ℓ_1 Approaches to PCA and ICA

S. Keeling and K. Kunisch

Fifth MOBIS SFB Status Seminar

November 14–16, 2012

Steierische Modellierungswoche 2012

Projekt: Signalverarbeitung

Tutorium: Trennung von Datenquellen in unkorrelierte und unabhängige Komponenten

a.o.Univ.Prof. Mag.Dr. Stephen Keeling

`http://math.uni-graz.at/keeling/`

Literatur:

`http://cis.legacy.ics.tkk.fi/aapo/papers/IJCNN99_tutorialweb/`

Dokumentation:

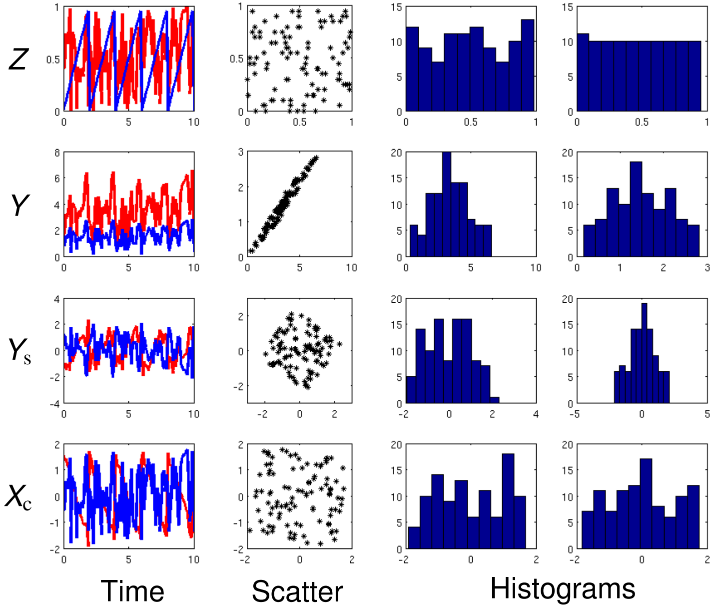
`http://math.uni-graz.at/keeling/skripten/Tutorium.pdf`

Dank an Herrn Dipl.-Ing. Dr. Gernot Reishofer
für seine Unterstützung für diese Arbeit!

Inhaltsverzeichnis

Graphical Demonstration of PCA/ICA

Sources Z, Measurements Y, sphered Y_s, separated X_c



Formulation of PCA/ICA

- Rows of Z are unknown samples of sources which are independent and not Gauß distributed.

$$Z = \begin{bmatrix} z_1(t_1) & z_1(t_2) & \cdots & z_1(t_n) \\ \vdots & \vdots & & \vdots \\ z_m(t_1) & z_m(t_2) & \cdots & z_m(t_n) \end{bmatrix}$$

Formulation of PCA/ICA

- Rows of Z are unknown samples of sources which are independent and not Gauß distributed.

$$Z = \begin{bmatrix} z_1(t_1) & z_1(t_2) & \cdots & z_1(t_n) \\ \vdots & \vdots & & \vdots \\ z_m(t_1) & z_m(t_2) & \cdots & z_m(t_n) \end{bmatrix}$$

- Rows of Y are measured samples of unknown mixtures of the sources

$$Y = AZ$$

no longer independent and now more Gauß distributed.

Formulation of PCA/ICA

- Rows of Z are unknown samples of sources which are independent and not Gauß distributed.

$$Z = \begin{bmatrix} z_1(t_1) & z_1(t_2) & \cdots & z_1(t_n) \\ \vdots & \vdots & & \vdots \\ z_m(t_1) & z_m(t_2) & \cdots & z_m(t_n) \end{bmatrix}$$

- Rows of Y are measured samples of unknown mixtures of the sources

$$Y = AZ$$

no longer independent and now more Gauß distributed.

- Goal is to undo the trend toward Gaußianity to recover the sources

$$X = WY$$

with $W = U\Lambda^{-\frac{1}{2}}V^{\mathrm{T}} \approx A^{-1}$ but unavoidable ambiguity $x_i(t) \approx \pm z_k(t)$.

