Kullback-Leibler Divergence Estimation Based on Iterative Gaussianization

V. Laparra, J. Malo, G. Camps-Valls, R. Santos-Rodriguez

{Valero.Laparra, Jesus.Malo, Gustau.Camps, Raul.Santos}@uv.es

Kullback-Leibler divergence

Non-symmetric measure of the difference between two probability distributions,

$$D_{KL}(p,q) = \int_{\mathbb{R}^d} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}$$

Problem: high dimensional KLD estimation is difficult

The proposed method is useful:

- Curse of dimensionality alleviated
- Low computational cost
- Non-parametric

Consider two distributions $p_1(\mathbf{x})$ and $q_1(\mathbf{x})$ in space X ($\mathbf{x} \in X$). Let $\mathcal{G} : X \to Y$ (linear or nonlinear) denote a differentiable and invertible mapping function that converts $\mathbf{x}$ into $\mathbf{y}$,

$$p_1(\mathbf{x}) = p_2(\mathcal{G}(\mathbf{x}))|\nabla_{\mathbf{x}}\mathcal{G}|$$

$$q_1(\mathbf{x}) = q_2(\mathcal{G}(\mathbf{x}))|\nabla_{\mathbf{x}}\mathcal{G}|$$

The KL divergence between two distributions is invariant under transformation $\mathcal{G}$,

$$D_{KL}(p_1, q_1) = D_{KL}(p_2, q_2).$$

Proposed method

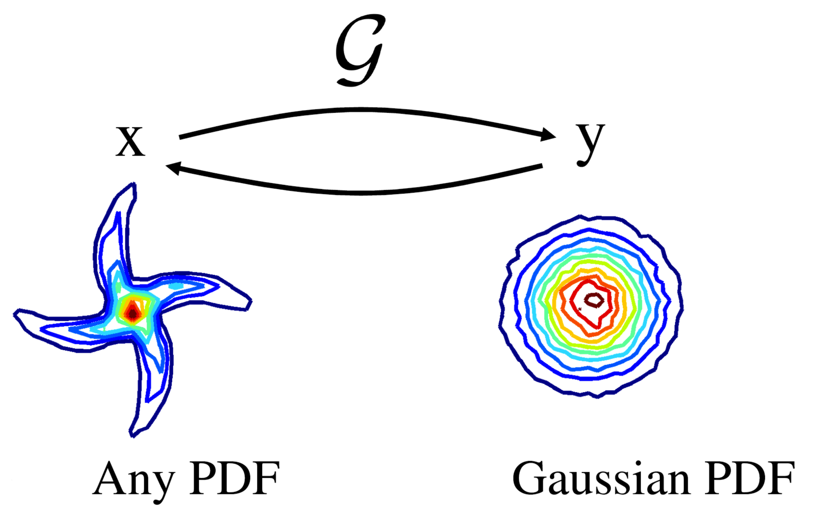
Initial objective: $D_{KL}(p_1, q_1)$

1. Find transform $\mathcal{G}_1$ that gaussianizes q_1 ($q_2 = \mathcal{N}(\mathbf{0}, \mathbf{I})$)
2. Apply $\mathcal{G}_1$ to p_1

New objective: $D_{KL}(p_2, \mathcal{N}(\mathbf{0}, \mathbf{I}))$

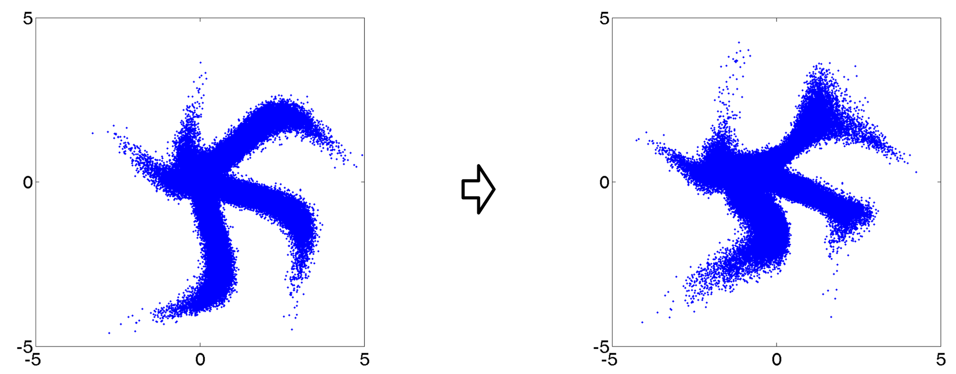
3. Find transform $\mathcal{G}_2$ that gaussianizes p_2
4. Compute objective

