Blind Signal Separation in the Presence of Gaussian Noise

Mikhail Belkin
Luis Rademacher
James Voss
Ohio State University

Cocktail Party Problem (Example)

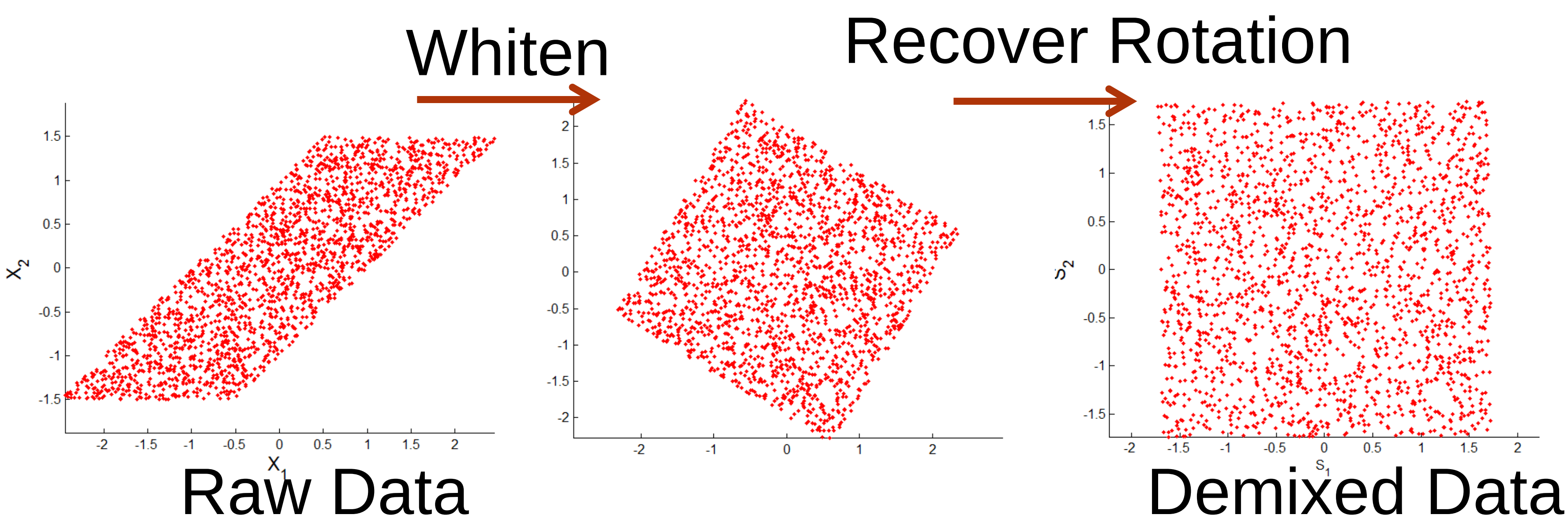
- Problem: n persons speaking in a room with n microphones.
- Microphones capture a superposition of the speech signals.
- Goal: Recover each persons' speech.

Independent Component Analysis (ICA)

- Observe samples from a random vector $\boldsymbol{X} \in \mathbb{R}^n$:
$$\boldsymbol{X} = A\boldsymbol{S} + \boldsymbol{\eta}$$
- $\boldsymbol{S} \in \mathbb{R}^n$ is a latent random vector with *independent coordinates*.
  - Coordinates variables are non-Gaussian.
  - Assumed $\mathrm{Cov}(\boldsymbol{S}) = I$.
- $A \in \mathbb{R}^{n \times n}$ and is full rank.
  - Columns form a spanning basis of $\mathbb{R}^n$.
  - $\boldsymbol{S}_i$ acts in the direction A_i.
- $\boldsymbol{\eta}$ is an additive noise.
  - Traditionally assumed 0.
  - We model as unknown, Gaussian noise independent of $\boldsymbol{S}$.
- Goal: Recover A.

Typical (noiseless) ICA Procedure

- Step 1: Whiten data to have covariance I.
  - Data is left multiplied by a matrix W such that $WA = R$ (a rotation).
  - Orthogonalizes the latent signals.
  - Recovers A up to a rotation matrix.
- Step 2: Find the rotation.

