Independent Subspace Analysis

Special Course in Information Science IV:
Unsupervised Learning

Juha Karvanen

November 9, 2000

Based on

Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces

by Aapo Hyvärinen and Patrik Hoyer

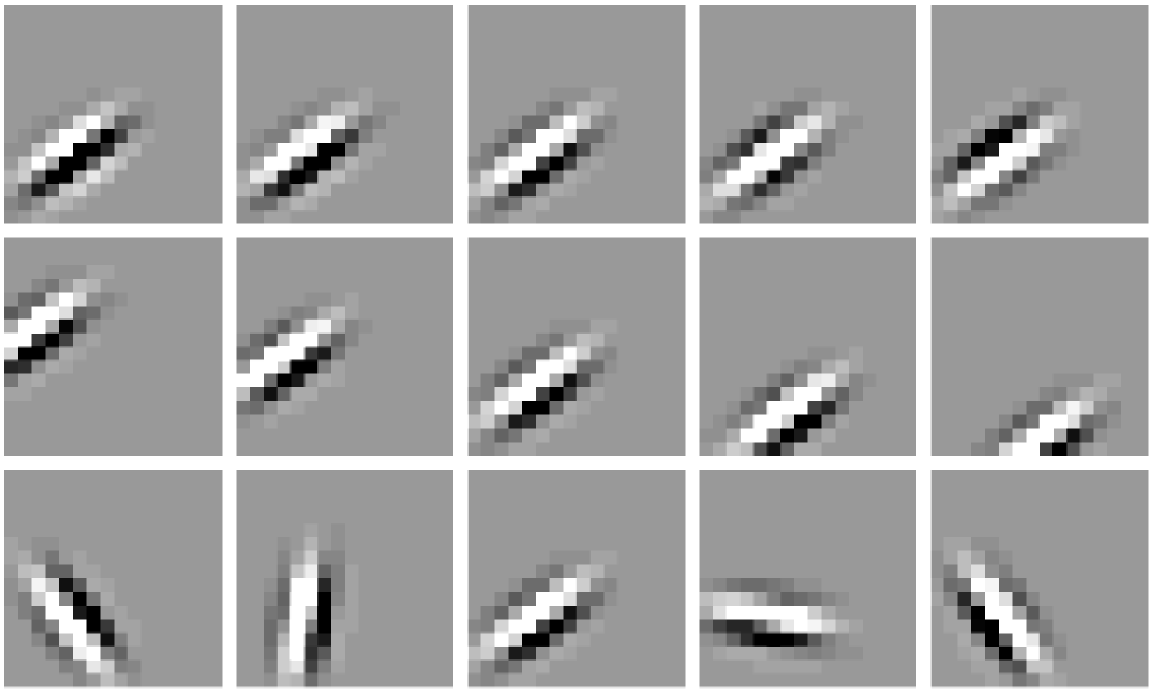
Neural Computation 12, 2000

Outline

- Introduction
- Models for image data
- Invariant feature subspaces
- Multidimensional ICA
- Combining invariant feature subspaces and multidimensional ICA
- Learning algorithm
- Examples

Introduction

- Goal: Model complex cells in visual cortex.
- Complex cells are phase- and shift-invariant.
- Example: varying phase (top row), location (shift) (middle row) and orientation (bottom row)



Model for image data

$$I_k(x,y) = \sum_{i=1}^{m} b_i(x,y) s_i, \quad k = 1, 2, \ldots, K$$

where the $b_i(x,y)$ are basis functions common for all images and the s_i are stochastic coefficients, different for each image $I_k(x,y)$.