Formulation of PCA/ICA

Steps:

- Centering:

$$Y_{\mathrm{c}} = Y - \overline{Y}$$

Formulation of PCA/ICA

Steps:

- Centering:

$$Y_{\mathrm{c}} = Y - \overline{Y}$$

- Sphering:

$$K = \frac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$

Formulation of PCA/ICA

Steps:

- Centering:
$$Y_{\mathrm{c}} = Y - \overline{Y}$$
- Sphering:
$$K = \frac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$
- Rotation:
$$X_{\mathrm{c}} = UY_{\mathrm{c}}, \quad U^{\mathrm{T}} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$$
where each $\boldsymbol{u}_k$ minimizes Gaußianity.

Formulation of PCA/ICA

Steps:

- Centering:

$$Y_{\mathrm{c}} = Y - \overline{Y}$$

- Sphering:

$$K = \frac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$

- Rotation:

$$X_{\mathrm{c}} = UY_{\mathrm{c}}, \quad U^{\mathrm{T}} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$$

where each $\boldsymbol{u}_k$ minimizes Gaußianity.

For example, Kurtosis

$$\mathcal{W}(\boldsymbol{x}) = M_4(\boldsymbol{x}) - 3M_2^2(\boldsymbol{x})$$

satisfies $\mathcal{W}(\boldsymbol{n}) = 3\sigma^4 - 3\sigma^4 = 0$ for $\boldsymbol{n} \sim N(\mu, \sigma^2)$.

Formulation of PCA/ICA

Steps:

- Centering:
$$Y_{\mathrm{c}} = Y - \overline{Y}$$
- Sphering:
$$K = \tfrac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$
- Rotation:
$$X_{\mathrm{c}} = UY_{\mathrm{c}}, \quad U^{\mathrm{T}} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$$
where each $\boldsymbol{u}_k$ minimizes Gaußianity.

  For example, Kurtosis
$$\mathcal{W}(\boldsymbol{x}) = M_4(\boldsymbol{x}) - 3M_2^2(\boldsymbol{x})$$
satisfies $\mathcal{W}(\boldsymbol{n}) = 3\sigma^4 - 3\sigma^4 = 0$ for $\boldsymbol{n} \sim N(\mu, \sigma^2)$.

  So $J(\boldsymbol{u}) = -\mathcal{W}^2(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u})$ may be minimizied with $\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{u}_l = \delta_{kl}$.

Formulation of PCA/ICA

Steps:

- Centering:

$$Y_\mathrm{c} = Y - \overline{Y}$$

- Sphering:

$$K = \frac{1}{n} Y_\mathrm{c} Y_\mathrm{c}^\mathrm{T}, \quad KV = V\Lambda, \quad Y_\mathrm{s} = \Lambda^{-\frac{1}{2}} V^\mathrm{T} Y_\mathrm{c}$$

- Rotation:

$$X_\mathrm{c} = UY_\mathrm{c}, \quad U^\mathrm{T} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$$

  where each $\boldsymbol{u}_k$ minimizes Gaußianity.

  For example, Kurtosis

$$\mathcal{W}(\boldsymbol{x}) = M_4(\boldsymbol{x}) - 3M_2^2(\boldsymbol{x})$$

  satisfies $\mathcal{W}(\boldsymbol{n}) = 3\sigma^4 - 3\sigma^4 = 0$ for $\boldsymbol{n} \sim N(\mu, \sigma^2)$.

  So $J(\boldsymbol{u}) = -\mathcal{W}^2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u})$ may be minimizied with $\boldsymbol{u}_k^\mathrm{T}\boldsymbol{u}_l = \delta_{kl}$.

- Shift:

$$X = X_\mathrm{c} + U\Lambda^{-\frac{1}{2}} V^\mathrm{T}\overline{Y} = U\Lambda^{-\frac{1}{2}} V^\mathrm{T} Y = WY$$

Formulation of PCA/ICA

(PCA) Let the data be so decomposed,

$$Y_{\mathrm{c}} = Y - \overline{Y}, \quad K = \frac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$

Let $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$ with $\lambda_1 \geq \cdots \geq \lambda_m$. With $P \in \mathbb{R}^{r \times m}$, $r < m$, $P_{i,j} = \delta_{i,j}$, the data Y are so projected to its r strongest principal components,

$$Y \approx Y_P = \overline{Y} + V\Lambda^{\frac{1}{2}} P^{\mathrm{T}} P Y_{\mathrm{s}} = \overline{Y} + \frac{1}{n} (P Y_{\mathrm{s}})^{\mathrm{T}} (P Y_{\mathrm{s}})$$

