11}$$

and the 2nd and 4th moments of u_n are defined as

$$\mathcal{R}_i = E\{u_n u_{n-i}\} \qquad \mathcal{R}_{\substack{i\\j\\k}} = E\{u_n u_{n-i} u_{n-j} u_{n-k}\}.$$

The corresponding equation for the skewness is found similarly by writing out (4) as a function of $\mathbf{f}$, taking the gradient with respect to the m^{th} filter coefficient, f_m, and equating to zero. We obtain the following; for $m, i = 0 \dots N$,

$$\sum_i f_i^2 \mathcal{R}_{\substack{0\\m-i}} + \sum_{i \neq j} f_i f_j \mathcal{R}_{\substack{j-i\\m-i}} - \sqrt{\sigma_v^2}\, S_v \sum_i f_i \mathcal{R}_{m-i} = 0, \tag{12}$$

where the 3rd moment of u_n is defined as

$$\mathcal{R}_{\substack{i\\j}} = E\{u_n u_{n-i} u_{n-j}\}.$$

We note that the kurtosis-based stationary points (10) consist of a system of $N+1$ polynomial equations in $N+1$ variables $(f_0, \dots, f_N)$. Each equation in the system has the same monomial support and a total degree of 5. The Bezout upper bound on the number of solutions (i.e. stationary points of the error surface) is then 5^{N+1}.

Similarly, the skewness system of equations consist of $N+1$ polynomials of total degree 4, yielding a Bezout upper bound on the number of stationary points of 4^{N+1}. Even for moderate filter lengths $(N+1)$, the number of possible stationary points is considerably smaller for the skewness error surface. This generally means faster convergence for gradient algorithms of the form (7).

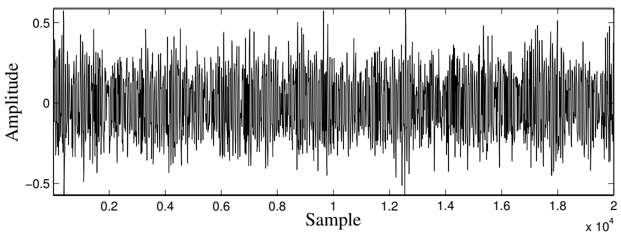

Figure 2: Measurement data from a sound recording of a diesel engine.

3. EXPERIMENTAL RESULTS

3.1. MED Algorithm Comparison

To support the view in Section 2, a simulation was done in which two block-mode versions of the same MED algorithms, one using kurtosis and the other using skewness as the objective function, were applied to real measurement data.

The data, shown in Fig. 2, consisted of a sound recording of a running diesel engine. Referring to Fig. 1, the source signal s_n is the explosions from the pistons. The transfer function $\mathbf{h}$ is the engine block and housing through which the source signal propagates. The source signal is thought to be impulsive, i.e. it has a sparse and asymmetric distribution, although the measured signal appears symmetric and Gaussian, as seen in Fig. 2, after passing through the transfer function. The measurement data consists of u_n plus added noise. The deconvolution filter length was chosen to be 2000.

The MED algorithm used in the experiment was based on the filter iteration (7). Although the stepsize, μ, can be varied during iteration in several ways, a fixed stepsize was used for simplicity. Each algorithm was run 35 times, using different unit-norm initializations. Each filter was initialized with one large center tap, and the rest of the taps picked randomly from a normal distribution, with a standard deviation of 2% of the center tap magnitude. This is a reasonable approximation to the 'customary center tap initialization' of blind deconvolution folklore.

4000 iterations were performed to allow both algorithms to converge. The stepsizes for the two algorithms cannot be directly compared. In order to make a fair comparison, the stepsize for skewness was chosen just small enough to keep almost all runs stable, while the kurtosis stepsize was chosen so that about half of the runs became unstable going into convergence. In this way, the convergence rate of the kurtosis algorithm was essentially maximized for fixed stepsize.

