A

Experimental Setups

There is still no universal agreement on the optimum placement of the microphones for the speech input in cars. Our experiments have been carried out in a Mercedes S320 vehicle with two different microphone arrays. An overview is shown in Fig. A.1.

A.1 The Four-Element Compact Array Mounted in the Rear-View Mirror

This four-element mirror array is mounted in the rear-view mirror. AKG directional microphones of type cardioid are used. This arrangement is known to be compatible with product design constraints, since it belongs to the standard speech input equipment in the Mercedes E class. The distance from the driver mouth to the mirror depends on the position of the driver seat. This distance can amount between 50 and 80 cm, a typical mouth–microphone distance being about 60 cm. The experimental setup with the driver and the codriver is depicted in Fig. A.2.

A.2 The Two-Element Distributed Array Mounted on the Car Ceiling

The two-element distributed array consists of directional microphones oriented to the driver and the codriver, respectively. The two microphones are placed on roof control panel with 17-cm spacing. In contrast to the four-element compact array mounted in the rear-view mirror, this microphone arrangement provides the interferer reference signal directly. The experimental setup is depicted in Fig. A.3. For this setup, we use PEIKER directional microphones of type cardioid.
Note that it is straightforward to extend this two-element array to a four-element array by placing directional microphone on the roof, close to each of the two back-seat passengers. This is of particular interest when separating the driver speech from the interfering back-seat left passenger, since the four-element compact array mounted in the rear-view mirror may fail in this case.
A.3 Acoustic Characteristics of the Car Cabin

Room impulse response

The car cabin is usually regarded as weakly reverberant. The reverberation time, which is denoted by $T_{60}$ and is defined as the time which is necessary for the sound energy to decrease by 60 dB [40], is about $T_{60} = 50$ ms.

Estimating the impulse response between the driver mouth and a given microphone can be done using an artificial head. To avoid the influence of the loudspeaker, we use a close-talk microphone mounted on the artificial head. Figure A.4 shows the impulse response between this close-talk microphone and the microphone $x_1$ of the four-element compact array mounted in the rear-view mirror.

Background noise

Car noise consists of motor noise and also results from the wind and from the contact between the tires and the road. It can be regarded as diffuse and has its main energy in low frequency bands [14]. The PSD of road noise that was recorded at 100 km h$^{-1}$ is shown in Fig. A.5.

A.4 Illustration of the Difficulty in the Design of a Reliable DTD

An experiment is conducted in the car interior with two male speakers recorded with a four-element microphone array mounted in the rear-view

---

Fig. A.3. Experimental setup for the two-element distributed array mounted on the car ceiling. The microphones are mounted on the roof control pad.

\[ \text{driver (target)} \]

\[ \text{codriver (interferer)} \]

\[ \begin{align*}
40 \text{ cm} & \\
17 \text{ cm} & \\
40 \text{ cm} & \\
\end{align*} \]
Fig. A.4. Impulse response estimated between a close-talk microphone mounted on the artificial head on the driver seat and the microphone $x_1$ of the four-element compact array mounted in the rear-view mirror.

Fig. A.5. PSD of a typical road noise recorded at 100 km h$^{-1}$. The PSD is estimated using the Welch periodogram averaged over 160,000 samples (FFT length $N_{FFT}=1024$, sample frequency $f_s=16,000$ Hz).

The speakers are positioned on the driver and codriver seats, respectively. We observe the power of the source signals, which indicates at what time each speaker is active. On the other hand, we examine the input SIR estimation:

$$SIR_{est}(p) = \frac{(M - 1)x_0^2(p)}{\sum_{m'=0}^{M-1} x_{B,m'}^2(p)}$$  \hspace{1cm} (A.1)

\footnote{We refer to Appendix A for a detailed description of the experimental setup, which is depicted in Figs. A.1 and A.2.}
We are interested in finding a threshold $\text{SIR}_{\text{th}}$ such that (1) $\text{SIR}_{\text{th}} < \text{SIR}_{\text{est}, \text{in}}(p)$ during target silences and interferer activity and (2) $\text{SIR}_{\text{th}} > \text{SIR}_{\text{est}, \text{in}}(p)$ during target activity. As can be see from Fig. A.6, such a threshold does not exist in general: The estimate $\text{SIR}_{\text{est}, \text{in}}(p)$ can reach similar levels during target and interferer activity. Hence, a trade-off must be found between

- robustness: choosing a low, conservative threshold that prevents target cancelation and
- accuracy in the detection of interferer activity: choosing a higher threshold that allows adaptation more often.

Obviously, this issue worsens with increasing overlap of the target and interferer speech. This simple example has illustrated the difficulties that arise with the design of a reliable DTD.
Far- and Free-Field Acoustic Propagation Model and Null Beamforming

B.1 Far- and Free-Field Model

We consider a source at position \((r, \phi, \theta)\) in spherical coordinates. Without loss of generality, the reference is taken at the position of the microphone \(x_1\). Next, we assume a uniform linear array (ULA) with intermicrophone spacing \(\Delta\) arranged along the \(z\)-axis as shown in Fig. B.1. Then the position of the source relative to the array is invariant by rotational symmetry of angle \(\theta\). Let us denote the distance from the source \(s\) to the second microphone \(x_2\) by \(r_2\). We have

\[
 r_2^2 = r^2 + \Delta^2 + 2\Delta r \sin \theta. 
\]  

(B.1)

We next compute the length \(r_2 - r\) of the path traveled by the sound wave when propagating from the reference microphone to the next one. Using (B.1), we find that this additional path is given by

\[
 r_2 - r = r \left( \sqrt{1 + \left( \frac{\Delta}{r} \right)^2 + 2 \frac{\Delta}{r} \sin \theta - 1} \right). 
\]  

(B.2)

The far-field model corresponds to the situation \(r \gg \Delta\). Using the first-order development \(\sqrt{1 + x} \approx 1 + \frac{x}{2} + o(x^2)\), we can compute the limit of (B.2) for \(r \to +\infty\):

\[
 \lim_{r \to +\infty} (r_2 - r) = \Delta \sin \theta. 
\]  

(B.3)

Therefore the time \(\tau_\theta\) needed by the sound wave to travel from one sensor to the next one if given for the far- and free-field model by

\[
 \tau_\theta = \frac{\Delta \sin \theta}{c}, 
\]  

(B.4)

where \(c\) denotes the speed of sound.
B.2 Null Beamforming

In the following we explain how to compute filter coefficients \( w \) with a unit response in the direction \( \theta_1 \) and a zero response in the direction \( \theta_2 \). Without loss of generality, we assume that \( \theta_1 = 0 \), which is equivalent to steering the array toward \( \theta_1 \).

We want to find the MISO separation system \( w \) that fulfills

\[
\begin{align*}
g(w, 0) &= \delta_D, \\
g(w, \theta_2) &= 0,
\end{align*}
\]

(B.5)

where \( g(w, \theta) \) denotes the spatial response of \( w \) in the direction \( \theta \), as defined in (2.27). To simplify the presentation, we assume that the delay \( \tau_{\theta_2} = f_s \frac{\Delta}{c} \sin \theta_2 \) is an integer. Then the pair of equations in (B.5) may be reformulated as follows:

\[
\begin{align*}
\sum_{m=1}^{M} w_{m,k} &= \delta_{D,k} \quad \text{for} \ k = 1, \ldots, L, \\
\sum_{m=1}^{M} w_{m,k+(m-1)\tau_{\theta_2}} &= 0 \quad \text{for} \ k = 1, \ldots, L,
\end{align*}
\]

(B.6)

where \( \delta_{D,k} = 1 \) for \( k = D \) and \( \delta_{D,k} = 0 \) otherwise. The lower equation in (B.6), \( \sum_{m} w_{m,k+(m-1)\tau_{\theta_2}} = 0 \), characterizes null beamformers which have a zero spatial response at \( \theta_2 \). The spatial response of our null beamformer is also specified in the target direction \( \theta_1 = 0 \). To obtain a matrix form of (B.6), we introduce the matrix \( R_{nn} \) as
$$\mathbf{R}_{nn} \triangleq \begin{bmatrix} \mathbf{I}_{L \times L} & \mathbf{D}_{\tau \theta_2} & \cdots & \mathbf{D}_{(M-1)\tau \theta_2} \\ \mathbf{D}_{-\tau \theta_2} & \mathbf{I}_{L \times L} & \cdots & \mathbf{D}_{(M-2)\tau \theta_2} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{D}_{-(M-1)\tau \theta_2} & \cdots & \cdots & \mathbf{I}_{L \times L} \end{bmatrix}.$$ (B.7)

\(\mathbf{D}_{\tau}\) is a square matrix containing zeros, except for the \(\tau\)th upper diagonal that contains ones. In fact \(\mathbf{R}_{nn}\) corresponds to the microphone correlation matrix \(\mathbf{E}\{\mathbf{x}(p)\mathbf{x}^T(p)\}\) when the source at \(\theta_2\) emits a stationary white noise.