Formulation of PCA/ICA

(PCA) Let the data be so decomposed,

$$Y_{\mathrm{c}} = Y - \overline{Y}, \quad K = \tfrac{1}{n} Y_{\mathrm{c}} Y_{\mathrm{c}}^{\mathrm{T}}, \quad KV = V\Lambda, \quad Y_{\mathrm{s}} = \Lambda^{-\frac{1}{2}} V^{\mathrm{T}} Y_{\mathrm{c}}$$

Let $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$ with $\lambda_1 \geq \cdots \geq \lambda_m$. With $P \in \mathbb{R}^{r \times m}$, $r < m$, $P_{i,j} = \delta_{i,j}$, the data Y are so projected to its r strongest principal components,

$$Y \approx Y_P = \overline{Y} + V\Lambda^{\frac{1}{2}} P^{\mathrm{T}} P Y_{\mathrm{s}} = \overline{Y} + \tfrac{1}{n}(PY_{\mathrm{s}})^{\mathrm{T}}(PY_{\mathrm{s}})$$

(ICA) Let the data be further so decomposed,

$$X_{\mathrm{c}} = UY_{\mathrm{s}}$$

With $Q \in \mathbb{R}^{r \times m}$, $r < m$, $Q_{i,j} = \delta_{q_i,j}$, the data Y are so projected to the r independent components $\{q_1, \ldots, q_r\}$,

$$Y \approx Y_Q = \overline{Y} + V\Lambda^{\frac{1}{2}} U^{\mathrm{T}} Q^{\mathrm{T}} Q X_{\mathrm{c}} = \overline{Y} + \tfrac{1}{n}(QX_{\mathrm{c}})^{\mathrm{T}}(QX_{\mathrm{c}})$$

Benefits of ℓ_1 Formulations

Centering. Given data $\boldsymbol{x} = \langle a, b, \ldots, b \rangle \in \mathbb{R}^m$, $a < b$,

Benefits of ℓ_1 Formulations

Centering. Given data $\boldsymbol{x} = \langle a, b, \ldots, b \rangle \in \mathbb{R}^m$, $a < b$,

$$\begin{aligned}
\mu_2(\boldsymbol{x}) &= \arg\min_{\mu} \sum_{i=1}^{m} (\mu - x_i)^2 \\
&= \arg\min_{\mu} \left[(\mu - a)^2 + (m-1)(\mu - b)^2 \right] = \frac{1}{m}[a + (m-1)b]
\end{aligned}$$

Benefits of ℓ_1 Formulations

Centering. Given data $\boldsymbol{x} = \langle a, b, \ldots, b \rangle \in \mathbb{R}^m$, $a < b$,

$$
\begin{aligned}
\mu_2(\boldsymbol{x}) &= \arg\min_{\mu} \sum_{i=1}^{m} (\mu - x_i)^2 \\
&= \arg\min_{\mu} \left[(\mu - a)^2 + (m-1)(\mu - b)^2 \right] = \frac{1}{m}[a + (m-1)b] \\
\mu_1(\boldsymbol{x}) &= \arg\min_{\mu} \sum_{i=1}^{m} |\mu - x_i| \\
&= \arg \min_{a \le \mu \le b} [(\mu - a) + (m-1)(b - \mu)] = b \quad \text{(robust!)}
\end{aligned}
$$

Benefits of ℓ_1 Formulations

Centering. Given data $\boldsymbol{x} = \langle a, b, \ldots, b \rangle \in \mathbb{R}^m$, $a < b$,

$$\begin{aligned} \mu_2(\boldsymbol{x}) &= \arg\min_{\mu} \sum_{i=1}^{m} (\mu - x_i)^2 \\ &= \arg\min_{\mu} \left[(\mu - a)^2 + (m-1)(\mu - b)^2 \right] = \frac{1}{m}[a + (m-1)b] \\ \mu_1(\boldsymbol{x}) &= \arg\min_{\mu} \sum_{i=1}^{m} |\mu - x_i| \\ &= \arg\min_{a \le \mu \le b} [(\mu - a) + (m-1)(b - \mu)] = b \quad \text{(robust!)} \end{aligned}$$

Best generalization for higher dimensional data, $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}^{\mathrm{T}} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{\mu}_1(Y) = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^n} \sum_{i=1}^{m} \|\boldsymbol{\mu} - \boldsymbol{y}_i\|_{\ell_2}$$

