

# Optimal convolutive filters for real-time detection and arrival time estimation of transient signals

Michal Natora, Felix Franke, and Klaus Obermayer

**Abstract**—Linear convolutive filters are fast in calculation and in application, and thus, often used for real-time processing of continuous data streams. In the case of transient signals, a filter has not only to detect the presence of a specific waveform, but to estimate its arrival time as well. In this study, a measure is presented which indicates the performance of detectors in achieving both of these tasks simultaneously. Furthermore, a new sub-class of linear filters within the class of filters which minimize the quadratic response is proposed. The proposed filters are more flexible than the existing ones, like the adaptive matched filter or the minimum power distortionless response beamformer, and prove to be superior with respect to that measure in certain settings. Simulations of a real-time scenario confirm the advantage of these filters as well as the usefulness of the performance measure.

**Index Terms**—Adaptive matched filter, minimum variance distortionless response, beamforming, Capon beamformer, linear filters, performance measure

## I. INTRODUCTION

**F**OR detection of signals in single data samples corrupted by Gaussian noise, linear filters, in particular the adaptive matched filter (AMF), have been proven to be powerful. Their performance is measured with respect to the probability of detection and of false alarm; see [1] for a performance analysis of the AMF and other filters. The AMF has been applied amongst others in radar and antenna systems [2]. In other applications, however, the incoming data stream does not consist of a few data samples, but of a continuous data stream, whereas the signal is present only in a few of the samples (transient signals). In this case, the signal must not only be detected, but also its arrival time must be estimated.

The research field of optimal simultaneous detection and estimation has been mainly initiated by the work presented in [3]. Based on this theory some detectors were developed [4]–[6], and most of these approaches rely on order statistics. In the work of [5], however, the authors mention, that in the case of long signals, linear convolutive filters prove to be superior to order statistics. Moreover, linear convolutive filters are computationally much more efficient, and thus, more suitable for real-time applications than order statistics.

This raises the question of which detectors should be used for the mentioned task, and how their performance should be compared. This study focuses in particular on the performance of linear filters, since they are easy to implement and are

optimal in the class of linear transformations [7]. Although the performance of various detectors for transient signals was compared (see [8]–[10]), these studies compared only the detection performance and linear convolutive filters were rarely used for comparison.

Linear convolutive filters, in the following abbreviated simply by the term linear filters, are a convenient approach for the task of simultaneous detection and arrival time estimation of transient signals, and, thanks to their computational efficiency, suitable for real-time applications. For example, they are used for extracting information from bio-medical data [11]–[13] or in speech processing (see [14] for a survey).

However, to the knowledge of the authors, no work exists to date which would propose a measure assigning a performance to detectors with respect to their ability of simultaneously detecting the presence as well as estimating the arrival time of transient signals.

This work is organized as follows: In Sec. II-B the general optimization problem is presented to which linear filters are the solution. By modifying the optimization criteria, a new class of linear filters is derived. In Sec. II-C a measure of performance of detectors with respect to simultaneous detection and arrival time estimation is presented. In Sec. III-A different linear filters are compared with respect to this measure. The results from simulations in Sec. III-B agree with the theoretical findings and demonstrate the usefulness of these new filters and of the performance measure. The work is summarized and discussed in Sec. IV and a brief outlook on further research directions is given.

## II. METHOD

### A. Notation

A notation is used in which symbols for scalar quantities are represented by lower case letters, vectorial quantities are represented by bold lower case letters, and matrices are represented by bold upper case letters. A vector  $\xi$  usually represents a set of sampled data points, the index set being centered around zero, i.e.  $\xi = (\xi_{-b}, \dots, \xi_b)^\top$ .  $T_\xi$  denotes the dimension of the vector  $\xi$ , i.e.  $T_\xi = 2b + 1$ .

The symbol  $\delta_y(x)$  denotes the usual Kronecker delta function, i.e.  $\delta_y(x) = 1$ , if  $x = y$ , and  $\delta_y(x) = 0$  otherwise.

The noncyclic cross-correlation between two vectors  $x$  and  $y$  is denoted as  $x \star y = z$ , where  $z_t = \sum_\tau x_\tau y_{\tau+t}$ . It is  $T_z = T_x + T_y - 1$ .