The model is assumed invertible

$$s_i = \langle w_i, I \rangle,$$

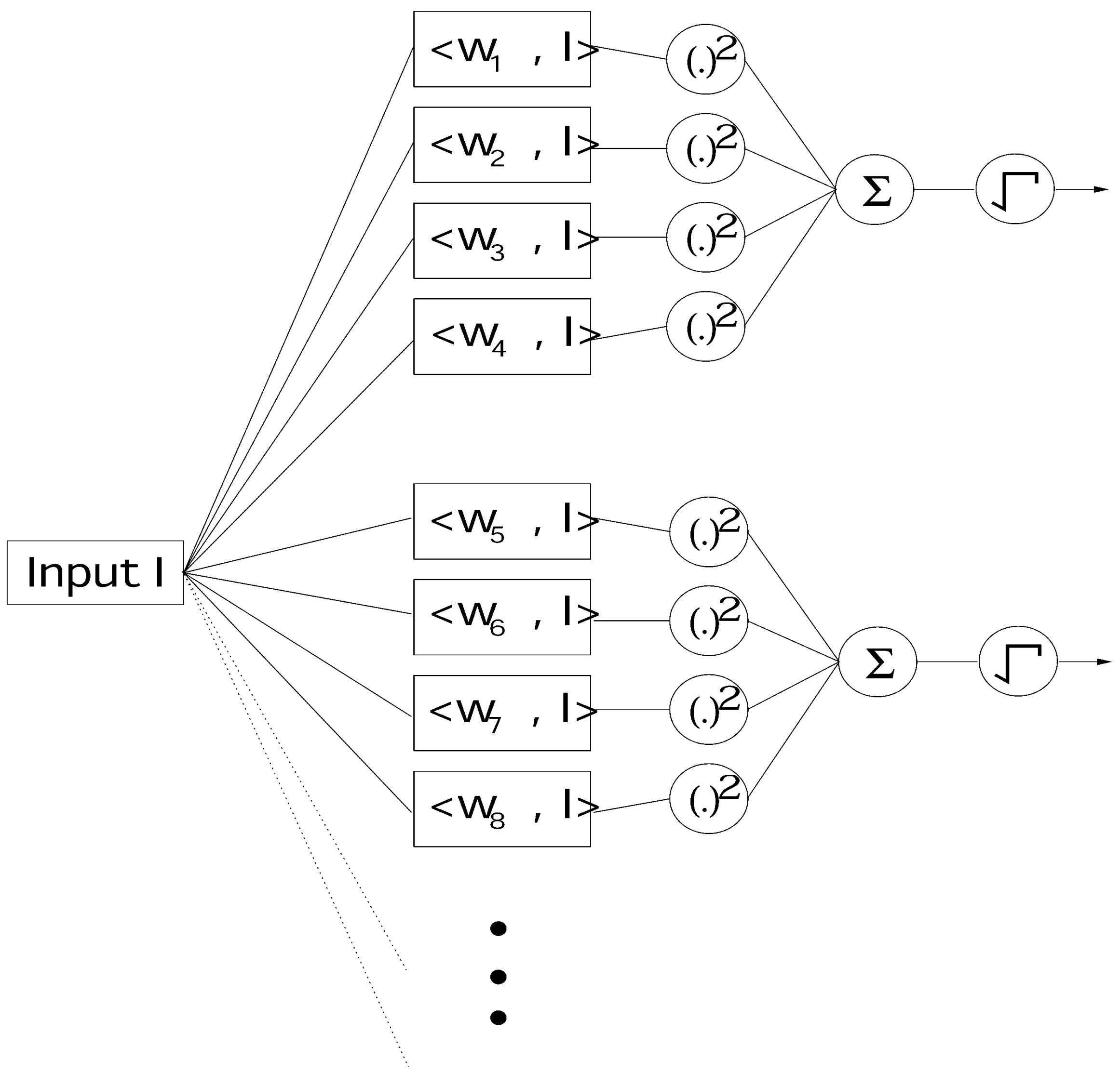
where the w_i denote the inverse filters and $\langle w_i, I \rangle = \sum_{x,y} w_i(x,y) I(x,y)$ denotes the dot product.

- s_i is the activity of w_i when I is presented or correlation between w_i and I

Invariant Feature Subspaces

- Invariant features emerges when the norm of basis vector subspace is used as response

$$F(I) = \sum_{i=1}^{n} s_i = \langle w_i, I \rangle^2 .$$



Multidimensional ICA

- Cardoso 1998
- Components s_i can be divided to the groups: The groups are mutually independent but components inside a group may be dependent on each other.
- Linear dependencies can be removed in whitening
- Likelihood of data $I_1, I_2, \ldots, I_K$ is expressed

  $$\log L\left(I_k(x,y), k = 1 \ldots K; w_i(x,y), i = 1 \ldots m\right) = \prod_{k=1}^{K} \left[|\det \mathbf{W}| \prod_{j=1}^{J} p_j(\langle w_i, I_k \rangle),\ i \in S_j \right]$$

  where
  J is the number of independent feature subspaces,
  S_j, $j = 1, \ldots, J$ is the set of the indices of the s_i belonging to the subspace of index j,
  $p_j(.)$ is the probability density inside the jth n-tuple of s_i,
  $\mathbf{W}$ is a matrix containing the filters $w_i(x,y)$ as its columns.