Benefits of ℓ_1 Formulations

Sphering. The ℓ_2 approach is obtained by minimizing

$$R_k(\boldsymbol{v}) = \frac{\frac{1}{n}\langle Y_k Y_k^{\mathrm{T}} \boldsymbol{v}, \boldsymbol{v}\rangle}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle} = \left[\frac{\| Y_k^{\mathrm{T}} \boldsymbol{v}\|_{\ell_2}}{\sqrt{n}\|\boldsymbol{v}\|_{\ell_2}}\right]^2$$

where

$$Y_k = (I - V_{k-1} V_{k-1}^{\mathrm{T}}) Y_{\mathrm{c}}, \quad k = 2, \ldots, m-1, \quad Y_1 = Y_{\mathrm{c}}$$

and setting

$\boldsymbol{v}_k = \mathrm{argmin}_{\boldsymbol{v}} R_k(\boldsymbol{v})$, $\lambda_k = R_k(\boldsymbol{v}_k)$, $V_k = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $V = V_m$.

Benefits of ℓ_1 Formulations

Sphering. The ℓ_2 approach is obtained by minimizing

$$R_k(\boldsymbol{v}) = \frac{\frac{1}{n}\langle Y_k Y_k^{\mathrm{T}} \boldsymbol{v}, \boldsymbol{v}\rangle}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle} = \left[\frac{\| Y_k^{\mathrm{T}} \boldsymbol{v}\|_{\ell_2}}{\sqrt{n}\|\boldsymbol{v}\|_{\ell_2}}\right]^2$$

where

$$Y_k = (I - V_{k-1} V_{k-1}^{\mathrm{T}}) Y_{\mathrm{c}}, \quad k = 2, \ldots, m-1, \quad Y_1 = Y_{\mathrm{c}}$$

and setting

$\boldsymbol{v}_k = \mathrm{argmin}_{\boldsymbol{v}} R_k(\boldsymbol{v})$, $\lambda_k = R_k(\boldsymbol{v}_k)$, $V_k = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $V = V_m$.

Best generalization for ℓ_1 is obtained by minimizing

$$F_k(\boldsymbol{v}) = \frac{\| Y_k^{\mathrm{T}} \boldsymbol{v}\|_{\ell_1}}{\sqrt{n}\|\boldsymbol{v}\|_{\ell_2}}$$

where

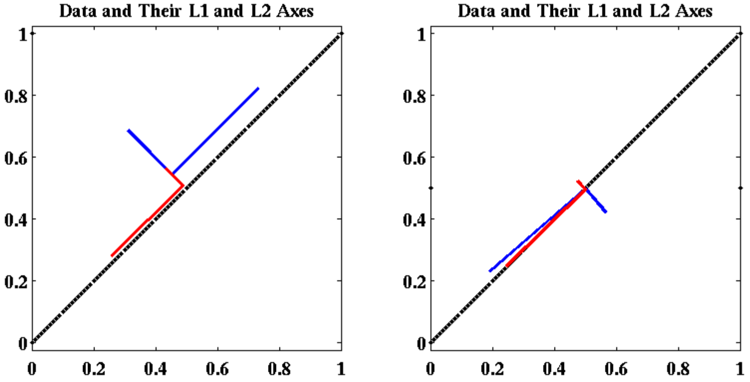
$$Y_k = (I - V_{k-1} V_{k-1}^{\mathrm{T}}) Y_{\mathrm{c}}, \quad k = 2, \ldots, m-1, \quad Y_1 = Y_{\mathrm{c}}$$

and setting

$\boldsymbol{v}_k = \mathrm{argmin}_{\boldsymbol{v}} F_k(\boldsymbol{v})$, $\lambda_k = F_k(\boldsymbol{v}_k)$, $V_k = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $V = V_m$.

Benefits of ℓ_1 Formulations

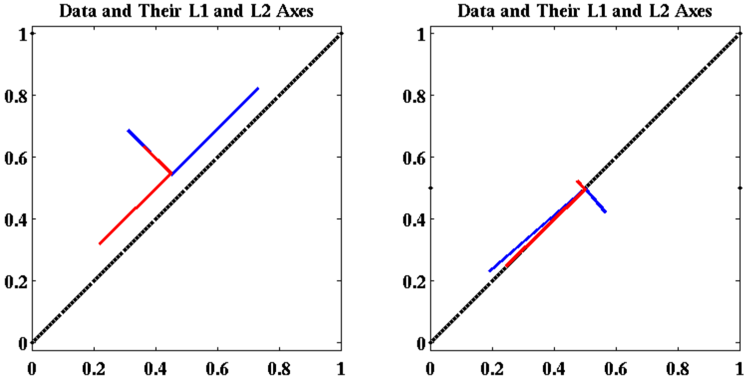
Outliers accumulated at $(0, 1)$, then at $(0, \frac{1}{2})$ and $(1, \frac{1}{2})$,



Blue is for ℓ_2, Red is for $\ell_1(\ell_2)$.

