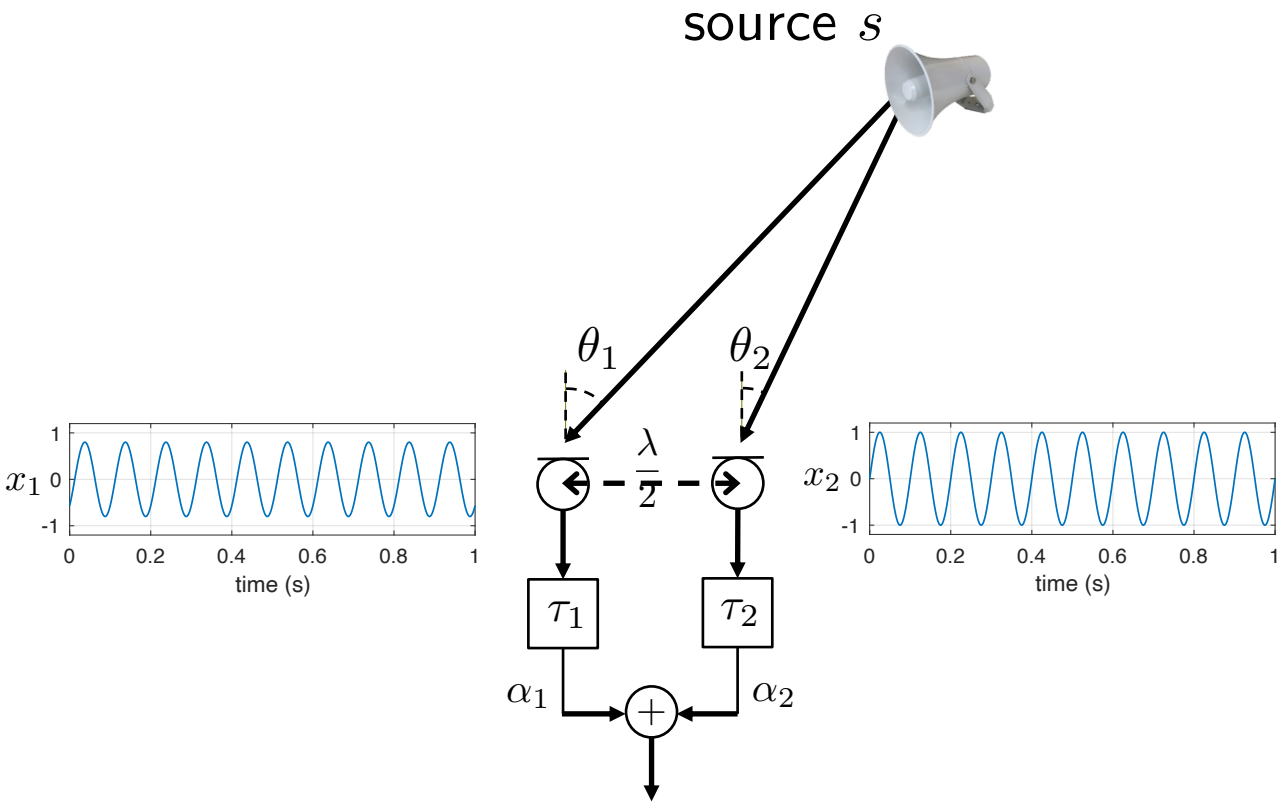
Clock-Offset and Microphone Gain Mismatch Invariant Beamforming

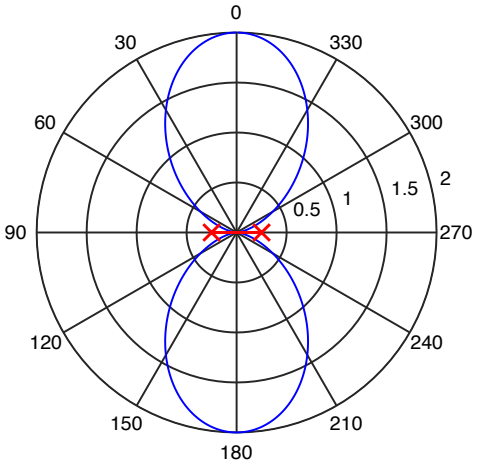
A Generalised Eigenvalue Decomposition Approach

Sofia-Eirini Kotti , Richard Heusdens and Richard C. Hendriks

Beamforming (narrow band, stationary)



$\theta_1 = \theta_2$ (far field)