As a comparison between the two algorithms, the kurtosis and skewness versus iteration number was recorded for each run and plotted in Fig. 3. In Fig. 4, the averages of all 35 runs are shown for both kurtosis and skewness. The two plots in Fig. 4 are normalized to the same final value, since the magnitudes of the two objective functions cannot be directly compared.

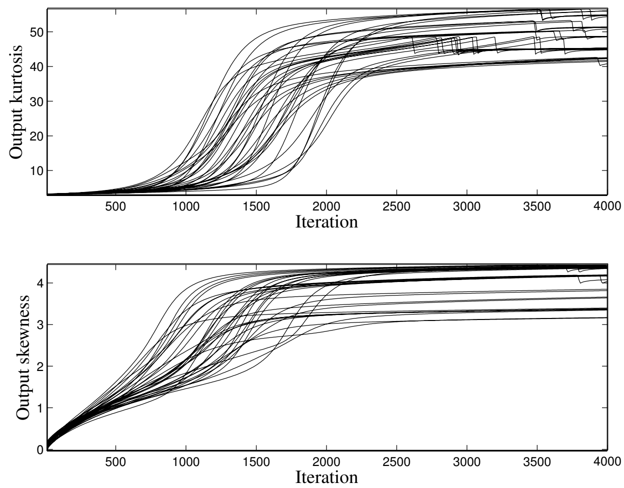

Figure 3: Kurtosis (top) and skewness (bottom) versus filter iteration.

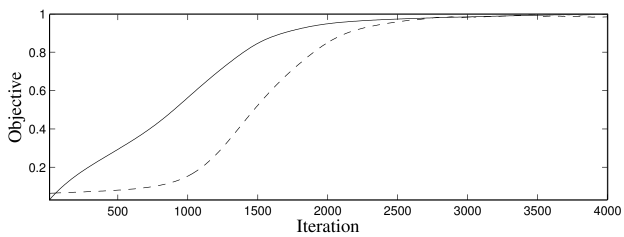

Figure 4: Kurtosis (dashed) and skewness (solid) versus filter iteration.

As seen from Fig. 4, the MED algorithm using skewness is initially steeper and reaches 50% of its final value considerably faster than the kurtosis algorithm.

The results shown in Fig. 4 provide support for the results in Section 2, namely that the error surface of skewness contains fewer stationary points, meaning less 'flat' regions at which the MED algorithm might get stalled. Fig. 5 shows the deconvolution filter outputs for one run of the kurtosis and skewness MED algorithms. Both algorithms have deconvolved the source signal and produced a sparsely distributed signal.

3.2. Error Surface Topology Comparison for a 3-Tap Filter

As an illustrative comparison, the error surfaces for skewness and kurtosis for a low-dimensional (3-tap) filter were compared visually. An impulsive signal was synthesized and filtered through a ARMA(1,1) low-pass filter. The error surfaces for skewness and kurtosis were then plotted over a set of unit-norm, three-tap deconvolution filters (i.e. the unit sphere). Figures 6 and 7 show contour plots of the error surfaces for kurtosis and skewness respectively.

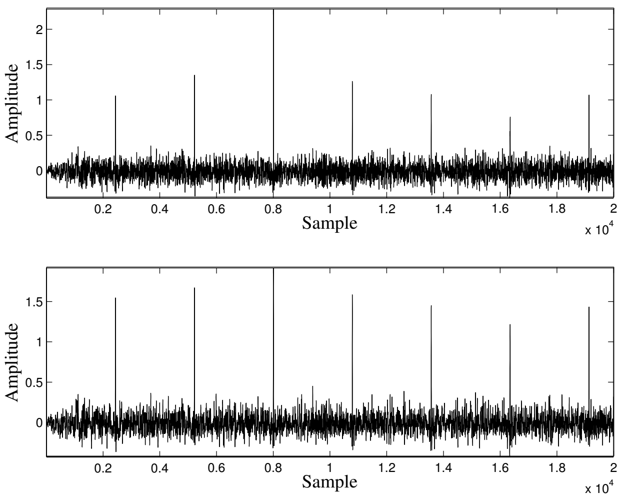

Figure 5: Kurtosis (top) and skewness (bottom) deconvolution filter outputs.