The notion of variance is slightly abused by attributing the variance to a probability density function (pdf)  $f(x)$  rather

M. Natora and K. Obermayer are with the Institute for Software Engineering and Theoretical Computer Science, Berlin Institute of Technology, 10623 Berlin, Germany. Email: natora@cs.tu-berlin.de

F. Franke is with the Bernstein Center for Computational Neuroscience, 10115 Berlin, Germany.

than to a random variable  $X$ , i.e.

$$\text{Var}_{f(x)} := \text{Var}_{f(x)}(X) = \sum_x x^2 f(x) - \left( \sum_x x f(x) \right)^2 \quad (1)$$

### B. Linear convolutive filters

The measured data  $x_t$  is a continuously sampled data stream which is a linear mixture of a signal source and a noise source  $\eta_t$ . The signal is assumed to be sparse, i.e. consisting only of a short waveform  $\xi$  at specific times. Formally, the data generating process is written as <sup>1</sup>:

$$x_t = \sum_{\tau} v_{t-\tau} \xi_{\tau} + \eta_t \quad (2)$$

The point process  $v_t$  defines the times at which the waveform  $\xi$  is present, and can be modeled for example by a Bernoulli process. The noise  $\eta_t$  is assumed to be Gaussian, with zero mean and covariance matrix  $C$  (not necessarily white). It is assumed that the amplitude distribution of  $v_t$  as well as of  $\xi$  does not change in time, hence, only the presence of the waveform and its arrival time has to be detected, but not its amplitude scaling. Further, it is assumed that the signal waveform  $\xi$  is known, and, due to its sparseness, the noise covariance matrix  $C$  can be estimated reliably.

A perfect detector should retrieve the underlying point process  $v_t$ , as, in this case, all signals were detected and all arrival times estimated correctly. In the following, the focus will be on detectors in the class of linear filters which minimize the quadratic response to the data, combined with a pointwise thresholding of the filter output. This class of filters has the advantage of having an analytical expression, which allows for fast calculation (see [15] for other classes of linear filters).

The optimization problem for this kind of filters is stated as follows:

$$f = \underset{f}{\text{argmin}} \left\{ |l|^2 + \alpha f^{\top} C f \right\} \quad \text{subject to } f^{\top} \cdot \xi = 1 \quad (3)$$

where  $l$  is the filter response to the waveform  $\xi$ , i.e.  $l := f \star \xi$ . The optimization criteria can be understood intuitively: The first term demands response of the filter to the signal to be minimal, except for the correct arrival time, in which case the filter should respond with a well defined response of 1 (which is ensured by the optimization constraint). The response of the filter to noise segments should be minimal as well. Since the noise was assumed to be Gaussian and zero mean, one has to minimize  $\text{Var}(f \star \eta)$ . A short calculation yields  $\text{Var}(f \star \eta) = f^{\top} C f$ . The  $\alpha$  parameter varies the ratio between minimization of the filter response to the signal and to noise.

The solution to the problem in Eq. 3 is given by

$$f = \frac{H^{-1} \xi}{\xi^{\top} H^{-1} \xi} \quad (4)$$

<sup>1</sup>For the sake of clarity and of simplicity, the analysis will be restricted to the case of single channel data. The entire method can be extended to multi-channel data in a straightforward manner.

where the matrix  $H$  is given by  $H := \Xi + \alpha C$ , and  $(\Xi)_{k,l} := (\xi \star \xi)_{k-l}$ .

In the limit of  $\alpha \rightarrow \infty$ , the filter reduces to  $f = C^{-1} \xi / (\xi^{\top} C^{-1} \xi)$ , which is the classical adaptive matched filter (AMF), see [16]<sup>2</sup>, also called minimum variance distortionless response (MVDR) beamformer, or simply Capone beamformer [7], [17]. This detector will be referred to as the “no suppression filter”.

On the other hand, for a particular choice of  $\alpha$  proportional to the occurrence frequency of the transient signal, the minimum power distortionless response (MPDR) beamformer is obtained [7]. This detector will be referred to as the “full suppression filter”.

The original optimization problem in Eq. 3 will be generalized in two ways:

1) Variable suppression matrix: Instead of either full suppression of the signal or no suppression at all, one can demand to suppress only specific shifts  $i$  of the waveform. In this case  $l$  is replaced by  $M \cdot l$ , where the suppression matrix  $M$  is a diagonal matrix with  $m_{i,i} = 1$  if the shift  $(f \star \xi)_i$  should be suppressed, and  $m_{i,i} = 0$  otherwise.

2) Variable target function: In the original optimization problem the response of the filter to the template had to be minimal, i.e. the least square distance to zero. Instead, one can minimize the distance to an arbitrary function  $s$ , which is expressed by the substitution of  $l$  with  $s - l$ .

