Independent Component Analysis Using Singular Value Decomposition

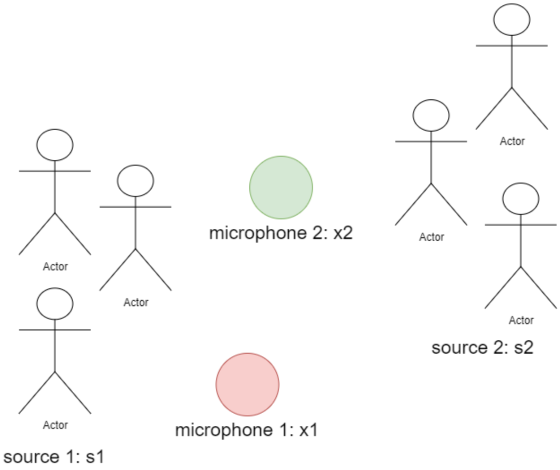
Department of Electrical & Computer Engineering,
University of Waterloo, ON, Canada

Data and Knowledge Modeling and Analysis (ECE 657A)

Course Instructor: Prof. Mark Crowley
TA and Presenter of Slides: Benyamin Ghojogh

Lecture Outline

Cocktail Party Effect & Blind Source Separation



Lecture Outline

Sources and Measured Signals

Assume we have two sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. Assume we have two measurements (sensors) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Each measurement is a linear combination of the sources:

$$\boldsymbol{x}_1 = a_{11}\boldsymbol{s}_1 + a_{12}\boldsymbol{s}_2, \tag{1}$$

$$\boldsymbol{x}_1 = a_{21}\boldsymbol{s}_1 + a_{22}\boldsymbol{s}_2. \tag{2}$$

If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are images, they are reshaped to column vectors, so as $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. We define:

$$\boldsymbol{x} := \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{s} := \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix} \tag{3}$$

Thus, we have:

$$\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \boldsymbol{s} = \boldsymbol{A}\,\boldsymbol{s} \tag{4}$$

$\boldsymbol{A}$ is called the **mixing matrix**.

Assumptions

We assume:

1. We center the measurements $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.
2. The sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are not Gaussian distributions. Usually natural images and data are not Gaussian.
3. The sources are statistically independent:
   $\boldsymbol{s}_1 \perp\!\!\!\perp \boldsymbol{s}_2 \implies \mathbb{P}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{P}(\boldsymbol{s}_1) \times \mathbb{P}(\boldsymbol{s}_2)$

Lecture Outline

Singular Value Decomposition

In $\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s}$, both $\boldsymbol{A}$ and $\boldsymbol{s}$ are unknown. Only $\boldsymbol{x}$ is known.

Let us apply Singular Value Decomposition (SVD) on the matrix $\boldsymbol{A}$:

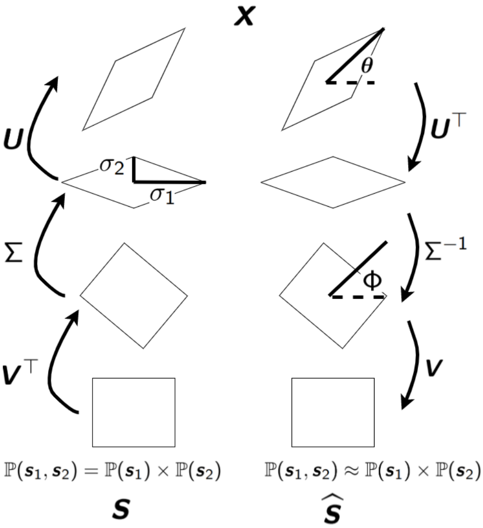
$$\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top \tag{5}$$

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} \implies \boldsymbol{x} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top\boldsymbol{s} \tag{6}$$

$$\widehat{\boldsymbol{s}} = \boldsymbol{A}^{-1}\boldsymbol{x} = (\boldsymbol{U}\Sigma\boldsymbol{V}^\top)^{-1}\boldsymbol{x} = \boldsymbol{V}^{-\top}\Sigma^{-1}\boldsymbol{U}^{-1}\boldsymbol{x} = \boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^\top\boldsymbol{x} \tag{7}$$

Singular Value Decomposition

$\boxed{\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top}$ and $\boxed{\widehat{\boldsymbol{s}} = \boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^\top\boldsymbol{x}}$ and $\boxed{\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} = \boldsymbol{U}\Sigma\boldsymbol{V}^\top\boldsymbol{s}}$



Lecture Outline

Step 1: on $\boldsymbol{U}^\top$

$\widehat{\boldsymbol{s}} = \boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^\top\boldsymbol{x}$, Let us consider $\boldsymbol{U}^\top\boldsymbol{x}$. We assume the statistical distributions $\mathbb{P}(\boldsymbol{s}_1)$ and $\mathbb{P}(\boldsymbol{s}_2)$ have zero mean. We minimize the second moment or the variance:

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\theta) &= \sum_{j=1}^{n}\Big([\boldsymbol{x}_1(j), \boldsymbol{x}_2(j)]\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix}\Big)^2 \\ &= \sum_{j=1}^{n}\big(\boldsymbol{x}_1(j)\cos(\theta) + \boldsymbol{x}_2(j)\sin(\theta)\big)^2 \\ &= \sum_{j=1}^{n}\big(\boldsymbol{x}_1^2(j)\cos^2(\theta) + 2\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)\sin(\theta)\cos(\theta) + \boldsymbol{x}_2^2(j)\sin^2(\theta)\big)\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathbb{V}\mathrm{ar}(\theta)}{\partial\theta} = \sum_{j=1}^{n}\Big(&-2\boldsymbol{x}_1(j)^2\cos(\theta)\sin(\theta) \\ &+ 2\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)\big(\cos^2(\theta) - \sin^2(\theta)\big) + 2\boldsymbol{x}_2^2\sin(\theta)\cos(\theta)\Big)\end{aligned}$$

Step 1: on $\boldsymbol{U}^\top$

$$\begin{aligned}
\frac{\partial \mathbb{V}\mathrm{ar}(\theta)}{\partial \theta} &= \sum_{j=1}^{n} \Big(-2\boldsymbol{x}_1(j)^2 \cos(\theta)\sin(\theta) \\
&\quad + 2\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)\big(\cos^2(\theta) - \sin^2(\theta)\big) + 2\boldsymbol{x}_2^2 \sin(\theta)\cos(\theta)\Big) \\
&= \sum_{j=1}^{n} \Big((\boldsymbol{x}_2^2(j) - \boldsymbol{x}_1^2(j))\sin(2\theta) + 2\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)\cos(2\theta)\Big) \overset{\text{set}}{=} 0 \\
&\implies \sin(2\theta)\sum_{j=1}^{n}(\boldsymbol{x}_2^2(j) - \boldsymbol{x}_1^2(j)) = -2\cos(2\theta)\sum_{j=1}^{n}\boldsymbol{x}_1(j)\boldsymbol{x}_2(j) \\
&\implies \tan(2\theta) = \frac{-2\sum_{j=1}^{n}\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)}{\sum_{j=1}^{n}(\boldsymbol{x}_2^2(j) - \boldsymbol{x}_1^2(j))} \\
&\implies \boxed{\theta = \frac{1}{2}\tan^{-1}\Big(\frac{-2\sum_{j=1}^{n}\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)}{\sum_{j=1}^{n}(\boldsymbol{x}_2^2(j) - \boldsymbol{x}_1^2(j))}\Big)} \qquad (8)
\end{aligned}$$

Step 1: on $\boldsymbol{U}^\top$

$$\boxed{\theta = \frac{1}{2}\tan^{-1}\left(\frac{-2\sum_{j=1}^{n}\boldsymbol{x}_1(j)\boldsymbol{x}_2(j)}{\sum_{j=1}^{n}(\boldsymbol{x}_2^2(j)-\boldsymbol{x}_1^2(j))}\right)} \tag{9}$$

In polar coordinate, we can represent measurements as $\boldsymbol{x}_1 = \boldsymbol{r}\cos(\psi)$ and $\boldsymbol{x}_2 = \boldsymbol{r}\sin(\psi)$, so:

$$\boxed{\theta = \frac{1}{2}\tan^{-1}\left(\frac{\sum_{j=1}^{n} r^2(j)\sin(2\psi_j)}{\sum_{j=1}^{n} r^2(j)\cos(2\psi_j)}\right)} \tag{10}$$

$$\boldsymbol{U} = \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix} \tag{11}$$

$$\implies \boxed{\boldsymbol{U}^\top = \begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix}} \tag{12}$$

Lecture Outline

Step 2: on Σ^{-1}

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{13}$$

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} \tag{14}$$

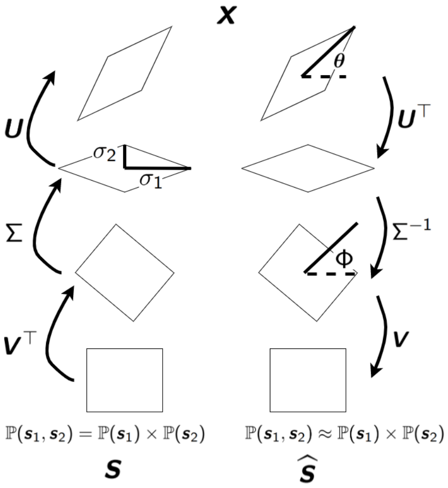
$$\boxed{\Sigma^{-1} \approx \begin{bmatrix} \frac{1}{\rho_1} & 0 \\ 0 & \frac{1}{\rho_2} \end{bmatrix}} \tag{15}$$

$$\boxed{\rho_1 = \sum_{j=1}^{n} \left([\boldsymbol{x}_1(j), \boldsymbol{x}_2(j)] \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \right)^2} \tag{16}$$