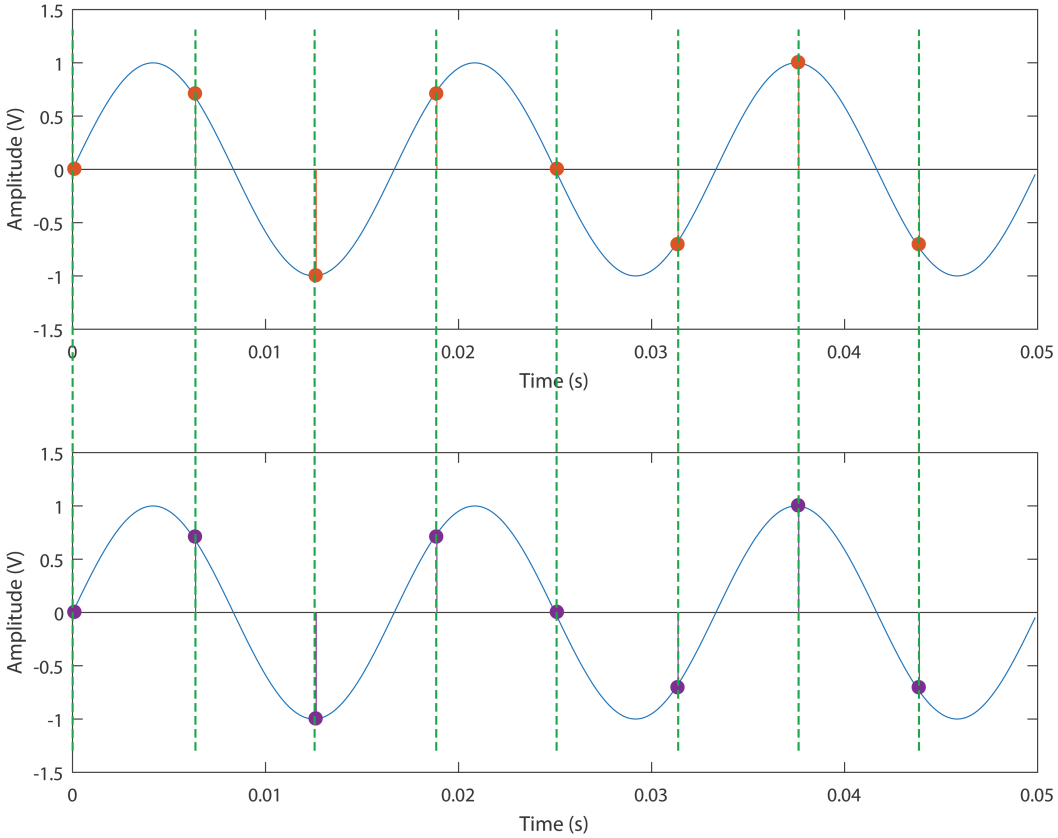


$$\tau_1 = \tau_2 = 0$$
$$\alpha_1 = \alpha_2 = 1$$

Clock synchronisation

Reference microphone

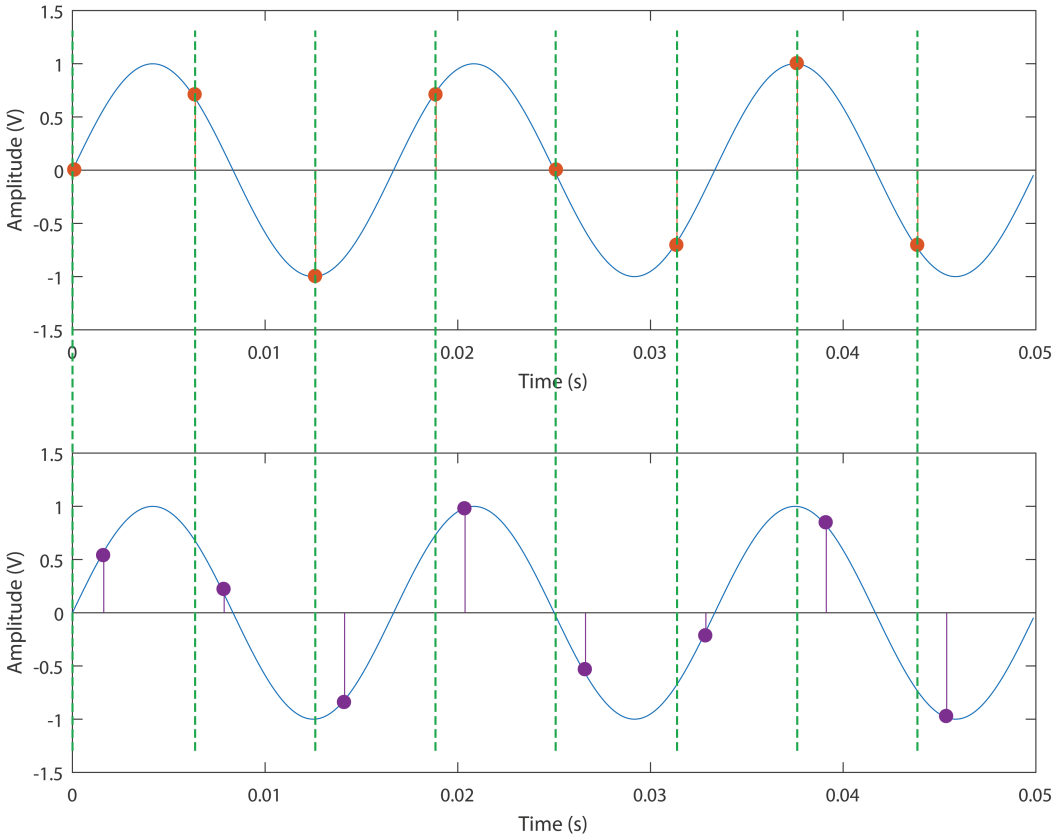
Second microphone



Clock-offset

Reference microphone

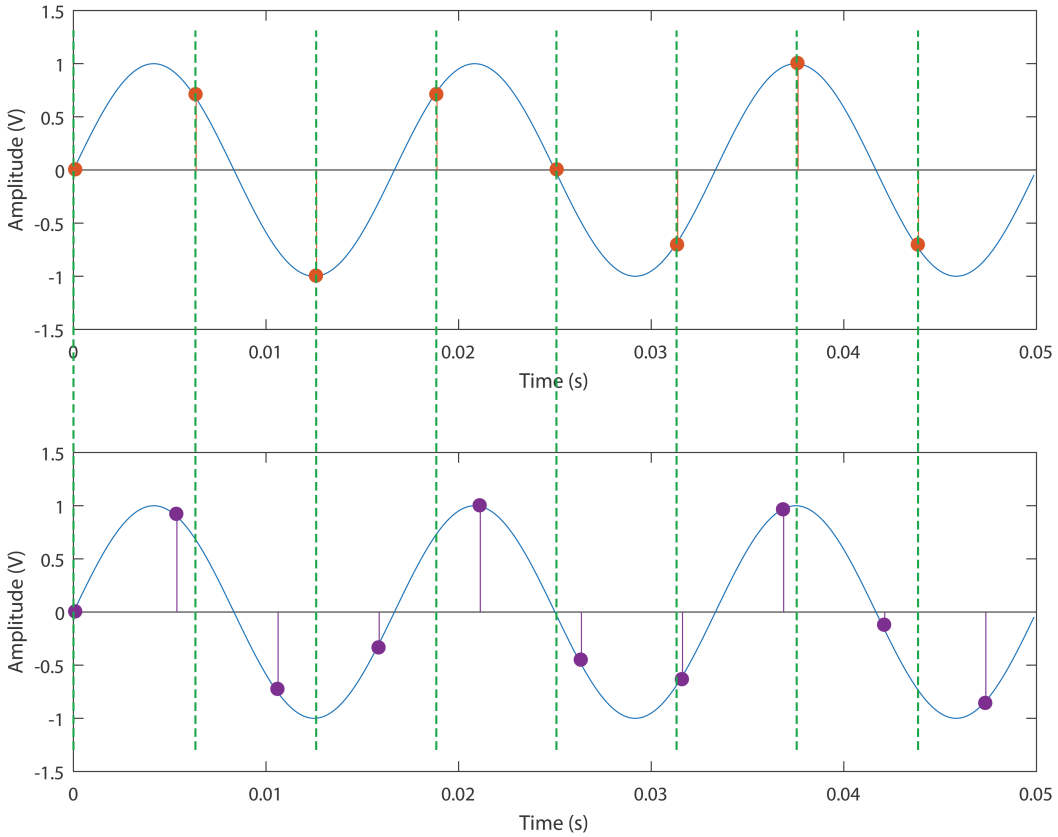
Second microphone



Clock-skew

Reference microphone

Second microphone