Combining both variations 1) and 2) this leads to a modified optimization problem stated as

$$f = \underset{f}{\text{argmin}} \left\{ |s - M \cdot l|^2 + \alpha f^{\top} C f \right\} \quad \text{s.t. } f^{\top} \cdot \xi = 1 \quad (5)$$

The solution to this modified optimization problem can still be obtained analytically.

**Proposition 1.** *The solution to the optimization problem stated in Eq. 5 is given by*

$$f = \left( G^{-1} - \frac{G^{-1} \xi \xi^{\top} G^{-1}}{\xi^{\top} G^{-1} \xi} \right) \cdot (s \star \xi)_{[-b,b]} + \frac{G^{-1} \xi}{\xi^{\top} G^{-1} \xi} \quad (6)$$

where  $G := \tilde{\Xi} + \alpha C$ , and  $(\tilde{\Xi})_{k,l} := \sum_{\tau} m_{\tau,\tau}^2 \xi_{k+\tau} \xi_{l+\tau}$

The proof is given in Appendix A. If  $s = \mathbf{0}$  or  $s_i = \delta_0(i)$  the first term in Eq. 6 disappears. Furthermore, if the suppression matrix  $M$  is the identity matrix,  $M = \mathbf{1}$ , the original formula in Eq. 4 is obtained, whereas for  $M$  being the zero matrix,  $M = \mathbf{0}$ , the no suppression filter is obtained.

### C. Performance measure

The processing flow of a detector consists of two consecutive steps: filtering, and an application of a threshold  $\gamma$  to the filter output. Hence, it is desired that after these two steps, the underlying point process  $v_t$  in Eq. 2 is obtained. If one

<sup>2</sup>Note that in [16] the filters were obtained under the constraint  $f^{\top} C f = 1$  instead, however, in terms of detection performance the filters are equivalent. Also, we will still refer to this filter as the *adaptive* matched filter, even if the *exact* noise covariance matrix is known.

achieves the correct estimation of this point process, the signal has been detected and the arrival times retrieved successfully.

Since a signal consisting of a unique waveform without amplitude variations was assumed, one can restrict itself to the analysis of detection and arrival time estimation of the waveform itself. Therefore, the output of a perfect detector  $D$  must always be  $D(\boldsymbol{\xi} + \boldsymbol{\eta}) = \delta_0(x)$ . As such, the perfect detector reconstructs the original point process  $v_t$  for all possible thresholds. Hence, one would like to have a measure which indicates the closeness of a detector output to the  $\delta_0(x)$  function. In contrast, the classical performance measure, which is the probability of detection  $P_D$  (see e.g. [1]), only indicates whether the waveform was detected at all, but does not measure the closeness of the detection probability to the correct arrival time.

Based on these observations, the following measure of performance  $P_{DE}$  (for a fixed, but arbitrary threshold  $\gamma$ ) for combined detection and arrival time estimation is proposed:

$$P_{DE} := \frac{\text{Var}_{1/2}(\bar{p}(x) + \delta_0(x)) - \text{Var}_{1/2}(p(x) + \delta_0(x))}{\text{Var}_{1/2}(\bar{p}(x) + \delta_0(x))} \quad (7)$$

where  $\bar{p}(x)$  is a pdf for which  $\text{Var}_{1/2}(p(x) + \delta_0(x))$  is maximal, i.e.

$$\bar{p}(x) := \underset{p(x)}{\text{argmax}} \{ \text{Var}_{1/2}(p(x) + \delta_0(x)) \}.$$

$p(x)$  is a detector dependent pdf which is at each point  $x$  in time proportional to the probability that the filter output is above the threshold  $\gamma$ , i.e.

$$p(x) := \frac{P_D(x)}{\sum_x P_D(x)} \quad (8)$$

where  $P_D(x)$  is the classical probability of detection. In the case of linear filters, one has  $P_D(x) = \text{Prob}[(\mathbf{f} \star (\boldsymbol{\xi} + \boldsymbol{\eta}))_x \geq \gamma]$ .

Two important properties of  $P_{DE}$  are stated in the following propositions.

**Proposition 2.** *In the case of a discrete pdf defined on the interval  $[-a, a]$ ,  $P_{DE}$  is given by*

$$P_{DE} = 1 - \frac{2}{a^2} \text{Var}_{1/2}(p(x) + \delta_0(x)). \quad (9)$$

The proof is given in Appendix B. In contrast to Eq. 7, the expression in Eq. 9 no longer depends on the unknown quantity  $\bar{p}(x)$ , and thus, allows for calculation of the performance measure in real applications.

