Generative Topographic Mapping for Nonlinear ICA

Richard Everson, Matt Glennen

Department of Computer Science, University of Exeter

R.M.Everson@exeter.ac.uk

www.dcs.ex.ac.uk/people/reverson

UNIVERSITY of EXETER

Linear mixing

Linear mixing

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

- Observations lie on a hyperplane

Sources $\mathbf{s}(t) \in \mathbb{R}^M$
Observations $\mathbf{x}(t) \in \mathbb{R}^N$
A mixing matrix
Noise $\mathbf{n}(t) \in \mathbb{R}^N$

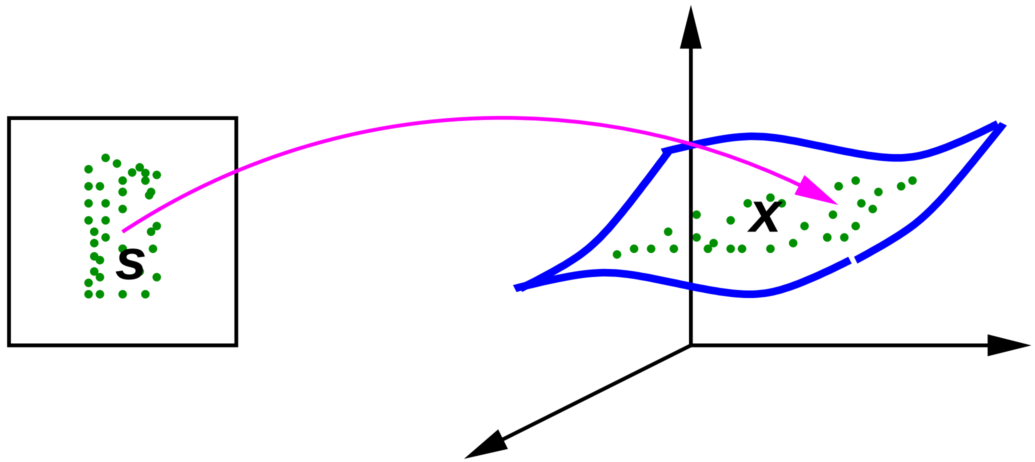
Estimated sources

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$$
$$\mathbf{W}\mathbf{A} \approx \mathbf{I}$$

Mutual information

$$\begin{aligned} I(\hat{\mathbf{s}}) &= \int p(\hat{\mathbf{s}}) \log \frac{p(\hat{\mathbf{s}})}{\prod_m p_m(\hat{s}_m)} d\hat{\mathbf{s}} \\ &= I(\mathbf{P}\mathbf{D}\hat{\mathbf{s}}) \end{aligned}$$

Nonlinear mixing



- Observations lie on a manifold
- **f** might not be smooth!
- **f** should at least be invertible

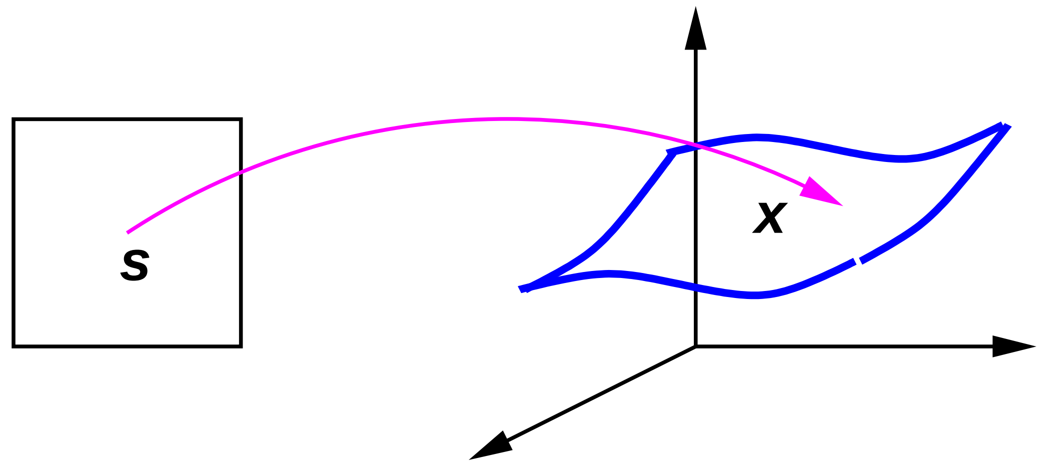
$$\mathbf{x} = \mathbf{f}(\mathbf{s}) + \mathbf{n}$$

Mutual information

$$I(\mathbf{g}(\hat{\mathbf{s}})) = I(\hat{\mathbf{s}}) \quad \text{for} \quad \mathbf{g}(\hat{\mathbf{s}}) = \begin{bmatrix} g_1(\hat{s}_1) \\ g_2(\hat{s}_2) \\ \vdots \\ g_M(\hat{s}_M) \end{bmatrix} \qquad g_m \text{ invertible}$$