Clock synchronisation

Clock skew:

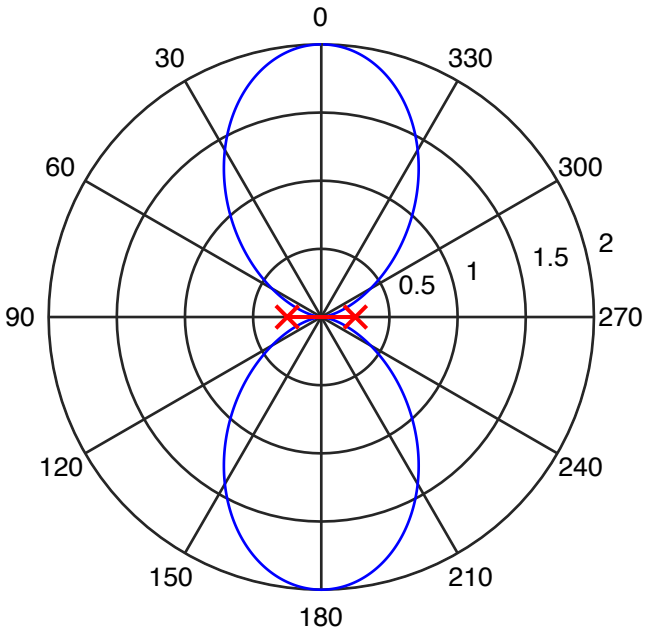
- Not a problem for uniform hardware in general
- Negligible when buffers are read out at regular time instances (no error aggregation)

Clock offset:

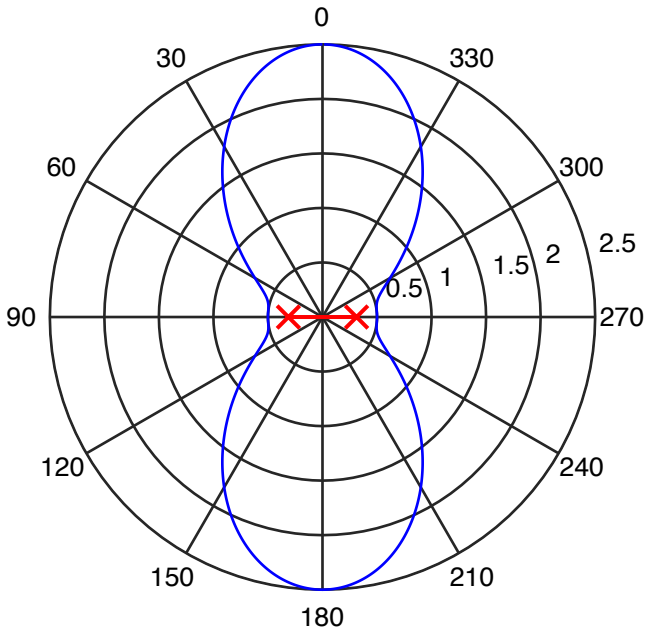
- Inherently present due to different onset times of the devices and/or internal sensor delays