Combining invariant feature subspaces and multidimensional ICA

- Probability density $p_j(.)$ is expressed as function of the sum of squares of the $s_i, i \in S_j$.
- Model reduces to

$$\log L\left(I_k(x,y), k = 1 \ldots K; w_i(x,y)\right) =$$

$$\sum_{k=1}^{K}\sum_{j=1}^{J} \log p\left(\sum_{i \in S_j} \langle w_i, I_k \rangle^2\right) + K \log|\det \mathbf{W}|,$$

  where $i = 1, \ldots, m$ and $p\left(\sum_{i \in S_j} s_i^2\right) = p_j(s_i, i \in S_j)$ gives the probability density inside the jth n-tuple of s_i.

- For example, (multidimensional) exponential distribution can be used

$$\log p\left(\sum_{i \in S_j} s_i^2\right) = -\alpha \left[\sum_{i \in S_j} s_i^2\right]^{1/2} + \beta$$

  where α and β are scaling and normalization constants.

Learning algorithm

- A stochastic gradient ascent of the log-likelihood is

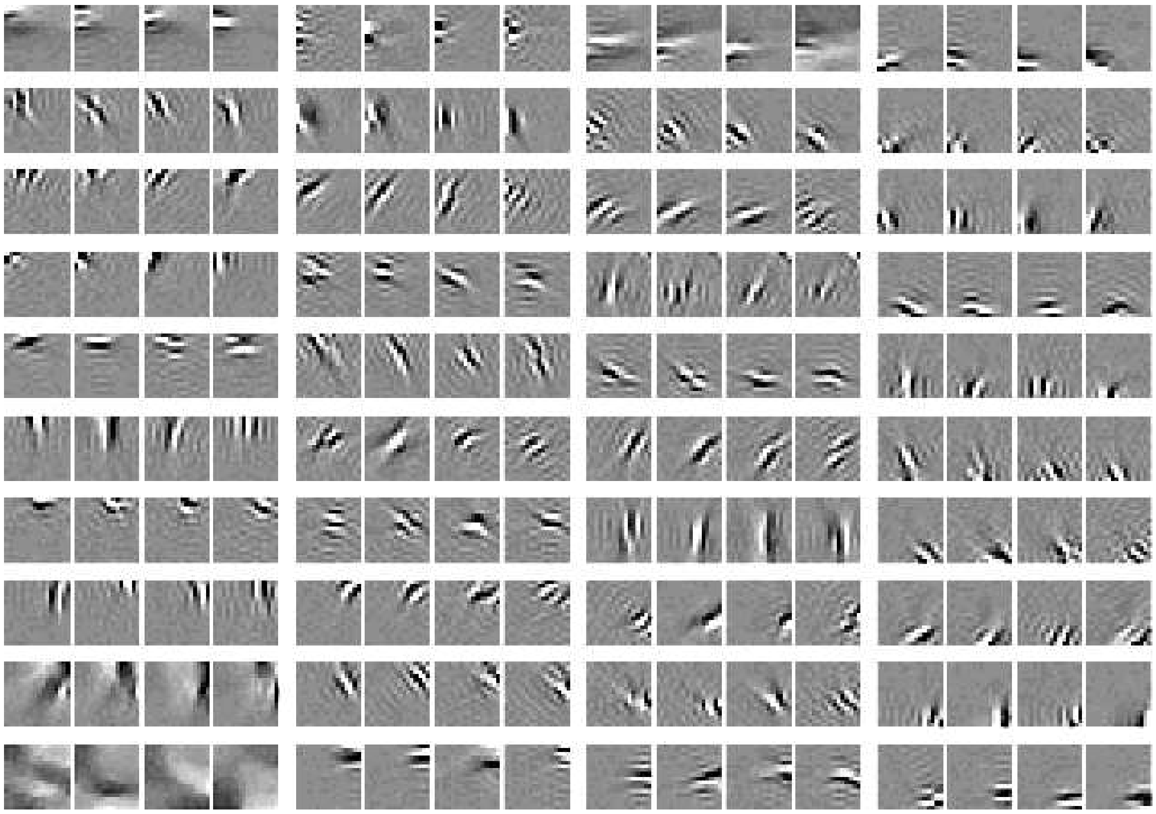
$$\Delta w_i(x,y) \propto I(x,y)\langle w_i, I\rangle g\left(\sum_{r\in S_{j(i)}} \langle w_r, I\rangle^2\right),$$