Generative Topographic Mapping

Bishop, Svensén, Williams



Generative model

$$p(\mathbf{x} \mid \mathbf{s}, \mathbf{W}, \beta) \; = \; \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left\{-\frac{\beta}{2} \parallel \mathbf{f}(\mathbf{s}; \mathbf{W}) - \mathbf{x} \parallel^2\right\}$$

Noise precision β

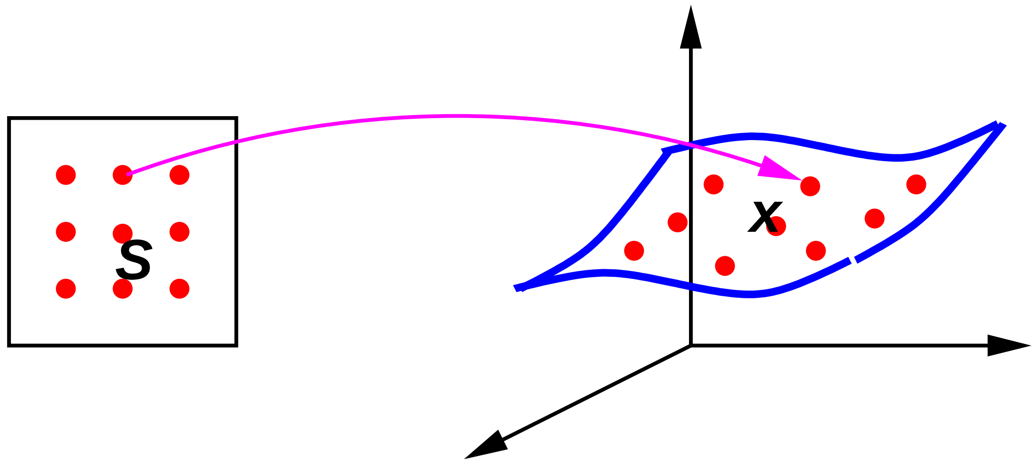
$$\mathbf{f}(\mathbf{s}; \mathbf{W}) \; = \; \mathbf{W}\phi(\mathbf{s})$$

Basis functions $\phi(\mathbf{s})$ located at a fixed grid of locations over latent space.

UNIVERSITY of EXETER

GTM for ICA

Pajunen & Karhunen



$$p(\mathbf{x} \,|\, \mathbf{W}, \beta, \Theta) \;=\; \int p(\mathbf{x} \,|\, \mathbf{s}, \mathbf{W}, \beta) p(\mathbf{s} \,|\, \Theta) \, d\mathbf{s}$$

Sources defined on a grid of K latent points, $\mathbf{s}_{ij}$

$$p(\mathbf{s} \,|\, \Theta) \;=\; \sum_{i,j}^{K} \delta(\mathbf{s} - \mathbf{s}_{ij}) p(s_1 \,|\, \theta_1) p(s_2 \,|\, \theta_2)$$

UNIVERSITY of EXETER

Likelihood & EM

Likelihood

$$\ln \mathcal{L}(\mathbf{W}, \beta, \Theta) \;=\; \sum_{n}^{N} \ln \left\{ \sum_{i,j}^{K} p(\mathbf{x}_n \,|\, \mathbf{s}_{i,j}, \mathbf{W}, \beta) p(s_{1,i} \,|\, \theta_1) p(s_{2,j} \,|\, \theta_2) \right\}$$

EM Responsibilities

$$\begin{aligned} R_{ijn}(\mathbf{W}_{\text{old}}, \beta_{\text{old}}, \Theta_{\text{old}}) &= p(\mathbf{s}_{i,j} \,|\, \mathbf{x}_n, \mathbf{W}_{\text{old}}, \beta_{\text{old}}, \Theta_{old}) \\ &\propto p(\mathbf{x}_n \,|\, \mathbf{s}_{ij}, \mathbf{W}_{\text{old}}, \beta_{\text{old}}) p(\mathbf{s}_{ij} \,|\, \Theta_{\text{old}}) \end{aligned}$$

Expected complete data log-likelihood

$$\langle \mathcal{L}_{\text{comp}}(\mathbf{W}, \beta, \Theta) \rangle \;=\; \sum_{n}^{N} \sum_{i,j}^{K} R_{ijn}^{\text{old}} \ln p(\mathbf{x}_n \,|\, \mathbf{s}_{ij}, \mathbf{W}, \beta) p(s_{1,i} \,|\, \theta_1) p(s_{2,j} \,|\, \theta_2)$$

