SUBOPTIMAL ROBUST ESTIMATION USING RANK SCORE FUNCTIONS

Christopher L. Brown

Dept. of Signals & Systems
Chalmers Univ. of Tech.
Göteborg, Sweden
chris.brown@ieee.org

Ramon F. Brcich

School of Elec. & Comp. Eng.
Curtin Univ. of Tech.
Perth, Australia
r.brcich@ieee.org

Anisse Taleb

Ericsson AB
Stockholm,Sweden
Anisse.Taleb@era.ericsson.se

ABSTRACT

A parameter estimation scheme based on the adaptive modelling of the score function of M-estimators is presented. The weights of basis functions are estimated from the data to match the empirical distribution. The bases utilise rank based score functions to remove dependence on scale from the basis selection process. While determination of appropriate bases for a distribution is shown to be possible, the robustness and adaptivity of the scheme means good results may be achieved regardless.

1. INTRODUCTION

Consider the well known scenario where a parametric signal, $s_t(\boldsymbol{\theta})$, is observed in additive i.i.d noise, x_n,

$$y_n \quad = \quad s_n(\boldsymbol{\theta}) + x_n, \qquad n = 1, \ldots, N.$$

Estimation of the unknown parameter vector $\boldsymbol{\theta}$ is required in numerous signal processing (and other) applications. Classical least squares solutions have been well studied for the case of Gaussian noise. Numerous non-Gaussian cases have also received attention when a parametric model for the distribution is assumed. Some examples are the K-distribution used in radar, the generalised Gaussian distribution or the various heavy tailed models [1], which have been proposed for wireless and underwater communications.

Of course when an estimator is based on a specific model, any model deviations may be problematic and require that the estimator be robust to it. To this end one may use M-estimators [2, 3] to implement sub-optimal estimators which are robust to changes in distribution. M-estimators use a robust suboptimal score function instead of the optimal influence function (as determined by the noise distribution).

In [4, 5] it was proposed that the score function be modelled as a linear combination of basis functions, whose weights are adaptively estimated from the observations. Under suitable constraints it was shown that this scheme resulted in improved small sample performance, with minimal large sample loss. However, for appropriate choice of non-linear bases, an estimate of the scale (variance) of the noise was required.

Here, a modification to this scheme is presented that uses rank based score functions. Whereas the previous estimator required a scale dependent nonlinearity, here this is replaced by a linear function and a scale independent rank based nonlinearity. The design of such nonlinearities is discussed and brief simulation results presented.

2. SIGNAL MODEL

Consider the general signal in additive noise model,

$$y_n = s_n(\boldsymbol{\theta}) + x_n \qquad n = 1, \ldots, N$$

where x_n is i.i.d. noise and the signal, s_n, is parameterised by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T$, $(\cdot)^T$ denoting transposition. The aim is to estimate $\boldsymbol{\theta}$ from N observations y_n. Given the noise density, $f(x)$, one obtains the ML solution as

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} -\log f\left(y_n - s_n(\boldsymbol{\theta})\right)$$

Alternatively, the solution to the P coupled equations

$$\sum_{n=1}^{N} \psi\left(y_n - s_n(\boldsymbol{\theta})\right) \frac{ds_n(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \mathbf{0}$$

can be found, where $\psi(x) = -f'(x)/f(x)$, $f'(x) = \frac{d}{dx} f(x)$, is the influence function of $f(x)$. It is clear that without *a priori* knowledge of $f(x)$ estimation of $\boldsymbol{\theta}$ cannot be optimal.

In an M-estimator [2] $-\log f(x)$ is replaced with a similarly behaved function, chosen to confer robustness on the estimator under deviations from a nominal density. Estimates for $\boldsymbol{\theta}$ are obtained by solving the P coupled equations

$$\sum_{n=1}^{N} \varphi\left(y_n - s_n(\boldsymbol{\theta})\right) \frac{ds_n(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \mathbf{0} \qquad .$$

In [4, 5], $\varphi(x)$ was not set in advance but estimated from the observations. Here, again $\varphi(x)$ will be determined by the observations, however under a different form to previous contributions.