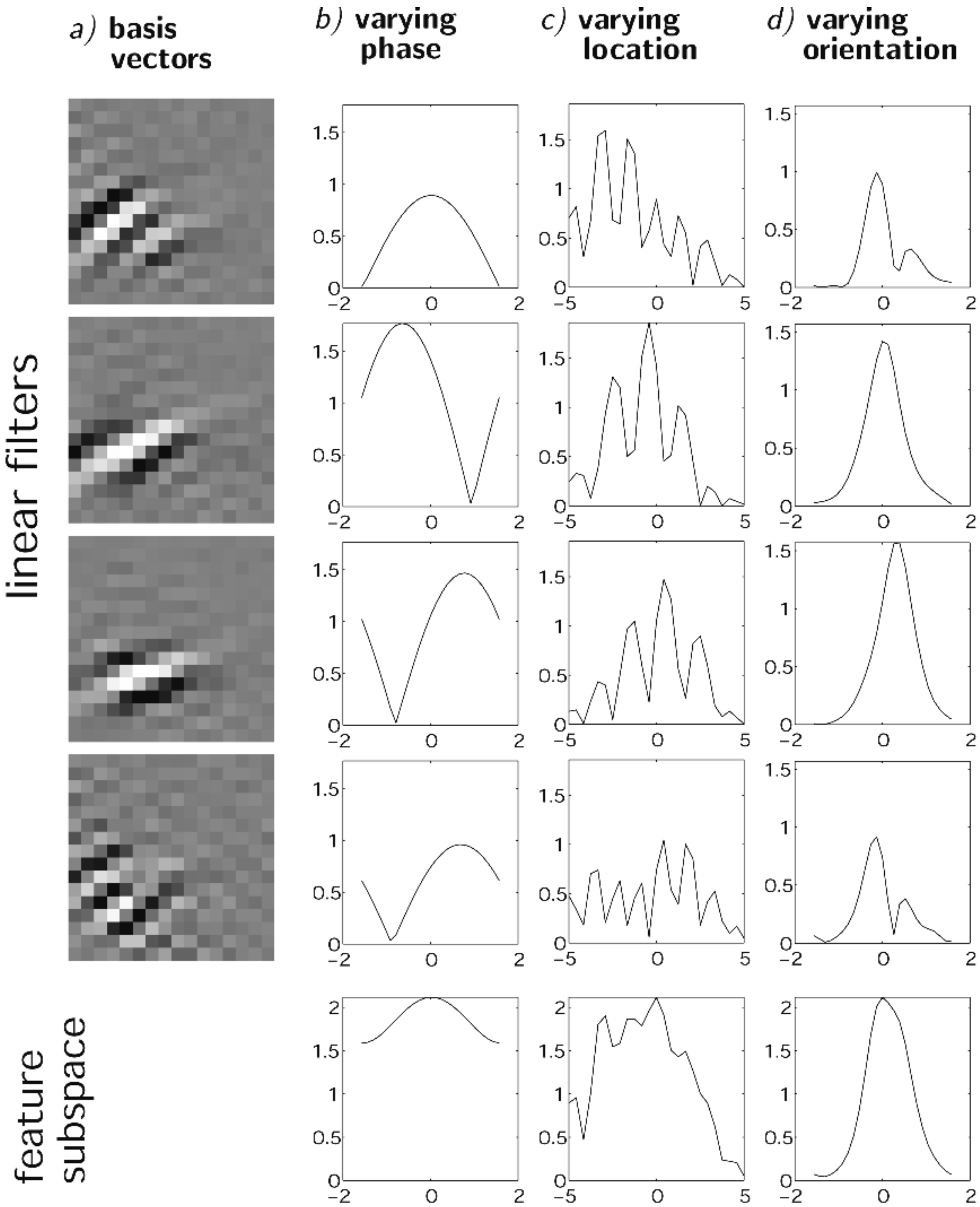
  where $j(i)$ is the index of subspace to which w_i belongs
  and $g = p'/p$ is the score function of p.