UNIVERSITY of EXETER

M step

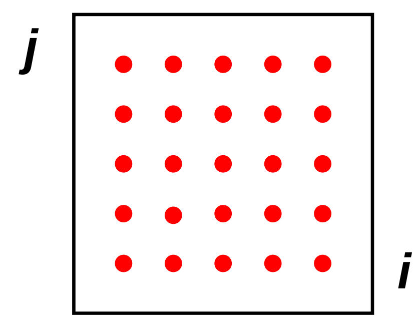
β, W Standard GTM update equations

Source parameters Factorial form $p(\mathbf{s} \mid \Theta) = p(s_1 \mid \theta_1)p(s_2 \mid \theta_2)$ leads to maximisation of:

$$\sum_i \sum_j \sum_n R_{ijn} \ln p(s_{1,j} \mid \theta_1) + \sum_j \sum_i \sum_n R_{ijn} \ln p(s_{2,i} \mid \theta_2)$$

$$= \sum_i r_i^1 \ln p(s_{1,i} \mid \theta_1) + \sum_j r_j^2 \ln p(s_{2,j} \mid \theta_2)$$

r_i^1 is average (over all data and s_2) responsibility for latent points with coordinate s_1, i.

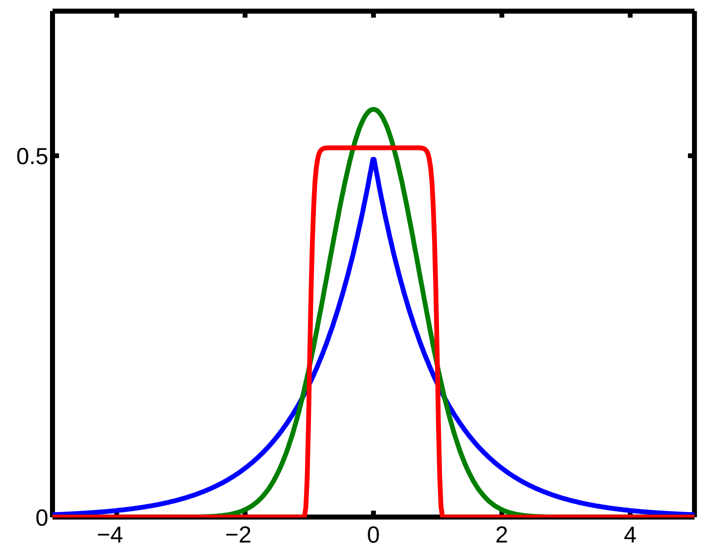


UNIVERSITY of EXETER

Source model

Generalised exponentials

$$p(s \mid \omega, \rho) = \frac{\rho}{2\omega\Gamma(1/\rho)} \exp\{-\left|\frac{s}{\omega}\right|^{\rho}\}$$



$\rho = 1$	Laplacian
$\rho = 2$	Gaussian
$\rho \to \infty$	Uniform

Parameters, ω & ρ, easily estimated from weighted maximum likelihood

$$\sum_i r_i \ln p(s_i \mid \rho, \omega)$$

UNIVERSITY of EXETER

Algorithm

Initialisation Linear ICA on the decorrelating manifold
Generalised exponential source model
β estimated from quality of linear fit

E-step Calculate responsibilities, R_{ijn}

M-step Maximise complete data log-likelihood wrt $\mathbf{W}$, β, ρ_m, ω_m

Regularisation Weight decay regularisation $\lambda \sum_{nl} W_{nl}^2$

Recovered sources

$$\hat{\mathbf{s}}_n = \langle \mathbf{s}_n \,|\, \mathbf{x}_n, \mathbf{W}^*, \beta^*, \Theta^* \rangle = \int p(\mathbf{s} \,|\, \mathbf{x}_n, \mathbf{W}^*, \beta^*, \Theta^*)\mathbf{s}\, d\mathbf{s}$$

$$= \sum_{ij}^{K} R_{ijn} \mathbf{s}_{ij}$$

UNIVERSITY of EXETER

Illustration

- Uniform and Laplacian distributed sources. $N = 5000$ observations.
- Nonlinear mixing:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \epsilon \tanh(\mathbf{Q}\mathbf{A}\mathbf{s})$$

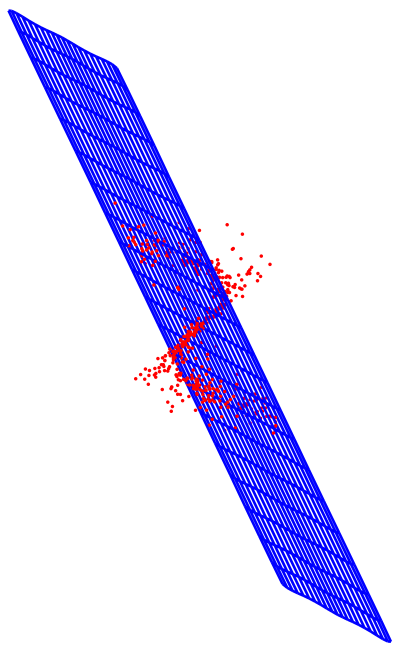
- $\mathbf{Q}$ strictly upper triangular, therefore a volume-preserving nonlinearity (Deco & Brauer)
- Elements of $\mathbf{A}$ and $\mathbf{Q}$ drawn from $\mathcal{N}(0, 1)$

$$\mathbf{A} = \begin{bmatrix} -0.096 & -1.336 \\ -0.832 & 0.714 \\ 0.294 & 1.624 \end{bmatrix}$$