3. ROBUST PARAMETER ESTIMATION UTILISING RANK SCORE FUNCTIONS

Consider a linear model for $\varphi(x)$ [4, 5],

$$\varphi(x) = \sum_{k=1}^{K} a_k g_k(x) = \boldsymbol{a}^T \boldsymbol{g}(x) \tag{1}$$

where $\boldsymbol{a} = (a_1, \ldots, a_K)^T$ are weights and the functions $\boldsymbol{g}(x) = (g_1(x), \ldots, g_K(x))^T$ are the appropriately chosen basis functions. Without loss of generality, the constraint $\sum_k a_k = 1$ is imposed. Then if it is possible to find appropriate weights such that $\varphi(\cdot)$

1. **Initialisation** Set $j = 0$. Obtain an initial estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_0$.
2. **Determine the residuals**
$$\hat{x}_n = y_n - s_n(\hat{\boldsymbol{\theta}}_j).$$
Rank the elements of $\{x_1, \ldots, x_N\}$ in ascending order. Denote the rank of $\hat{x}_n$ in $\{x_1, \ldots, x_N\}$ as r_n. If $r_n = i$, then $x_n = x_{[i]}$.
3. **Estimate the score function** Estimate the weights by,
$$\hat{\boldsymbol{a}} = \left(\sum_{n=1}^{N} \hat{x}_n^2 \boldsymbol{h}\left(\frac{r_n}{N}\right)\boldsymbol{h}^T\left(\frac{r_n}{N}\right)\right)^{-1}\left(\sum_{i=1}^{N} \frac{\hat{x}_{[i+1]}\boldsymbol{h}(\frac{i+1}{N}) - \hat{x}_{[i-1]}\boldsymbol{h}(\frac{i-1}{N})}{x_{[i+1]} - x_{[i-1]}}\right) \quad .$$
The score function is then $\varphi(x_n) = \hat{\boldsymbol{a}}^T \boldsymbol{h}\left(\frac{r_n}{N}\right) x_n$.
4. **Update the parameter estimate** Solve,
$$\sum_{n=1}^{N} \varphi(y_n - s_n(\hat{\boldsymbol{\theta}}_j))\frac{ds_n(\hat{\boldsymbol{\theta}}_j)}{d\boldsymbol{\theta}} = \boldsymbol{0}$$
given the initial estimate $\hat{\boldsymbol{\theta}}_j$, assign the solution to $\hat{\boldsymbol{\theta}}_{j+1}$.
5. **Check for convergence** If $||\hat{\boldsymbol{\theta}}_{j+1} - \hat{\boldsymbol{\theta}}_j|| < \epsilon$ stop, otherwise set $j \to j+1$ and go to step 2.

Table 1. Algorithm for M-estimation with rank based score function estimation.

approximates the unknown $\psi(\cdot)$, the corresponding estimator will be near optimal.

Given that $\lim_{x\to\pm\infty} g_k(x)f(x) = 0$, one obtains the optimal least squares solution to minimising the MSE between $\varphi(\cdot)$ and $\psi(\cdot)$ as
$$\boldsymbol{a} = \mathsf{E}\left[\boldsymbol{g}(x)\boldsymbol{g}^T(x)\right]^{-1}\mathsf{E}\left[\boldsymbol{g}'(x)\right] \qquad (2)$$

Now let $g_k(x_n; \boldsymbol{x}) = x_n h_k\left(\frac{r_n}{N}\right)$ where $\boldsymbol{x} = \{x_1, \ldots, x_N\}$ and r_n is the rank of x_n in $\boldsymbol{x}$. While "conventional" non-linear score functions are scale dependent, i.e. scaling the observations requires appropriate scaling of the score functions (in x) to maintain performance, rank score functions are intrinsically scale independent [6].

If all the noise observations are scaled by a positive value, their *relative* positions, and hence their ranks, will not change. Therefore, for a being a positive scalar,
$$\begin{aligned} g_k(ax_n; a\boldsymbol{x}) &= ax_n h_k\left(\frac{r_n}{N}\right) \\ g_k(ax_n) &= ag_k(x_n) \end{aligned}$$

A scaling of the observations will only result in a linear scaling of the rank based score function. The same cannot be said of a general non-linear function.