Equation (B.6) may now be written as

$$\mathbf{w}^T [\mathbf{C} \mathbf{R}_{nn}] = \begin{bmatrix} \delta_0^T & 0_{1 \times ML} \end{bmatrix}$$ (B.8)

in matrix form. (The matrix \(\mathbf{C}\) was given in (3.24).) If there are more than two microphones, the \(2L\) equations (B.6) form an under determined set of constraints for the \(ML\) filter coefficients. If we simultaneously constrain \(\mathbf{w}\) to minimize the white-noise gain, the filters \(\mathbf{w}\) are given by the pseudoinverse

$$\mathbf{w}^T = [\delta_0^T \ 0_{1 \times ML}] [\mathbf{C} \mathbf{R}_{nn}]^+.$$ (B.9)

In general, \([\mathbf{C} \mathbf{R}_{nn}]\) has neither full column rank nor full row rank. The computation of the pseudoinverse involves its singular value decomposition and it is difficult to find a general closed formula for \(\mathbf{w}\). For this reason, the solution of (B.9) may rather be found numerically.
The RGSC According to Hoshuyama et al.

This appendix describes the robust GSC (RGSC) proposed by Hoshuyama et al. [53]. Using an adaptation control mechanism, the RGSC can be an efficient LCMV beamformer because it adapts the constraint of distortionless transmission of the desired source to the current acoustic environment [47]. The RGSC includes an adaptive blocking matrix to minimize the leakage of the target signal. The adaptive blocking matrix consists of a set of “interference cancelers” implemented as adaptive filters. As shown by Herbold et al. [47], the RGSC is an LCMV beamformer with a relaxed constraint and may yield an improved suppression of the interference signal relative to the original GSC with a fixed blocking matrix [42]. Section C.1 describes the RGSC which can be used in conjunction with the four-element compact array mounted in the rear-view mirror. In Sect. C.2, we modify this RGSC for use with the two-element distributed array mounted on the car ceiling.

C.1 RGSC for the Four-Element Compact Array Mounted in the Rear-View Mirror

The blocking matrix is square with $M$ outputs instead of $M - 1$ in the original GSC, and has a particular form. To obtain an expression of its outputs, we need to define the filters $b_m, m = 1, \ldots, M$ of length\(^1\) $L$. The coefficients of the filter $b_m$ are stacked in the $L \times 1$ vector:

$$b_m \triangleq (b_{m,0}, \ldots, b_{m,L-1})^T.$$  \hspace{1cm} (C.1)

\(^1\) The filter length of $b_m$ is not necessarily equal to the filter length of the interference canceler. Let us temporarily denote the length of $b_m$ by $L_B$. Maximum robustness against leakage is obtained for $L_B \geq L$, since if $L_B \geq L$ then the interference canceler is not long enough to access any target component. To minimize the complexity, we set $L_B = L$. 
Fig. C.1. Robust generalized sidelobe canceler (RGSC) according to Hoshuyama et al.

We also need to define the vector $x_0(p)$ for the target reference signal:

$$x_0(p) \triangleq (x_0(p), \ldots, x_0(p-L+1))^T.$$  \hfill (C.2)

The outputs of the blocking matrix are then given by

$$x_{B,m}(p) = x_m(p-D_B) + b_m^T x_0(p),$$  \hfill (C.3)

as depicted in Fig. C.1. As shown in (C.8), the optimum filters $b_m$ have to model the inverse of a sum of room impulse responses. Since the room impulse responses may be nonminimum-phased, modeling it inverse requires an acausal delay [50]. Therefore, the delay $D_B$ is typically set to $D_B = L/2$. It must also be included in the interference canceler, which yields $D = L/2 + D_B = L$.

The filters $b_m$ are adapted so that the target leakage at the output of the blocking matrix is minimum. In practice, the adaptation is carried out when the desired source is dominant in such a way that the variance $\mathbb{E}\{x_{B,m}^2(p)\}$ is minimized. This is achieved with the Wiener solution:

$$b_m = - \left( \mathbb{E}\{x_0(p)x_0^T(p)\} \right)^{-1} \mathbb{E}\{x_m(p-D_B)x_0(p)\}.$$  \hfill (C.4)

In an adaptive context, the filters $b_m$ can be adapted with the NLMS algorithm as follows:

$$b_m(p+1) = b_m(p) - \mu_{B,NLMS} \frac{x_{B,m}(p)x_0(p)}{\|x_0(p)\|^2}.$$  \hfill (C.5)

For our experiments with speech signals, the step-size has been set to $\mu_{B,NLMS} = 0.1$. The delay in the blocking matrix path is $D_B = L/2$. The delay in the fixed beamformer path is $D = D_B + L/2 = L$. 
Representation in the DTFT domain

Let us denote by $H_m(\omega)$ the transfer function from the desired source $S(\omega)$ to the $m$th microphone $X_m(\omega)$ and set $H(\omega) = (H_1(\omega), \ldots, H_M(\omega))^T$. In the noiseless case, the source–microphone relationship is written in the DTFT domain as:

$$X(\omega) = H(\omega)S(\omega). \quad (C.6)$$

We denote the DTFT of the filter $b_m$ by $B_m(\omega)$ and we define the vector $B(\omega) \triangleq (B_1(\omega), \ldots, B_M(\omega))^T$. The filters $B_m(\omega)$ are adapted so that the target leakage at the blocking matrix output $X_{B,m}(\omega) = e^{-i\omega D} X_m(\omega) + B_m(\omega)X_0(\omega)$ is minimum. It can be shown that this leads to a signal-dependent spatial constraint $W^H(\omega) H(\omega) S(\omega) = S(\omega)$ [47]. If the desired signal has no energy in a given frequency band, the constraint vanishes, yielding $B(\omega) = I$. This allows the filters $W(\omega) = 0$ in that band. This would not be possible if the spatial constraint was maintained across the entire spectrum. The filters $B_m(\omega)$ that minimize $E \{ |X_{B,m}(\omega)|^2 \}$ are obtained by the Wiener solution:

$$B_m(\omega) = -e^{-i\omega D} \Phi_{X_mX_0}(\omega)/\Phi_{X_0}(\omega). \quad (C.7)$$

$\Phi_{X_mX_0}$ denotes the CPSD of the signals $x_m$ and $x_0$. Likewise, $\Phi_{X_0}$ denotes the PSD of the signal $x_0$. Combining (C.7) and (C.6) for $S_m(\omega) \neq 0$ yields the optimal solution in the minimum mean square error (MMSE) sense

$$B_m(\omega) = -e^{-i\omega D} H_m(\omega)/W_0(\omega)H(\omega). \quad (C.8)$$

Equation (C.8) shows that the optimal blocking matrix filters $B_m(\omega)$ involve the inverse of the sum of delayed acoustic transfer functions, since $W_0^H(\omega) H(\omega) = \sum_{m=1}^{M} e^{i\omega(m-1)\tau_0} H_m(\omega)$.

C.2 RGSC for the Two-Element Distributed Array Mounted on the Car Ceiling

A data-dependent “blocking matrix” can be considered for the two-element distributed array mounted on the car ceiling, too. Using a set of interference cancelers for the blocking matrix as in Sect. C.1, we can incorporate an additional adaptive filter in the GSC structure. This transforms the AIC shown in Fig. C.2 a into the RGSC shown in Fig. C.2 b. The outputs $x_{B,m}(p)$ of the blocking matrix are defined as

$$x_{B,m}(p) \triangleq x_{m+1}(p) - b_m^T x_1(p) \quad \text{for } m = 1, \ldots, M-1. \quad (C.9)$$

The filters $b_m$ should be adapted using interference-free input signal. They can be computed with (C.4) or (C.5) using $x_0(p) = x_1(p)$ and $x_{B,m}(p) = x_{m+1}(p)$. 
The RGSC According to Hoshuyama et al.

$\alpha_1(\theta)\ x_1(t) \rightarrow z^{-D} \rightarrow y(t)$

$\alpha_2(\theta)\ x_2(t) \rightarrow a \ (w_2)$

$\alpha_1(\theta)\ x_1(t) \rightarrow z^{-D} \rightarrow y(t)$

$\alpha_2(\theta)\ x_2(t) \rightarrow a \ (w_1)$

Fig. C.2. Beamformer architecture for two directional microphones. (a) Adaptive interference canceler (AIC). (b) RGSC according to [53] modified for use with the two-element distributed array mounted on the car ceiling.