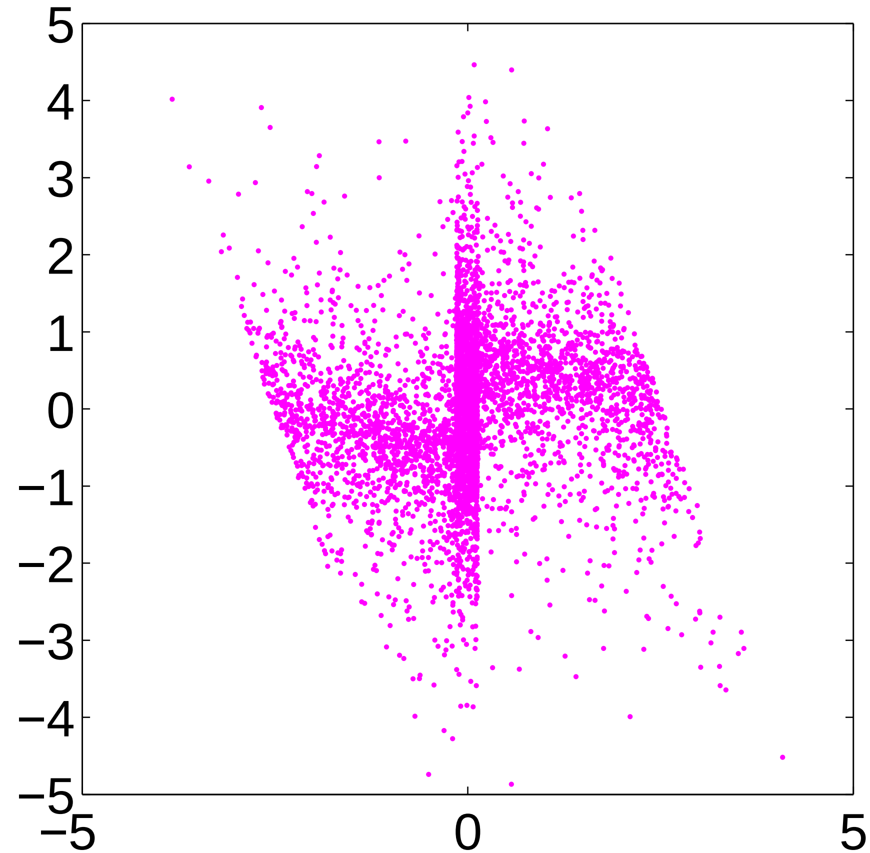
- $K = 21^2$ latent grid points
- $L = 16$ basis functions