Assuming that expectations under the unknown distribution can be replaced by sample averages using the observations, the matrix to be inverted in (2) becomes
$$\begin{aligned} \sum_{n=1}^{N} \boldsymbol{g}(\hat{x}_n)\boldsymbol{g}^T(\hat{x}_n) &= \sum_{n=1}^{N} \hat{x}_n \boldsymbol{h}\left(\frac{r_n}{N}\right)\boldsymbol{h}^T\left(\frac{r_n}{N}\right)\hat{x}_n \\ &= \sum_{n=1}^{N} \hat{x}_n^2 \boldsymbol{h}\left(\frac{r_n}{N}\right)\boldsymbol{h}^T\left(\frac{r_n}{N}\right) \quad . \end{aligned}$$

The vector $\mathsf{E}\left[\boldsymbol{g}'(x)\right]$ can be approximated by
$$\begin{aligned} \sum_{n=1}^{N} \boldsymbol{g}'(x_n) &= \sum_{i=1}^{N} \boldsymbol{g}'(x_{[i]}) \\ &\approx \sum_{i=1}^{N} \frac{\boldsymbol{g}(x_{[i+1]}) - \boldsymbol{g}(x_{[i-1]})}{x_{[i+1]} - x_{[i-1]}} \\ &= \sum_{i=1}^{N} \frac{x_{[i+1]}\boldsymbol{h}(\frac{i+1}{N}) - x_{[i-1]}\boldsymbol{h}(\frac{i-1}{N})}{x_{[i+1]} - x_{[i-1]}} \end{aligned}$$
where $x_{[i]}$ is the i'th order statistic, i.e. the i'th largest element of $\boldsymbol{x}$, $x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[N]}$. (Note that, in practice, this estimate of the rate of change of $\boldsymbol{g}(x)$ may be smoother when using larger "steps" around $\boldsymbol{g}(x_{[i]})$, i.e. using the gradient of the function between $\boldsymbol{g}(x_{[i\pm\delta]})$ where $\delta > 1$.)

The estimate of the score function, from (1), is then incorporated into an M-estimation algorithm. The resulting scheme iterates between three main steps, finding the residuals, estimating the weights used in the score function and updating the parameter estimates. The algorithm is summarised in Table 1.

4. RANK BASED SCORE FUNCTIONS

As with their non-rank counterparts, the choice of rank based score functions should be made using available knowledge regarding the unknown distribution. Having said that, rank based score functions have been shown to exhibit greater robustness to deviations from distributional assumptions [7]. See the discussion in Appendix A regarding the relationship between the true noise distribution and the chosen basis functions.

The following example finds the appropriate rank based score function for Cauchy distributed noise. The pdf of the standard

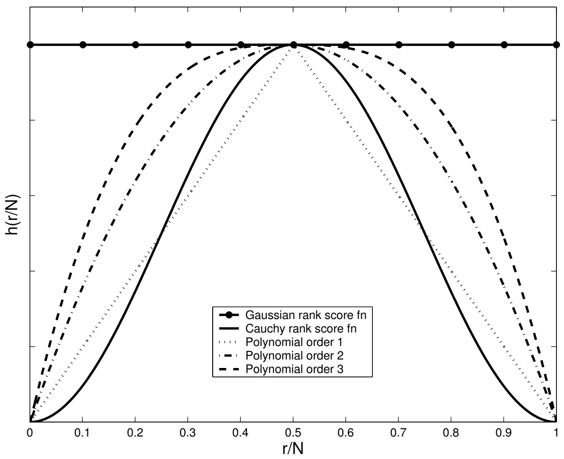

Fig. 1. Sample rank score basis functions.

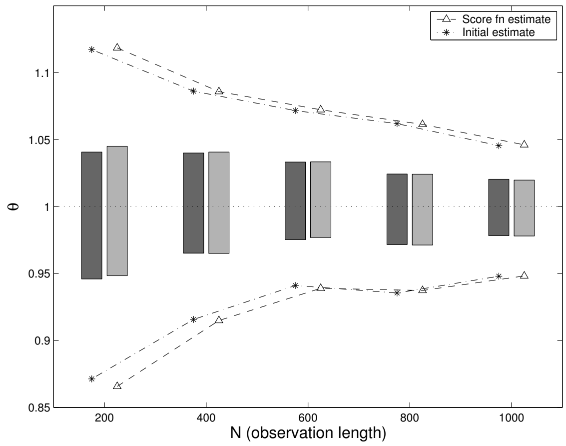

Fig. 2. Spread of estimates of DC signal in Gaussian noise (dark box – initial estimate, light box – score function estimate).