C.3 Experimental Comparison: GSC vs. RGSC

Offline experiments in stationary conditions

This section examines how the RGSC mitigates the target-cancelation problem. In the following, the performance of the RGSC is compared with that of the GSC in stationary conditions with nonadaptive, off-line beamformers. That way, the comparison should remain independent of the implementation. By contrast, the beamformer performance in adaptive mode depends highly on implementation parameters like the step-size and the DTD thresholds.

The source signals are white-noise signals emitted by artificial heads to ensure stationary conditions. The microphone signals are sampled at $f_s = 16$ kHz. The filter length is set to $L = 256$ in the case of the four-element compact array mounted in the rear-view mirror and to $L = 512$ in the case of the distributed array. The experimental setups are described in Appendix A in greater detail.

Varying parameters

To evidence how the adaptive blocking matrix mitigates the power-inversion effect, we let vary the SIR of the input signals that are used to adapt the
interference canceler, which is denoted by SIR_{in}. To define SIR_{in}, we decompose the input signals as the sum of the contribution of the desired source and of that of the interferers, as given in Sect. 2.4:

\[ x_m(p) = x_{\text{sig},m}(p) + x_{\text{int},m}(p). \]  

(2.61)

In offline mode with stationary signals, we use a batch estimate of the expectation operator:

\[ \hat{E}\{f(A(p))\} \triangleq \frac{1}{T} \sum_{p=1}^{T} f(A(p)). \]  

(2.65)

Using (2.61), we may define SIR_{in} and SIR^d_{in} for compact and distributed arrays respectively, as follows:

\[ \text{SIR}_{in} \triangleq \frac{\sum_{m=1}^{M} \hat{E}\{x_{\text{sig},m}^2(p)\}}{\sum_{m=1}^{M} \hat{E}\{x_{\text{int},m}^2(p)\}}, \quad \text{SIR}^d_{in} \triangleq \frac{\hat{E}\{x_{\text{sig},1}^2(p)\}}{\hat{E}\{x_{\text{int},1}^2(p)\}}. \]  

(C.10)

For a given blocking matrix \( B \), we also decompose the interference reference signals as the sum of the contribution of the desired source and of that of the interferers:

\[ x_{B,\text{sig},m}(p) \triangleq B^T x_{\text{sig},m}(p), \quad x_{B,\text{int},m}(p) \triangleq B^T x_{\text{int},m}(p). \]  

(C.11)

Using (C.11), the SIR at the input of the interference canceler, SIR_{AIC,in}, (that is, the SIR at the output of the blocking matrix) may be defined as

\[ \text{SIR}_{AIC,in} \triangleq \frac{\sum_{m=1}^{M} \hat{E}\{x_{B,\text{sig},m}^2(p)\}}{\sum_{m=1}^{M} \hat{E}\{x_{B,\text{int},m}^2(p)\}}, \quad \text{SIR}_{AIC,in} \triangleq \frac{\hat{E}\{x_{B,\text{sig},1}^2(p)\}}{\hat{E}\{x_{B,\text{int},1}^2(p)\}}. \]  

(C.12)

SIR_{AIC,in} may be obtained by subtracting a fixed offset \( a \) from SIR_{in} in the dB scale:

\[ \text{SIR}_{AIC,in} = \text{SIR}_{in} - a. \]  

(C.13)

The offset \( a \) depends on the blocking matrix \( B \). According to the power-inversion effect described in Sect. 3.4, the SIR at the output of the beamformer should be zero for SIR_{in} = a.

Filter adaptation

The filter adaptation and the performance evaluation are carried out in three steps:

(i) First the RGSC filters \( b_m \) for \( m = 1, \ldots, M \) of the data-dependent blocking matrix are computed according to (C.4). These filters \( b_m \) are trained using the desired source input signals \( d(p) \), which corresponds to a best-case scenario. For the GSC, the blocking matrix is fixed.
(ii) Then the interference canceler $a$ is adapted on the input signals having a prescribed $SIR_{in}$ (SIR at the input of the beamformer). The adaptation of the interference canceler $a$ is based on the Wiener solution (3.22). However, the inversion of $R_{x_B x_B}$ in (3.22) for $L = 256$ is numerically badly conditioned. Therefore, we regularize the inversion with a regularization factor $\epsilon > 0$, as follows:

$$a^{(0)} = (R_{x_B x_B} + \epsilon I)^{-1} E \{x_0(p - D)x_B(p)\}. \quad (C.14)$$

This yields a biased first estimate $a^{(0)}$ of the optimal $a$. For our experiments, we set $\epsilon = 0.01$ which led to the best results. Then, we improve $a^{(0)}$ using the gradient descent for the cost function in (3.1). The gradient descent does not necessitate matrix inversion. Using the gradient in (3.21) yields

$$a^{(n+1)} = a^{(n)} - \mu \frac{\partial J}{\partial a}$$

$$= a^{(n)} - 2\mu \sum_{p=1}^{T} y(p)x_B(p). \quad (C.15)$$

The step-size $\mu$ is set to $\mu = 0.05$ and the iterations are carried out until $J$ does not decrease by more than 0.1 dB per iteration.

(iii) After (i) and (ii), the filters are fixed and the performances of the full GSC and RGSC structures are evaluated using unit power input signals from the target and from the interferer.

C.3.1 Experiments with the Four-Element Compact Array Mounted in the Rear-View Mirror

We consider first the four-element directional microphone array mounted in the rear-view mirror (see Sect. A.1) for more details). The delay $D$ is set to $D = L/2$. The RGSC filters $b_m$ for $m = 1, \ldots, M$ of the data-dependent blocking matrix are trained using the desired source input signals $d(p)$. It could be measured that this data-dependent blocking matrix brings a 11.5 dB reduction of the SIR, that is, $a = 11.5$ dB in (C.13).

For our implementation of the GSC, the output signals of the fixed blocking matrix $x_{B,m}(p)$ are given by

$$x_{B,m}(p) = x_{m+1}(p) - x_{0}(p) \quad \text{for } m = 1, \ldots, M - 1,$$

$$= x_{m+1}(p) - \frac{1}{M} \sum_{m=1}^{M-1} x_{M-1}(p). \quad (C.17)$$

We could measure that the signal $x_{B,m}(p)$ exhibits a SIR which is about 3 dB lower that the input SIR, that is, $a = 3$ dB in (C.13). The results are shown in Fig. C.3.
Fig. C.3. Performances of the GSC (a) and the RGSC (b) as a function of $\text{SIR}_{\text{in}}$ (SIR at the beamformer input), for the four-element compact array mounted in the rear-view mirror. The RGSC beamformer structure is shown in Fig. C.1. For the RGSC, the blocking matrix is trained with interferer-free input signals, which is the best-case scenario. The experimental setup is depicted in Fig. A.2. The steered DOA is $\theta = 20^\circ$. The filter length is set to $L = 256$.

It can be observed that the GSC does not bring any SIR improvement for $\text{SIR}_{\text{in}}$ around 3 dB. This observation fits well with the power-inversion effect. Moreover we can see that the level of the desired signal is significantly reduced for large $\text{SIR}_{\text{in}}$. On the other hand, the data-dependent blocking-matrix attenuates the desired source components at the outputs $x_{B,m}$ of the blocking matrix. This protects efficiently against cancelation of the target signal: We can see that the reduction of the desired signal level is much smaller than that of the GSC. However, in accordance with the power-inversion effect, the SIR improvement vanishes for $\text{SIR}_{\text{in}} = 11.5$ dB: The signals $x_{B,m}$ still contain reflected paths of the desired source and are correlated to the output of the fixed beamformer $x_0$. Since the optimal interference canceler minimizes this correlation, an important degradation of the interference reduction results for $\text{SIR}_{\text{in}} > -5$ dB.

C.3.2 Experiments with the Two-Element Distributed Array Mounted on the Car Ceiling

We compare the AIC structure and the RGSC structure with a data-dependent blocking matrix. Both structures are depicted in Fig. 3.2. We could measure that the SIR at microphone $x_2$ is about 5 dB below the SIR at microphone $x_1$, that is, $a = 5$ dB in (C.13). The blocking matrix filter $b_1$ is adapted with the desired source signals $d(p)$, which represents a best-case scenario. This data-dependent blocking matrix brings an additional 7 dB reduction of
the SIR with respect to the SIR at microphone $x_2$, that is, $a = 12$ dB in (C.13). Figure C.4 shows the SIR improvement for various input SIR. The RGSC is slightly more robust against signal cancelation, but both structures exhibit similar $SR^c$ curves. Although these results do not account for the distortion of the target signal, they indicate that the target signal cancelation problem is not significantly better relieved by the RGSC than by the AIC.

This may be explained by the causality constraints than can be set on the AIC (that is, the AIC structure with $D = 0$). For the free-field propagation model, the target signal with positive DOA reaches the microphone $x_2(p)$ after $x_1(p)$, i.e., after a positive delay. Then, causal filtering $a^T x_2(p)$ cannot compensate this delay to suppress the target at the output $y(p) = x_1(p) + a^T x_2(p)$. This may prevent the desired source to be canceled. As a consequence, the AIC is chosen for the two-element distributed array mounted on the car ceiling.