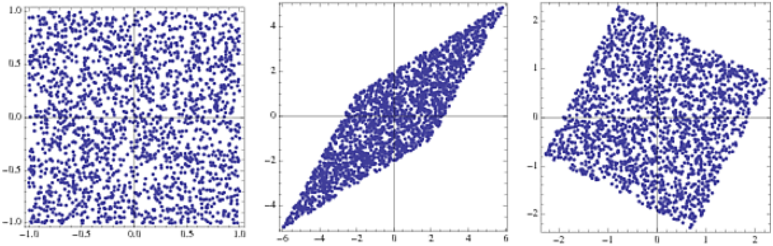
$$\boxed{\rho_2 = \sum_{j=1}^{n} \left([\boldsymbol{x}_1(j), \boldsymbol{x}_2(j)] \begin{bmatrix} \cos(\theta - \frac{\pi}{2}) \\ \sin(\theta - \frac{\pi}{2}) \end{bmatrix} \right)^2} \tag{17}$$

Whitening

Recall $\widehat{\boldsymbol{s}} = \boldsymbol{V}\Sigma^{-1}\boldsymbol{U}^\top \boldsymbol{x}$. The steps 1 and 2 result in whitening: $\Sigma^{-1}\boldsymbol{U}^\top \boldsymbol{x}$. After whitening, the covariance matrix of measurements becomes the identity matrix (see below figure).



Whitening Example



The credit of this image is for https://dsp.stackexchange.com/questions/80/

what-are-the-proper-pre-processing-steps-to-perform-independent-component-analys

Lecture Outline

Step 3: on $\boldsymbol{V}$

The first moment was assumed to be zero. The second moment was taken care of in step 1. Usually natural images or data are not much skewed so we ignore the third moment (skewness). The fourth moment is Kurtosis which is a statistical measure that defines how heavily the tails of a distribution differ from the tails of a Gaussian distribution. We have assumed $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ do not have Gaussian distributions so their Kurtosis is not zero.

$$K(\phi) = \sum_{j=1}^{n} \left([\boldsymbol{x}_1, \boldsymbol{x}_2] \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix} \right)^4 \tag{18}$$

$$\frac{\partial K(\phi)}{\partial \phi} \overset{\text{set}}{=} 0 \implies \boxed{\phi = \frac{1}{4} \tan^{-1} \left(\frac{\sum_{j=1}^{n} r^2(j) \sin(4\psi_j)}{\sum_{j=1}^{n} r^2(j) \cos(4\psi_j)} \right)} \tag{19}$$

$$\boxed{\boldsymbol{V} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}} \tag{20}$$

Lecture Outline

Examples

Blind source separation for:

- **music** separation:
  `https://cnl.salk.edu/~tewon/Blind/blind_audio.html`
- **speech** separation:
  `https://cnl.salk.edu/~tewon/Blind/blind_audio.html`
- **EEG** separation: [1]
- **fMRI** separation [2]
- learning independent **filters** on natural **images**:
  `https://pydeep.readthedocs.io/en/latest/tutorials/ICA_natural_images.html`

Useful Resources To Read

- ICA using **Singular value decomposition**: Tutorial YouTube videos by Prof. J. Nathan Kutz: [Click here] and [Click here] and [Click here]. **These slides are based on his videos.**
- ICA using **maximum likelihood estimation**: Tutorial YouTube videos by Prof. Andrew Ng at the Stanford University: [Click here]
- Survey paper: "Survey on independent component analysis" [3]
- Tutorial paper: "A tutorial on independent component analysis" [4]
- Tutorial paper: "Independent component analysis: A tutorial" [5]
- Survey paper: "An overview of independent component analysis and its applications" [6]
- A book on ICA: [7]

References

[1] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in neural information processing systems*, pp. 145–151, 1996.

[2] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human brain mapping*, vol. 6, no. 3, pp. 160–188, 1998.

[3] A. Hyvärinen, "Survey on independent component analysis," 1999.

[4] J. Shlens, "A tutorial on independent component analysis," *arXiv preprint arXiv:1404.2986*, 2014.

[5] A. Hyvärinen and E. Oja, "Independent component analysis: A tutorial," 1999.

[6] G. R. Naik and D. K. Kumar, "An overview of independent component analysis and its applications," *Informatica*, vol. 35, no. 1, 2011.

[7] J. V. Stone, *Independent component analysis: a tutorial introduction*. MIT press, 2004.