Cauchy distribution is $f(x) = \frac{1}{\pi(1+x^2)}$, thus

$$\psi(x) = -\frac{f'(x)}{f(x)} = \frac{2x}{1+x^2} \qquad .$$

To find the rank score function that is best for $f(x)$, then set $\psi(x) = xh\left(\frac{r}{N}\right)$. Since $\lim_{N\to\infty} \frac{r(x)}{N} = F(x)$, where $F'(x) = f(x)$, then

$$\begin{aligned} h(F(x)) &= \frac{2}{1+x^2} \\ h(u) &= \frac{2}{1+\left(F^{-1}(u)\right)^2} \\ &= \frac{2}{1+\tan^2(\pi(u-\frac{1}{2}))} \\ &= 1-\cos(2\pi u) \qquad . \end{aligned}$$

By contrast, and unsurprisingly, under the Gaussian case the procedure is trivial

$$\begin{aligned} \psi(x) &= x \\ h(F(x)) &= 1 \\ h(u) &= 1 \end{aligned}$$

Of course the optimal score function under Gaussianity is a linear function, thus any rank score function is "unnecessary".

Generic bases using, for example, polynomials are easily implemented, and have been used in the results that follow in this paper. Plots of the bases are shown in Figure 1. Note that the symmetry of the basis functions is a consequence of the assumption of symmetry of $f(x)$.

5. RESULTS

Brief simulation results are presented here to demonstrate the performance of the described technique. The basis functions used included the Cauchy and Gaussian influence functions and polynomials up to order 3 (as described in section 4). Observation sizes were $N = 200$, 400, 600, 800 and 1000. The first example concerned the simple task of estimating a DC level in noise,

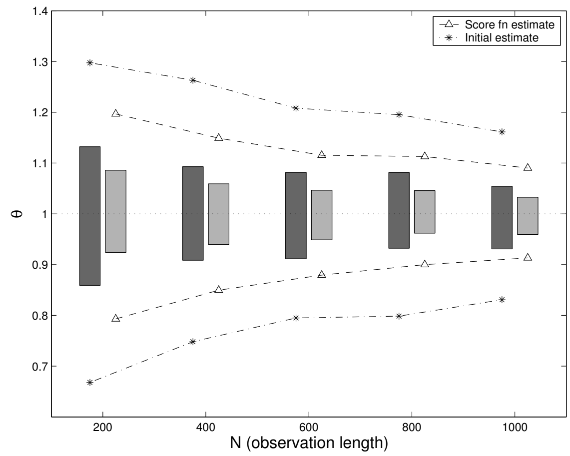

Fig. 3. Spread of estimates of DC signal in t_2 noise (dark box – initial estimate, light box – score function estimate).

$y_n = \theta + x_n$. The initial estimate was taken to be the sample average. Box-whisker plots of the results are shown in Figure 2 for Gaussian noise and Figure 3 for t_2 noise. The boxes enclose the interquartile range, i.e. they join the 25'th and 75'th percentiles of the estimates found by repeated Monte Carlo simulation. Also shown by a $*$ or $\triangle$ are the 5'th (lower mark) and 95'th (upper mark) percentiles. The true parameter value is shown by the dotted line. As expected, under the Gaussian case, no advantage can be gained, however, for a highly non-Gaussian case, significant improvement is evident.

In the second example shown here, the frequency of a sinusoid in noise is estimated. Shown in Figure 4 is the spread of frequency estimates when Laplacian (double exponential) noise is added. The initial frequency estimate is taken as the peak of the zero-padded periodogram. There appears to be a significant improvement in the estimate obtained by the rank score function based technique. This was achieved even without optimisation of

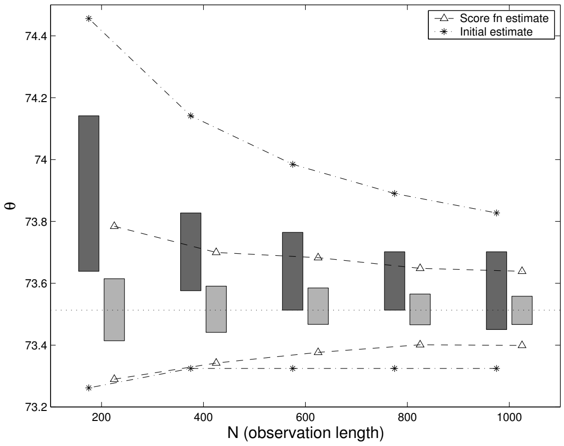

Fig. 4. Spread of estimates of frequency of a sinusoid in Laplacian noise (dark box – initial estimate, light box – score function estimate).

the score function choice.

