Convex Optimization for Blind Source Separation on Statistical Manifolds

Simon Luo[1,2], Lamiae Azizi[1,2] and Mahito Sugiyama[3,4]

[1] School of Mathematics and Statistics, The University of Sydney [2] Data Analytics for Resources and Environments (DARE), Australian Research Council

[3] National Institute of Informatics [4] JST, PRESTO

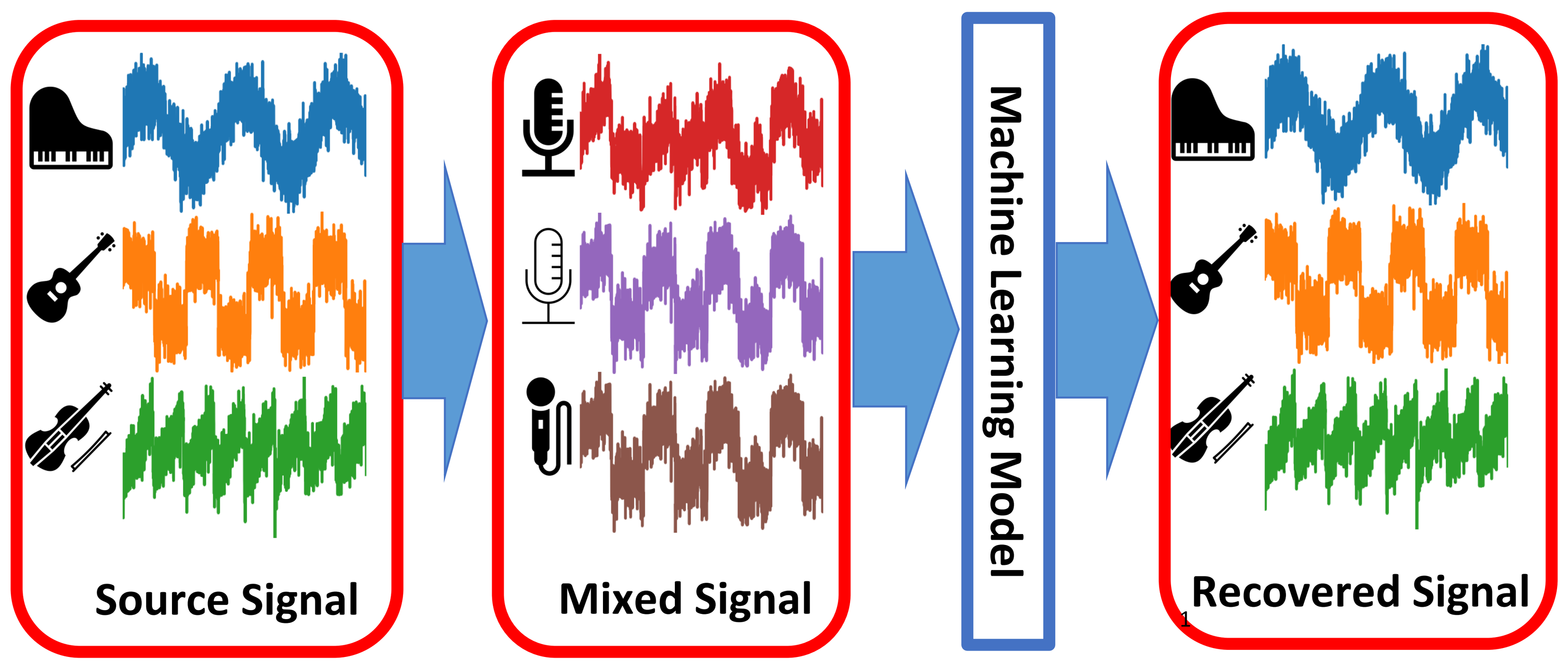
Thirty-fourth Annual Conference on Neural Information Processing Systems (NeurIPS 2020) Workshop on Differential Geometry meets Deep Learning

NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- Formulated Blind Source Separation as a Convex Optimization
- Uses graph structure to represent interactions between source signals
- Graph Structure (poset) allows for more complex interactions between source signals (e.g., higher-order interaction effects between source signals)
- Recovers the sign of the signal
- Minimizes KL-Divergence from a set of received signals to our poset structure

Cocktail Party Problem



Independent Component Analysis (ICA)

$$X = AZ$$

- X – Received "mixed" signal
- A – Mixing matrix (constrained to be an orthogonal matrix)
- Z – Source signal or Reconstructed signal

ICA attempts to learn W and Z, where W is constrained to be an orthonormal matrix

$$Z = A^{-1}X = WX$$

Such that $\mathbf{W}^{\mathrm{T}}\mathbf{W} = I$

Log-Linear Model on a Partially Ordered Set (poset)

The log-linear model is defined over a partial order set (poset) $(S, \preccurlyeq)$

Dual coordinate system (θ, η) of a statistical manifold

- η Expectation parameter
- θ Natural parameter in the exponential family

$$\eta_\omega = \sum_{s \in \Omega} \mathbf{1}_{s \succcurlyeq \omega} p(s), \qquad \log p(\omega) = \sum_{s \in \Omega} \mathbf{1}_{s \preccurlyeq \omega} \theta_s .$$

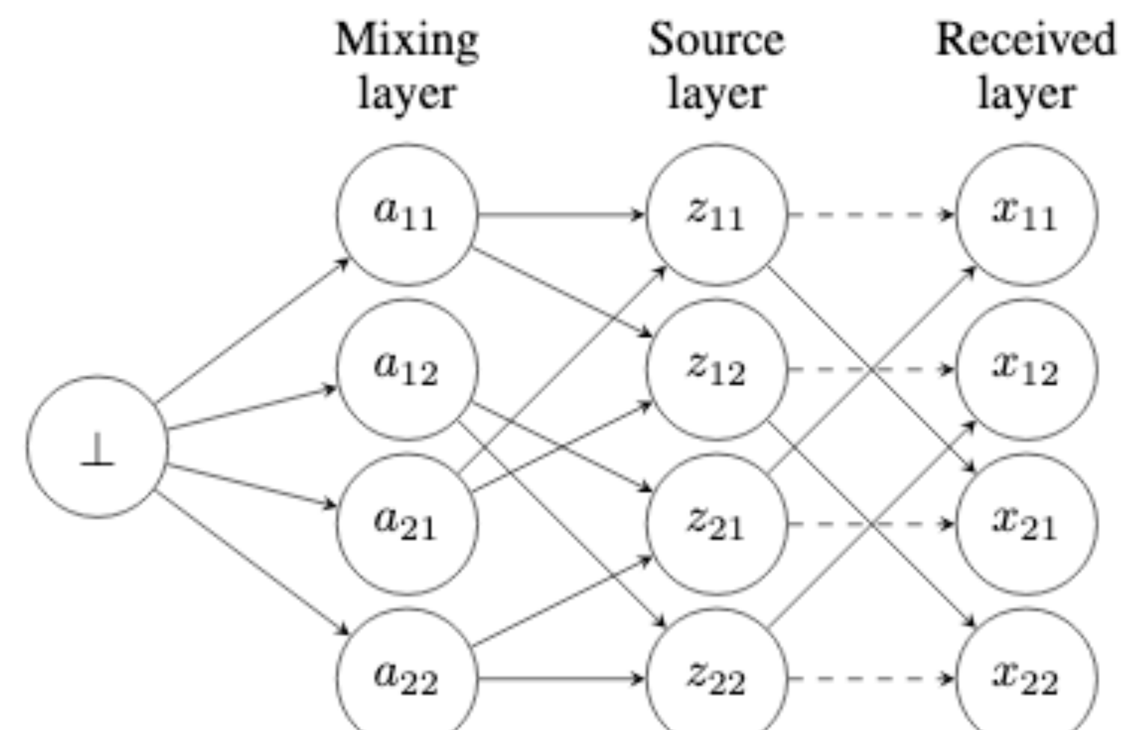
Solving a convex optimization by minimizing KL divergence with gradient,

$$\frac{\partial}{\partial \theta_s} D_{KL}(\hat{P} \parallel P) = \eta - \hat{\eta} = \Delta\eta$$