We will focus on clock-offset only

Gain mismatch



$$g_1 = g_2 = 1$$



$$g_1 = 1, \;\; g_2 = 1.5$$

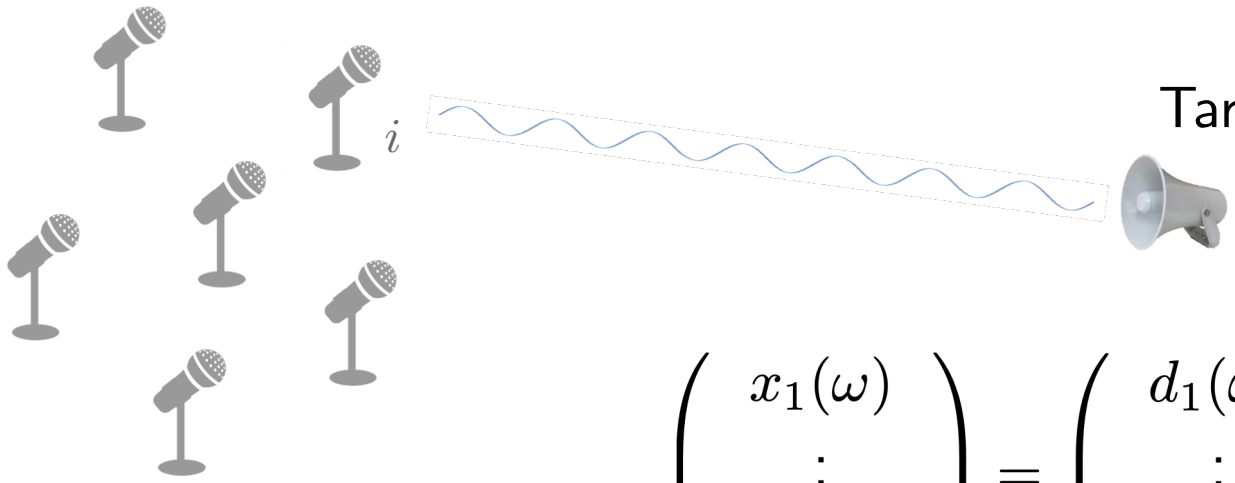
Outline

- Signal model
- Generalised eigenvalue decomposition (GEVD)
- GEVD-based optimal beamformers
- Clock-offset and gain invariance
- Experimental results

Signal model

Received signal:

$$y_i(\omega) = x_i(\omega) + v_i(\omega)$$



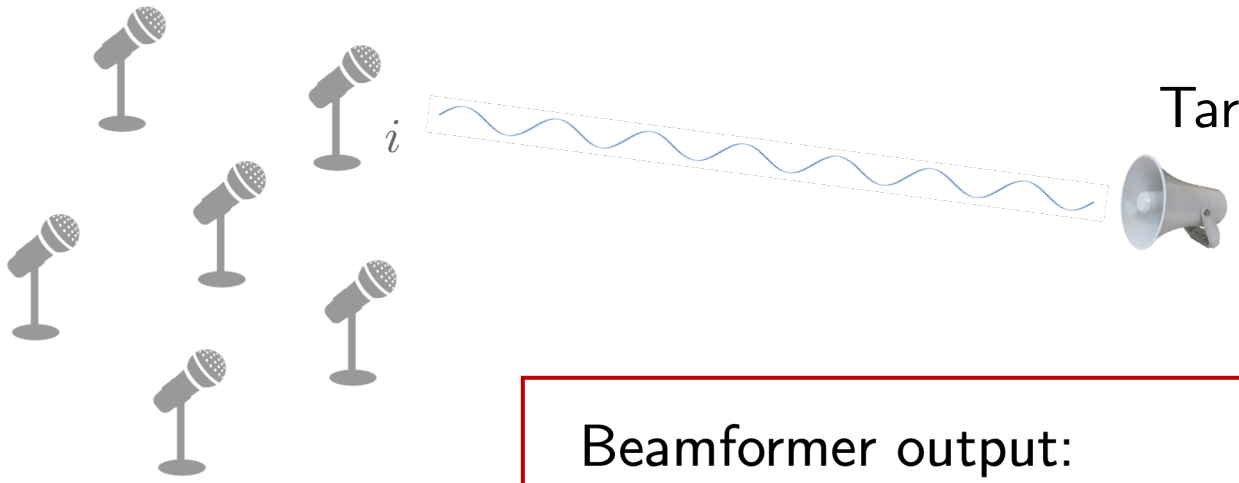
Target signal $s(\omega)$

$$\underbrace{\begin{pmatrix} x_1(\omega) \\ \vdots \\ x_m(\omega) \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} d_1(\omega) \\ \vdots \\ d_m(\omega) \end{pmatrix}}_{d} \cdot \; s(\omega)$$