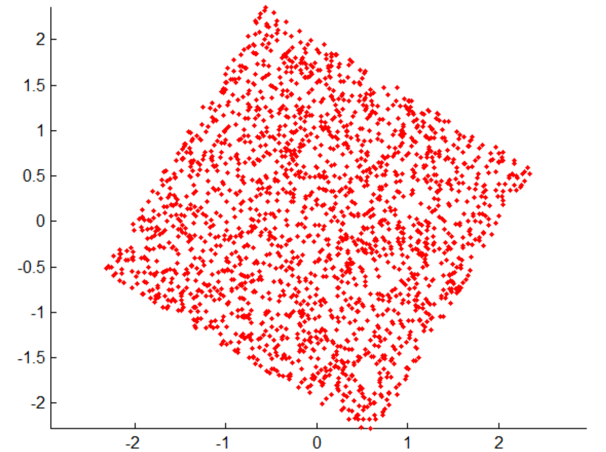


Data mixed from Uniform Distributions

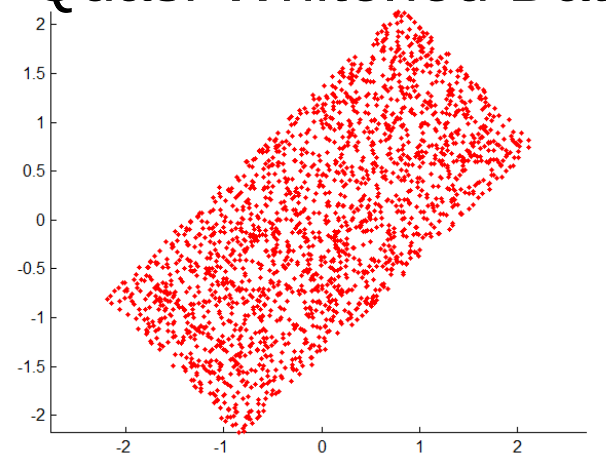
Relaxing Step 1

- Whitening the latent signal $A\boldsymbol{S}$ is impossible in the noisy case.
  - Let $\boldsymbol{\xi}$ be a white, Gaussian r.v.
  - $A(\boldsymbol{S}+\boldsymbol{\xi})+\boldsymbol{\eta}$ vs. $A\boldsymbol{S}+(A\boldsymbol{\xi}+\boldsymbol{\eta})$.
- **Definition:** If $WA = RD$ where R is a rotation matrix, and D is a diagonal scaling matrix, then W is a **quasi-whitening** matrix.
  - Recall model: $\boldsymbol{X} = A\boldsymbol{S}+\boldsymbol{\eta}$
- Applying W to data orthogonalizes the latent components.

Whitened Data



Quasi-Whitened Data



Related Work (efficient noisy ICA)

- Aapo Hyvärinen (1999) discusses noisy ICA when the **noise covariance is known**.
- Arie Yeredor (2000) provides a one-step solution to noisy ICA using the Hessian of the directional 2^{nd} Characteristic Function.
- Arora, Ge, Moitra, and Sachdeva (2012) introduced quasi-whitening and provide an **efficient** noisy ICA algorithm for the special case where all latent signals have fourth cumulant of the same sign.
- Hsu and Kakade (2012) state a one-step solution to noisy ICA using the Hessian of the directional fourth cumulant.

Our Contribution

- We introduce an **efficient** quasi-whitening algorithm for noisy ICA with latent signals of non-zero fourth cumulants (possibly of mixed sign).
  - Relies on multivariate cumulant tensors.
- Compatible with variations of existing methods for Step 2.
  - $\kappa_4(y) \coloneqq \mu_4(y) - 3\mu_2(y)^2$ (central moments).
  - Restricting $\boldsymbol{v}$ to the unit sphere, the local maxima of
    $$\boldsymbol{v} \mapsto |\kappa_4(\boldsymbol{v} \cdot (RD\boldsymbol{S} + W\boldsymbol{\eta}))|$$
    give the columns of R.

What are Cumulants?

- Cumulants are functions of random variables, similar to moments.
  - Respect independence more nicely than moments.
  - Low order cumulants: mean, variance
  - Have natural sample versions (k-statistics).
- Let $Cum(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l)$ denote the cross-cumulant between random variables $\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k$, and $\boldsymbol{Y}_l$.
- Let $Q_{\boldsymbol{Y}}$ denote the 4D cumulant tensor with entries:

$$(Q_{\boldsymbol{Y}})_{ijkl} = Cum\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right)$$

- $\kappa_4(\boldsymbol{Y}_i) = Cum(\boldsymbol{Y}_i, \boldsymbol{Y}_i, \boldsymbol{Y}_i, \boldsymbol{Y}_i)$ gives the fourth univariate cumulant.

Properties of Multivariate Cumulants
(stated for fourth cumulant)

- (Symmetry) $Cum(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l)$ is invariant under permutation of indices.
- (Multilinearity) Let $\boldsymbol{Z}_i$ be a random variable, and let $\alpha \in \mathbb{R}$.

$$Cum\left(\boldsymbol{Y}_i + \boldsymbol{Z}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right)$$
$$= Cum\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right) + Cum\left(\boldsymbol{Z}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right),$$
$$Cum\left(\alpha \boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right) = \alpha Cum\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right).$$

- (Independence) If $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ are independent:

$$Cum\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{Y}_k, \boldsymbol{Y}_l\right) = 0.$$

  - Implies for $\boldsymbol{Y}$ and $\boldsymbol{Z}$ independent, $Q_{\boldsymbol{Y}+\boldsymbol{Z}} = Q_{\boldsymbol{Y}} + Q_{\boldsymbol{Z}}$.
- (Vanishing Gaussians) $Q_{\boldsymbol{\eta}} = 0$.
  - Cumulants order $r \geq 3$ are 0 for Gaussian random variables.