Converged once $\frac{\partial}{\partial \theta_s} D_{KL}(\hat{P} \parallel P) = 0$, that means $\eta = \hat{\eta}$

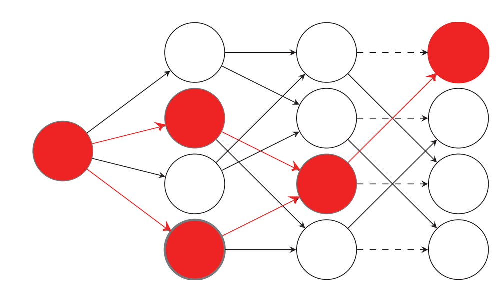
Information Geometric Blind Source Separation

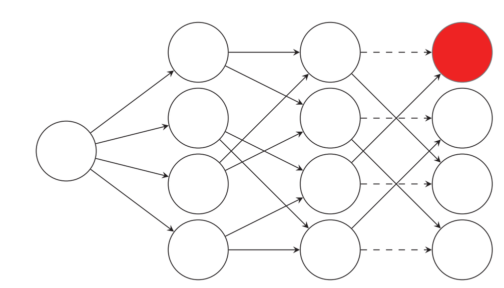
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$



Updating the Received Layer

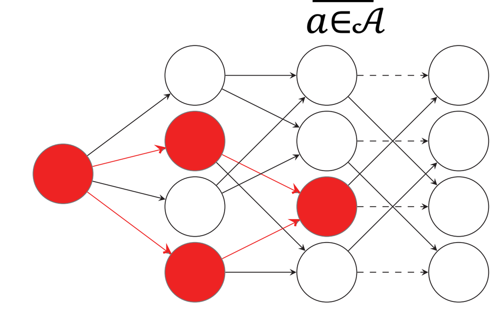
$$\log p(x) = \sum_{z \in Z} \mathbf{1}_{z \preccurlyeq x} \theta_z + \sum_{a \in \mathcal{A}} \mathbf{1}_{a \preccurlyeq x} \theta_a + \theta_\perp, \qquad \eta_x = \sum_{x' \in X} \mathbf{1}_{x \preccurlyeq x'} p(x') = p(x)$$

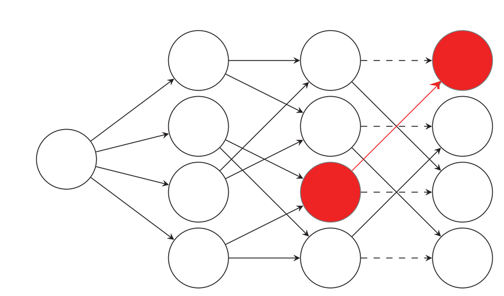




Updating the Source Layer

$$\log p(z) = \theta_z + \sum_{a \in \mathcal{A}} \mathbf{1}_{a \preccurlyeq z} \theta_a + \theta_\perp, \qquad \eta_z = \sum_{x' \in X} \mathbf{1}_{z \preccurlyeq x} p(x) + p(z)$$