Signal model

Received signal:

$$y_i(\omega) = x_i(\omega) + v_i(\omega)$$



Target signal $s(\omega)$

Beamformer output:

$$w^H y = w^H x + w^H v$$

Signal model

We will consider the received signals y, x, v and s as realisations of zero mean, wide-sense stationary processes Y, X, V, and S, respectively. We have

$$Y = X + V.$$

Assuming the noise and target are uncorrelated, the cross-power spectral density (CPSD) matrix of the received process Y is given by

$$R_Y = R_X + R_V,$$

where $R_Y = \mathrm{E}(YY^H)$ and R_X and R_V are defined similarly. The operator $\mathrm{E}(\cdot)$ denotes the expectation operator

Signal model

With this we have

$\text{rank}(R_X) = 1$

$$R_X = \mathrm{E}(XX^H) = \mathrm{E}(dSS^H d^H) = dd^H \mathrm{E}|S|^2 = \sigma_S^2 dd^H$$

If the noise is spatially uncorrelated ($R_V = \sigma_V^2 I_m$), we have

$$R_Y = R_X + R_V = U\Lambda U^{-1} + \sigma_V^2 I_m = U(\Lambda + \sigma_V^2 I_m)U^{-1}$$

Eigenvalue decomposition (EVD) of R_X

An estimate of R_X can be obtained by truncating the last $m-1$ eigenvalues of R_Y. What if $R_V \neq \sigma_V^2 I_m$?

Generalised eigenvalue decomposition

Given the Hermitian matrices $A, B \in \mathbb{C}^{n \times n}$ with $B \succ 0$, there exists a non-singular $U = (u_1, \ldots, u_n), u_i \in \mathbb{C}^n$, such that

$$U^H A U = \operatorname{diag}(a_1, \ldots, a_n) \quad \text{and} \quad U^H B U = \operatorname{diag}(b_1, \ldots, b_n).$$

Hence, we have $BU = U^{-H} \Lambda_B$ so that

$$AU = U^{-H} \Lambda_A = U^{-H} \Lambda_B \Lambda_B^{-1} \Lambda_A = BU\Lambda$$

That is, $Au_i = \lambda_i B u_i$ for $i = 1, \ldots, n$ where $\lambda_i = a_i / b_i$.

This decomposition is known as the *generalised eigenvalue decomposition (GEVD)*.

Generalised eigenvalue decomposition

Note that since $B \succ 0$ (B is invertible), we have

$$B^{-1}Au_i = \lambda_i u_i$$

Hence, the generalised eigenvalues and eigenvectors of (A, B) are the (ordinary) eigenvalues and eigenvectors of the matrix $B^{-1}A$.

Application to R_X and R_V, and setting $b_i = 1$ for all i, we have

$$U^H R_X U = \Lambda \quad \text{and} \quad U^H R_V U = I_M,$$

where $\Lambda \succeq 0$. Hence, the pair (Λ, U) are the eigenvalues/vectors of the matrix $R_V^{-1}R_X$.

Generalised eigenvalue decomposition

Again, since $R_Y = R_X + R_V$, we have

$$U^H R_Y U = \Lambda + I_M \quad \Leftrightarrow \quad R_Y = U^{-H}(\Lambda + I_M)U^{-1}$$

Note that the matrix $R_V^{-1} R_X$ is *not* Hermitian, and as such $U^{-1} \neq U^H$.

Similar to what we did before, we can estimate R_X by truncating the last $m-1$ generalised eigenvalues of R_Y

Optimal beamformers

Consider the mean squared-error (MSE) between the beamformer output and the desired target signal at the reference microphone, which we will assume, without loss of generality, to be microphone 1.

We have

$$\begin{aligned}\mathrm{E}|w^H Y - X_1|^2 &= \mathrm{E}|w^H X + w^H V - X_1|^2 \\ &= \mathrm{E}|w^H X - X_1|^2 + \mathrm{E}|w^H V|^2,\end{aligned}$$

where we used the property $\mathrm{E}(XV^H) = 0$. The term $\mathrm{E}|w^H X - X_1|^2$ represents the *signal distortion*, whereas the term $\mathrm{E}|w^H V|^2$ represents the *residual noise variance*

Optimal beamformers

We can compromise between signal distortion and noise reduction by defining the constraint optimisation problem

$$\begin{aligned} \text{minimise} \quad & \mathrm{E}|w^H X - X_1|^2 \\ \text{subject to} \quad & \mathrm{E}|w^H V|^2 \leq c, \end{aligned}$$

where $0 \leq c \leq \sigma_{V_1}^2$ and $\sigma_{V_1}^2$ the noise variance at the reference microphone before beamforming.

Optimal beamformers

Express the beamformer weights in terms of the generalised eigenvectors:

$$w = Ua \quad \text{with} \quad a \in \mathbb{C}^m.$$

Let $e_1 = (1, 0, \ldots, 0)^T \in \mathbb{C}^m$. With this we have $x_1 = e_1^H x$ so that we can express the objective function as

$$\begin{aligned} \mathrm{E}|w^H X - X_1|^2 &= \mathrm{E}|a^H U^H X - e_1^H X|^2 \\ &= a^H \Lambda a + \sigma_{X_1}^2 - 2\mathrm{Re}\{a^H U^H R_X e_1\}, \end{aligned}$$

and the feasible set becomes $\{a \in \mathbb{C}^m : a^H a \leq c\}$.