6. ACKNOWLEDGMENT

The authors would like to thank Prof. Mats Viberg and Prof. Abdelhak Zoubir for valuable discussions and comments.

7. CONCLUSIONS

A robust estimation scheme with adaptive estimation of the unknown influence function using a linear combination of basis functions has been presented. The scale invariance of rank based functions removes a significant design consideration as compared to a previous scheme. The choice of basis functions remains an important consideration in the implementation of the technique, however, as shown, finding bases suited to a particular class of distributions can be done. Preliminary work has begun on improved basis selection using the findings in Appendix A as well as Akaike's Information Criterion and similar methods, to facilitate improved performance.

A. NOISE DISTRIBUTION AND THE CHOICE OF BASES

The addition of an "inappropriate" choice of basis function should, in theory, have no effect on the implemented influence function, $\varphi(x)$, since its weight can be set to 0. However, it is still worthwhile to investigate how the the implemented influence function is affected by the choice of bases.

Consider a score function, $g_k(x)$, which is the influence function (i.e. the "optimal" score function) for a distribution with pdf $f_k(x)$, then

$$g_k(x) = -\frac{d}{dx}\log f_k(x) \tag{3}$$

and, consequently,

$$f_k(x) \propto e^{-G_k(x)}$$

where $G'_k(x) = g_k(x)$.

Now consider a function, $\varphi(x)$, that is the linear combination of other basis functions, as in (1). It is now desired to find the distribution, $f(x)$, for which this will be the optimal choice, i.e. the distribution for which $\varphi(x)$ is equal to the true influence function $\psi(x)$.

$$\begin{aligned}
\varphi(x) = \sum_{k=1}^{K} a_k g_k(x) &= -\frac{d}{dx}\log f(x) \\
\sum_{k=1}^{K} -a_k G_k(x) &\propto \log f(x) \\
\prod_{k=1}^{K} \left[e^{-G_k(x)}\right]^{a_k} &\propto f(x) \\
\prod_{k=1}^{K} \left[f_k(x)\right]^{a_k} &\propto f(x)
\end{aligned}$$

Hence it can be seen that if each basis function $g_k(x)$ is the influence function for a particular distribution with pdf $f_k(x)$, then the score function formed by a linear combination of basis functions is the influence function of a distribution whose pdf is the weighted geometric mean of the $f_k(x)$, $k = 1, \ldots, K$.

This relationship is of particular interest when considering the rate of decay of the tails of distributions. Any non-zero weighting given to $g_k(x)$ corresponding to a bounded pdf, $f_k(x)$, will result in a bounded $f(x)$. Similarly, if all $f_k(x)$ pdfs have slow tail decay rates, $f(x)$ will also inherit this property.

In the above discussion, no form was imposed on $g_k(x)$ and, hence, these results apply equally to rank based score functions as they do to non-rank based equivalents.

B. REFERENCES

[1] J. Friedmann, H. Messer, and J.-F. Cardoso, "Robust parameter estimation of a deterministic signal in impulsive noise," *IEEE Trans. on Signal Processing*, vol. 48, no. 4, April 2000.

[2] P. Huber, *Robust Statistics*, Wiley, 1981.

[3] X. Wang and H. V. Poor, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. on Signal Processing*, vol. 47, no. 2, pp. 289–305, February 1999.

[4] A. Taleb, R. Brcich, and M. Green, "Suboptimal robust estimation for signal plus noise models," in *Conference Record of The Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, 2000.

[5] R. F. Brcich and A.M. Zoubir, "Robust estimation with parametric score function estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2002*, Orlando, FL, USA, May 2002.

[6] T. P. Hettmansperger, *Statistical Inference Based on Ranks*, John Wiley, New York, 1984.

[7] C. L. Brown and A. M. Zoubir, "A nonparametric approach to signal detection in impulsive interference.," *IEEE Trans. on Signal Processing*, vol. 48, no. 9, pp. 2665–2669, September 2000.