Benefits of ℓ_1 Formulations

Outliers accumulated at $(0, 1)$, then at $(0, \frac{1}{2})$ and $(1, \frac{1}{2})$,

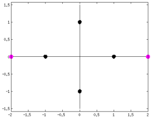


Blue is for ℓ_2, Red is for $\ell_1(\ell_1)$.

Robust Objective Function for ICA

Test data $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2\}^{\mathrm{T}} \in \mathbb{R}^{2\times n}$, each pair in $\{(\pm 1, 0), (0, \pm 1)\}$ except for outliers

$$
\begin{aligned}
(\boldsymbol{y}_1)_1 &= \alpha, \quad (\boldsymbol{y}_2)_1 = 0 \\
(\boldsymbol{y}_1)_2 &= -\alpha, \quad (\boldsymbol{y}_2)_2 = 0
\end{aligned}
$$

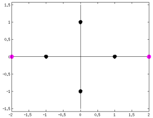


Then $\overline{Y} = (0, 0)$ and $V = I$.

Robust Objective Function for ICA

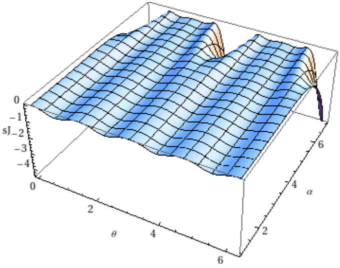
Test data $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2\}^{\mathrm{T}} \in \mathbb{R}^{2\times n}$, each pair in $\{(\pm 1, 0), (0, \pm 1)\}$ except for outliers

$$(\boldsymbol{y}_1)_1 = \alpha, \quad (\boldsymbol{y}_2)_1 = 0$$
$$(\boldsymbol{y}_1)_2 = -\alpha, \quad (\boldsymbol{y}_2)_2 = 0$$



Then $\overline{Y} = (0, 0)$ and $V = I$.

The Kurtosis objective function $J(\boldsymbol{u}) = -\mathcal{W}^2(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u})$ has the following landscape for the test data:



$\boldsymbol{u} = \{\cos(\theta), \sin(\theta)\}$ with $\theta = \frac{\pi}{4}$ is the robust solution.

This solution is obtained for $\alpha \approx 0$, but not for α moderately larger.

Robust Objective Function for ICA

An alternative objection function is based on the ℓ_1 moment,

$$\mathcal{M}(\boldsymbol{x}) = M_1(\boldsymbol{x}) - \sqrt{M_2(\boldsymbol{x})}\sqrt{\frac{2}{\pi}}$$

where $\mathcal{M}(\boldsymbol{n}) = \sigma\sqrt{2/\pi} - \sigma\sqrt{2/\pi} = 0$ for $\boldsymbol{n} \sim N(\mu, \sigma^2)$.

Robust Objective Function for ICA

An alternative objection function is based on the ℓ_1 moment,

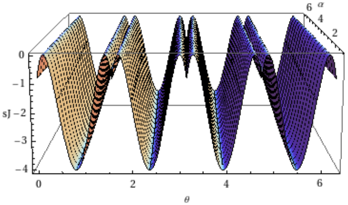
$$\mathcal{M}(\boldsymbol{x}) = M_1(\boldsymbol{x}) - \sqrt{M_2(\boldsymbol{x})}\sqrt{\frac{2}{\pi}}$$

where $\mathcal{M}(\boldsymbol{n}) = \sigma\sqrt{2/\pi} - \sigma\sqrt{2/\pi} = 0$ for $\boldsymbol{n} \sim N(\mu, \sigma^2)$.

The new objective function

$$J(\boldsymbol{u}) = -\mathcal{M}^2(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u})$$

has the following landscape for the test data:



$\boldsymbol{u} = \{\cos(\theta), \sin(\theta)\}$ with $\theta = \frac{\pi}{4}$ is the robust solution.

This solution is obtained for a large range of $\alpha > 0$.

Minimizing the Robust Objective Function for ICA

The robust objective function

$$J(\boldsymbol{u}) = -\mathcal{M}^2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = -[M_1(Y_\mathrm{s}\boldsymbol{u}) - \sqrt{2/\pi}]^2$$

$(M_2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = 1)$ is minimized under the condition $\boldsymbol{u}^\mathrm{T}\boldsymbol{u} = 1$.