Quasi-Whitening Algorithm

- **Definition**: Let $M \in \mathbb{R}^{n\times n}$ be a matrix. Then, define an operation of tensors on matrices:

$$(Q_{\boldsymbol{X}} \circ M)_{ij} := \sum_{k,l=1}^{n} Cum(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l) m_{lk}$$

- The proposed algorithm:

Let $\mathbf{x}^{(i)}$ denote the i^{th} sample of $\mathbf{X}$.
Let $\hat{Q}_{\mathbf{X}}$ give the sample estimate of $Q_{\mathbf{X}}$.
function QuasiWhiten($\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}$)
    $C \leftarrow \hat{Q}_{\mathbf{X}} \circ I$
    $M \leftarrow \hat{Q}_{\mathbf{X}} \circ C^{-1}$
    Let $\hat{B}\hat{B}^T$ be a decomposition of M.
    return $\hat{B}^{-1}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(N)})$.
end function

Algorithm's Validity

Lemma: Let $M \in \mathbb{R}^{n\times n}$. Then $Q_X \circ M = ADA^T$ where D is diagonal, $d_{qq} = \kappa_4(\boldsymbol{S}_q)A_q^T M A_q$.

Proof:

- $(Q_{\boldsymbol{X}} \circ M)_{ij}$
$$= \left(Q_{A\boldsymbol{S}+\boldsymbol{\eta}} \circ M\right)_{ij} = (Q_{A\boldsymbol{S}} \circ M)_{ij}$$
$$= \sum_{k,l=1}^{n} Cum\left((A\boldsymbol{S})_i, (A\boldsymbol{S})_j, (A\boldsymbol{S})_k, (A\boldsymbol{S})_l\right) m_{lk}$$
$$= \sum_{k,l=1}^{n} \sum_{q=1}^{n} A_{iq}A_{jq} Cum\left(\boldsymbol{S}_q, \boldsymbol{S}_q, \boldsymbol{S}_q, \boldsymbol{S}_q\right) A_{kq}A_{lq}\, m_{lk}$$
- $Q_{\boldsymbol{X}} \circ M = ADA^T$

Algorithm's Validity

Theorem: Let $M = Q_X \circ (Q_X \circ I)^{-1}$. Then,

1. $M = ADA^T$ where D is diagonal, $d_{qq} = \frac{1}{\|A_q\|^2}$.
2. Let BB^T be a decomposition of M, then B^{-1} is a quasi-whitening matrix.

Proof sketch (of 2):

- $D^{\frac{1}{2}}$ is real valued.
- $BB^T = ADA^T \Rightarrow I = (B^{-1}AD^{\frac{1}{2}})(D^{\frac{1}{2}}A^T {B^{-1}}^T)$.
- $R = B^{-1}AD^{\frac{1}{2}}$ for R a rotation.
- $B^{-1}A = RD^{-\frac{1}{2}}$, giving B^{-1} is a quasi-whitening matrix.

Main Result

Theorem: Let $1-\delta$ give the probability of success, and let $\epsilon > 0$ be an error parameter. Let $\hat{Q}_{\boldsymbol{X}}$ give the sample estimate of $Q_{\boldsymbol{X}}$. Let $\hat{B}\hat{B}^T$ be a decomposition of $\hat{Q}_{\boldsymbol{X}} \circ (\hat{Q}_{\boldsymbol{X}} \circ I)^{-1}$. Given polynomial samples, with probability $1-\delta$, $\hat{B}^{-1}$ is an approximate quasi-matrix such that:

1. The latent coordinates are approximately orthogonalized. For i $\neq j$,
$$-\epsilon < \frac{\left\langle \hat{B}^{-1}A\boldsymbol{e}_i, \hat{B}^{-1}A\boldsymbol{e}_j \right\rangle}{\left\| \hat{B}^{-1}A\boldsymbol{e}_i \right\| \left\| \hat{B}^{-1}A\boldsymbol{e}_j \right\|} < \epsilon$$
2. The latent coordinates are scaled:
$$(1-\epsilon)\|A_i\|^2 \leq \left\| \hat{B}^{-1}A\boldsymbol{e}_i \right\|^2 \leq (1+\epsilon)\|A_i\|^2$$

Quasi-Whitening Algorithm Restated

Let $\mathbf{x}^{(i)}$ denote the i^{th} sample of $\mathbf{X}$.
Let $\hat{Q}_{\mathbf{X}}$ give the sample estimate of $Q_{\mathbf{X}}$.
function QuasiWhiten($\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}$)
    $C \leftarrow \hat{Q}_{\mathbf{X}} \circ I$
    $M \leftarrow \hat{Q}_{\mathbf{X}} \circ C^{-1}$
    Let $\hat{B}\hat{B}^T$ be a decomposition of M.
    return $\hat{B}^{-1}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(N)})$.
end function

- Provably efficient
- Performs the relaxed Step 1 of noisy ICA (quasi-whitening).
- Compatible with small variations on existing algorithms for Step 2 of ICA.

Thank You

Any Questions?