**Proposition 3.**  *$P_{DE}$  takes values in the interval  $[0, 1]$ . The maximal value of 1 is attained if and only if  $p(x) = \delta_0(x)$ .*

The proof is given in Appendix C. This last proposition establishes bounds on the range in which the values of  $P_{DE}$  fall. A value close to 1 indicates a good performance, whereas a value close to 0 indicates a poor performance of the detector. Moreover, it states that only the perfect detector can achieve the best possible performance.

As in the calculation of the quantity  $p(x)$  a normalization is involved in order to obtain a pdf (see Eq. 8), even a single small value exceeding the threshold will be normalized to a

pdf. If the threshold is increased towards infinity, the measure will indicate a better and better performance, although the real probability of detection will become arbitrarily small. Hence, in contrast to the classical measures, one has to restrict the range of possible thresholds. A reasonable choice is to set  $\gamma_{max} = \max_t \{(\mathbf{f} \star \boldsymbol{\xi})_t\}$ , and  $\gamma_{min} = E[\mathbf{f} \star \boldsymbol{\eta}]$ . The upper threshold is justified by the fact that in the noise-free case, a threshold greater than the maximal value of the filter response to the waveform would lead to zero detections. The lower bound of the threshold is also justified, since a threshold below the average response to a noise segment would always lead to detection of the signal, except when the detector is meaningless.

### III. RESULTS

#### A. Numerical evaluation

The measure in Eq. 9 indicates performance of a filter for one fixed (but arbitrary) threshold  $\gamma$ . In order to assign an overall performance to a detector, however, a total measure is needed. As such, slightly modified receiver operating characteristics (ROC) and the area under these ROC curves (AUC) were used [18]. The  $x$ -axis of the ROC curve corresponded to the probability of false alarm  $P_{FA}$  [1], i.e. the probability that a data segment containing only noise will be incorrectly detected as signal. Instead of  $P_D$ , the  $y$ -axis corresponded to the proposed  $P_{DE}$  measure. According to the properties of  $P_{DE}$  in Sec. II-C, a larger value of the AUC indicates a better performance of the corresponding filter.

In this evaluation setting, three different linear filters were compared, namely the no suppression filter, the full suppression filter and a particular case of the proposed filter class. The waveform of the signal had a length of  $T_\xi = 7$ , whereas the noise covariance matrix was set to  $\mathbf{C} = 0.025 \cdot \mathbf{1}$ , resulting in a SNR of 14.0 db.

In the case of zero mean Gaussian noise the probability of detection is given by the expression

$$P_D(x) = 0.5 \cdot \left( 1 - \text{erf} \left( \frac{\gamma - l(x)}{\sqrt{2 \cdot \mathbf{f}^\top \mathbf{C} \mathbf{f}}} \right) \right),$$

where  $\text{erf}(x)$  denotes the standard error function, and  $l(x) = (\mathbf{f} \star \boldsymbol{\xi})_x$ .  $P_{FA}$  is obtained by  $P_{FA} = P_D(l(x) = 0)$ .  $P_{DE}$  was then calculated according to Eq. 9 with  $a = 0.5 \cdot (T_{\mathbf{f} \star \boldsymbol{\xi}} - 1) = 6$ .

For a linear filter, the average response to zero mean noise is zero, i.e.  $E[\mathbf{f} \star \boldsymbol{\eta}] = 0$ . It turned out that for this particular evaluation setting one has  $\max_t \{(\mathbf{f} \star \boldsymbol{\xi})_t\} = 1$  for all considered filters. Hence, the threshold  $\gamma$  was varied in the interval  $[0, 1]$  (in steps of 0.002).

Recall, that the linear filters depended on the trade-off parameter  $\alpha$ , see Eq. 4 and Eq. 6. The AUC was computed for all  $\alpha$  values starting from  $\alpha = 0$  in steps of 0.005 up to a value for which the performance started to converge to the performance of the no suppression filter; see Sec. II-B for explanation. The results are shown in Fig. 1.

Although the filters attain their best performance at different  $\alpha$  values (see Fig. 1), the proposed filter, called partial suppression filter, achieved the highest AUC.