UNIVERSITY of EXETER

Linear ICA $\epsilon = 1$

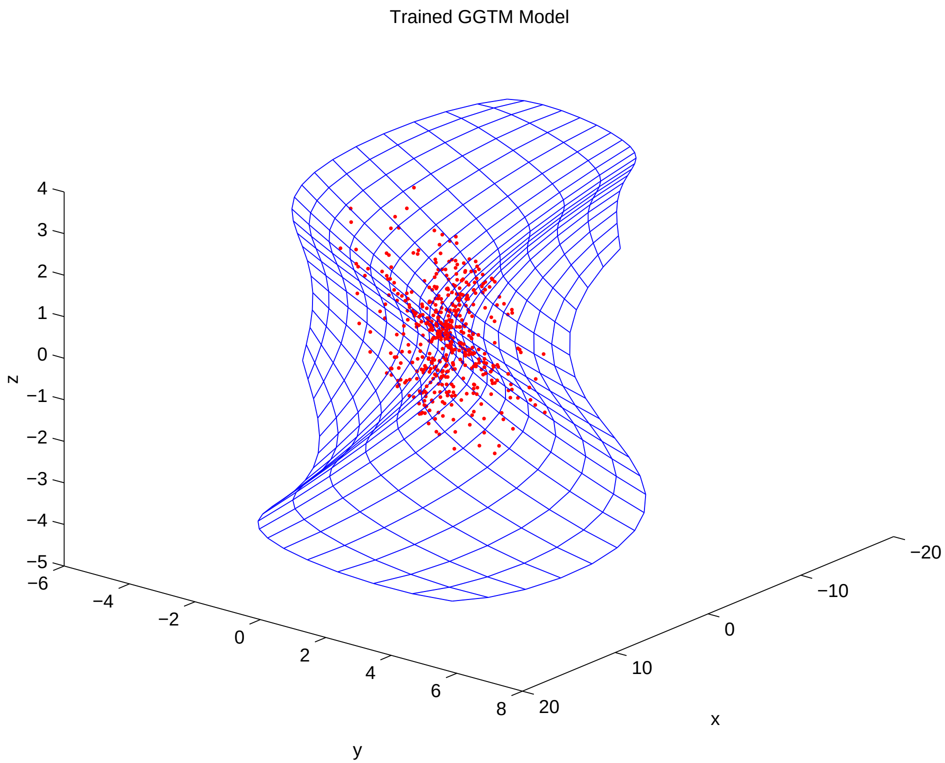


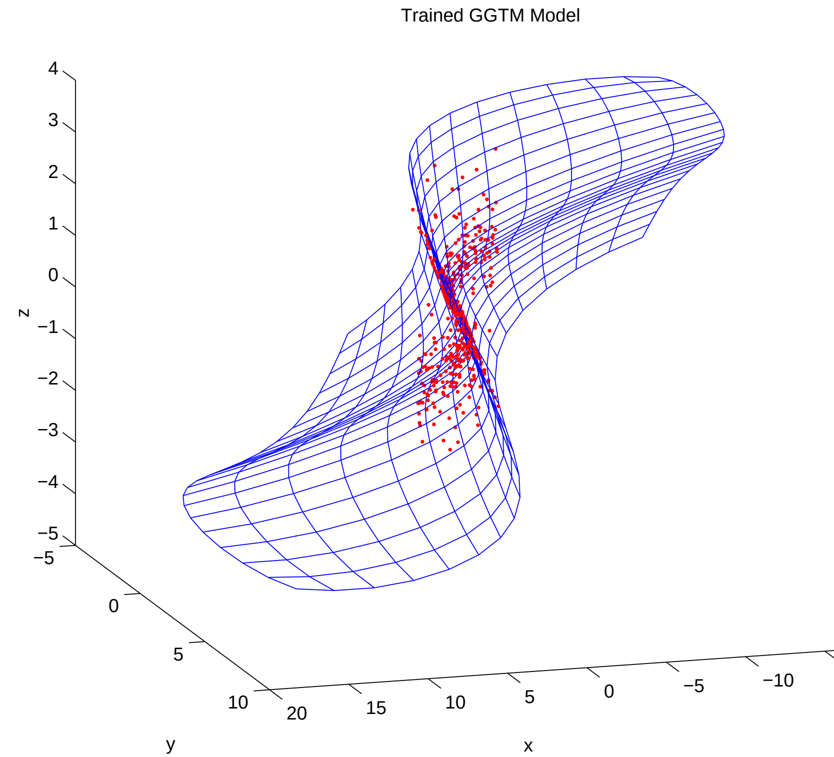
ICA hyperplane



Estimates vs sources

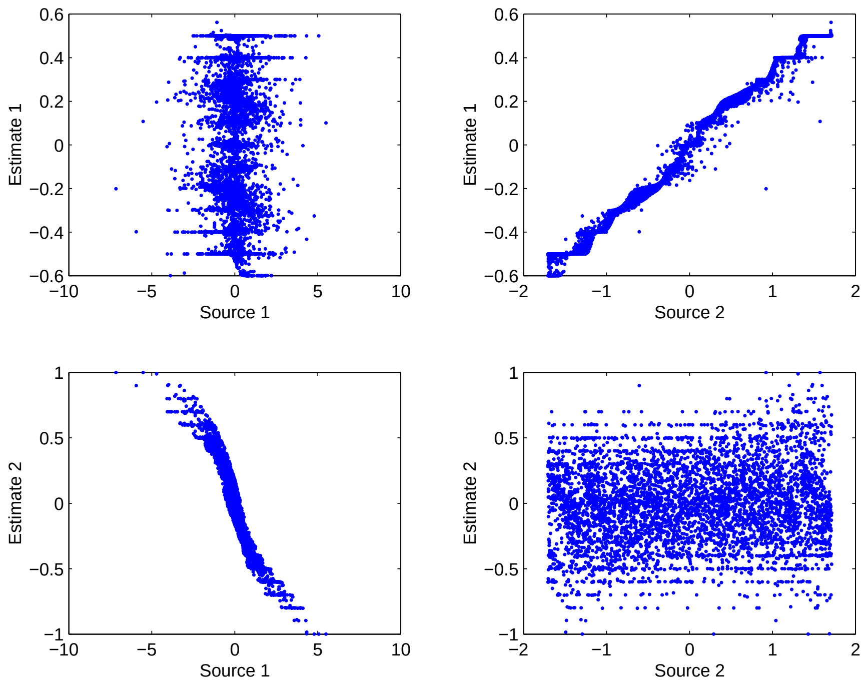
GTM-ICA $\epsilon = 1$



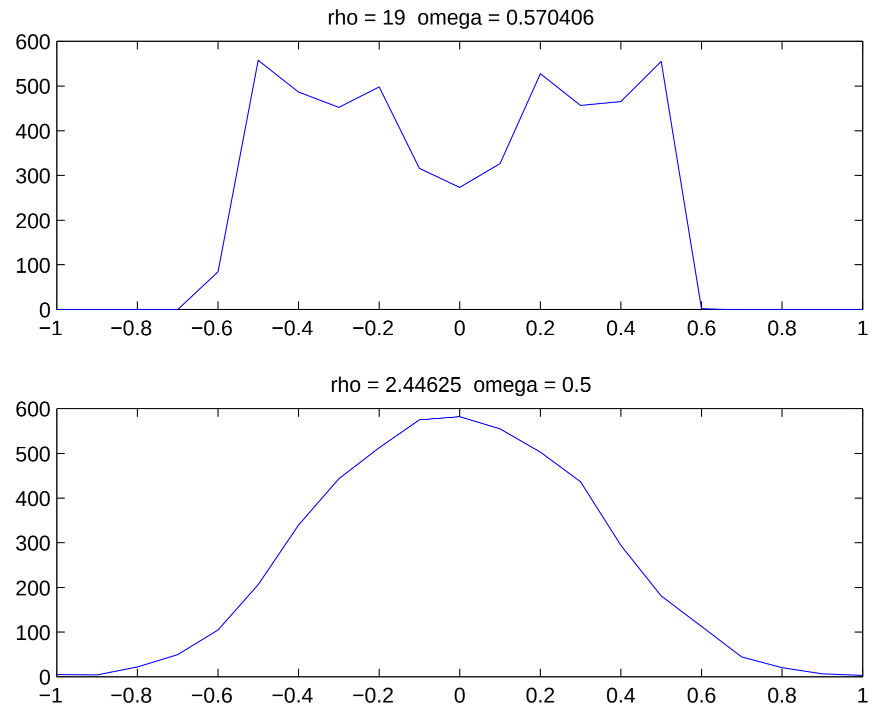