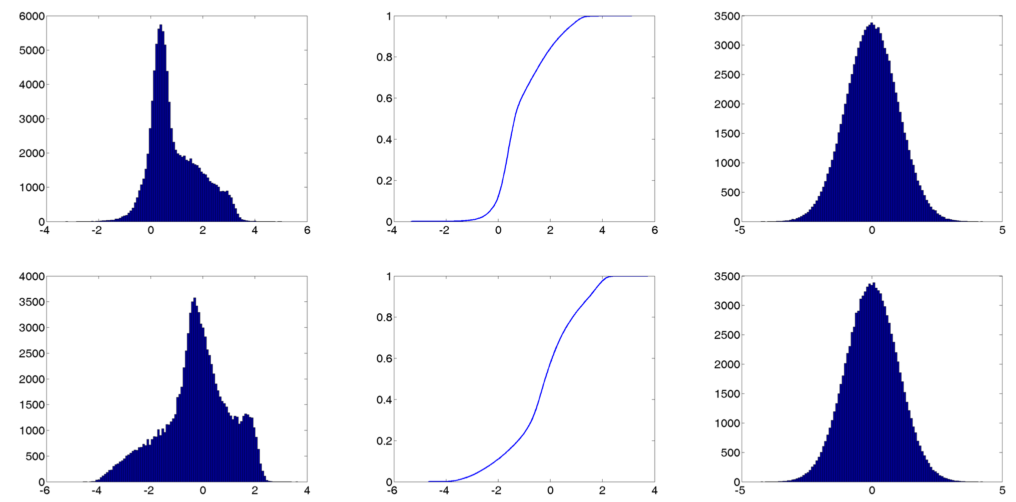
Iterative gaussianization



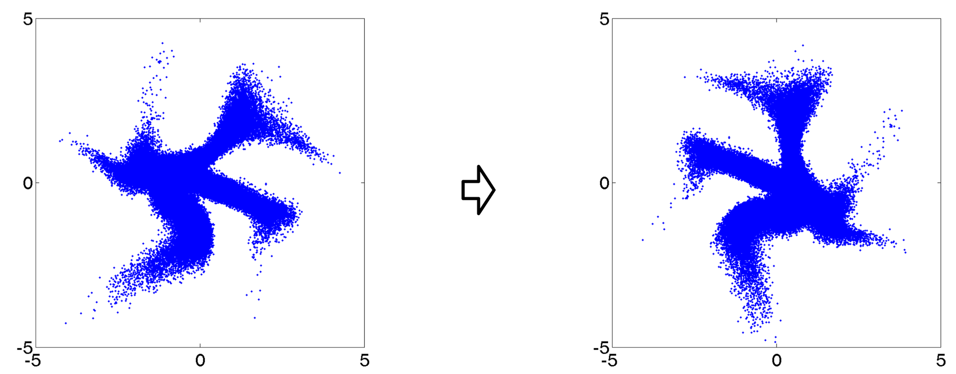
Iterative algorithm with 2 steps per iteration: $\mathbf{x}^{(k+1)} = \mathbf{R}_{(k)}\mathbf{\Psi}_{(k)}(\mathbf{x}^{(k)})$

1. Marginal Gaussianization ($\mathbf{\Psi}$)





2. Rotation ($\mathbf{R}$)



Divergence estimation

$D_{KL}(p(\mathbf{x}), \mathcal{N}(\mathbf{0}, \mathbf{I})) = I(\mathbf{x}) + D_{KLm}(\mathbf{x})$