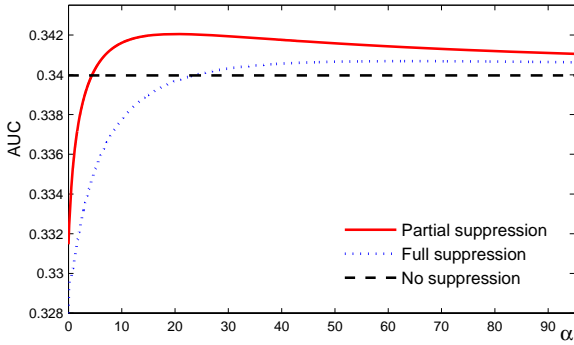


Fig. 1. AUC of various filters for different  $\alpha$  values. The partial suppression filter was calculated by Eq. 6 in which the target function was set to  $s_i = \delta_0(i)$ ,  $\forall i \in [-6, 6]$  and the diagonal matrix entries of  $\mathbf{M}$  were set such that  $m_{-1,-1} = m_{0,0} = m_{1,1} = 0$ , and  $m_{i,i} = 1 \forall i \notin \{-1, 0, 1\}$ . This proposed filter achieved the highest AUC score. The optimal performance for the partial suppression filter was achieved at  $\alpha = 20.220$ , for the full suppression filter at  $\alpha = 63.685$ , whereas, by construction, the no suppression filters had constant performance for all  $\alpha$  values.

### B. Simulations

The results from the previous section indicate that partial suppression filters are advantageous in comparison to the full and no suppression filters. To verify this result in a realistic setting, Monte Carlo simulations were performed. In particular, a single simulation consisted of a data stream containing 1000 signal segments and twice as many noise segments. The identical waveform and also the same noise statistics as the ones described in the previous section were used. The implementation was realized in MATLAB<sup>®</sup>.

For performance comparison the previously calculated filters were used, with the  $\alpha$  parameter set at specific values for which the respective filter achieved best performance (see Fig. 1).

As scope the area of realtime applications was chosen. In such a setting, at time  $t_0$  only data  $x_t$  from precedent times  $t \leq t_0$  are available. Nevertheless, the decision about signal presence has to be made already at time  $t_0$ . Consequently, every threshold crossing is immediately accounted for a signal presence, and every detection, which does not correspond to the exact signal arrival time, is counted as a false positive detection (FP). Accordingly, only successful detections at the exact arrival time of a signal are counted as true positive detections (TP). By varying the threshold (in steps of 0.0025) the corresponding ROC curves were obtained, see Fig. 2.

For the assessment of the overall performance, the AUC was computed and considered only up to the smallest (common for all filters) relative FP value for which  $\text{rel. TP} = 1$ , in order to avoid redundant computations. The AUCs of all filters averaged over 10 independent simulations are shown in Tab. I, and the variance across the simulations was of the order of  $10^{-7}$ .

The partial suppression filter achieved the best score, followed by the full suppression filter and lastly the no suppression filter. This is the same ranking as predicted in Sec. III-A.

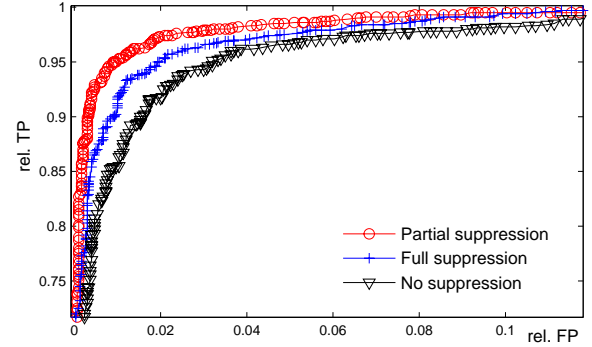


Fig. 2. ROC curves for various filters based on a single simulation. The number of true positive detections was normalized by the total number of signal segments, whereas the number of false positive detections was normalized by the total number of data segments. Clearly, the proposed partial suppression filter outperforms the other filters.

TABLE I  
AVERAGE AUC FOR VARIOUS FILTERS.

	Partial	Full	No
av. AUC	0.396487	0.395009	0.392696

### IV. DISCUSSION AND CONCLUSION

In contrast, the classical performance measure  $P_D/P_{FA}$  would not have predicted the correct ranking:  $P_D = \text{Prob}[\mathbf{f}^T \cdot (\boldsymbol{\xi} + \boldsymbol{\eta}) \geq \gamma]$  is largest for the no suppression filter and smallest for the full suppression filter (and vice-versa for  $P_{FA}$ ).

To sum up, a measure was proposed which assigns a performance to a detector with respect to simultaneous detection and arrival time estimation of transient signals. Although the proposed measure is general and suitable for most detectors, the detector class of linear filters is of particular interest. In the popular sub-class of minimal quadratic response filters, the existing filters were modified by introducing a suppression matrix and a target function. The proposed filters have the advantage of still being analytically computable, but offer more flexibility than the existing filters. The widely used adaptive matched filter, the Capone filter and the minimum power distortionless response beamformer are all particular realizations within the proposed filter class.