UNIVERSITY of EXETER

GTM-ICA $\quad \epsilon = 1$



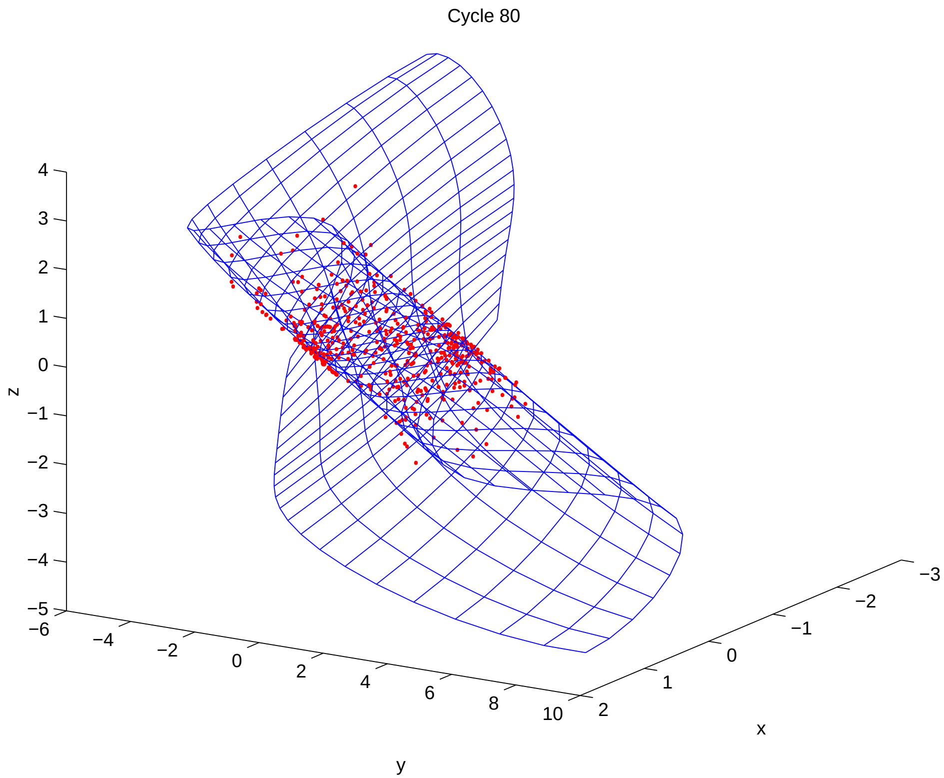
Estimates vs sources

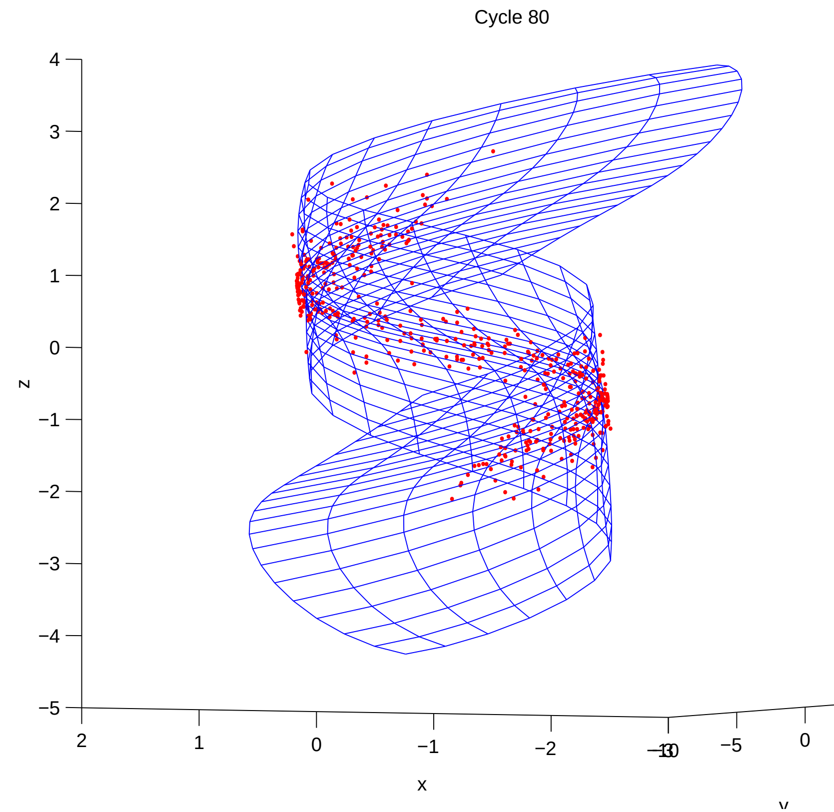


Source densities

UNIVERSITY of EXETER

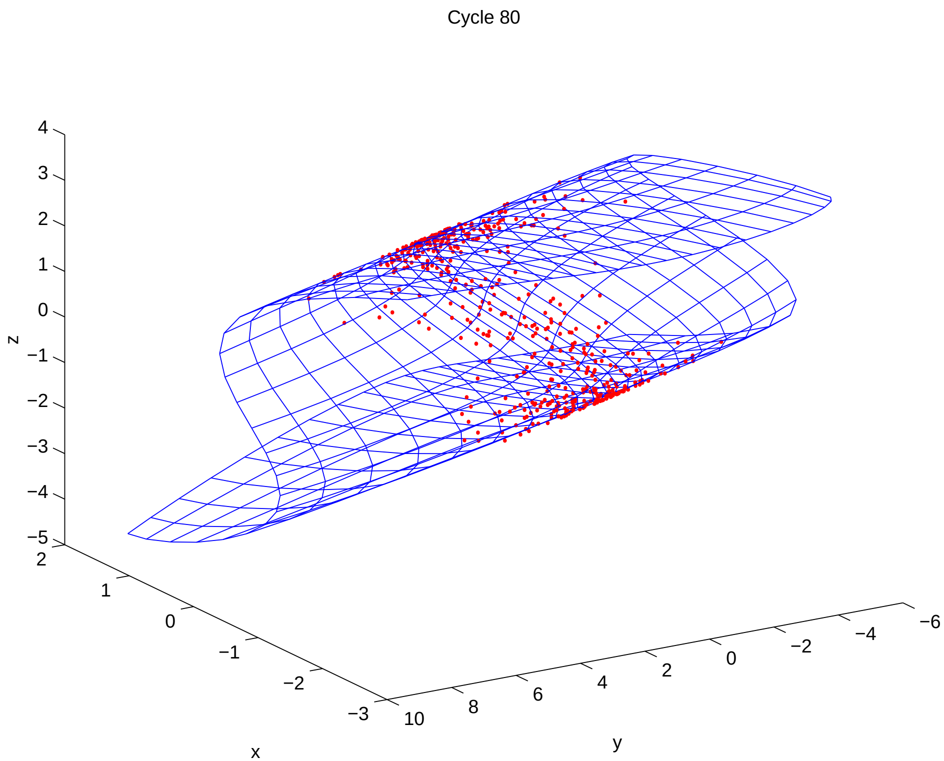
GTM-ICA $\epsilon = 1.5$

