

sample from. Next we sample the real-valued observable variables given the factors:

$$\mathbf{x} = \mathbf{W}\mathbf{h} + \mathbf{b} + \text{noise} \quad (13.2)$$

where the noise is typically Gaussian and diagonal (independent across dimensions). This is illustrated in figure 13.1.

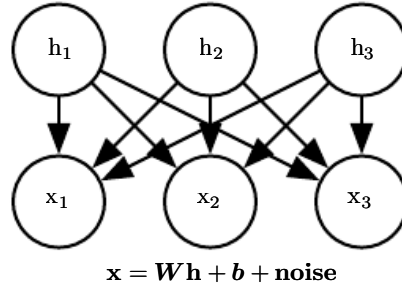


Figure 13.1: The directed graphical model describing the linear factor model family, in which we assume that an observed data vector  $\mathbf{x}$  is obtained by a linear combination of independent latent factors  $\mathbf{h}$ , plus some noise. Different models, such as probabilistic PCA, factor analysis or ICA, make different choices about the form of the noise and of the prior  $p(\mathbf{h})$ .

## 13.1 Probabilistic PCA and Factor Analysis

Probabilistic PCA (principal components analysis), factor analysis and other linear factor models are special cases of the above equations (13.1 and 13.2) and only differ in the choices made for the noise distribution and the model's prior over latent variables  $\mathbf{h}$  before observing  $\mathbf{x}$ .

In **factor analysis** (Bartholomew, 1987; Basilevsky, 1994), the latent variable prior is just the unit variance Gaussian

$$\mathbf{h} \sim \mathcal{N}(\mathbf{h}; \mathbf{0}, \mathbf{I}) \quad (13.3)$$

while the observed variables  $x_i$  are assumed to be **conditionally independent**, given  $\mathbf{h}$ . Specifically, the noise is assumed to be drawn from a diagonal covariance Gaussian distribution, with covariance matrix  $\boldsymbol{\psi} = \text{diag}(\boldsymbol{\sigma}^2)$ , with  $\boldsymbol{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2]^\top$  a vector of per-variable variances.

The role of the latent variables is thus to *capture the dependencies* between the different observed variables  $x_i$ . Indeed, it can easily be shown that  $\mathbf{x}$  is just a multivariate normal random variable, with

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}; \mathbf{b}, \mathbf{W}\mathbf{W}^\top + \boldsymbol{\psi}). \quad (13.4)$$