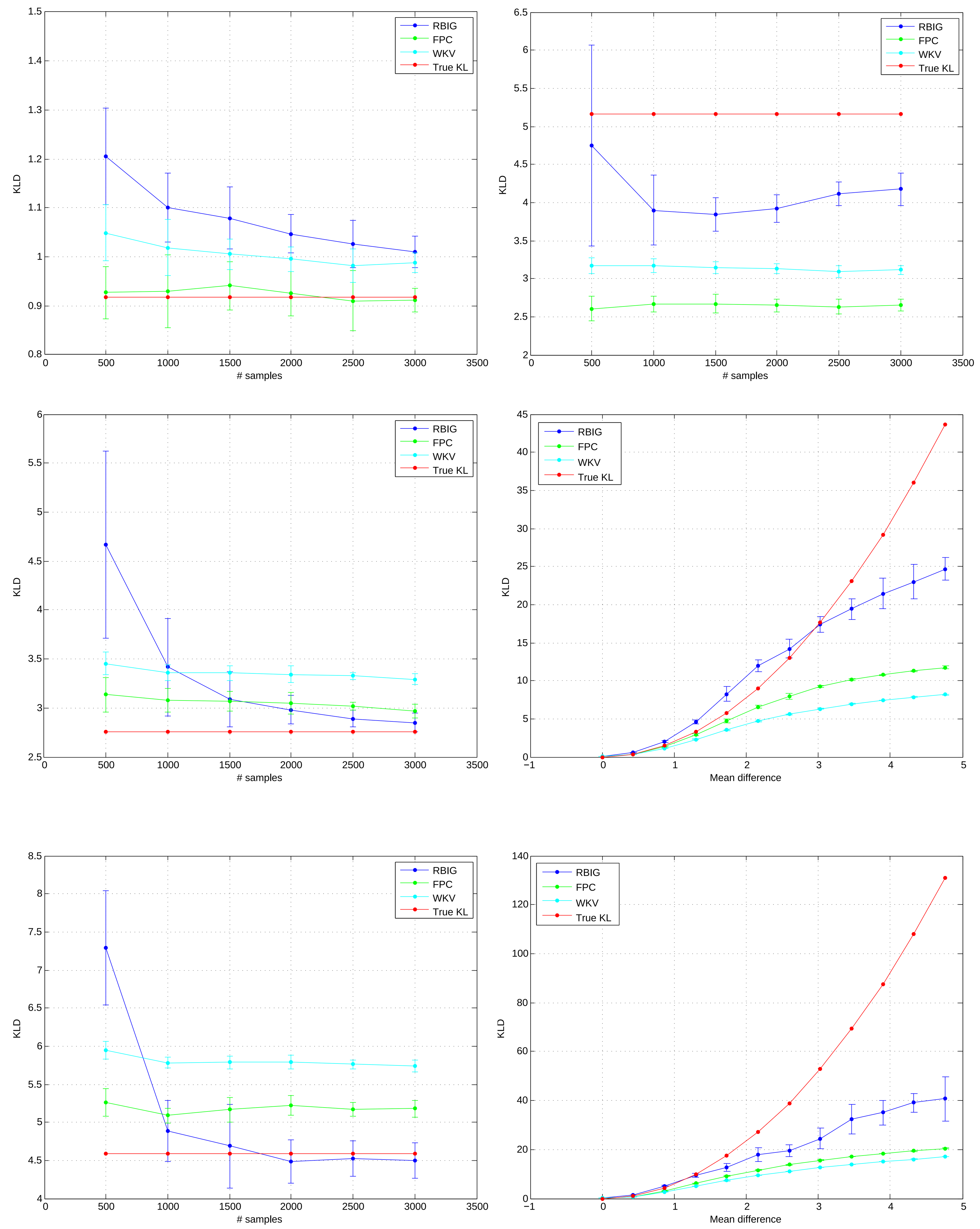
$I(\mathbf{x}) \to$ multi-information

$D_{KLm}(\mathbf{x}) \to$ marginal KL divergence

Estimation: compute $\Delta D_{KL(k)} = D_{KLm}(\mathbf{x}^{(k)}) = \Delta I_{(k-1)}$ at each iteration

Results

Estimation of KL divergence between two Gaussian densities in various shapes and with various dimensionalities. Left figures depict the experimental results for isotropic-isotropic Gaussians of 2, 6 and 10 dimensions. Top right presents isotropic-correlated Gaussians with increasing numbers of samples. Bottom right figures show how the estimation results increase with the increase of the mean difference for a fixed number of data samples.



References

V. Laparra, G. Camps-Valls, J. Malo, Iterative Gaussianization: From ICA to Random Rotations. IEEE Transactions on Neural Networks, 2011

F. Liese, I. Vajda, On Divergences and Informations in Statistics and Information Theory, IEEE Transactions on Information Theory, 2006

Q. Wang, S. R. Kulkarni, S. Verdu, Divergence estimation for multidimensional densities via k-nearest-neighbor distances. IEEE Transactions on Information Theory, 2009

F. Perez-Cruz, Estimation of Information Theoretic Measures for Continuous Random Variables. NIPS, 2008.

Universitat de València