Autoencoders

Lecture slides for Chapter 14 of *Deep Learning*

www.deeplearningbook.org

Ian Goodfellow

2016-09-30

Structure of an Autoencoder

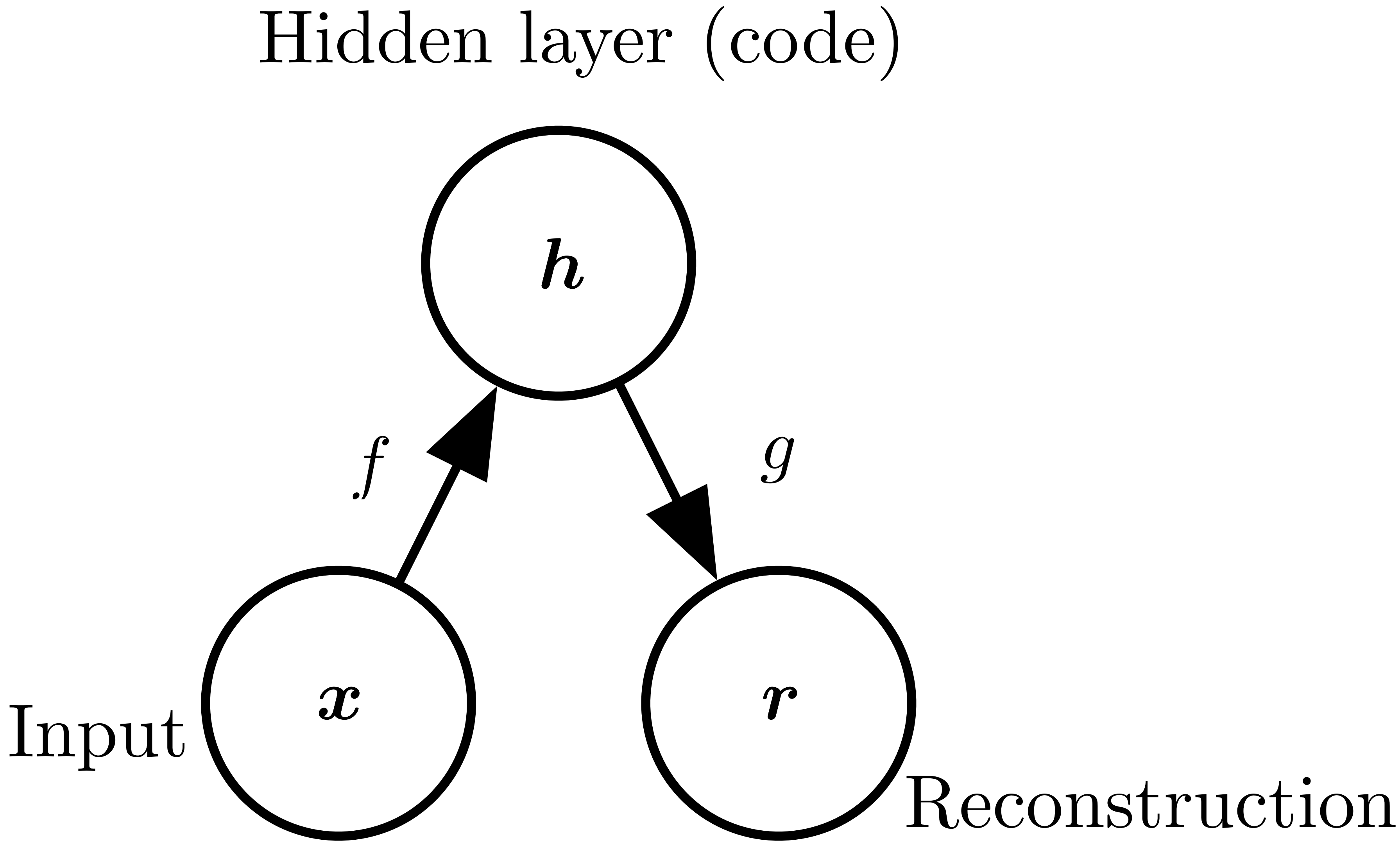


Figure 14.1

Stochastic Autoencoders

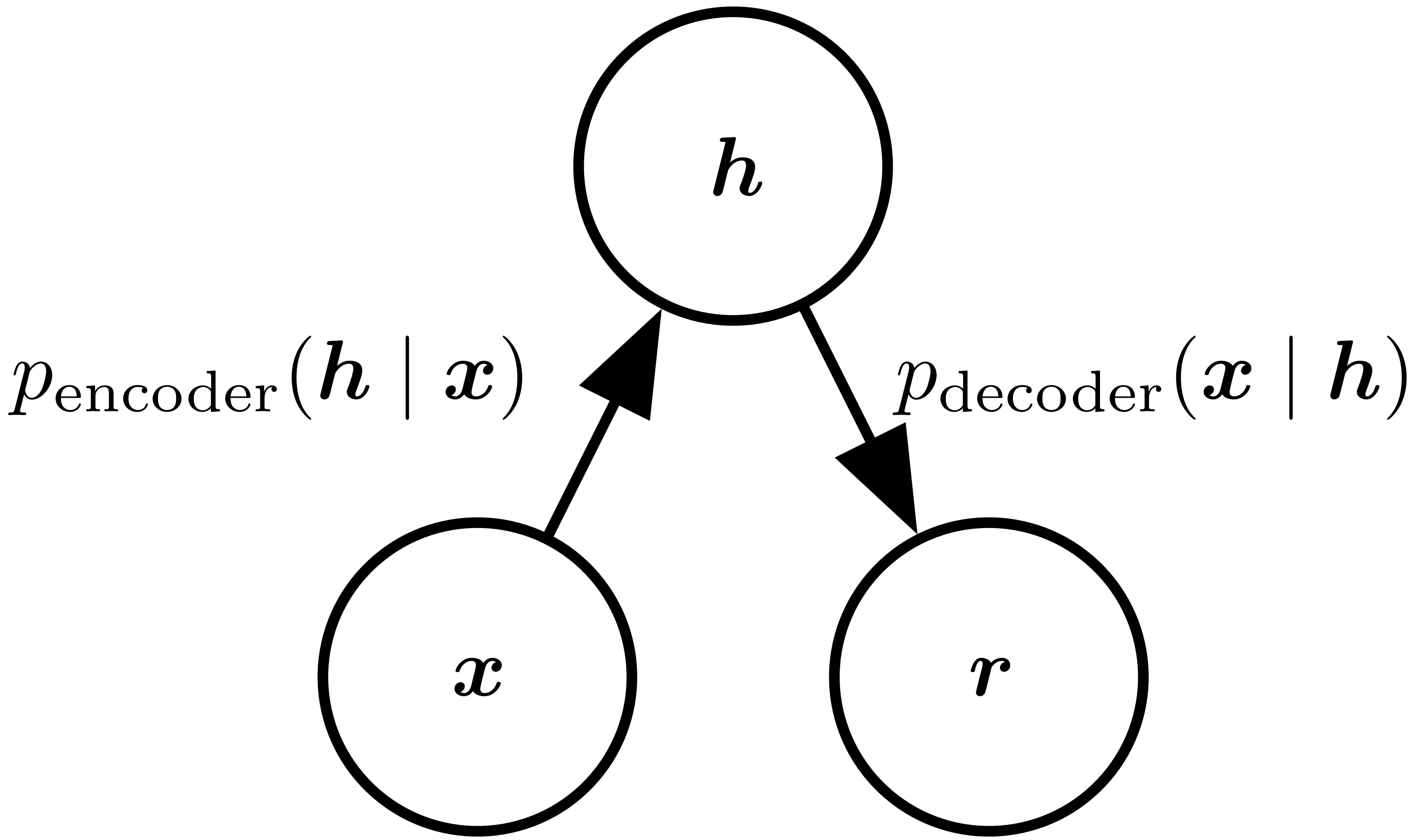


Figure 14.2

(Goodfellow 2016)

Avoiding Trivial Identity

- Undercomplete autoencoders
  - h has lower dimension than x
  - f or g has low capacity (e.g., linear g)
  - Must discard some information in h
- Overcomplete autoencoders
  - h has higher dimension than x
  - Must be regularized