**C.4 Conclusion**

The RGSC proposed by Hoshuyama et al. [53] has been described. A modification of this RGSC for use in conjunction with the two-element distributed array mounted on the car ceiling has been proposed.

In comparison to the original GSC [42] for the four-element compact array mounted in the rear-view mirror, the $M$ outputs of the adaptive blocking matrix may provide more degrees of freedom to the interference canceler, depending on the target signal spectrum. This may yield an improved
suppression of the interferer signal at the beamformer output. Moreover, the RGSC provides a significant improvement over the GSC in terms of robustness against target signal cancelation. Therefore, in our experiments, the RGSC is chosen to determine the performance of controlled beamforming.

For the two-element distributed array mounted on the car ceiling, the advantage of the RGSC over the AIC is less significant than for the four-element compact array mounted in the rear-view mirror. The increased computational demand of the RGSC over the AIC may not be justified. For this reason, we prefer to use the AIC to determine the performance of controlled beamforming in our experiments.
As explained in Sect. 7.1.1, the analysis of the global stability for convolutive BSS is hardly tractable. This chapter examines the local stability in the vicinity of a particular equilibrium point. The local stability of BSS algorithms has been investigated in the vicinity of the inverse of the mixing system [8, 26, 39, 52]. There, the sources are not only separated but also deconvolved, possibly using a feedback filter architecture, as in [26]. However, in many practical cases, only the separation of the sources is desired (or achievable). Therefore, we will study the stability around an equilibrium point that separates but does not deconvolve the sources.

This chapter constitutes a generalization of the analysis for white source signals presented in [16]. It is organized as follows: Section D.1 defines the mixing and separation models and the notations. Section D.2 derives the linearization of the learning rules in the vicinity of the equilibrium. The conditions for the local stability are derived in Sect. D.3.

D.1 Mixing and Separation Models

We consider a $2 \times 2$ mixing and separation scenario with mixing and separation filters of the same lengths ($L_m = L$). We assume that the diagonal channels of the mixing and separation systems are delayed unit responses, that is,

$$h_{11} = h_{22} = \delta_d, \quad (D.1)$$

for some delay $0 \leq d \leq L/2$. Note that this assumption is only a normalization of the mixing process. The off-diagonal channels are denoted by $h_{12}$ and $h_{21}$. We constrain accordingly the diagonal filters of the separation system

$$w_{11} = w_{22} = \delta_d. \quad (D.2)$$

This model is depicted in Fig. 2.6. It approximates a scenario where each source $s_n(p)$ is positioned close to microphone $x_n(p)$. It is also motivated by the following remarks:
(i) Let us consider a small deviation $\varepsilon$ around the equilibrium point $W_{opt}$, that is, $W = W_{opt} + \varepsilon$. We parametrize the deviation as $\varepsilon^C(n) = \varepsilon(n)H$. $\varepsilon^C(n)$ corresponds to an additive deviation for the global system $C(n) = W(n)H$, since $C(n) = W_{opt}H + \varepsilon^C(n)$. It can be shown that the diagonal components $\varepsilon_{ii}(n)$ of $\varepsilon^C(n)$ are constant at first order [33]. Therefore, the diagonal elements of the global system $C_{ii}(n)$ are not going to change significantly and $W(n)$ will not reach the vicinity of $W_{opt}$. The constraint $w_{ii} = \delta_d$ for $i = 1, \ldots, N$ provides additional information that enable convergence toward the vicinity of $W_{opt}$.

(ii) The constraints $w_{ii} = \delta_d$ for $i = 1, 2$ ensure the uniqueness of the equilibrium point, which is $w_{ij} = -h_{ij}$ for $i, j = 1, 2$ and $i \neq j$.

(iii) Under the constraint $w_{ii} = \delta_d$ for $i = 1, 2$, the self-closed learning rules (6.51) are equal to their self-closed counterparts (6.53). This mixing system is separated with the following separation system:

$$
\begin{align*}
    w_{11} &= \delta_d, \\
    w_{12} &= -h_{12}, \\
    w_{21} &= -h_{21}, \\
    w_{22} &= \delta_d.
\end{align*}
$$

For later reference, we define the vector $w_{eq}$ of length $2L - 1$ as

$$
    w_{eq} = \delta_{2d} - h_{12} * h_{21}.
$$

In fact, $w_{eq} = (w_{eq,0}, \ldots, w_{eq,2L-2})^T$ is the source–output response at the solution (D.3). The sources $s_1$ and $s_2$ are assumed stationary within time blocks of length $L$ and their self-correlation function is denoted by

$$
    r_{n,\tau}(p) = E\{s_n(p)s_n(p-\tau)\}.
$$

D.2 Linearization of the NG-SOS-BSS Updates

We examine the local stability of the separation algorithm around the separating solution (D.3). To this end, we set

$$
\begin{align*}
    w_{12}(n) &= -h_{12} + \varepsilon_{12}(n), \\
    w_{21}(n) &= -h_{21} + \varepsilon_{21}(n),
\end{align*}
$$

with the deviations

$$
\begin{align*}
    \varepsilon_{12} &= (\varepsilon_{12,0}, \ldots, \varepsilon_{12,L-1})^T, \\
    \varepsilon_{21} &= (\varepsilon_{21,0}, \ldots, \varepsilon_{21,L-1})^T.
\end{align*}
$$

The expectation of algorithm (6.51) can be written as:

$$
\begin{align*}
    w_{12}(n+1) &= w_{12}(n) - \mu \sum_{k=1}^{K} r_{y_1y_2}(kL)/\sigma_1^2(kL), \\
    w_{21}(n+1) &= w_{21}(n) - \mu \sum_{k=1}^{K} r_{y_2y_1}(kL)/\sigma_2^2(kL),
\end{align*}
$$

where

$$
\begin{align*}
    r_{y_1y_2}(kL) &= E\{y_1(kL)y_2(kL)\}, \\
    r_{y_2y_1}(kL) &= E\{y_2(kL)y_1(kL)\}.
\end{align*}
$$
where

\[ r_{y_i,y_j}(p) = (r_{y_i,y_j,-d(p)}, \ldots, r_{y_i,y_j,L-1-d(p)}) \]  
\[ r_{y_i,y_j,\tau}(p) = E\{y_i(p)y_j(p-\tau)\} \]  
\[ \sigma_i^2(p) = E\{y_i^2(p)\} \]

for \( i, j = 1, 2 \). The BSS algorithm is driven by the output cross correlation \( r_{y_i,y_j}(p) \). Neglecting the quadratic terms in \( \varepsilon_{ij,k}(n) \), this cross correlation is first-order approximated by

\[ r_{y_1,y_2,\tau}(p) = \sum_{u=0}^{2L-2} \sum_{v=0}^{L-1} w_{eq,u} \varepsilon_{21,v}(n) r_{1,v+d+\tau-u}(p) \]
\[ + \sum_{u=0}^{L-1} \sum_{v=0}^{2L-2} w_{eq,v} \varepsilon_{12,u}(n) r_{2,v+\tau-u-d}(p). \]  

(D.15)

The first-order approximation for \( r_{y_2,y_1,\tau}(p) \) is simply given by

\[ r_{y_2,y_1,\tau}(p) = r_{y_1,y_2,-\tau}(p). \]  

(D.16)

Substituting \( w_{ij}(n) \) in (D.6) and (D.7) into (D.10) and (D.11), and replacing \( r_{y_i,y_j,\tau}(p) \) by its value in (D.15), the deviations \( \varepsilon_{12}(n+1) \) and \( \varepsilon_{21}(n+1) \) are given as a linear function of \( \varepsilon_{12}(n) \) and \( \varepsilon_{21}(n) \). Thus, the first-order approximation of (D.10) and (D.11) can be written in terms of \( \varepsilon_{12}(n) \) and \( \varepsilon_{21}(n) \) in matrix form:

\[ \varepsilon_{12}(n+1) = \varepsilon_{12}(n) - \mu \sum_k \left( \frac{1}{\sigma_1^2(kL)} A_{11}(kL) \varepsilon_{12}(n) + \frac{1}{\sigma_1^2(kL)} A_{12}(kL) \varepsilon_{21}(n) \right), \]  

\[ \varepsilon_{21}(n+1) = \varepsilon_{21}(n) - \mu \sum_k \left( \frac{1}{\sigma_2^2(kL)} A_{21}(kL) \varepsilon_{12}(n) + \frac{1}{\sigma_2^2(kL)} A_{22}(kL) \varepsilon_{21}(n) \right) \]

(D.17)