Minimizing the Robust Objective Function for ICA

The robust objective function

$$J(\boldsymbol{u}) = -\mathcal{M}^2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = -[M_1(Y_\mathrm{s}\boldsymbol{u}) - \sqrt{2/\pi}]^2$$

$(M_2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = 1)$ is minimized under the condition $\boldsymbol{u}^\mathrm{T}\boldsymbol{u} = 1$.

The solution is obtained from a stationary point of

$$L(\boldsymbol{u}, \lambda) = -[M_1(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) - \sqrt{2/\pi}]^2 + \lambda(\boldsymbol{u}^\mathrm{T}\boldsymbol{u} - 1)/2$$

Minimizing the Robust Objective Function for ICA

The robust objective function

$$J(\boldsymbol{u}) = -\mathcal{M}^2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = -[M_1(Y_\mathrm{s}\boldsymbol{u}) - \sqrt{2/\pi}]^2$$

($M_2(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) = 1$) is minimized under the condition $\boldsymbol{u}^\mathrm{T}\boldsymbol{u} = 1$.

The solution is obtained from a stationary point of

$$L(\boldsymbol{u}, \lambda) = -[M_1(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) - \sqrt{2/\pi}]^2 + \lambda(\boldsymbol{u}^\mathrm{T}\boldsymbol{u} - 1)/2$$

We have $D_{\boldsymbol{u}}J(\boldsymbol{u}) = -\phi(\boldsymbol{u})G(\boldsymbol{u})\boldsymbol{u}$ with

$$\phi(\boldsymbol{u}) = 2[M_1(Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}) - \sqrt{2/\pi}] \quad \text{and} \quad G(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_\mathrm{s}\boldsymbol{e}_i\boldsymbol{e}_i^\mathrm{T}Y_\mathrm{s}^\mathrm{T}}{|\boldsymbol{e}_i^\mathrm{T}Y_\mathrm{s}^\mathrm{T}\boldsymbol{u}|}$$

Minimizing the Robust Objective Function for ICA

The robust objective function

$$J(\boldsymbol{u}) = -\mathcal{M}^2(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}) = -[M_1(Y_{\mathrm{s}}\boldsymbol{u}) - \sqrt{2/\pi}]^2$$

$(M_2(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}) = 1)$ is minimized under the condition $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$.

The solution is obtained from a stationary point of

$$L(\boldsymbol{u}, \lambda) = -[M_1(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}) - \sqrt{2/\pi}]^2 + \lambda(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} - 1)/2$$

We have $D_{\boldsymbol{u}}J(\boldsymbol{u}) = -\phi(\boldsymbol{u})G(\boldsymbol{u})\boldsymbol{u}$ with

$$\phi(\boldsymbol{u}) = 2[M_1(Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}) - \sqrt{2/\pi}] \quad \text{and} \quad G(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_{\mathrm{s}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{s}}^{\mathrm{T}}}{|\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}|}$$

A stationary point $(\boldsymbol{u}^\star, \lambda^\star)$ satisfies $-D_{\boldsymbol{u}}J(\boldsymbol{u}^\star) = \lambda^\star\boldsymbol{u}^\star$ or with $\lambda^\star = \mu^\star(\boldsymbol{u}^\star)\phi(\boldsymbol{u}^\star)$ the nonlinear eigenspace problem,

$$G(\boldsymbol{u}^\star)\boldsymbol{u}^\star = \mu^\star(\boldsymbol{u}^\star)\boldsymbol{u}^\star, \quad \boldsymbol{u}^{\star\mathrm{T}}\boldsymbol{u}^\star = 1$$

Minimizing the Robust Objective Function for ICA

The nonlinear eigenspace problem is solved by a vector iteration.

Let $\boldsymbol{u}_l \approx \boldsymbol{u}^\star$ with $\|\boldsymbol{u}_l\| = 1$ and an update $\boldsymbol{u}_{l+1}$ is determined by,

$$\boldsymbol{u} = G(\boldsymbol{u}_l)\boldsymbol{u}_l, \quad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

After convergence

$$\boldsymbol{u}^\star = \lim_{l\to\infty} \boldsymbol{u}_l$$

is the first column of U^{T}.

Minimizing the Robust Objective Function for ICA

The nonlinear eigenspace problem is solved by a vector iteration.