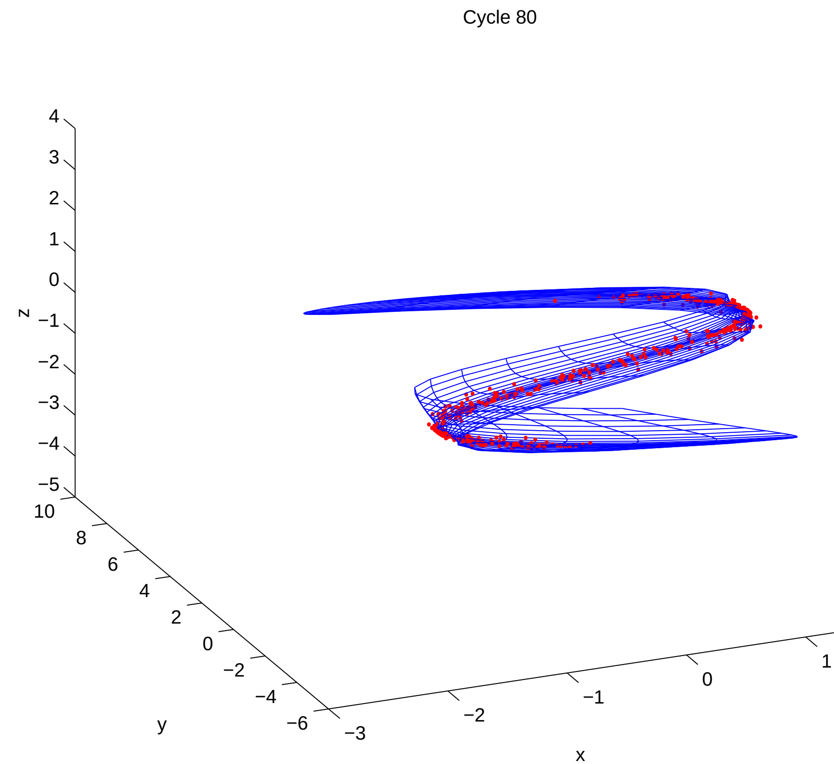




UNIVERSITY of EXETER

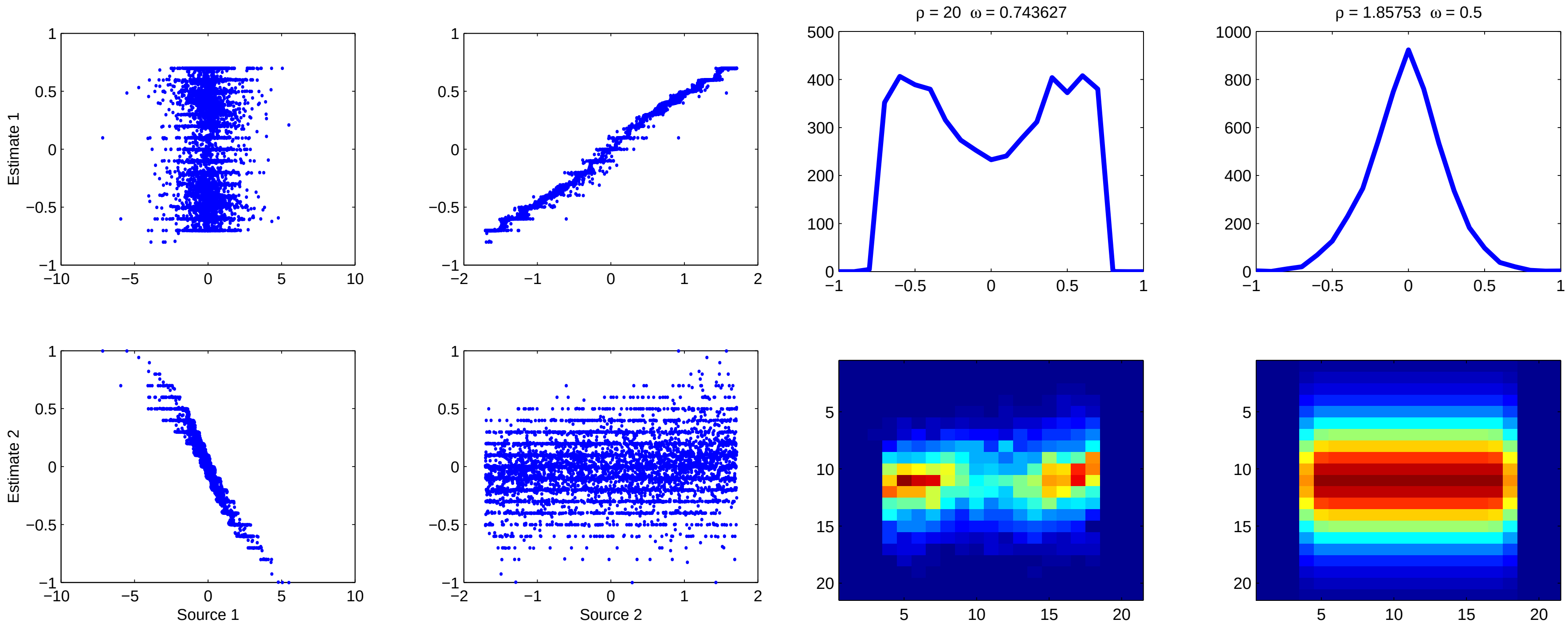
GTM-ICA $\epsilon = 1.5$





UNIVERSITY *of* EXETER

GTM-ICA $\epsilon = 1.5$



Estimates vs sources

Source densities

UNIVERSITY of EXETER

Conclusions

Summary

- Generative model for nonlinear mixing
- EM straight-forwardly modified to learn sources

Problems

- Sensitive to initial linear hyperplane
- Nonlinearity requires careful control

Future

- Reversible jump MCMC methods for GTM, GTM-ICA
- Temporal source models: generalised AR models

UNIVERSITY of EXETER