(D.18)

with

\[ A_{11}(p) = \begin{pmatrix} w_{eq,0}^{(2)}(p) & w_{eq,1}^{(2)}(p) & \ldots & w_{eq,L-1}^{(2)}(p) \\ w_{eq,-1}^{(2)}(p) & w_{eq,0}^{(2)}(p) & \ldots & w_{eq,L-2}^{(2)}(p) \\ \vdots & \vdots & \ddots & \vdots \\ w_{eq,-L+1}^{(2)}(p) & \ldots & \ldots & w_{eq,0}^{(2)}(p) \end{pmatrix}, \]  

(D.19)
\[
A_{12}(p) = \begin{pmatrix}
  w_{\text{eq},0}(p) & w_{\text{eq},1}(p) & \cdots & w_{\text{eq},L-1}(p) \\
  w_{\text{eq},1}(p) & w_{\text{eq},2}(p) & \cdots & w_{\text{eq},L}(p) \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{\text{eq},L-1}(p) & \cdots & \cdots & w_{\text{eq},2L-2}(p)
\end{pmatrix}, \quad (D.20)
\]

\[
A_{21}(p) = \begin{pmatrix}
  w_{\text{eq},0}(p) & w_{\text{eq},1}(p) & \cdots & w_{\text{eq},L-1}(p) \\
  w_{\text{eq},1}(p) & w_{\text{eq},2}(p) & \cdots & w_{\text{eq},L}(p) \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{\text{eq},L-1}(p) & \cdots & \cdots & w_{\text{eq},2L-2}(p)
\end{pmatrix}, \quad (D.21)
\]

\[
A_{22}(p) = \begin{pmatrix}
  \tilde{w}_{\text{eq},0}(p) & \tilde{w}_{\text{eq},1}(p) & \cdots & \tilde{w}_{\text{eq},L-1}(p) \\
  \tilde{w}_{\text{eq},1}(p) & \tilde{w}_{\text{eq},2}(p) & \cdots & \tilde{w}_{\text{eq},L}(p) \\
  \vdots & \vdots & \ddots & \vdots \\
  \tilde{w}_{\text{eq},L+1}(p) & \cdots & \cdots & \tilde{w}_{\text{eq},0}(p)
\end{pmatrix}, \quad (D.22)
\]

The coefficients \(w_{\text{eq},k}(p)\) come from the convolution of \(r_n(p)\) and \(w_{\text{eq}}\) and are given by

\[
w_{\text{eq},k}(p) = \sum_{u=0}^{2L-2} w_{\text{eq},u} r_{n,k-u}(p). \quad (D.23)
\]

The coefficients \(\tilde{w}_{\text{eq},k}(p)\) are similarly defined, as follows: Let us introduce the response \(\tilde{w}_{\text{eq}}\), which is a shifted version of \(w_{\text{eq}}\):

\[
\tilde{w}_{\text{eq},k} = w_{\text{eq},k+2d}. \quad (D.24)
\]

Then, according to (D.4), we have

\[
\tilde{w}_{\text{eq},k} = \delta_k - \sum_{u=0}^{L-1} h_{12,u+d} h_{21,k+d-u}, \quad (D.25)
\]

and we define \(\tilde{w}_{\text{eq},k}(p)\) as the convolution of \(\tilde{w}_{\text{eq},k}\) and \(r_{n,k}(p)\):

\[
\tilde{w}_{\text{eq},k}(p) = \sum_{u=0}^{2L-2} \tilde{w}_{\text{eq},u} r_{n,k-u}(p). \quad (D.26)
\]

Now, we consider the situation where each source is silent over a certain time block \(T\), say, \(T_k = [kL - 3L + 3, kL]\), while the other sources are not silent over this interval:

\[
\forall n = 1, \ldots, N \quad \exists k \text{ such that } s_n(kL) = 0 \text{ and } \forall m \neq n, s_m(kL) \neq 0. \quad (D.27)
\]
Without loss of generality, suppose \( s_1 \) is silent on the first time block \( T_1 \). Let \( \sigma_1^2(L) \) tend to zero in (D.17). One finds that the second term on the right-hand side of (D.17), namely

\[
\frac{1}{\sigma_1^2(kL)} A_{11}(kL) \varepsilon_{12}(n),
\]

(D.28)

tends to \(+\infty\). By contrast, the last term of (D.17), namely

\[
\frac{1}{\sigma_1^2(kL)} A_{12}(kL) \varepsilon_{21}(n),
\]

(D.29)

remains bounded. Since we are interested in the direction of the update and not in its norm, we can consider a small step-size \( \mu \) proportional to \( \sigma_1^2(L) \) so that the limit of the second term on the right-hand side (D.17) exists. Then (D.29) vanishes at this limit, as well as the contributions of the arguments of the sum in (D.17) for \( k = 2, \ldots, K \). One obtains the following relation on \( \varepsilon_{12}(n) \):

\[
\varepsilon_{12}(n+1) = (I - \tilde{\mu} A_{11}(L)) \varepsilon_{12}(n),
\]

(D.30)

where \( \tilde{\mu} \sigma_1^2(L) = \mu \). The same reasoning for source \( s_2 \) silent at the second time block \( T_2 \) leads to a similar equation for \( \varepsilon_{21}(n) \):

\[
\varepsilon_{21}(n+1) = (I - \tilde{\mu} A_{22}(2L)) \varepsilon_{12}(n).
\]

(D.31)

Therefore, the local stability depends on the positiveness of the eigenvalues of \( A_{11}(L) \) and \( A_{22}(2L) \).

D.3 Local Stability Conditions

Causal mixing, white sources

Suppose that the source \( s_2 \) is white in \( T_1 \), then its correlation function \( r_{2,\tau} \) is a Dirac impulse, \( r_{2,\tau} = \delta(\tau) \). If additionally \( d = 0 \), then the definitions (D.23) and (D.24) yield

\[
\tilde{w}_{\text{eq}}^{(n)}(p) = w_{\text{eq}}^{(n)}(p) = w_{\text{eq}}.
\]

(D.32)

If \( d = 0 \), i.e., if the mixing is causal, then \( w_{\text{eq},k} = 0 \) for \( k < 0 \). Consequently, according to (D.19), the matrices \( A_{11}(L) \) and \( A_{22}(2L) \) in (D.30) and (D.31) become upper triangular. The eigenvalues are the diagonal elements, \( w_{\text{eq},0} = 1 - h_{12,0} h_{21,0} \). Therefore, a necessary and sufficient condition for the local stability is

\[
1 - h_{12,0} h_{21,0} > 0.
\]

(D.33)
General case

The matrix $A_{11}(L)$ in (D.30) is a Toeplitz matrix and can be made circulant using the elements of its first row and of its first column. Let us define

$$c_1 = \left( \tilde{w}_{eq,0}^{(2)}(L), \ldots, \tilde{w}_{eq,L-L+1}^{(2)}(L), \tilde{w}_{eq,L}^{(2)}(L), \ldots, \tilde{w}_{eq,1}^{(2)}(L) \right)^T.$$

The time argument $L$ appears because the local behavior is dominated by the source statistics at time block $T_1$, as in (D.30). We denote the $2L-1 \times 2L-1$ circulant matrix with first column $c_1$ by $C$. Similarly, we double the size of $\varepsilon_{12}$:

$$\tilde{\varepsilon}_{12} = \begin{bmatrix} \varepsilon_{12} \\ 0_{L-1 \times 1} \end{bmatrix}$$

and we define the projection

$$\Omega = \begin{bmatrix} I_{L \times L} & 0_{L \times L-1} \\ 0_{L-1 \times L} & 0_{L-1 \times L-1} \end{bmatrix}.$$

Then (D.30) becomes

$$\tilde{\varepsilon}_{12}(n+1) = \Omega (I - \tilde{\mu}C) \tilde{\varepsilon}_{12}(n).$$

Equation (D.40) shows that a sufficient condition for the local stability is that the diagonal elements of $\Lambda$, i.e., the DFT values of $\tilde{w}_{eq}^{(2)}$, have positive realparts. Let us denote by $\tilde{W}_{eq}^{(2)}(k)$ the $k$th frequency bin of the DFT of $\tilde{w}_{eq}^{(2)}$, for $k = 0, \ldots, 2L-2$:

$$\tilde{W}_{eq}^{(2)}(k) = \sum_{\tau = -L+1}^{L-1} \tilde{w}_{eq,\tau}^{(2)}(L) e^{2i\pi\tau k/(2L-1)}.$$

