QRD–based optimal filtering for acoustic noise reduction

Geert Rombouts, Marc Moonen
KULeuven/ESAT–SISTA
Kasteelpark Arenberg10
3001 Heverlee
Belgium
geert.rombouts@esat.kuleuven.ac.be, marc.moonen@esat.kuleuven.ac.be

Abstract

We present a new QRD-based technique for optimal multichannel filtering and its application to multi–channel acoustic noise reduction. A recursively calculated adaptive filter optimally estimates the speech component in a noisy signal. The complexity of this algorithm is an order of magnitude than that of existing optimal filtering based algorithms which are mainly based upon SVD–decompositions, while performance is kept at the same level.

1 Introduction

In teleconferencing, hands–free telephony or voice controlled systems, a microphone array may be used instead of a single microphone in an attempt to reduce unwanted disturbances (e.g. car noise, computer noise, background speakers) based on spatio–temporal filtering. A typical setup is shown schematically in **Figure 1** for an array with 4 microphones. We aim at designing an optimal filter which will use all available information (e.g. reflections) in order to optimally reconstruct the signal of interest.

In [1] [2] the optimal filtering problem was solved by means of a GSVD (Generalised Singular Value Decomposition)–approach. In this framework, a filter which estimates the signal of interest from the microphone signals can be chosen from a set of such filters after computing the GSVD.

In this paper we will describe a QRD–based optimal filtering approach to the problem. The QRD–decomposition is already less complex than the SVD as such, and one can make the selection of the filter that will be calculated in advance, introducing an extra cost reduction. The complexity of this algorithm is an order of magnitude lower than that of existing optimal filtering based algorithms which are based upon SVD–decompositions, while performance is kept at the same level.

The paper is organized as follows. In sections 2 and 3 we will introduce optimal filtering based noise reduction. Then we describe how noise reduction can be implemented by means of QRD–based filtering in section 4.

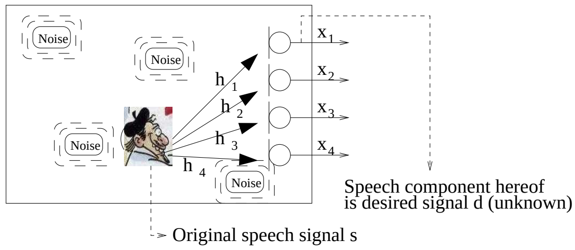

Figure 1: Adaptive optimal filtering in the acoustic noise reduction context.

Then the complexity figures and the simulation results are given in sections 5 and 6.

2 Optimal filtering based noise reduction

The speech component in the i'th microphone at time k is $d_i(k) = h_i(k) \bigotimes s(k) \quad i = 1 \ldots M$, where M is the number of microphones, $s(k)$ is the speech signal and $h_i(k)$ represents the room response path from the speech source to microphone i. The i'th microphone signal is $x_i(k) = d_i(k) + v_i(k) \quad i = 1 \ldots M$, where $v_i(k)$ is the noise component (sum of the contributions of all noise sources at microphone i). We define the input vector

$$\mathbf{x}(k) = \begin{pmatrix} \mathbf{x}_1(k) \\ \mathbf{x}_2(k) \\ \vdots \\ \mathbf{x}_M(k) \end{pmatrix} \quad \mathbf{x}_1(k) = \begin{pmatrix} x_1(k) \\ x_1(k-1) \\ \vdots \\ x_1(k-N) \end{pmatrix}$$

The noise input vector $\mathbf{v}(k)$ and the desired signal vector $\mathbf{d}(k)$ are defined in a similar way with $\mathbf{x}(k) = \mathbf{d}(k) + \mathbf{v}(k)$. The following standard assumptions are made :

- The noise signal is uncorrelated with the speech signal. This results in

$$\begin{aligned} \varepsilon\{\mathbf{x}\mathbf{x}^T\} &= \varepsilon\{\mathbf{v}\mathbf{v}^T\} + \underbrace{\varepsilon\{\text{cross terms}\}}_{=0} + \varepsilon\{\mathbf{d}\mathbf{d}^T\} \\ &= \varepsilon\{\mathbf{v}\mathbf{v}^T\} + \varepsilon\{\mathbf{d}\mathbf{d}^T\} \\ \varepsilon\{\mathbf{d}\mathbf{d}^T\} &= \varepsilon\{\mathbf{x}\mathbf{x}^T\} - \varepsilon\{\mathbf{v}\mathbf{v}^T\} \end{aligned}$$

Here $\varepsilon\{\cdot\}$ is the expectation operator.