Optimal beamformers

Hence, the constraint optimisation problem can be expressed as

$$\begin{aligned} \text{minimise} \quad & a^H \Lambda a - 2\mathrm{Re}\{a^H U^H R_X e_1\} \\ \text{subject to} \quad & a^H a \leq c. \end{aligned}$$

The optimal solution, say a^*, is given by

$$a^* = (\Lambda + \mu I_m)^{-1} U^H R_X e_1,$$

where $\mu \geq 0$ is a Lagrange multiplier. As a consequence, we have

$$w^* = U(\Lambda + \mu I_m)^{-1} U^H R_X e_1.$$

Low-rank multichannel Wiener filter

Let $U^{-H} = Q = (q_1, \ldots, q_m)$, $q_i \in \mathbb{C}^m$. With this, we can express R_X as

$$R_X = U^{-H} \Lambda U^{-1} = Q \Lambda Q^H,$$

so that

$$w^* = U(\Lambda + \mu I_m)^{-1} \Lambda Q^H e_1.$$

since $U^H Q = I_m$.

If $\text{rank}(R_X) = r < m$ $(\lambda_{r+1}, \ldots, \lambda_m = 0)$, we can select the first r vectors $q_1, \ldots, q_r$. That is, $R_X = Q_r \Lambda_r Q_r^H$, and the optimal filters become

$$w^* = U_r(\Lambda_r + \mu I_r)^{-1} \Lambda_r Q_r^H e_1.$$

Optimal beamformers

Remarks:

- The case $\mu = 1$ and $r = m$ gives the classical *multi-channel Wiener filter*.
- If $r = 1$, we have $w^* = \alpha u_1$, $\alpha \neq 0$, which leads to the *maximum SNR beamformer*. This holds for every $\mu \geq 0$.
- If $\text{rank}(R_X) = r \leq m$ and we choose $\mu = 0$, the result will be the *MVDR beamformer*.

Clock offset and gain mismatch

Let τ_i denote the clock offset of the ith microphone with respect to the reference microphone, so that $\tau_1 = 0$. Moreover, let g_i denote the gain of microphone i and assume, without loss of generality, that the gain of the reference microphone is $g_1 = 1$.

With this, the received microphone signals can be expressed as $\tilde{y} = Ty$ where

$$T = \mathrm{diag}(1, g_2 e^{j\omega\tau_2}, \cdots, g_m e^{j\omega\tau_m}).$$

As a consequence, since $y = x + v$, we have $\tilde{y} = T(x + v) = \tilde{x} + \tilde{v}$.

Clock offset and gain mismatch

Let $\tilde{R}_X$ and $\tilde{R}_V$ denote the CPSD matrices of the *unsynchronised* target and noise process, respectively. Since $\tilde{X} = TX$ and $\tilde{V} = TV$, we have

$$\tilde{R}_X = \mathrm{E}(\tilde{X}\tilde{X}^H) = T\mathrm{E}(XX^H)T^H = TR_XT^H,$$

and similarly we find $\tilde{R}_V = TR_VT^H$. Hence,

$$\tilde{R}_V^{-1}\tilde{R}_X = (TR_VT^H)^{-1}(TR_XT^H) = T^{-H}R_V^{-1}R_XT^H,$$

and we conclude that $\tilde{R}_V^{-1}\tilde{R}_X$ is similar to $R_V^{-1}R_X$ and, therefore, share the same eigenvalues.

Clock offset and gain mismatch

We have the following result.

Proposition 1. Let $\tilde{U}^H \tilde{R}_X \tilde{U} = \tilde{\Lambda}$ and $\tilde{U}^H \tilde{R}_V \tilde{U} = I_m$ be the GEVD of $(\tilde{R}_V, \tilde{R}_X)$. Then $\tilde{\Lambda} = \Lambda$ and $\tilde{U} = T^{-H} U B$, where $B = \mathrm{diag}(B_1, \cdots, B_k)$, $B_i \in \mathbb{C}^{m_i \times m_i}$ unitary, and m_i denotes the algebraic multiplicity of λ_i and k the number of distinct eigenvalues $(\sum_{i=1}^{k} m_i = m)$.

Clock offset and gain mismatch

Proof:

Since $\tilde{R}_V^{-1}\tilde{R}_X = \tilde{U}\tilde{\Lambda}\tilde{U}^{-1}$ is similar to $R_V^{-1}R_X = U\Lambda U^{-1}$, we conclude that $\tilde{\Lambda} = \Lambda$. In addition, since $\tilde{R}_V^{-1}\tilde{R}_X = T^{-H}R_V^{-1}R_X T^H = T^{-H}U\Lambda(T^{-H}U)^{-1}$ and the fact that eigenvectors associated to λ_i are unique up to an invertible transform $B_i \in \mathbb{C}^{m_i \times m_i}$, we conclude that $\tilde{U} = T^{-H}UB$ where $B = \mathrm{diag}(B_1, \cdots, B_k)$, $B_i \in \mathbb{C}^{m_i \times m_i}$ invertible. Moreover, since

$$I_m = \tilde{U}^H \tilde{R}_V \tilde{U} = (B^H U^H T^{-1})(T R_V T^H)(T^{-H} U B) = B^H B,$$

we conclude that B is unitary, which completes the proof. □

Clock offset and gain mismatch

We calculate the optimal beamformers based on *unsynchronised* data. Application of Proposition 1 then yields

$$\begin{aligned}\tilde{w}^* &= \tilde{U}_r(\tilde{\Lambda}_r + \mu I_r)^{-1}\tilde{\Lambda}_r\tilde{Q}_r^H e_1 \\ &= T^{-H}U_r B_r(\Lambda_r + \mu I_r)^{-1}\Lambda_r B_r^H Q_r^H T^H e_1 \\ &\overset{(a)}{=} T^{-H}U_r(\Lambda_r + \mu I_r)^{-1}\Lambda_r Q_r^H e_1 \\ &= T^{-H}w^*,\end{aligned}$$

where (a) follows from the fact that B_r and $(\Lambda_r + \mu I_r)^{-1}\Lambda_r$ commute and $T^H e_1 = e_1$.