We similarly define $\tilde{W}_{eq}(k)$ and $R_2(k)$, the $k$th frequency bin of $\tilde{w}_{eq}$ and $r_2(L)$ (the correlation function of the source $s_2$ at time block $T_1$), respectively. According to (D.26), $\tilde{W}_{eq}^{(2)}(k)$ can be factorized as
provided the response \( \tilde{w}_{eq} \) of length \( L_w \) and the correlation function \( r_2 \) of length \( L_r \) fulfill \( L_w + L_r - 1 \leq 2L - 1 \) (which is true if the filters \( h_{12} \) and \( h_{21} \) are short enough). Since \( r_2 \) is symmetric, its DFT \( R_2(k) \) is real-valued. It is further assumed that \( R_2(k) \) is positive. According to (D.25), we have

\[
\tilde{W}_{eq}(k) = 1 - e^{-4i\pi dk/(2L-1)} H_{12}(k) H_{21}(k). \tag{D.43}
\]

The positiveness of \( \tilde{W}_{eq}^{(2)}(k) \) is obtained if \(|e^{-4i\pi dk/(2L-1)} H_{12}(k) H_{21}(k)| < 1\), that is, if

\[
|H_{12}(k) H_{21}(k)| < 1 \tag{D.44}
\]

which is the result used in Chap. 7. Note that the same reasoning applied to \( \varepsilon_{21}(n) \) leads to the same sufficient local stability conditions.
Notations

Conventions

\( x \) \hspace{1cm} \text{real scalar}
\( \mathbf{x} \) \hspace{1cm} \text{vector}
\( \mathbf{X} \) \hspace{1cm} \text{matrix or vector for a frequency-domain variable}
\( \mathbf{X}^H \) \hspace{1cm} \text{matrix Hermitian (complex conjugate) transpose}
\( \mathbf{X}^T \) \hspace{1cm} \text{matrix transpose}
\( \mathbf{X}^{-T} \) \hspace{1cm} \text{inverse of } \mathbf{X}^T
\( \triangleq \) \hspace{1cm} \text{definition (as opposed to assignment or assertion)}

Abbreviations and acronyms

s.t. \hspace{1cm} \text{subject to}
AIC \hspace{1cm} \text{Adaptive Interference Canceler}
BGSC \hspace{1cm} \text{Blind Generalized Sidelobe Canceler}
BM \hspace{1cm} \text{Blocking Matrix}
BSS \hspace{1cm} \text{Blind Source Separation}
DOA \hspace{1cm} \text{Direction of Arrival}
DTD \hspace{1cm} \text{Double-Talk Detector}
DFT \hspace{1cm} \text{Discrete Fourier Transform}
DTFT \hspace{1cm} \text{Discrete-Time Fourier Transform}
FD-SOS \hspace{1cm} \text{Frequency-Domain Second-Order Statistics BSS algorithm}
FD-HOS \hspace{1cm} \text{Frequency-Domain Higher-Order Statistics BSS algorithm}
FFT \hspace{1cm} \text{Fast Fourier Transform}
FIR \hspace{1cm} \text{Finite Impulse Response}
GSC \hspace{1cm} \text{Generalized Sidelobe Canceler}
GSD \hspace{1cm} \text{Generalized Sidelobe Decorrelator}
IFFT \hspace{1cm} \text{Inverse Fast Fourier Transform}
ILMS \hspace{1cm} \text{Implicit LMS}
LCMV \hspace{1cm} \text{Linearly Constrained Minimum Variance}
LS \hspace{1cm} \text{Least-Mean Square}
LTI \hspace{1cm} \text{Linear Time Invariant}
MIMO  Multiple-Input Multiple-Output
MISO  Multiple-Input Single-Output
NG-SOS-BSS  Natural Gradient Second-Order Statistics Blind Source Separation algorithm
PSD  Power Spectral Density
QIC  Quadratic Inequality Constraint
RGSC  Robust Generalized Sidelobe Canceler
STFT  Short-Time Discrete Fourier Transform
SOS-BSS  Second-Order Statistics Blind Source Separation
ULA  Uniform Linear Array
WER  Word Error Rate

Scalar variables

$a_{\text{ILMS}}$  ILMS convergence contraction factor
$a_{\text{NLMS}}$  NLMS convergence contraction factor
$a_{\text{QIC}}$  constant upper limit in the quadratic inequality constraint
$b(p)$  target signal at the output when $a(p) = a_{\text{opt}}(p)$
$b_{\text{ILMS}}$  ILMS divergence contraction factor
$b_{\text{NLMS}}$  NLMS divergence contraction factor
$d$  number of acausal coefficients which are treated as acausal in the convolution operator $\star_d$
$f_s$  sampling frequency
$h_{mn,k}$  $k$th tap of the discrete-time acoustic channels from the $n$th source to the $m$th receiver
$h_{r,\theta}$  continuous time acoustic channel from the source position $\theta$ to the receiver position $r$
$p$  discrete time
$s_n(p)$  $n$th source signal (speech source)
$s_{\theta}(t)$  continuous time signal emitted at position $\theta$
$t$  continuous time
$x(r,t)$  continuous time signal received at position $r$
$x_n(p)$  $n$th observed signal (microphone signal)
$w_{m,k}$  $k$th tap of the separation filter for the $m$th input of a MISO separation system
$w_{nm,k}$  $k$th tap of the separation filter for the $m$th input and $n$th output of a MIMO separation system
$x_0(p)$  target reference signal (fixed beamformer output signal)
$x_{B,m'}(p)$  blocking matrix $m'$th output signal
$x_m(p)$  $m$th received signal (microphone signal)
$x_{\text{int},m}(p)$  contribution of the interference signal in the input signal $x_m(p)$
$x_{\text{sig},m}(p)$  contribution of the desired signal in the input signal $x_m(p)$
y$(p)$  output signal for MISO separation system
$y_n(p)$  $n$th output signal for MIMO separation system
$y_{\text{int}}(p)$  contribution of the interferences at the output
\( C \) dimension of the constraint for LCMV beamformer
\( C_A \) complexity of an addition or subtraction
\( C_{\text{batch}} \) complexity of a batch algorithm
\( C_{\text{block-wise}} \) complexity of a block-wise batch algorithm
\( C_{\text{FFT}}^{\text{conv}} \) complexity for convolution in the DFT domain with FFT of length \( L_{\text{FFT}} \)
\( C_M \) complexity of a multiplication or division
\( D \) delay in the fixed beamformer path
\( G_{\mathbf{w}}(\omega, \theta) \) space–frequency response
\( H(\mathbf{y}) \) entropy of \( \mathbf{y} \)
\( H_{mn}(k) \) in the local stability analysis \( H_{mn}(k) \) is the DFT of \( h_{mn} \) padded with \( L - 1 \) zeros at frequency bin \((-k)\)
\( I(\mathbf{y}) \) mutual information of \( \mathbf{y} \)
\( \text{IR}(p) \) reduction of the interference signal level for compact arrays
\( \text{IR}_{[t_0, t_1]} \) reduction of the interference signal level, averaged between times \( t_0 \) and \( t_1 \), for compact arrays
\( \text{IR}_d(p) \) reduction of the interference signal level for distributed arrays
\( \text{IR}_d^{[t_0, t_1]} \) reduction of the interference signal level, averaged between times \( t_0 \) and \( t_1 \), for distributed arrays
\( J \) cost function (or “criterion”)
\( J|_S \) restriction of \( J \) to the Sylvester subspace \( S \)
\( J_{\text{LS}}(\mathbf{w}) \) least-square cost function
\( J_{\text{LMS}} \) least-mean-square criterion
\( J_{\text{BSS}}(\mathbf{W}) \) BSS cost function
\( J_{\text{BSS},\text{LS}}(\mathbf{W}) \) least-square BSS cost function
\( J_{\text{geo}}(\mathbf{W}) \) geometric cost function
\( K \) number of jointly diagonalized output correlation matrices in SOS-BSS
\( L \) length of the separation filters
\( L_m \) length of the mixing channels
\( M \) number of microphones
\( M' \) number of interference reference signals
\( N \) number of sources
\( N_{\text{iter}} \) number of iterations in the gradient descent
\( Q \) quality measure to determine the constants \( \mu_{\text{NLMS}}, \mu_0 \), and \( a_{\text{QIC}} \)
\( Q_{[t_0, t_1]} \) signal-to-interference ratio improvement, averaged between times \( t_0 \) and \( t_1 \), for compact arrays
\( Q_d^{[t_0, t_1]} \) signal-to-interference ratio improvement, averaged between times \( t_0 \) and \( t_1 \), for distributed arrays
\( \text{SIR}_{\text{imp}}(p) \) signal-to-interference ratio improvement for distributed arrays
210 E Notations

SIR\textsubscript{imp}(p) signal-to-interference ratio improvement for compact arrays
SR(p) reduction of the target signal level for compact arrays
SR\textsubscript{[t_0,t_1]} reduction of the target signal level, averaged between times \( t_0 \) and \( t_1 \), for compact arrays
SR\textsubscript{d}(p) reduction of the target signal level for distributed arrays
SR\textsubscript{d}[t_0,t_1] reduction of the target signal level, averaged between times \( t_0 \) and \( t_1 \), for distributed arrays
\( T \) number of samples