In fact, the target function can be used in order to adjust the smoothness of the filter response. This might be helpful in cases when the post processing consists not just of a pointwise thresholding, but of a more complex operation; e.g. when the data contains more than one signal source and a simultaneous detection and classification task has to be performed [19], [20].

On the other hand, the suppression matrix allows for the selective suppression of specific filter responses. This can be useful for incorporating prior knowledge about the signal into the filter design, as for example a refractory period or dead time.

Using the proposed measure, two existing filters (AMF/MVDR and MPDR) were compared with a particular filter of the just proposed filter class. The measure indicated

a favorable performance of the proposed filter, which was confirmed in simulations. In particular, the proposed filter was superior in a real-time detection and arrival time estimation task.

In the performed evaluation the target function and the suppression matrix were set manually. As an outlook for further investigations one might think of an online adaption scheme: The filtering is started with the classical adaptive matched filter, while in the background an optimization problem is solved, which aims at finding an optimal target function and suppression matrix. Once such a solution is found, the filter is adapted accordingly.

## APPENDIX

### A. Proof of proposition 1

The objective function of the optimization problem in Eq. 5 is convex and since the optimization constraint is linear, one can use the Lagrange multiplier method for solving it. The corresponding Lagrangian  $L$  is given by

$$L = |s - M \cdot l|^2 + \alpha \mathbf{f}^\top \mathbf{C} \mathbf{f} + \lambda (\mathbf{f}^\top \boldsymbol{\xi} - 1)$$

where  $\lambda$  is the Lagrange multiplier. The derivatives in respect to  $\mathbf{f}$  and  $\lambda$  can be calculated as

$$\begin{aligned} \frac{\delta L}{\delta f_{t_i}} &= 2 \left( \sum_{\tau} m_{\tau, \tau} \xi_{t_i + \tau} \sum_t m_{\tau, \tau} \xi_{t + \tau} f_t - s_{\tau} m_{\tau, \tau} \xi_{t_i + \tau} \right) \\ &\quad + 2\alpha \sum_t \mathbf{C}_{t_i, t} f_t + \lambda \xi_{t_i} \\ \frac{\delta L}{\delta \lambda} &= \left( \sum_t \xi_t f_t \right) - 1 \end{aligned}$$

The calculation of the second derivatives leads to

$$\begin{aligned} \frac{\delta L}{\delta f_{t_i} \delta f_{t_I}} &= 2 \sum_{\tau} m_{\tau, \tau}^2 \xi_{t_i + \tau} \xi_{t_I + \tau}^j + 2 \cdot \alpha \cdot \mathbf{C}_{t_i, t_I} \\ \frac{\delta L}{\delta \lambda \delta \lambda} &= 0 \quad \frac{\delta L}{\delta f_{t_i} \delta \lambda} = \xi_{t_i} \end{aligned}$$

The second derivatives of  $L$  are independent of  $\mathbf{f}$  and of  $\lambda$ . Therefore, the Taylor expansion of the first derivative of  $L$  around zero consists only of two terms and the solution can be obtained by solving

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} \frac{\delta L}{\delta f_{t_i}} \\ \frac{\delta L}{\delta \lambda} \end{pmatrix}_{f_{t_i}=0, \lambda=0} \\ &+ \begin{pmatrix} \sum_{t_I} \frac{\delta L}{\delta f_{t_i} \delta f_{t_I}}_{f_{t_i}=0, \lambda=0} \cdot f_{t_I} \\ \sum_{t_I} \frac{\delta L}{\delta \lambda \delta f_{t_I}}_{f_{t_i}=0, \lambda=0} \cdot f_{t_I} \end{pmatrix} + \begin{pmatrix} \frac{\delta L}{\delta f_{t_i} \delta \lambda}_{f_{t_i}=0, \lambda=0} \cdot \lambda \\ \frac{\delta L}{\delta \lambda \delta \lambda}_{f_{t_i}=0, \lambda=0} \cdot \lambda \end{pmatrix} \end{aligned}$$