- The noise signal is stationary as compared to the speech signal (by which we mean that its statistics change slower). This assumption allows us to estimate $\varepsilon\{\mathbf{v}\mathbf{v}^T\}$ during periods in which only noise is present.

We will use the speech component in one (or each) of the microphone signals as the desired signal input of the adaptive filter, i.e. $d_i(k)$ or $\mathbf{d}(k)$. This signal is obviously unknown . We can then write the Wiener solution for the optimal filtering problem with $\mathbf{x}$ the filter input and $\mathbf{d}$ the (unknown) desired filter output as

$$\begin{aligned} W_{\text{wf}} &= (\varepsilon\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\})^{-1}\varepsilon\{\mathbf{x}\mathbf{d}^T\} \qquad (1)\\ &= (\varepsilon\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\})^{-1}\varepsilon\{(\mathbf{d}+\mathbf{v})\mathbf{d}^T\}\\ &= (\varepsilon\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\})^{-1}\varepsilon\{\mathbf{d}\mathbf{d}^T\}\\ &= (\varepsilon\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\})^{-1}(\varepsilon\{\mathbf{x}\mathbf{x}^T\}-\varepsilon\{\mathbf{v}\mathbf{v}^T\}) \end{aligned}$$

Here, $\varepsilon\{\mathbf{x}\mathbf{x}^T\}$ is estimated from speech+noise–data, while $\varepsilon\{\mathbf{v}\mathbf{v}^T\}$ is estimated from noise–only data (during speech pauses). This W_{wf} is a matrix of which each column is an optimal MN–taps filter estimating the corresponding component of $\mathbf{d}$. One of these columns can then be chosen (arbitrarily) to optimally estimate the speech part in the corresponding entry of $\mathbf{x}^T(k)$, i.e. filter out the noise in one specific microphone signal.

3 Data driven approach

A data–driven approach will be based on data matrices $X(k)$ and $V(k)$, defined as

$$X(k) = \begin{pmatrix} \mathbf{x}^T(k) \\ \lambda_x \mathbf{x}^T(k-1) \\ \lambda_x^2 \mathbf{x}^T(k-2) \\ \vdots \end{pmatrix} V(k) = \begin{pmatrix} \mathbf{v}^T(k) \\ \lambda_v \mathbf{v}^T(k-1) \\ \lambda_v^2 \mathbf{v}^T(k-2) \\ \vdots \end{pmatrix}$$

We aim at tracking any changes in the environment by introducing a weighting scheme in order to reduce the impact of the contributions from the remote past. Let λ_x denote the forgetting factor for the speech+noise data, which can be different from λ_v, the forgetting factor for the noise only data. Since the noise is assumed to be stationary as compared to the speech one could choose $0 \ll \lambda_x < \lambda_v < 1$. We want $X^T(k)X(k)$ to be an estimate of $\varepsilon\{\mathbf{x}(k)\mathbf{x}^T(k)\}$. This is realised by

$$X^T(k+1)X(k+1) = \lambda_x^2 X^T(k)X(k) + (1-\lambda_x^2)\mathbf{x}(k+1)\mathbf{x}^T(k+1)$$

The same goes for the noise correlation matrix estimate

$$V^T(k+1)V(k+1) = \lambda_n^2 V^T(k)V(k) + (1-\lambda_n^2)\mathbf{v}(k+1)\mathbf{v}^T(k+1) \qquad (2)$$

Using the QR–decomposition [3] $X(k) = Q(k)R(k)$ with $Q(k)$ orthogonal and $R(k)$ upper triangular, hence $R^TR = X^TX$, the Wiener–solution is now estimated as

$$R^T(k)R(k)W(k) = R^T(k)R(k) - \underbrace{V^T(k)V(k)}_{\equiv P(k)}$$

$$W(k) = I - R^{-1}(k)R^{-T}(k)P(k)$$

Where I is the identity matrix. Due to the second assumption (the noise being stationary), $P(k)$ can be kept fixed during *speech + noise* periods and updated during *noise only* periods. $R^T(k)R(k)$ is fixed during *noise only* periods and can be updated [3] during *speech+noise* periods.

