Alternating projection for independent component analysis

Michael Mandel

LabROSA, Columbia University

April 24, 2007

Outline

Outline

1. Independent component analysis
2. Alternating projection
3. Brute force and perceptron solutions
4. Evaluation

Independent component analysis (ICA)

- D-dimensional random variable x, unobserved
- $p(x) = \prod_{d=1}^{D} p_d(x_d)$
- Unknown mixing matrix A
- Observations: N IID samples from $y = Ax$
- Want to recover x_n, $W = A^{-1}$
- Can show that W is orthogonal, $W^T W = I$

Previous work: Jebara (2002)

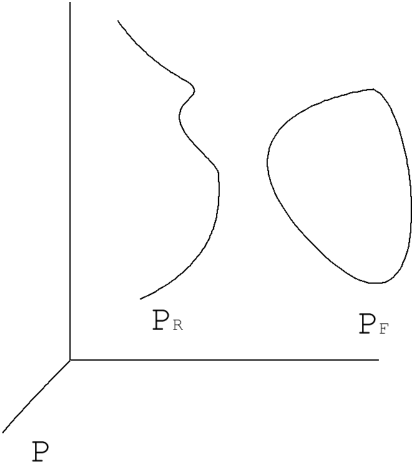
- Describes alternating projection solution to ICA
- Makes no assumptions about $p_d(x_d)$
- Nonparametric
- Uses approximate iterated variational singular value decomposition

This work

- Replaces variational SVD with brute force and perceptron solutions
- Compares all three to Kernel ICA, (Bach & Jordan, 2003) on 18 different 2D distributions
- Shows that these two new methods more accurately estimate W than the variational method, with accuracy close to Kernel ICA

Outline

Alternating projection

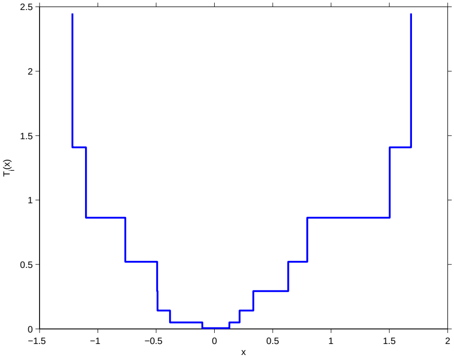


- Two manifolds in probability distribution space
- P_R: space of all rotated versions of empirical distribution
- P_F: space of all product-of-marginal distributions
- Alternate projecting from one to the point on the other with minimum KL divergence

Projecting from rotated to factored

- P^* is the distribution in P_R
- P is any product of marginals, i.e. in P_F
- $\tilde{P}$ is the product of the marginals of P^*
- $KL(P^*||P) = KL(P^*||\tilde{P}) + KL(\tilde{P}||P)$
  (Cover & Thomas, 1991)
- Therefore $KL(P^*||P)$ is minimized when $P = \tilde{P}$

Projecting from factored to rotated



- Dimensions are independent
- Gaussianize *i*th dimension using transfer function $T_i(x)$ based on empirical CDF
- Find W that minimizes cost

$$C(W) = \sum_{in} T_i^2(W_i^T y_n)$$

Outline

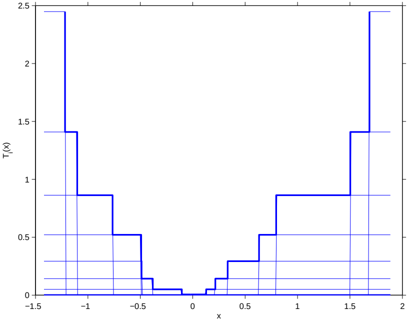
Brute force solution

Repeat until fix point

- Build $T_i(x)$ from the current rotation of data
- Construct W_θ for 500 angles from $-\pi$ to π
- Find $\hat{W}_\theta$ minimizing $C(W_\theta) = \sum_{in} T_i(W_\theta y_n)$
- Rotate data by $\hat{W}_\theta$

- Only works in 2D

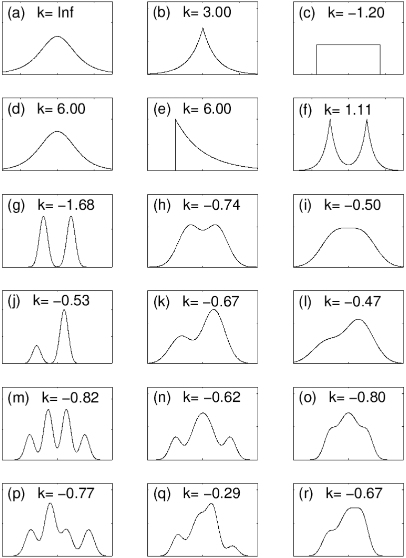
Perceptron solution



- Break cost function into a set of steps, positive and negative
- Treat each step as a classifier
- End up with loss: $L(W) = tr(W^T Q)$
- $Q_i = \sum_n y_n T_i^2(y_n)$
- Solve with SVD
- Works in any dimensionality

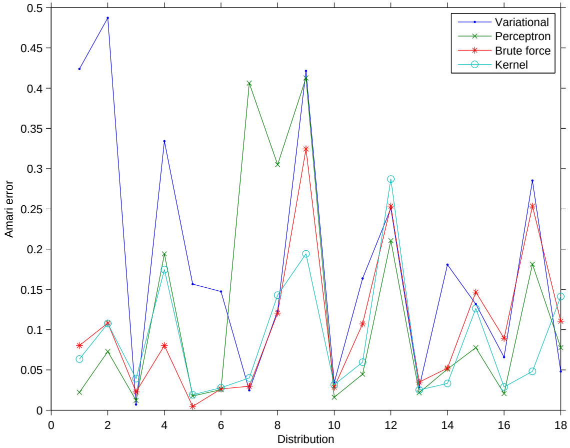
Outline

Distributions



- 18 marginal distributions, 1D
- k = kurtosis
- Same distribution for two independent dimensions
- Draw $N = 250$ data points
- Rotate data by random A matrix
- Compare estimated $\hat{W}$ to actual A matrix using Amari distance, invariant to permutations and reflections

Results: average error over 5 trials



Method	Avg
Variational	0.184
Perceptron	0.121
Brute force	0.104
Kernel	0.088

Summary

- Nonparametric ICA, making no assumptions about source distributions
- Improves upon Jebara (2002)
- Accuracy close to that of kernel ICA
- Still working on convexifying of P_R

Thank you

Any questions?