Let $\boldsymbol{u}_l \approx \boldsymbol{u}^\star$ with $\|\boldsymbol{u}_l\| = 1$ and an update $\boldsymbol{u}_{l+1}$ is determined by,

$$\boldsymbol{u} = G(\boldsymbol{u}_l)\boldsymbol{u}_l, \quad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

After convergence

$$\boldsymbol{u}^\star = \lim_{l\to\infty} \boldsymbol{u}_l$$

is the first column of U^{T}.

The next column of U^{T} is determined by a modified vector iteration.

For this, the projected data

$$Y_{\mathrm{p}} = (I - \boldsymbol{u}^\star \boldsymbol{u}^{\star\mathrm{T}}) Y_{\mathrm{s}}$$

have columns which are linearly independent from $\boldsymbol{u}^\star$.

Minimizing the Robust Objective Function for ICA

With the modified matrix,

$$\tilde{G}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_{\mathrm{p}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}}{|\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{u}|}$$

the modified vector iteration is,

$$\boldsymbol{u} = (I - \boldsymbol{u}^{\star}\boldsymbol{u}^{\star\mathrm{T}})\tilde{G}(\boldsymbol{u}_l)\boldsymbol{u}_l, \qquad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

Minimizing the Robust Objective Function for ICA

With the modified matrix,

$$\tilde{G}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_{\mathrm{p}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}}{|\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{u}|}$$

the modified vector iteration is,

$$\boldsymbol{u} = (I - \boldsymbol{u}^{\star}\boldsymbol{u}^{\star\mathrm{T}})\tilde{G}(\boldsymbol{u}_l)\boldsymbol{u}_l, \qquad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

The remaining columns of U^{T} are determined similarly, where $\boldsymbol{u}^{\star}$ above is replaced with the matrix $[\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}]$, when k columns $\{\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}\}$ of U^{T} have already been calculated.

Minimizing the Robust Objective Function for ICA

With the modified matrix,

$$\tilde{G}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_{\mathrm{p}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}} Y_{\mathrm{p}}^{\mathrm{T}}}{|\boldsymbol{e}_i^{\mathrm{T}} Y_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{u}|}$$

the modified vector iteration is,

$$\boldsymbol{u} = (I - \boldsymbol{u}^{\star}\boldsymbol{u}^{\star\mathrm{T}})\tilde{G}(\boldsymbol{u}_l)\boldsymbol{u}_l, \qquad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

The remaining columns of U^{T} are determined similarly, where $\boldsymbol{u}^{\star}$ above is replaced with the matrix $[\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}]$, when k columns $\{\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}\}$ of U^{T} have already been calculated.

Observation: The vector iterations converge very robustly to the global constrained minimum.

Minimizing the Robust Objective Function for ICA

With the modified matrix,

$$\tilde{G}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_{\mathrm{p}}\boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}}{|\boldsymbol{e}_i^{\mathrm{T}}Y_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{u}|}$$

the modified vector iteration is,

$$\boldsymbol{u} = (I - \boldsymbol{u}^{\star}\boldsymbol{u}^{\star\mathrm{T}})\tilde{G}(\boldsymbol{u}_l)\boldsymbol{u}_l, \qquad \boldsymbol{u}_{l+1} = \boldsymbol{u}/\|\boldsymbol{u}\|, \quad l = 1, 2, \ldots$$

The remaining columns of U^{T} are determined similarly, where $\boldsymbol{u}^{\star}$ above is replaced with the matrix $[\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}]$, when k columns $\{\boldsymbol{u}_1^{\star}, \ldots, \boldsymbol{u}_k^{\star}\}$ of U^{T} have already been calculated.

Observation: The vector iterations converge very robustly to the global constrained minimum.

Claim: At least convergence to a local constrained minimum can be proved with adequate step size control.

Application to DCE-MRI sequences

For each time $t = 1, \ldots, T$, the matrix of pixel values,

$$B(t) = \{B_{i,j}(t)\}_{1 \leq i,j \leq N}$$

is an image in the [Video].

Application to DCE-MRI sequences

For each time $t = 1, \ldots, T$, the matrix of pixel values,

$$B(t) = \{B_{i,j}(t)\}_{1 \leq i,j \leq N}$$

is an image in the [Video].