Regularized Autoencoders

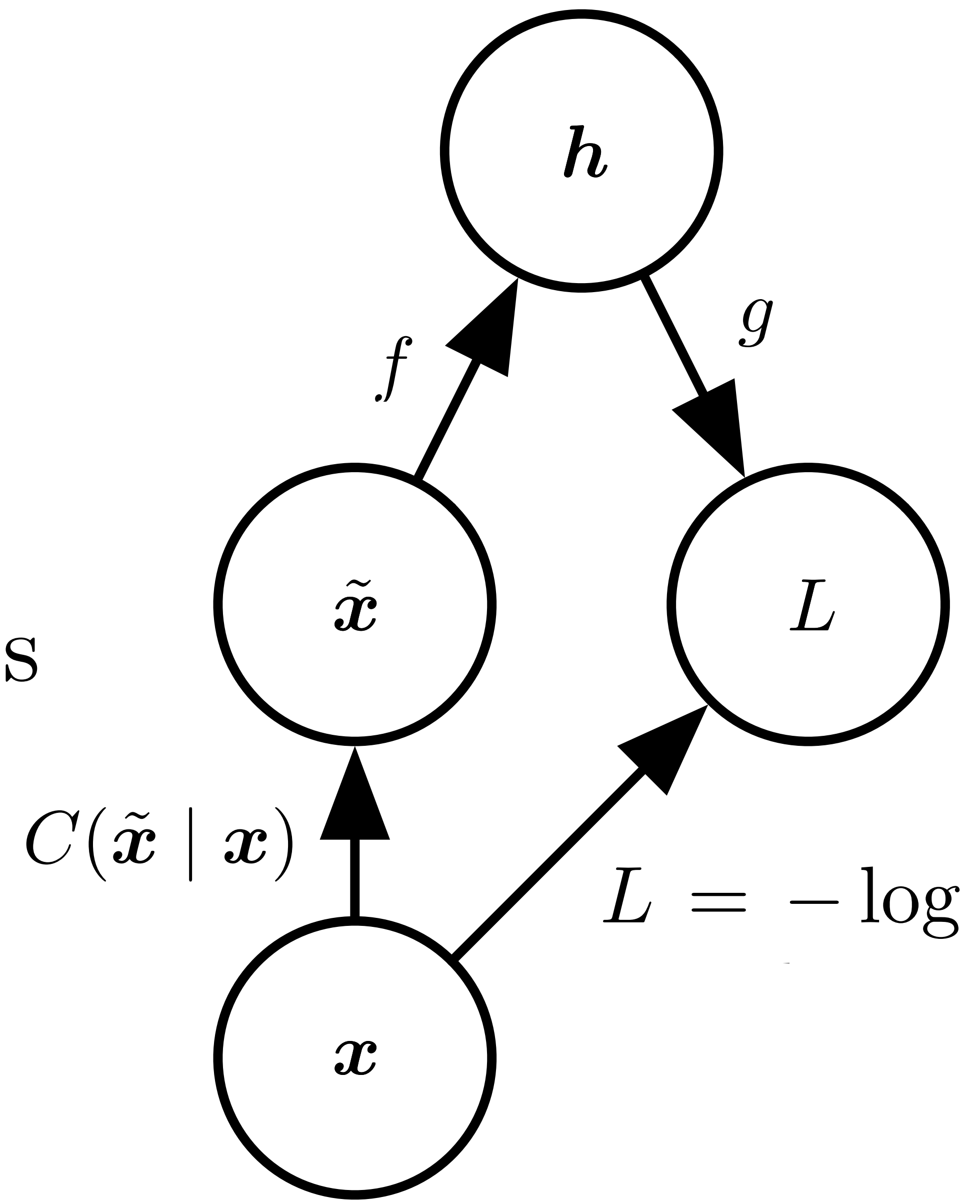
- Sparse autoencoders
- Denoising autoencoders
- Autoencoders with dropout on the hidden layer
- Contractive autoencoders

Sparse Autoencoders

- Limit capacity of autoencoder by adding a term to the cost function penalizing the code for being larger
- Special case of variational autoencoder
  - Probabilistic model
  - Laplace prior corresponds to L1 sparsity penalty
  - Dirac variational posterior

Denoising Autoencoder

C: corruption process
(introduce noise)



$L = -\log p_{\text{decoder}}(\boldsymbol{x} \mid \boldsymbol{h} = f(\tilde{\boldsymbol{x}}))$