In matrix notation, the above equation becomes

$$\mathbf{0} = \begin{pmatrix} -2(\mathbf{s} \star \boldsymbol{\xi})_{[-b, b]} \\ -1 \end{pmatrix} + \tilde{\mathbf{H}} \cdot \begin{pmatrix} \mathbf{f} \\ \lambda \end{pmatrix} \quad (10)$$

where one defined

$$(\mathbf{s} \star \boldsymbol{\xi})_{[-b, b]} := ((\mathbf{s} \star \boldsymbol{\xi})_{-b}, \dots, (\mathbf{s} \star \boldsymbol{\xi})_b)^\top$$

and

$$\tilde{\mathbf{H}} := \begin{pmatrix} 2(\tilde{\boldsymbol{\Xi}} + \alpha \cdot \mathbf{C}), & \boldsymbol{\xi} \\ \boldsymbol{\xi}^\top, & 0 \end{pmatrix}$$

and  $(\tilde{\boldsymbol{\Xi}})_{k, l} := \sum_{\tau} m_{\tau, \tau}^2 \xi_{k + \tau} \xi_{l + \tau}$ .

Define  $\mathbf{G} := \tilde{\boldsymbol{\Xi}} + \alpha \mathbf{C}$ . The inverse of  $\tilde{\mathbf{H}}$  is then given by [21]

$$\tilde{\mathbf{H}}^{-1} = \begin{pmatrix} \frac{1}{2} \left( \mathbf{G}^{-1} - \frac{\mathbf{G}^{-1} \boldsymbol{\xi} \boldsymbol{\xi}^\top \mathbf{G}^{-1}}{\boldsymbol{\xi}^\top \mathbf{G}^{-1} \boldsymbol{\xi}} \right) & \frac{\mathbf{G}^{-1} \boldsymbol{\xi}}{\boldsymbol{\xi}^\top \mathbf{G}^{-1} \boldsymbol{\xi}} \\ \frac{\boldsymbol{\xi}^\top \mathbf{G}^{-1}}{\boldsymbol{\xi}^\top \mathbf{G}^{-1} \boldsymbol{\xi}} & -\frac{2}{\boldsymbol{\xi}^\top \mathbf{G}^{-1} \boldsymbol{\xi}} \end{pmatrix}$$

The left multiplication of Eq. 10 with  $\tilde{H}^{-1}$  yields the solution for  $\mathbf{f}$ , which is given by

$$\mathbf{f} = \left( \mathbf{G}^{-1} - \frac{\mathbf{G}^{-1} \boldsymbol{\xi} \boldsymbol{\xi}^{\top} \mathbf{G}^{-1}}{\boldsymbol{\xi}^{\top} \mathbf{G}^{-1} \boldsymbol{\xi}} \right) (\mathbf{s} * \boldsymbol{\xi})_{[-b, b]} + \frac{\mathbf{G}^{-1} \boldsymbol{\xi}}{\boldsymbol{\xi}^{\top} \mathbf{G}^{-1} \boldsymbol{\xi}}$$

□

### B. Proof of proposition 2

If one can show that  $\text{Var}_{1/2(\bar{p}(x)+\delta_0(x))} = a^2/2$ , the proposition simply follows from Eq. 7. It is

$$\text{Var}_{1/2(p(x)+\delta_0(x))} = \text{Var}_{1/2p(x)} + \text{Var}_{1/2\delta_0(x)} = \text{Var}_{1/2p(x)}.$$

Strictly speaking  $1/2p(x)$  is not a pdf, but  $\text{Var}_{1/2p(x)}$  is still defined as in Eq. 1.

It is

$$\begin{aligned} \text{Var}_{1/2p(x)} &= \sum_x x^2 1/2p(x) - \left( \sum_x x 1/2p(x) \right)^2 \\ &\leq 1/2 \sum_x x^2 p(x) = 1/2 \text{Var}_{q(x)}, \end{aligned}$$

where  $q$  is a pdf of a discrete random variable with zero mean. The variance of any pdf  $q(x)$  on the interval  $[-a, a]$  is bounded by  $a^2$  [22]. Hence,  $\text{Var}_{1/2p(x)} \leq a^2/2$ .

Now, one can show that this upper bound is attained. Define  $\bar{p}(x) = 1/2(\delta_{-a}(x) + \delta_a(x))$ . Then, a straightforward calculation yields  $\text{Var}_{1/2(\bar{p}(x)+\delta_0(x))} = a^2/2$ . □

### C. Proof of proposition 3

It was already shown in the proof of proposition 2 that the lower bound is attained. It remains to show that the upper bound is attained, i.e. that  $\text{Var}_{1/2(p(x)+\delta_0(x))} = 0 \Leftrightarrow p(x) = \delta_0(x)$ .

“ $\Leftarrow$ ”:  $\text{Var}_{\delta_0(x)} = 0^2 \cdot 1 - (0 \cdot 1)^2 = 0$ .