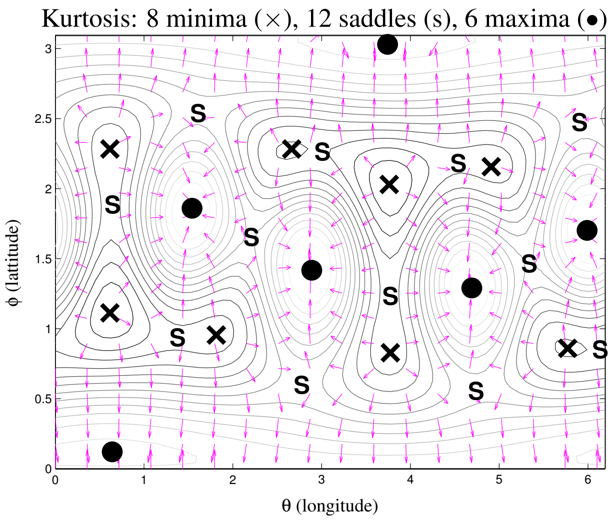

Figure 6: Kurtosis error surface.

Small arrows indicate the direction of the gradient, and the stationary points are marked and classified as minima (×), saddle points (s) or maxima (•). The figures show that the error surface of kurtosis has more stationary points than the skewness error surface. As a check, it was verified that the vector fields satisfied the Euler Characteristic of the sphere [6].

4. CONCLUSIONS

The use of skewness instead of kurtosis as the objective function for minimum entropy deconvolution of impulsive sources has the benefit of an error surface with fewer saddle points, allowing better convergence behaviour for simple, gradient-based methods. This has been demonstrated using both analytical and experimental results.

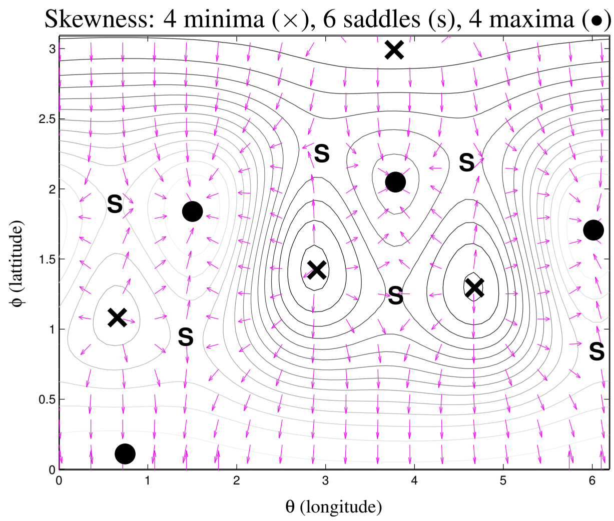

Figure 7: Skewness error surface.

Acknowledgements

This work has been partially supported by VITAL Project, Mål 1 EU Program.

REFERENCES

[1] R. A. Wiggins, "Minimum entropy deconvolution" *Geoexploration*, 16, 1978, pp. 21-35.

[2] D. L. Donoho, "On minimum entropy deconvolution", *Applied Time Series Analysis*, D.F. Findley, Ed., New York: Academic Press, 1981.

[3] H. Mathis and S. C. Douglas, "Bussgang blind deconvolution for impulsive signals", *IEEE Trans. on Signal Processing*, vol. 51, no. 7, July 2003.

[4] Zhi Ding and Tuan Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming", *IEEE Trans. on Signal Processing*, vol. 48, no. 6, June 2000.

[5] J. P. LeBlanc, I. Fijalkow, C. R. Johnson, Jr. "CMA fractionally spaced equalizers: stationary points and stability under IID and temporally correlated sources", *International Journal of Adaptive Control and Signal Processing*, vol. 12, 1998.

[6] V. Guillemin, A. Pollack, *Differential Topology*. Englewood Cliff, NJ, Prentice-Hall, 1974.