Figure 14.3

Denoising Autoencoders Learn a Manifold

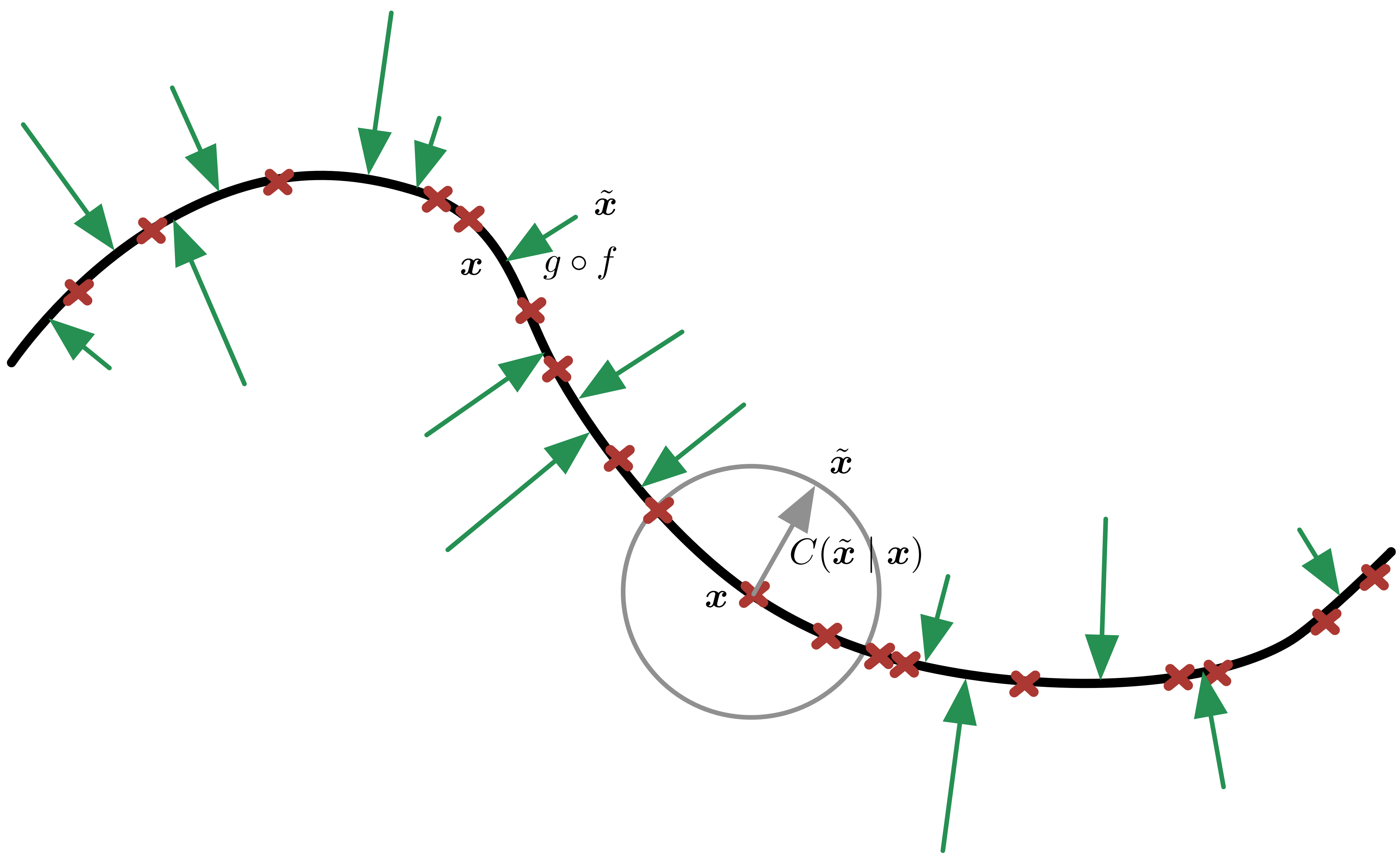


Figure 14.4

Score Matching

- Score: $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}).$ (14.15)

- Fit a density model by matching score of model to score of data

- Some denoising autoencoders are equivalent to score matching applied to some density models

Vector Field Learned by a Denoising Autoencoder

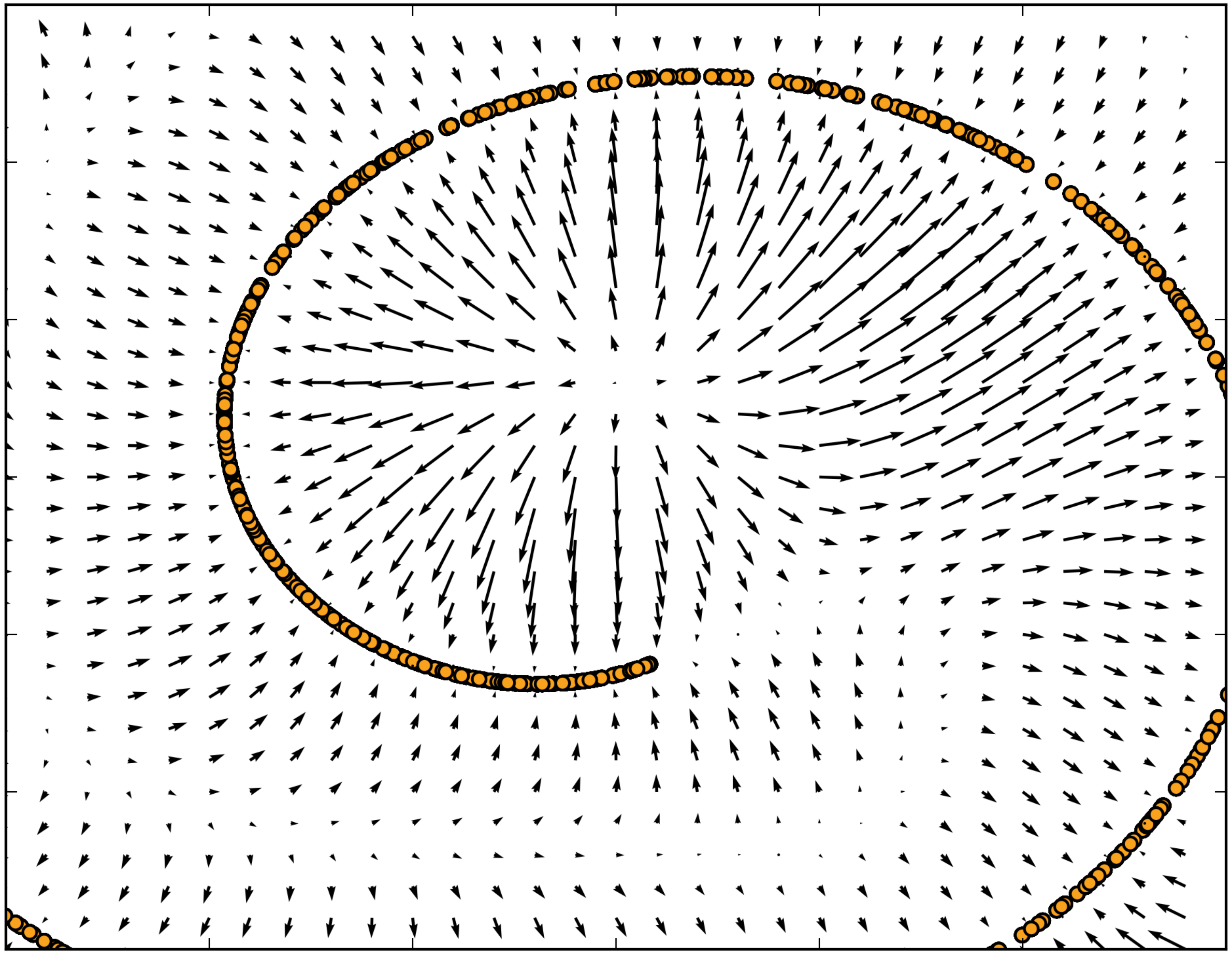


Figure 14.5

(Goodfellow 2016)