4 QRD–based realisation

Note that

$$W^N(k) = I - W(k) = R^{-1}(k)\underbrace{R^{-T}(k)P(k)}_{\equiv B(k)} \qquad (3)$$

defines a set of filters that optimally estimate the noise components in the microphone signals. The only storage required for the computation of $W^N(k)$ or $W(k)$ will be the matrix $R(k) \in \Re^{MN\times MN}$ and for $B(k) \in \Re^{MN\times MN}$. In fact, only one column of $B(k)$ has to be stored and updated, thus providing a signal or noise estimate for the corresponding component of $\mathbf{d}$. We will distinguish between a "speech+noise"– mode and an "noise–only"–mode.

4.1 Speech+noise – mode

Whenever a signal segment is identified as a speech+noise–segment, P(k) is not updated (second assumption), but $E\{X^T(k)X(k)\}$ needs to be updated. As we do not store $X(k)$ but $R(k)$ instead , the update formula for $R(k)$ is the standard QR–updating formula

$$\begin{pmatrix} 0 \\ R(k+1) \end{pmatrix} = \overline{Q}^T(k+1)\begin{pmatrix} \tilde{\mathbf{x}}^T(k) \\ \lambda_x R(k) \end{pmatrix}$$

with

$$\tilde{\mathbf{x}}(k) = \sqrt{1-\lambda_x^2}\,\mathbf{x}(k) \qquad (4)$$

$\overline{Q}^T(k+1)$ is orthogonal, consisting of a series of rotations over angles $\theta(k)$ and $R(k+1)$ is again upper triangular[1]. Updating $R(k)$ also implies a change in $B(k) = R^{-T}(k)P(k)$. In order to derive this update, we need an expression for the update of $R^{-1}(k)$. It is well known [4] that

$$\begin{pmatrix} * \\ R^{-T}(k+1) \end{pmatrix} = \overline{Q}^T(k+1)\begin{pmatrix} 0 \\ \frac{1}{\lambda}R^{-T}(k) \end{pmatrix}$$

Hence we have

$$\begin{pmatrix} * \\ B(k+1) \end{pmatrix} = \overline{Q}^T(k+1)\begin{pmatrix} 0 \\ \frac{1}{\lambda_x}B(k) \end{pmatrix}$$

[1] $Q(k) = \overline{Q}(k)\overline{Q}(k-1)...\overline{Q}(0)$ does not need to be stored.

The complete update can then be written as one single matrix update equation :

$$\begin{pmatrix} 0 & \mathbf{r}^T(k+1) \\ R(k+1) & B(k+1) \end{pmatrix} = \tag{5}$$
$$\overline{Q}^T(k+1) \begin{pmatrix} \tilde{\mathbf{x}}^T(k+1) & 0 \\ \lambda_x R(k) & \frac{1}{\lambda_x} B(k) \end{pmatrix}$$

The updated least squares solution can now be computed by backsubstitution, see formula 3.

4.2 Noise only–mode.

In the noise–only case, one needs to update $B(k) = R^{-T}(k)P(k) = R^{-T}(k)V^T(k)V(k)$, while $R(k)$ is obviously kept fixed. Using equation (2), we define $\tilde{\mathbf{v}}(k) = \sqrt{1-\lambda_n^2}\mathbf{v}(k)$. From equation (2) and the fact that in noise–only mode $R(k+1) = R(k)$, we find that

$$B(k+1) = \lambda_n^2 B(k) + (R^{-T}(k+1)\tilde{\mathbf{v}}(k+1))\tilde{\mathbf{v}}^{\mathbf{T}}(k+1)\})$$

Given $R(k+1)$, we can compute $(R^{-T}(k+1)\tilde{\mathbf{v}}(k+1))$ by a backsubstitution. By using an intermediate vector $\mathbf{u}(k+1)$: $R^T(k+1)\mathbf{u}(k+1) = \tilde{\mathbf{v}}(k+1)$. The simple multiplication $\mathbf{u}(k+1)\tilde{\mathbf{v}}^T(k+1)$ gives the update for all columns of $B(k+1)$, $B(k+1) = \lambda_n^2 B(k) + \mathbf{u}(k+1)\tilde{\mathbf{v}}^T(k+1)$. Again, the updated least squares solution can be computed by backsubstitution.