Clock offset and gain mismatch

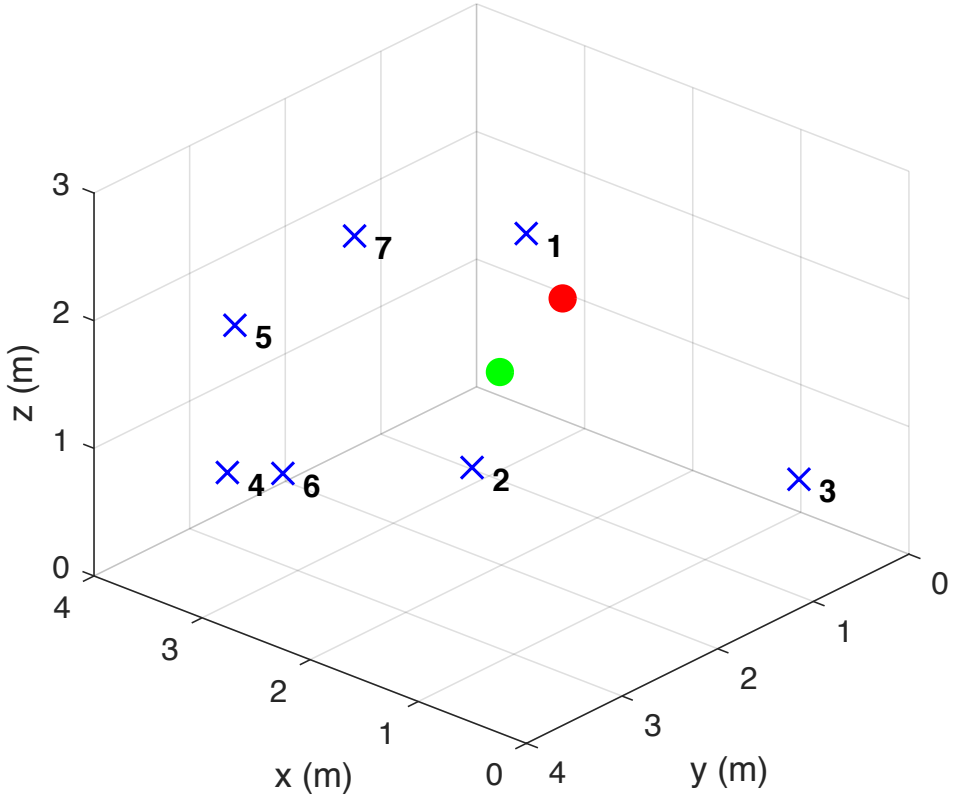
The output of the beamformer, $\tilde{w}^{*H}\tilde{y}$, then becomes

$$\tilde{w}^{*H}\tilde{y} = w^{*H}T^{-1}Ty = w^{*H}y,$$

and we conclude that *the GEVD-based beamformers are invariant to clock offsets and gain variations* and produce the same target estimate as if the clocks were perfectly synchronised and gains were perfectly equalised.

There is no need for clock-offset compensation and microphone gain equalisation !!!

Experimental results



Experimental results

- Target and interferer (both speech) signals, sampled at a sampling frequency of 16 kHz, were taken form the TIMIT database.
- The signals had a duration of 5 seconds and the signal-to-interferer ratio (SIR) at the reference microphone ($i = 1$) was set to 0 dB.
- The microphone-self noise was white Gaussian noise with 40 dB SNR.
- Processing of the signals was done on a frame-by-frame basis using a 30 ms, 50% overlap, Hann window.
- The covariance matrices R_Y and R_V were estimated by their sample covariance matrix.