Tangent Hyperplane of a Manifold

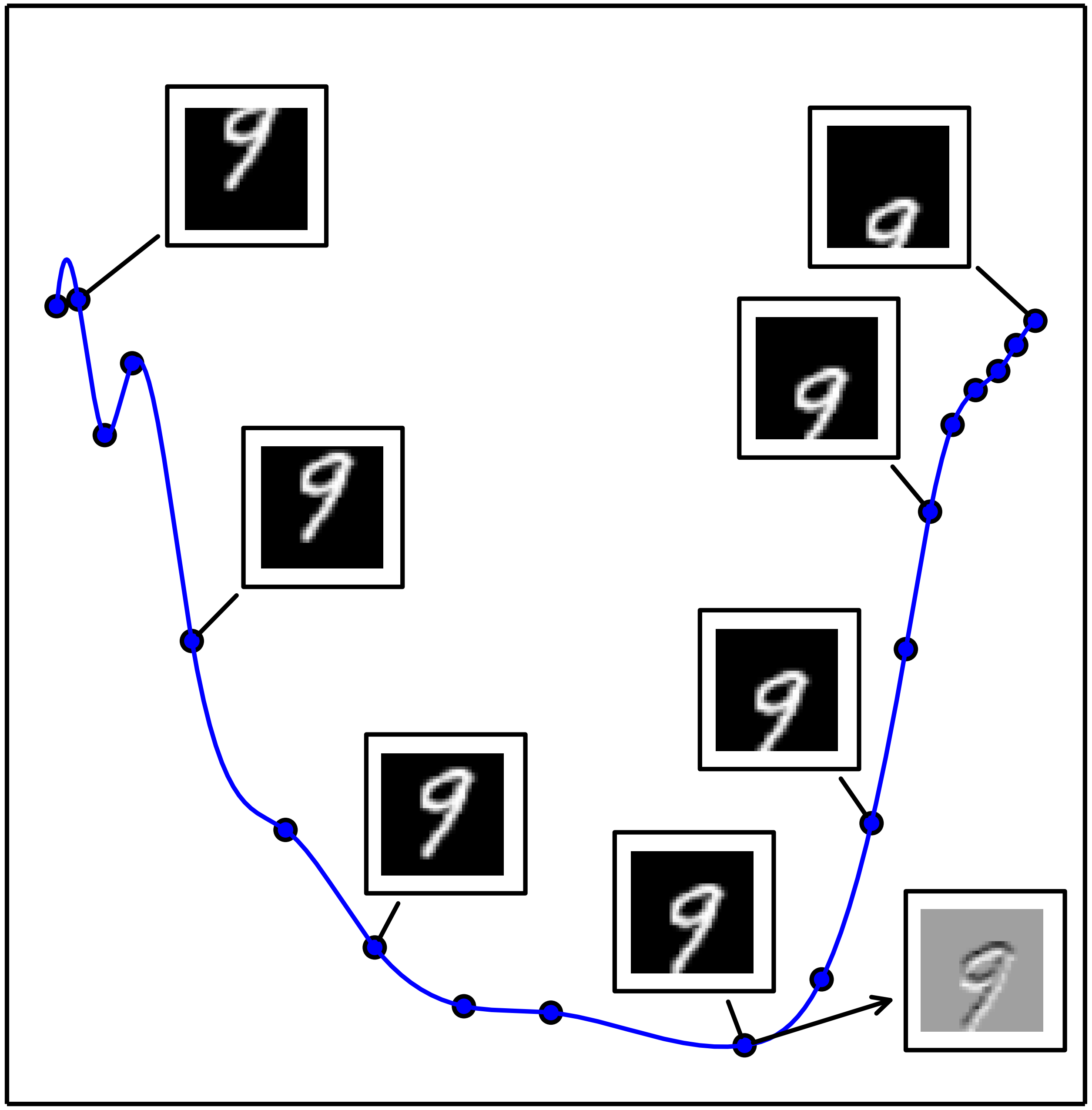


Figure 14.6

Learning a Collection of 0-D Manifolds by Resisting Perturbation