**Vectors and matrices**

\( a(p) \) adaptive interference canceler
\( a_{\text{opt}}(p) \) optimal adaptive interference canceler (Wiener solution)
\( c \) \( C \times 1 \) response vector
\( c_{\text{int}} \) global interference–output response for \textit{MISO} separation systems
\( c_n \) global \( n \)th source–output response for \textit{MISO} separation systems
\( d(p) \) contribution of the desired signal in \( x(p) \)
\( d(\theta) \) \( M(2L - 1) \times 1 \) vector used to define the spatial response \( g(W_n, \theta) \)
\( g(W_n, \theta) \) spatial response defined with the Sylvester matrix \( W_n \)
\( g(w, \theta) \) spatial response of the \textit{MISO} separation system \( w \) for the DOA \( \theta \)
\( g(w, \theta) \) spatial response of the \textit{MISO} separation system \( w \) for the 3D position \( \theta \)
\( h_{mn} \) \( L_m \times 1 \) mixing channel vector for the \( n \)th source and \( m \)th receiver
\( m(p) \) mismatch between the actual adaptive interference canceler and the Wiener solution
\( n(p) \) contribution of the local interference signal in \( x(p) \)
\( n^{(\text{road})}(p) \) contribution of the road noise signal \( x(p) \)
\( s_{\text{int}}(p) \) \((N - 1)(L + L_m - 1) \times 1 \) interference source signal vector
\( s_n(p) \) \( L + L_m - 1 \times 1 \) \( n \)th source vector
\( w \) \( ML \times 1 \) multichannel filter vector for \textit{MISO} separation system
\( w_0 \) fixed beamformer in a GSC beamformer
\( w_m \) \( L \times 1 \) separation filter for the \( m \)th input of a \textit{MISO} separation system
\( w_{nm} \) \( L \times 1 \) separation filter vector for \textit{MIMO} separation system
\( x(p) \) \( ML \times 1 \) multichannel input signal vector for \textit{MISO} separation system
\( x_B(p) \) interferer reference signal (blocking matrix output signal)
\( x_m(p) \) \( L \times 1 \) time-reversed \( m \)th input signal vector
\( y(p) \) \( NL \times 1 \) vector representing the outputs of a MIMO separation system
\( y_{[i,...,j]}(p) \) \((j - i + 1)L \times 1 \) vector representing the output signals \( y_i(p), \ldots, y_j(p) \)
\( y_n(p) \) \( L \times 1 \) vector representing the \( n \)th output signal of a MIMO separation system
\( B \) blocking matrix in a GSC beamformer
\( C \) \( ML \times C \) constraint matrix, or block Sylvester mixing matrix of size \( NL \times N(L_m + 2L - 2) \) representing the global source–output MIMO system
\( C_{nn'} \) Sylvester mixing matrix of size \( L \times (L_m + 2L - 2) \) representing a source–output channel
\( D \) matrix used to define a geometric constraint
\( D(\omega, \theta) \) steering vector for the source position \( \theta \)
\( G \) matrix defining a general geometric constraint
\( G_0 \) constraint of zero spatial response at the interference DOAs
\( G_1 \) constraint of unit spatial response at the target position
\( G_2 \) constraint of unit spatial response at the target position and of zero spatial response at the interference DOA
\( H \) block Sylvester mixing matrix of size \( M(2L-1) \times N(L_m + 2L - 2) \)
\( H_{\text{int}} \) \((N - 1)(L + L_m - 1) \times ML \) interference mixing matrix
\( H_{mn} \) \( L + L_m - 1 \times L \) matrix representing the mixing channel \( h_{mn} \) used in \( H_{\text{int}} \)
\( H_{mn} \) \((2L-1) \times (L_m + 2L - 2) \) Sylvester matrix representing the mixing channel \( h_{mn} \)
\( I \) identity matrix
\( R_{\text{int}}(p) \) \( s_{\text{int}}(p) \) correlation matrix
\( R_{xx}(p) \) \( ML \times ML \) correlation matrix for the input signal vector \( x(p) \)
\( R_{x_B x_B}(p) \) correlation matrix for \( x_B(p) \)
\( \hat{R}_{x_B x_B}(p) \) estimation of \( R_{x_B x_B}(p) \)
\( R_{yy}(p) \) correlation matrix for the output vector \( y(p) \)
\( \hat{R}_{yy}(p) \) estimated correlation matrix for the output vector \( y(p) \)
\( R_{y_n y_n}(p) \) regularized correlation matrix for the output vector \( y_n(p) \) in the \( n \)th channel
\( S_{y_i y_j}^{(n)} \) regularized output correlation matrix
\( W \) block Sylvester matrix representing the entire MIMO separation system
\( W(\omega) \) DTFT of the separation system \( w \)
\( W(p) \) in sample-wise adaptive mode, the separation matrix \( W \) at time \( p \)
\( W(n) \) in batch mode, the separation matrix \( W \) at the \( n \)th iteration
\( W_n \) \( n \)th row of \( W \)
\( W(n, p) \) in block-wise batch adaptive mode, the separation matrix \( W \) at the \( n \)th iteration in the block at time \( p \)
\[ \Delta W(n) \text{ update for the matrix } W(n), \text{ that is, } W(n+1) = W(n) - \mu(n) \Delta W(n) \]

\[ \Delta W^{(BSS)} \text{ BSS update} \]

\[ \Delta W^{(geo)} \text{ geometric update} \]

\[ W_{nm} \text{ } L \times (2L - 1) \text{ Sylvester matrix representing the separation filter } w_{nm} \]

\[ W_{opt} \text{ optimum separation matrix (equilibrium point)} \]

\[ W_{opt} \text{ optimum separation matrix (defined as the minimum of } \xi_{BSS}(W) \]

\[ \hat{W}_{opt} \text{ estimate of } W_{opt} \]

\[ W \text{ multichannel } z\text{-transform of the Sylvester matrix } W \]

**Functions and operators**

\[ \text{bdiag} A \text{ for a block matrix } A, \text{ operator that sets the off-diagonal submatrices to } 0 \]

\[ \text{bdiag}_{1,[2,\ldots,M]} A \text{ operator setting the off-diagonal submatrices } A_{1m} \text{ and } A_{m1} \text{ for } m = 2,\ldots,M \text{ to } 0 \]

\[ \text{boff} A \text{ for a block matrix } A, \text{ operator that sets the diagonal submatrices to } 0 \]

\[ \det(A) \text{ determinant of the matrix } A \]

\[ D_g(W_{opt})(\varepsilon(n)) \text{ derivative of } g \text{ at point } W_{opt}, \text{ as a linear function of a small deviation} \]

\[ E\{\} \text{ expectation operator (ensemble average)} \]

\[ \hat{E}\{\} \text{ estimation of } E\{\} \]

\[ f^{(C)}(\cdot) \text{ } C\text{th composition of a function } f \]

\[ S(A) \text{ operator which transforms a general matrix } A \text{ into a Sylvester matrix by summing the redundant terms} \]

\[ S_d(A) \text{ operator which transforms a general matrix } A \text{ into a Sylvester matrix using the } d\text{th row as reference} \]

\[ S_L(A) \text{ operator which transforms a general matrix } A \text{ into a Sylvester matrix using the } L\text{th column as reference} \]

\[ S_{approx}(A) \text{ generic approximation of } S(A) \text{ representing either } S_d(A) \text{ or } S_L(A) \]

\[ \text{tr}(A) \text{ trace of the matrix } A \]

\[ \|x\| \sqrt{x^H x} \text{ (vector norm)} \]

\[ \|X\| \sqrt{\text{tr}(X^H X)} \text{ (matrix norm)} \]

\[ * \text{ convolution operator} \]

\[ *_{d} \text{ convolution where } d \text{ coefficients are treated as acausal and where the result is truncated on } L \text{ coefficients} \]

\[ [\overline{B}]_S(z) \text{ operator that truncates a } z\text{-transform } \overline{B} \text{ to a support } S \subset [-L + 1, L - 1] \]

\[ \langle \cdot, \cdot \rangle \text{ scalar product associated to the Euclidean metric} \]

\[ [x] \text{ smallest integer larger than } x \]

\[ [x] \text{ largest integer smaller than } x \]
Sets

N natural numbers
Z integers
R real numbers
C complex numbers
S space of the block Sylvester matrices of size $NL \times M(2L - 1)$
S set of the single-sided $z$-transforms of length $L$
T space of the block Toeplitz matrices of size $NL \times NL$
T set of the multichannel two-sided $z$-transforms