4.3 Additional assumption

For multi–channel noise reduction, the forgetting factor λ_s should be chosen close enough to 1 for two reasons. First, if the window is taken too short, the system of equations which is solved by the RLS–algorithm will become ill–conditioned because the input signal (which is mostly speech) may not be persistantly exciting. Second, by imposing a large window, one avoids that the system performs a signal modelling, instead of modelling the spatial characteristics of the desired signal.

It is found that better noise reduction results can be obtained if we update $R(k)$ also during noise–only mode (with the same scheme as in section 4.1). For many types of noise, λ_n can be chosen smaller ($\lambda_n = 0.9997$ for 8000 Hz sampling rate) so that convergence is very fast. Experiments show that this approach delivers superior results concerning noise reduction, although in the beginning of a speech period, the signal sounds a bit 'muffled'.

4.4 Residual extraction

It has been shown in [5] that for a QR–updating scheme with a right hand side $Z(k)$ and an update of this right hand side with $\mathbf{d}(k+1)$, i.e.

$$\begin{pmatrix} 1 & 0 \\ 0 & Q(k) \end{pmatrix} \begin{pmatrix} \mathbf{x}^T(k+1) & \mathbf{d}^T(k+1) \\ R(k) & Z(k) \end{pmatrix} =$$
$$\underbrace{\begin{pmatrix} 1 & 0 \\ 0 & Q(k) \end{pmatrix} \overline{Q}(k+1)}_{[*|Q(k+1)]} \begin{pmatrix} 0 & \mathbf{r}^T(k+1) \\ R(k+1) & Z(k+1) \end{pmatrix}$$

one can write

$$\mathbf{d}^T(k+1) - \mathbf{x}^T(k+1) \underbrace{W(k+1)}_{R^{-1}(k+1)Z(k+1)} = \mathbf{r}^T(k+1) \prod_{i=1}^{MN} \cos\theta_i(k+1) \tag{6}$$

Where $\theta_i(k+1)$ are the angles of all the rotations that $\overline{Q}(k+1)$ consists of. This means that we can extract the least squares residuals without having to calculate the filter coefficients $W(k+1)$ first, which is remarkable. From [5], (4) and (5) it can be shown that this can be applied to our scheme too, since ($\mathbf{d}(k) = 0$ and $Z(k) = B(k)$)

$$\mathbf{y}_n^T(k+1) = 0 - X^T(k+1)W^N(k+1)\sqrt{1-\lambda_x^2}$$

The (speech) signal estimate during speech+noise–periods is obtained by subtracting $y_n(k+1)$ from the input signal $x(k+1)$. If we want to generate an output signal in noise–only mode too, we can execute a residual extraction procedure as in the speech+noise–mode, be it *without* updating the $R(k)$ and $B(k)$ ('frozen mode'). This will of course increase the complexity in noise–only mode, but since the updates need not to be calculated (only the rotation parameters $\overline{Q}(k)$ and the outputs), the extra complexity will be about half the complexity in speech+noise–mode. The complexity in noise–only mode will become roughly equal to the complexity in speech+noise mode.

5 Complexity

The SVD–based optimal filtering approach has a complexity of $O(M^3N^3)$ where M is the number of microphones and N is the number of filter taps per microphone channel. A reduced complexity approximation is possible for the GSVD–approach (based on SVD–tracking), leading to $O(27.5M^2N^2)$ [2] . These figures can be compared to the complexity of this algorithm : in noise-only mode, the complexity is $(MN)^2 + 3MN + M$ flops per sample (if no output signal is generated during noise periods). In speech+noise mode, the number of flops per sample is $3.5(MN)^2 + 15.5MN + M + 2$. In these calculations, one flop is one addition *or* one multiplication. These figures apply when only one filter output is calculated. For a typical setting of $N = 20$ and $M = 5$, we would obtain 36 557 flops per sample for the QR–based method as compared to 275 000 flops per sample for the GSVD–based method, which amounts to a 7 to 8–fold complexity reduction.