“ $\Rightarrow$ ”: Let  $q(x)$  be an arbitrary pdf. Without loss of generality, one can assume that  $q(x)$  is zero mean. Hence,  $0 = \text{Var}_{q(x)} = \sum_x x^2 q(x)$ , so  $q(x) = 0 \forall x \neq 0$ . Since it must be that  $\sum_x q(x) = 1$ , it follows that  $q(0) = 1$ , i.e.  $q(x) = \delta_0(x)$ . By plugging in  $q(x) = 1/2(p(x) + \delta_0(x)) = \delta_0(x)$ , it follows that  $p(x) = \delta_0(x)$ . □

### ACKNOWLEDGMENT

This research was supported by the Federal Ministry of Education and Research (BMBF) with the grants 01GQ0743 and 01GQ0410.

### REFERENCES

- [1] S. Z. Kalson, “An adaptive array detector with mismatched signal rejection,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, pp. 195–207, 1992.
- [2] W. Melvin, “A stap overview,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 19–35, 2004.
- [3] D. Middleton and R. Esposito, “Simultaneous optimum detection and estimation of signals in noise,” *IEEE Transactions on Information Theory*, vol. 14, 1968.
- [4] B. Baygun and A. HeroIII, “Optimal simultaneous detection and estimation under a false alarm constraint,” *IEEE Transaction on Information Theory*, vol. 41 (3), pp. 688 – 703, 1995.
- [5] E. Fishler and H. Messer, “Detection and parameter estimation of a transient signal using order statistics,” *IEEE Transactions on Signal Processing*, vol. 48, 2000.
- [6] F. Nicolls, “Constraints and invariance in target detection,” Ph.D. dissertation, University of Cape Town, 2000.
- [7] H. L. V. Trees, *Detection, Estimation, and Modulation Theory Part IV - Optimum Array Processing*. JOHN WILEY & SONS, 2002.
- [8] B. Friedlander and B. Porat, “Performance analysis of transient detectors based on linear data transformations,” *IEEE Transactions on Information Theory*, vol. 38 (2), pp. 665–673, 1992.
- [9] B. Porat and B. Friedlander, “Performance analysis of a class of transient detection algorithms - a unified framework,” *IEEE Transactions on Signal Processing*, vol. 40 (10), 1992.
- [10] Z. Wang and P. Willett, “A performance study of some transient detectors,” *IEEE Transactions on Signal Processing*, vol. 48 (9), 2000.
- [11] A. Yamazaki, T. Tajima, and K. Matsuoka, “Convolutional independent component analysis of eeg data,” *SICE 2003 Annual Conference*, vol. 2, pp. 1227– 1231, 2003.
- [12] P. H. Thakur, H. Lu, S. S. Hsiao, and K. O. Johnson, “Automated optimal detection and classification of neural action potentials in extra-cellular recordings,” *Journal of Neuroscience Methods*, vol. 162, no. 1-2, pp. 364–376, 2007.
- [13] M. Natora, F. Franke, M. Munk, and K. Obermayer, “Blind source separation of sparse overcomplete mixtures and application to neural recordings,” *Lecture Notes in Computer Science - Independent Component Analysis and Signal Separation*, vol. 5441, pp. 459–466, 2009.
- [14] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, *A Survey of Convolutional Blind Source Separation Methods*, ser. Springer Handbook of Speech Processing. Springer Press, Nov 2007.
- [15] D. Scholnik, “Mixed-norm fir filter optimization using second-order cone programming,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, pp. 1525–1528, 2002.
- [16] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, “A cfar adaptive matched filter detector,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28 (1), pp. 208–216, 1992.
- [17] O. Besson and F. Vincent, “Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors,” *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 452–459, 2005.
- [18] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27 (8), pp. 861–874, 2006.
- [19] R. Vollgraf and K. Obermayer, “Improved optimal linear filters for the discrimination of multichannel waveform templates for spike-sorting applications,” *IEEE Signal Processing Letters*, vol. 13, no. 3, pp. 121–124, 2006.
- [20] F. Franke, M. Natora, C. Boucsein, M. Munk, and K. Obermayer, “An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes,” *Journal of Computational Neuroscience*, 2009, in press.
- [21] H. Luetkepohl, *Handbook of Matrices*. John Wiley and Sons, 1996.
- [22] M. Ageel, “Variance upper bounds and a probability inequality for discrete alpha-unimodality,” *Aplicaciones Mathematicae*, vol. 27, pp. 403 – 410, 2000.