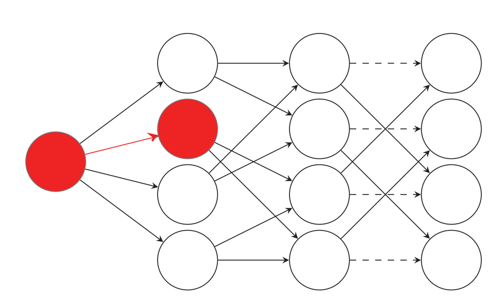


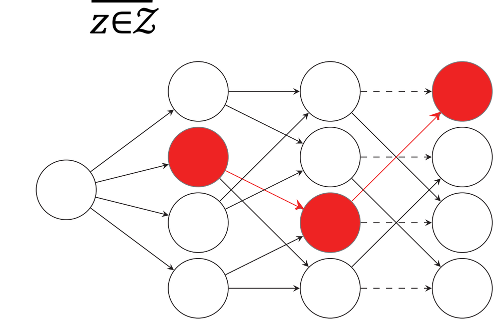


$$\frac{\partial}{\partial \theta_z} D_{KL}(\hat{p} \| p) = \sum_{x \in X} \mathbf{1}_{z \preccurlyeq x} \big(p(x) - \hat{p}(x)\big) + p(z)$$

Updating the Mixing Layer

$$\log p(a) = \theta_a + \theta_\perp, \qquad \eta_a = \sum_{x' \in X} \mathbf{1}_{a \preccurlyeq x'} p(x) + \sum_{z \in Z} \mathbf{1}_{a \preccurlyeq z} p(z) + p(a')$$





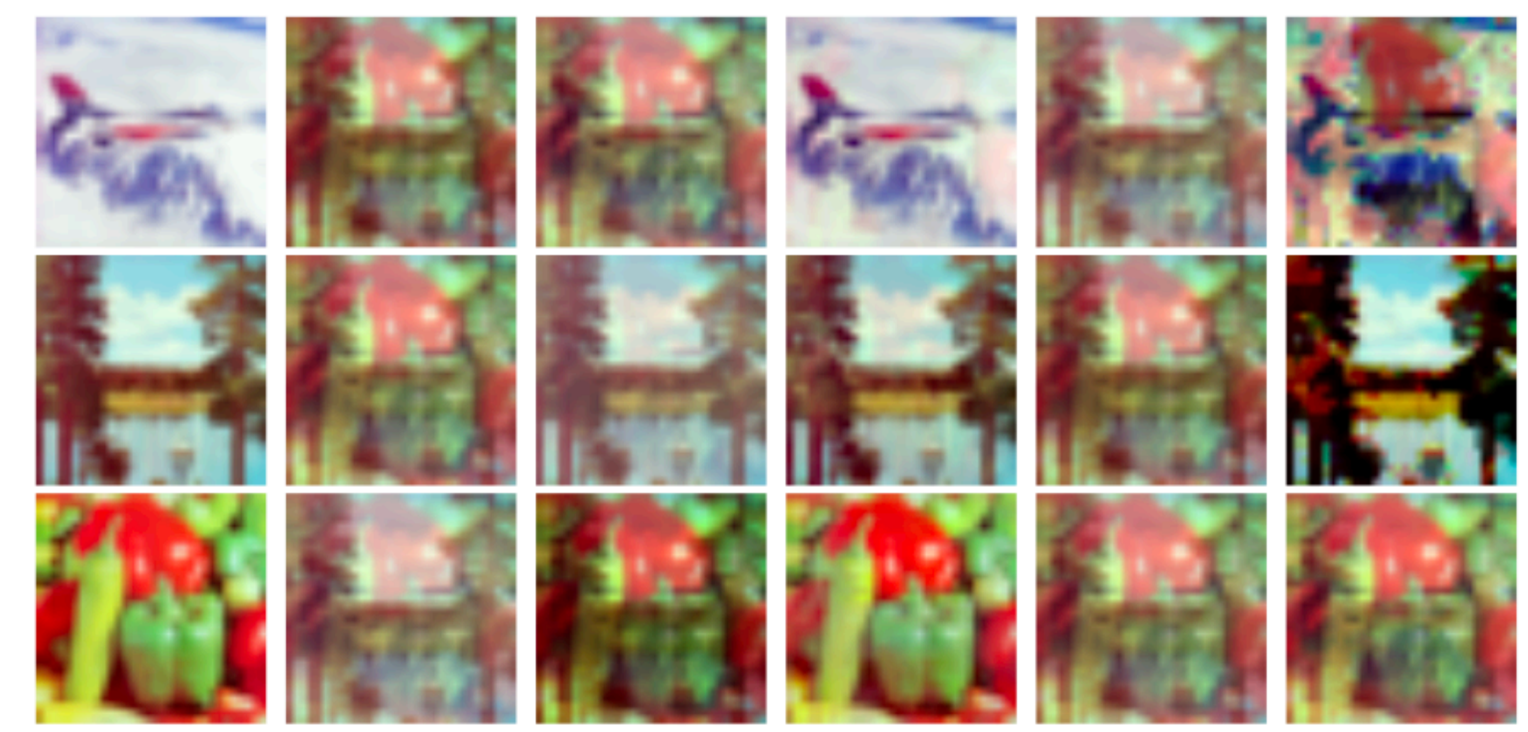
$$\frac{\partial}{\partial \theta_a} D_{KL}(\hat{p} \| p) = \sum_{x \in X} \mathbf{1}_{a \preccurlyeq x} \big(p(x) - \hat{p}(x)\big) + \sum_{z \in Z} \mathbf{1}_{a \preccurlyeq z} p(z) + p(a)$$