Greek letters

$\alpha$ regularization parameter
$\alpha(p)$ general contraction factor
$\beta$ $\beta L$ is the number of samples in the most recent input block for block-wise batch algorithms
$\delta$ fixed regularization term
$\delta_d$ vector representing a delay of $d$ taps
$\Delta$ interelement spacing for a uniform linear array
$\varepsilon_{\text{mismatch}}$ factor representing the interferer signal power at the beamformer output
$\varepsilon_{\text{leakage}}$ factor representing the amount of target leakage into the interference reference
$\varepsilon(n)$ small deviation around the equilibrium point $W_{\text{opt}}$
$\lambda$ geometric weight
$\lambda_i$ eigenvalue of $R_{yy}$
$\tilde{\lambda}_i$ error on the eigenvalue $\lambda_i$
$\lambda_i^{(n)}$ eigenvalue of $R_{yy}^{(n)}$, the output correlation matrix at iteration $n$
$\lambda_{\text{max}}$ largest eigenvalue of $\hat{R}_{\mathbf{x}B\mathbf{x}B}$ or $R_{\mathbf{x}B\mathbf{x}B}$
$\mu$ step-size
$\tilde{\mu}$ normalized step-size
$\mu_0$ step-size for the ILMS algorithm
$\mu_{\text{LMS}}$ step-size for the LMS algorithm
$\mu_{\text{max}}$ maximal step-size for the stability of ILMS
$\mu_{\text{NLMS}}$ step-size for the NLMS algorithm
$\theta$ direction of arrival
$\theta$ 3D source position
$\theta_1$ direction of arrival for the source of interest
$\sigma_1^2$ variance of the target signal at the beamformer output
$\sigma_2^2$ variance of the interference signal at the interference reference
$\tau_{m, \theta}$ delay needed for a sound wave emitted at $\theta$ to travel from the $m$th sensor to the next
$\omega$ angular frequency
E Notations

\( \xi_{BSS}(W) \) BSS cost function involving the expectation operator
\( \xi_{LMS}(w) \) least-mean-square cost function involving the expectation operator

**Special symbols**

◊ symbol denoting an unconstrained spatial response


References


Index

acoustic mixing, 7
  linear model, 7, 18, 149
physical parameters, 8
adaptive interference canceler, 30, see AIC
  ILMS-adapted, 42
adaptive interference canceller, 163, 165
  ILMS-adapted, 165, 166, 172, 175
array steering, 31
assumption
  far- and free-field propagation, 15, 32
  gaussian p.d.f., 71
  independence, 40, 65
instantaneous mixtures, 17, 74, 114, 116
linear model, 149
narrowband signal model, 108
no target leakage, 45, 46
sparse signals, 148
square systems, 75
stationarity, 12, 19, 39, 69
steered microphone array, 31
time-invariance, 10
two sources, 78
unit diagonal mixing channel, 21, 122
white signal, 46, 47
white signals, 123
zero-mean signals, 10

background noise, 4, 55, 168, 176
  SNR, 168
beamforming, 3, 125, 142
  concept, 27
  prior knowledge, 27, 38
beampattern, 16
blind beamforming, 125
blind source separation, 3
  ambiguities, 3, 66
  blind, 63
  combined with beamforming, 147, 163
  compared to beamforming, 125
convolutive mixtures, 120
cost function, 113
deflation approach, 77
delay-and-sum beamformer, 154
frequency domain, 108, 148, 175
parameter settings, 101, 102
block-wise adaptation, 40, 97, 134, 143
blocking matrix, 165
  BSS-adapted, 166
car environment, 2, 149, 157, 175
causality of the mixing/separation
  system, 60, 95, 102, 111, 123, 151, 159, 163
  as prior information, 150
delay-and-sum beamformer, 154
permutation ambiguity, 96, 126
source-microphone arrangement, 96, 175
cocktail-party problem, 1
compact microphone array, 2, 15, 23
complexity, 74, 78, 79, 91, 99, 133, 138
continuous adaptation, 39, 147, 165, 167, 172, 173
  implicit adaptation control, 42, 51, 61, 62
convergence, 113
  analysis, 113, 115, 116, 119, 127, 128, 175, 176
  global stability, 120, 121
  shape of the cost function, 126
convergence speed, 175
convolutive mixtures, 3, 66
cost function, 145
  BSS, 72, 108, 126, 130
  local minimum, 72, 126
  minimum, 11, 128
  mutual information, 70
  spurious minimum, 155
cross-talk, 122, 166, 175
decorrelation, 69, 116, 123
degrees of freedom, 17, 21, 25, 55, 60, 68, 122
  and prior information, 149
  and separation criterion, 155
delay-and-sum beamformer, 16, 33, 154, 167, 175
  unit spatial response, 153
DFT, 134
  block convolution, 134
Discrete-Time Fourier Transformation, 15
distortion, 2, 19, 58, 105, 108, 195
distributed microphone array, 2, 23
double-talk detector, 3, 36, 39, 53, 125, 133, 147, 165, 172, 173
eigenvalues, 41, 117, 128
  spread, 41
equilibrium point, 114, 121, 123
equivariance, 95, 114, 175
estimation variance, 129
experimental setup, 2
  source signals, 55

generalized sidelobe canceler, 30, 40
  adaptive interference canceler, 30
  blocking matrix, 30
  fixed beamformer, 30
generalized sidelobe canceller, 133
generalized sidelobe decorrelator, 161
gradient descent, 73, 126
holonomicity, 75, 116
human-machine interfaces, 1

ILMS, 42
  convergence, 48
  convergence vs. divergence, 47
  parameter settings, 57
  robustness against varying noise conditions, 57, 59
  stability, 50
  time-variant step-size, 44
  independence measure, 70
  instantaneous estimate, 13, 22, 100
  instantaneous mixtures, 3, 17, 66, 74, 114, 145
  interference signal, 1

  joint block diagonalization, 70

  Linearly Constrained Minimum Variance, 27
  closed-form solution, 29, 31, 127
  linear constraint, 28
  spatial constraint, 29

  LMS, 12, 40, 64, 148
  combined with BSS, 163
  convergence, 40
  cost function, 12, 130
  NLMS, 41

  microphone array, 2
  imperfections, 150

  Multiple-Input Multiple Output system (MIMO), 13, 66
  as a set of N MISO systems, 10, 152

  Multiple-Input Signal Output system (MISO, 10
  mutual information, 70

narrowband
  frequency-domain BSS, 3
  signal model, 4

  natural gradient, 74, 111
  application to BSS, 91
  approximation, 174, 176
  for convolutive mixtures, 174
  non-self-closed, 89, 103, 114, 174
  proportionality, 84
  self-closed, 87, 93, 103, 114, 174
  simplification, 91
NLMS
parameter settings, 57
normalization, 42, 119, 131, 140
null-steering beamformer, 33, 125
initialization with, 154

open issue, 96, 123, 176

partial separation, 76, 164, 174
performance
for a speech recognizer, 168
IR, 22
measures, 22
SIR, 22
SR, 22
permutation ambiguity, 68, 96, 126, 148, 151, 163, 175
prior information, 149, 175, 177
acoustic mixing model, 149
at the initialization, 150
multiple constraints, 162
preprocessing, 30, 160, 165
soft constraint, 157
strictness, 149, 154

quadratic inequality constraint, 52
regularization, 42, 100, 119, 120, 123, 175, 176
room acoustics, 1, 8, 20

SAD, 14, 63, 95, 140
caveats, 65
sample-wise adaptation, 40, 99, 134, 144
second-order statistics, 65, 69, 79
separation filters, 10
constraints, 17
length, 20
single channel noise reduction, 2
source
point source, 9
source silences, 128, 132, 175
and statistical super-efficiency, 140, 146

spatial response, 14
and causality, 159
as a constraint, 28, 152
spatial separation, 19, 20, 128, 142, 145
spectral separation, 19, 142, 148
speech recognition, 167, 175
speed of convergence, 40, 42, 47, 50, 98, 101, 110, 163
stability, 120, 121
global, 41–43, 113, 121
local, 121, 175
statistical efficiency, 131
and normalization, 132
super-efficiency, 130
step-size, 12, 24, 40, 42, 56, 61, 110, 113, 116, 119, 120, 128, 141, 157, 163, 175
Sylvester matrix, 68, 73, 81
z-transform, 85
redundancy, 81
Sylvester subspace, 82
gradient, 82
Symmetric Adaptive Decorrelation, see SAD
system identification, 46
target signal cancelation problem, 35, 37, 41, 63
energy-inversion effect, 36
target signal cancellation problem, 3, 164, 173
target leakage, 172
Toeplitz matrix
z-transform, 85
tracking capability, 36, 37, 61, 98, 105, 111, 138, 144, 163
unsupervised techniques, 3

Wiener solution, 31, 39, 127
word accuracy, 168
word-error rate (WER), 168