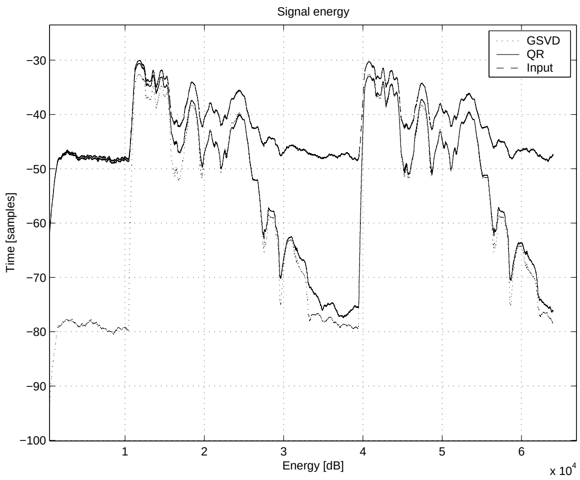

Figure 2: Performance comparison of GSVD-based optimal filtering (dotted) and QR–based optimal filtering (full line) versus the original signal (upper signal in the graph). The performance is roughly equal for both approaches.

6 Simulation results

Figure 2 compares the QRD–based optimal filtering method to the GSVD–based optimal filtering method. It shows that the performance of both methods is comparable. Here the QRD–method *does* generate an output when in 'noise only'–mode.

7 Algorithm description

In this algorithm, we choose only the first column $z(k)$ out of the matrix $B(k)$ described above. It corresponds to the autocorrelation vector of the first input channel. The newest sample received in this channel is called $u_1(k)$ in the algorithm description.

```
initialize R with unity matrix
loop (sample by sample):
  form new input vector u
  if (signal+noise)
    u *= $\sqrt{(1-\lambda_x^2)}$
    R-update(input=u, weighting=$\lambda_x$)
    z-update(input=0, weighting=1/$\lambda_x$)
      (use rotations from R-update)
      (gives residual)
    output ($u_1(k)$+residual)/$\sqrt{(1-\lambda_x^2)}$
  else
    u *= $\sqrt{(1-\lambda_n^2)}$
    optionally update R(u, $\lambda_x$)
                      z(0, 1/$\lambda_x$)
     (see section 4.3)
    b=backsubstitution($R^T$, u)
    z *= $\lambda_n^2$
    z += b*$u_1(k)$*$\sqrt{(1-\lambda_n^2)}$
```

8 Conclusion

We have derived a new recursive algorithm for optimal multichannel filtering with an unknown desired signal, applied to adaptive acoustic noise suppression. Experiments show that the performance is equal to similar SVD–based optimal filtering algorithms from the literature, but that the complexity of the QRD–based optimal filtering technique is significantly lower. If speech statistics information is allowed to be 'forgotten' during noise–only periods, the noise suppresion performance even increases drastically while the speech signal is not too much distorted.

9 Acknowledgements

Geert Rombouts is a Research Assistant with the I.W.T (Flemish Institute for Scientific and Technological Research in Industry). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme - IUAP P4-02 (1997-2001): Modeling, Identification, Simulation and Control of Complex Systems, the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology)of the Flemish Government, IWT project : MUSETTE-II ('Multimicrophone Signal Enhancement Techniques for handsfree telephony and voice controlled systems'). The scientific responsibility is assumed by its authors.

References

[1] S. Doclo and M. Moonen. Svd–based optimal filtering with applications to noise reduction in speech signals. In *Proc. of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99), New Paltz NY, USA,*, pages 143–146. IEEE, Oct 1999.

[2] S. Doclo and M. Moonen. Noise reduction in multimicrophone speech signals using recursive and approximate gsvd–based optimal filtering. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS-2000), Hilvarenbeek, The Netherlands*, March 2000.

[3] Golub and Van Loan. *Matrix Computations*, chapter 12. Johns Hopkins, 1996.

[4] F. Ling M. Moonen J. G. Proakis, C. M. Rader and C. L. Nikias I. K. Proudler. *Algorithms for Statistical Signal Processing.* Prentice–Hall, ISBN: 0-13-062219-2, 1/e edition, 2002.

[5] J. G. McWhirter. Recursive least squares minimisation using a systolic array. In *Proc. SPIE Real Time Signal Processing IV*, volume 431, pages 105–112, 1983.