Experimental results

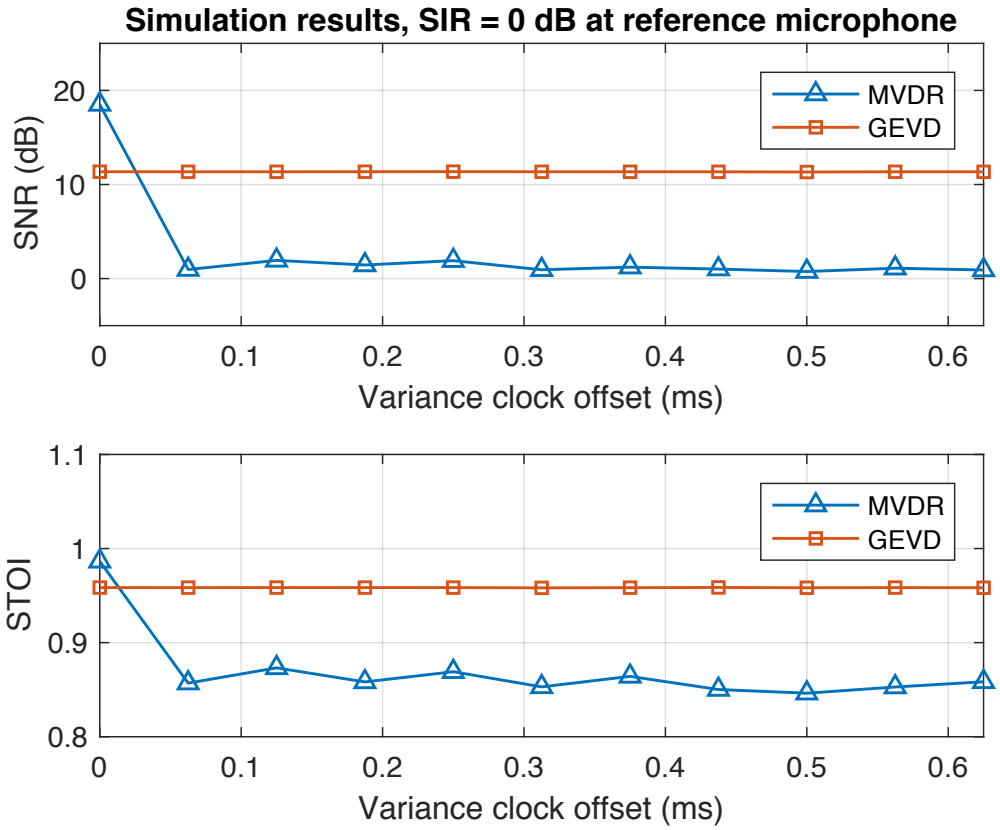
- The beamformer parameters were set to $r = 1$ and $\mu = 0$ which corresponds to the MVDR beamformer:

$$w_{\mathrm{MVDR}} = \frac{R_V^{-1} d}{d^H R_V^{-1} d}.$$

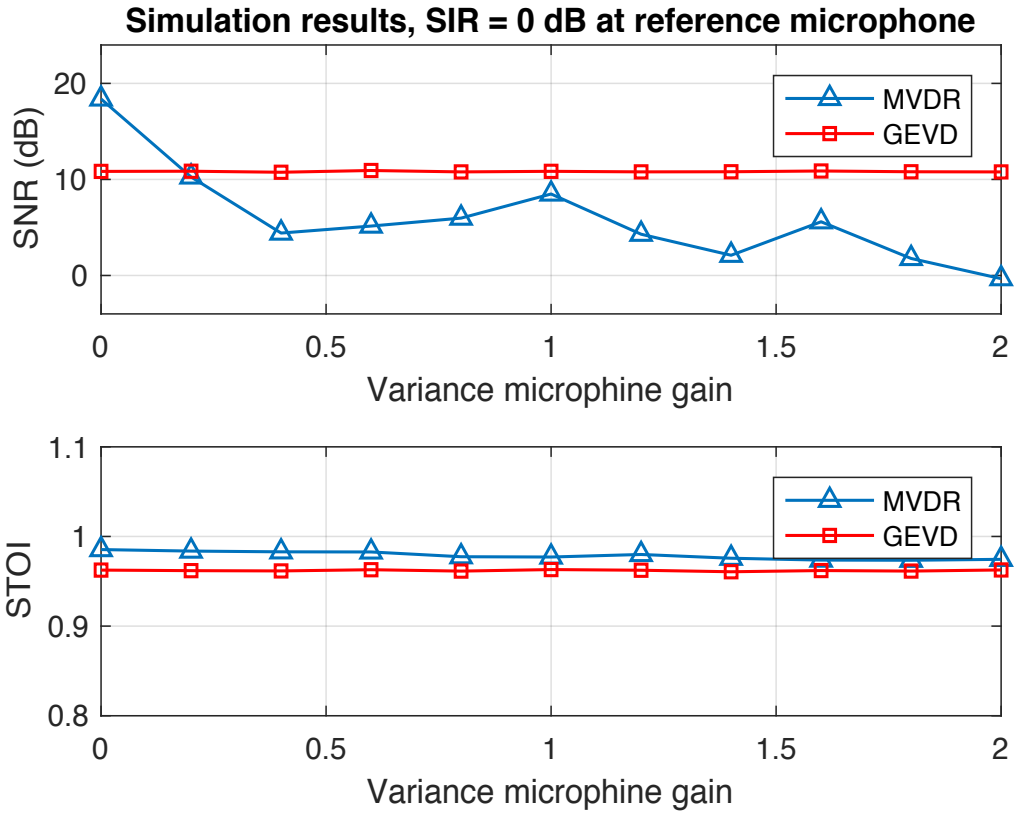
- The beamformer performance is evaluated in terms of both SNR and STOI scores as a function of the variance of the clock offset, where the SNR is defined as

$$\mathrm{SNR} = 10 \log \left(\frac{\|x_1\|_2^2}{\|w^H y - x_1\|_2^2} \right) \quad (\mathrm{dB}).$$

Experimental results



Experimental results



Conclusions

- GEVD-based beamformers are invariant to clock offset and gain mismatch
- There is no need to calibrate the microphones
- The steering vector is implicitly estimated!
- Don't need to know the sensor nor target locations!
- Seems to be a good tool for ad-hoc networks