With $m = T = 134$ and $n = N^2 = 400^2$ the images are represented as long vectors:

$$\boldsymbol{y}_t^{\mathrm{T}} = \{ \ B_{1,1}(t), \ldots, B_{N,1}(t), B_{1,2}(t), \ldots, B_{N,2}(t), \ldots, B_{1,N}(t), \ldots, B_{N,N}(t) \ \}$$

and PCA/ICA is carried out with $Y^{\mathrm{T}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$.

($Y \leftarrow Y^{\mathrm{T}}$ is equivalent but expensive.)

Application to DCE-MRI sequences

For each time $t = 1, \ldots, T$, the matrix of pixel values,

$$B(t) = \{B_{i,j}(t)\}_{1 \le i,j \le N}$$

is an image in the [Video].

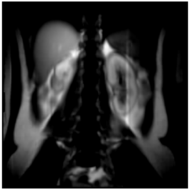
With $m = T = 134$ and $n = N^2 = 400^2$ the images are represented as long vectors:

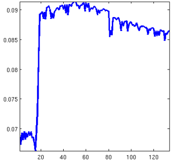
$$\boldsymbol{y}_t^{\mathrm{T}} = \{\ B_{1,1}(t), \ldots, B_{N,1}(t), B_{1,2}(t), \ldots, B_{N,2}(t), \ldots, B_{1,N}(t), \ldots, B_{N,N}(t)\ \}$$

and PCA/ICA is carried out with $Y^{\mathrm{T}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$.

($Y \leftarrow Y^{\mathrm{T}}$ is equivalent but expensive.)

To the left is the first row of Y_{s} (displayed as image),

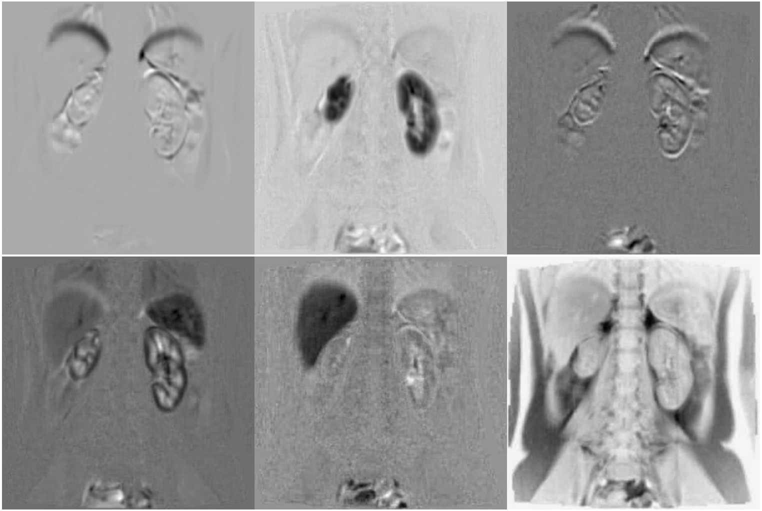




To the right is the first column of V.

Application to DCE-MRI sequences

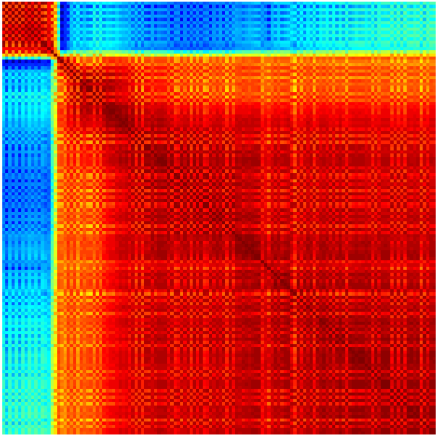
Top 6 independent components:



Eliminate motion, keep anatomy and contrast agent: [Video]

Application to DCE-MRI sequences

Virtual Gating, through segmentation of correlations:



Three groups [Video], stabilized further by PCA/ICA [Video].

Formulation in Function Space

Based upon the imaging examples:

- Sampling occurs continuously in time . . . ?
- Same number of sources as pixels,
  which refine to a continuum . . . ?

Formulation in Function Space

Based upon the imaging examples:

- Sampling occurs continuously in time ... ?
- Same number of sources as pixels,
  which refine to a continuum ... ?

Claim: That the sources be statistically independent requires that that they be countable.

Formulation in Function Space

Based upon the imaging examples:

- Sampling occurs continuously in time . . . ?
- Same number of sources as pixels,
  which refine to a continuum . . . ?

Claim: That the sources be statistically independent requires that that they be countable.

Consequence: The function space setting resembles the finite dimensional setting but with infinite matrices operating between bases in separable spaces.