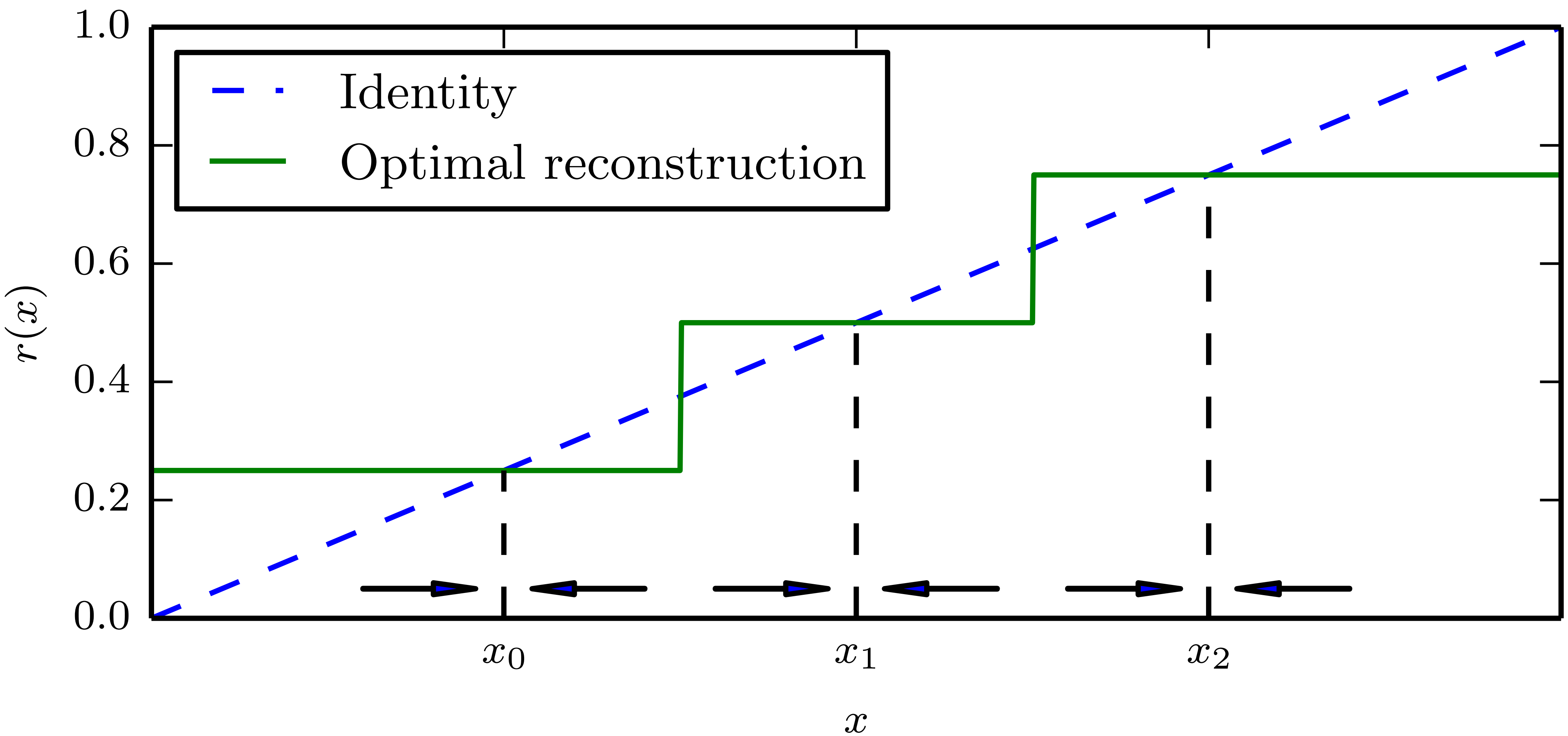


Figure 14.7

Non-Parametric Manifold Learning with Nearest-Neighbor Graphs

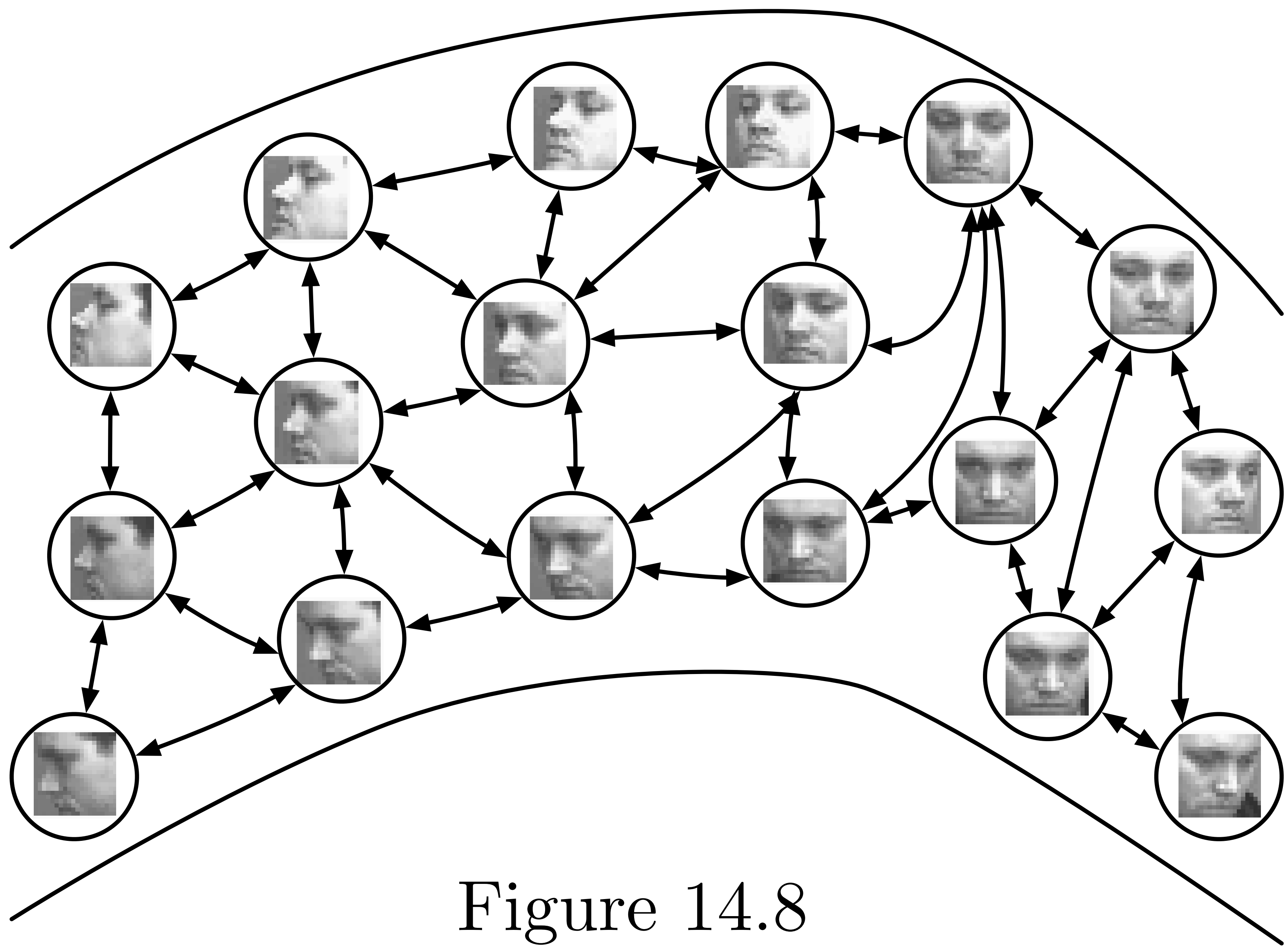


Figure 14.8