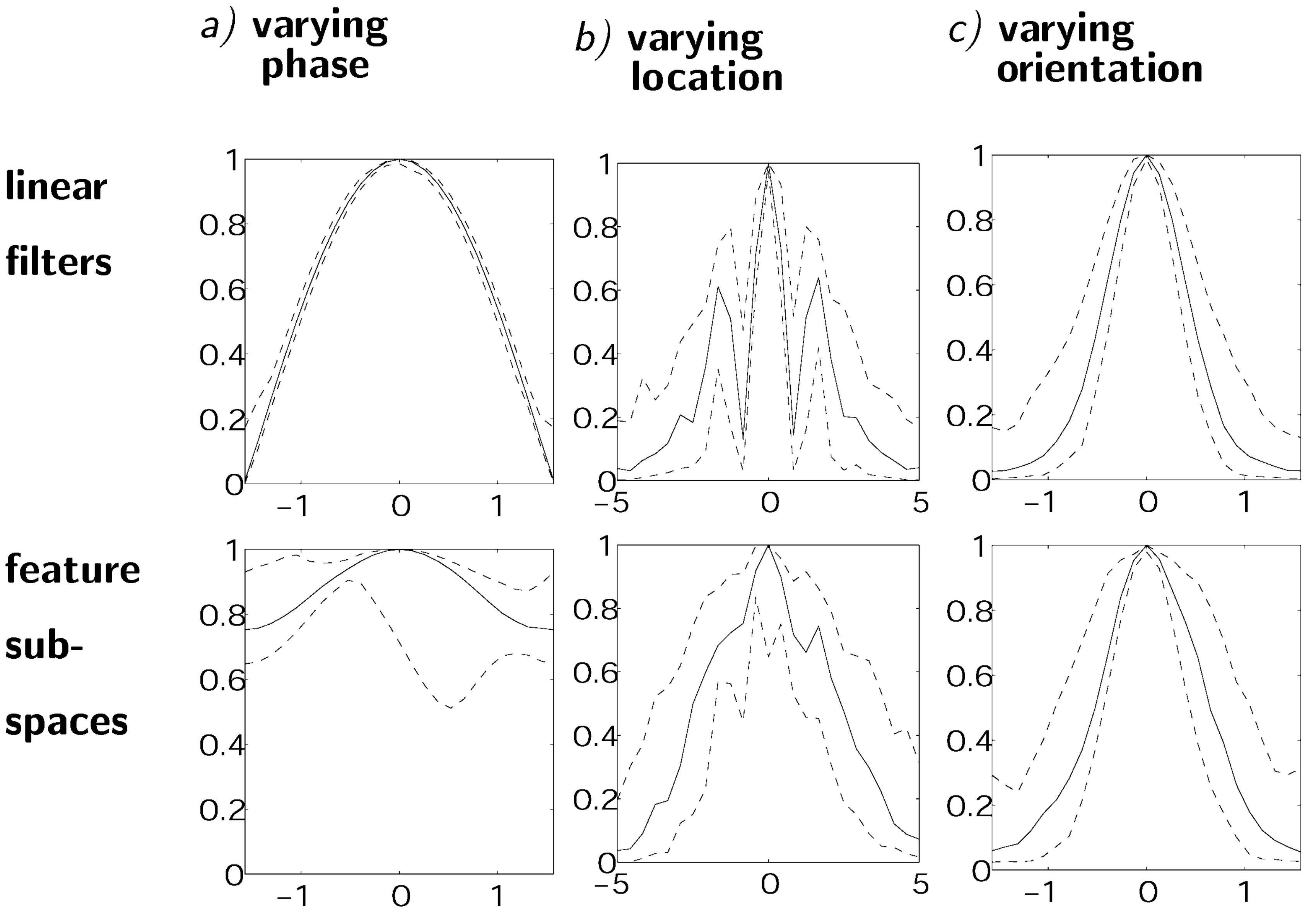
- for exponential distribution $g(u) = -\frac{1}{2}\alpha u^{-1/2}$

Example

- 16×16 pixel image patches from wildlife monochrome photographs were used.
- Resulting linear filters for subspace dimension 4:




a) basis vectors
b) varying phase
c) varying location
d) varying orientation
linear filters
feature subspace


a) varying phase
b) varying location
c) varying orientation
linear filters
feature sub- spaces