Including Higher-Order Feature Interactions

$$x_{lm} = \sum_n a_{ln} z_{nm} + \sum_{n_1} \sum_{n_2 > n_1} a_{\ln_1 n_2} z_{n_1 m} z_{n_2 m} + \sum_{n_1} \sum_{n_2 > n_1} \sum_{n_3 > n_2} a_{\ln_1 n_2 n_3} z_{n_1 m} z_{n_2 m} z_{n_3 m} + \cdots + \sum_{n_1} \cdots \sum_{n_k > n_{k-1}} a_{\ln_1 \ldots n_k} z_{n_1 m} \cdots z_{n_k m}$$

Results

Experiment with Third Order Mixing



Ground Truth | Mixed Signal | IGBSS | FastICA | DictLearn | NMF

Root Mean Squared Error Between Reconstructed Signal and Source Signal

Exp	Order	IGBSS	FastICA	DictLearn	NMF
USC-SPIP Dataset 1	1	$\mathbf{0.252 \pm 0.000}$	0.300 ± 0.089	0.394 ± 0.041	0.622 ± 0.000
	2	$\mathbf{0.260 \pm 0.000}$	0.285 ± 0.096	0.441 ± 0.080	0.662 ± 0.000
	3	$\mathbf{0.252 \pm 0.000}$	0.260 ± 0.111	0.362 ± 0.030	0.612 ± 0.000
USC-SPIP Dataset 2	1	$\mathbf{0.133 \pm 0.000}$	0.284 ± 0.064	0.474 ± 0.067	0.591 ± 0.000
	2	$\mathbf{0.256 \pm 0.000}$	0.263 ± 0.066	0.576 ± 0.008	0.684 ± 0.000
	3	0.282 ± 0.000	$\mathbf{0.239 \pm 0.056}$	0.593 ± 0.007	0.665 ± 0.000
USC-SPIP Dataset 3	1	$\mathbf{0.155 \pm 0.000}$	0.699 ± 0.047	0.478 ± 0.121	0.628 ± 0.000
	2	$\mathbf{0.200 \pm 0.000}$	0.280 ± 0.049	0.515 ± 0.007	0.709 ± 0.000
	3	$\mathbf{0.203 \pm 0.000}$	0.239 ± 0.056	0.536 ± 0.006	0.682 ± 0.000

References

- Mahito Sugiyama, Hiroyuki Nakahara, and Koji Tsuda. **Tensor balancing on statistical manifold**, ICML 2017.
- Simon Luo and Mahito Sugiyama, **Bias-variance trade-off in hierarchical probabilistic models using higher-order feature interactions**, AAAI 2019.
- http://sipi.usc.edu/database/