Tiling a Manifold with Local Coordinate Systems

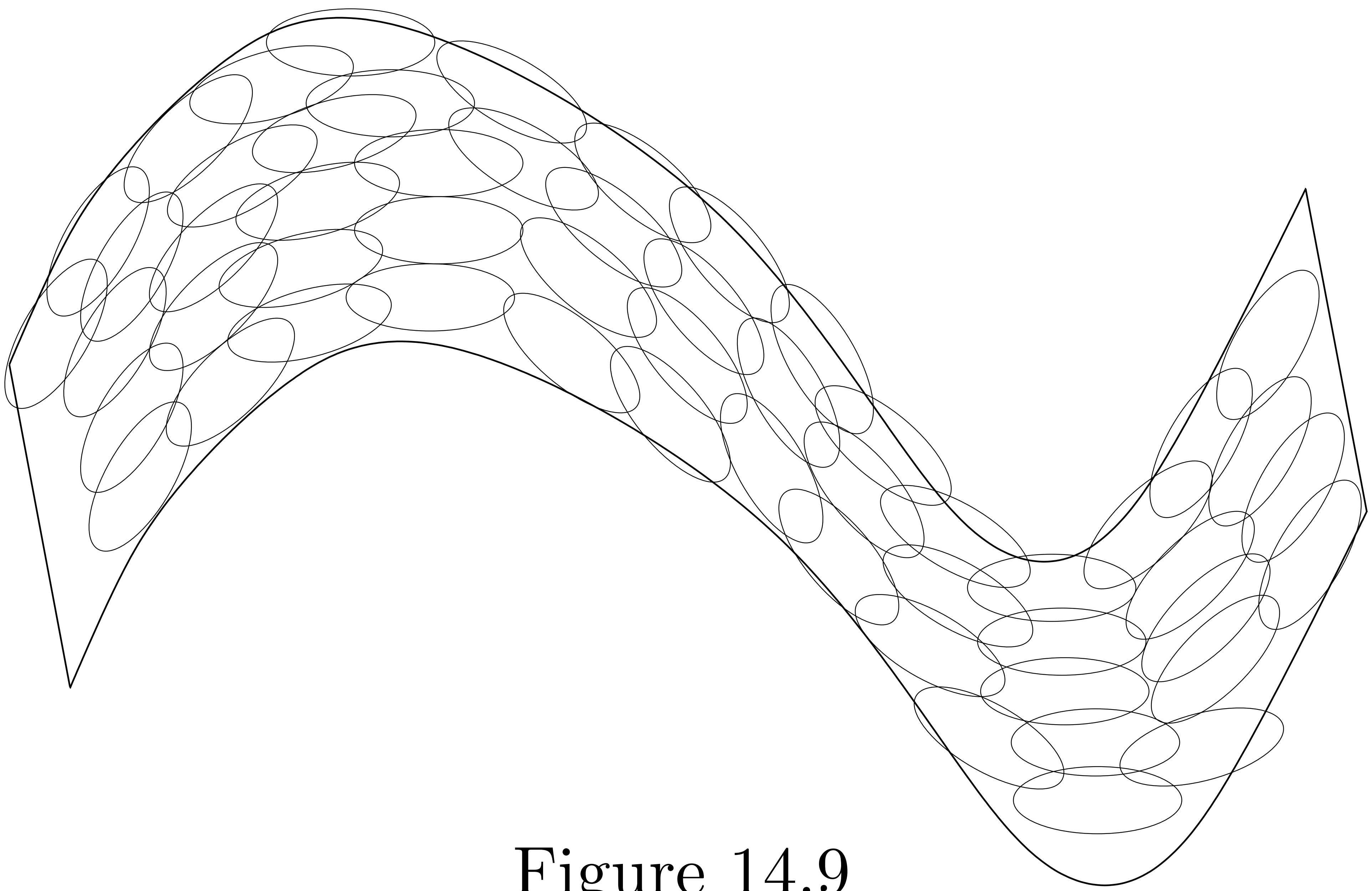


Figure 14.9

(Goodfellow 2016)

Contractive Autoencoders

$$\Omega(\boldsymbol{h}) = \lambda \left\| \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right\|_F^2 . \tag{14.18}$$

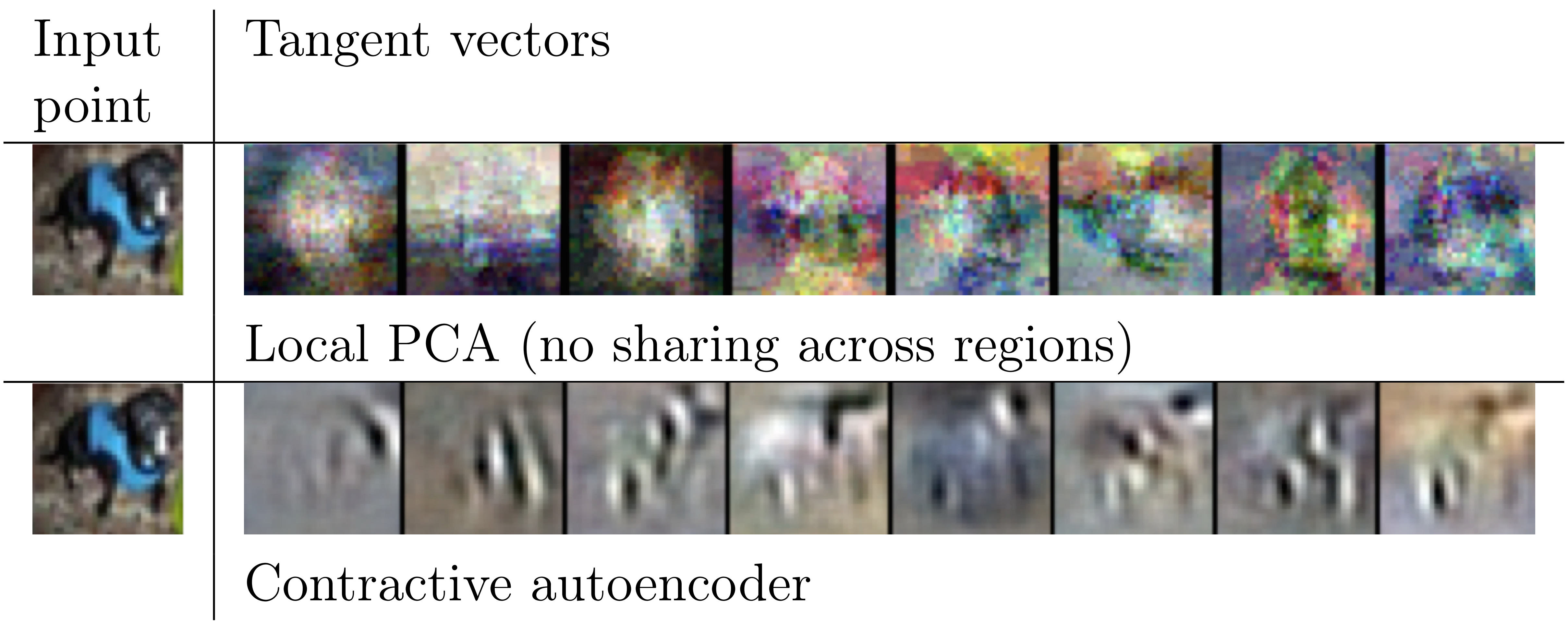


Figure 14.10

(Goodfellow 2016)