EEG582: Artificial Neural Networks

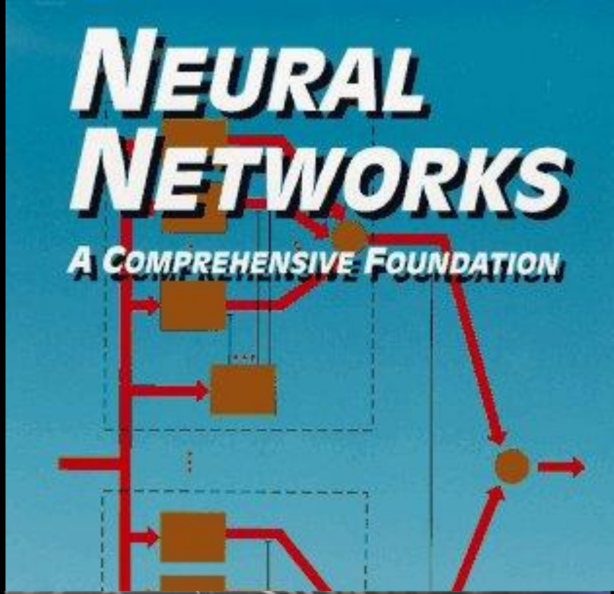
Neural Networks, A Comprehensive Foundation
by S. Haykin
Neural Networks and Deep Learning, by Michael Nielsen
http://neuralnetworksanddeeplearning.com
Introduction to Convolutional Neural Networks, by Vicky Kalogeiton

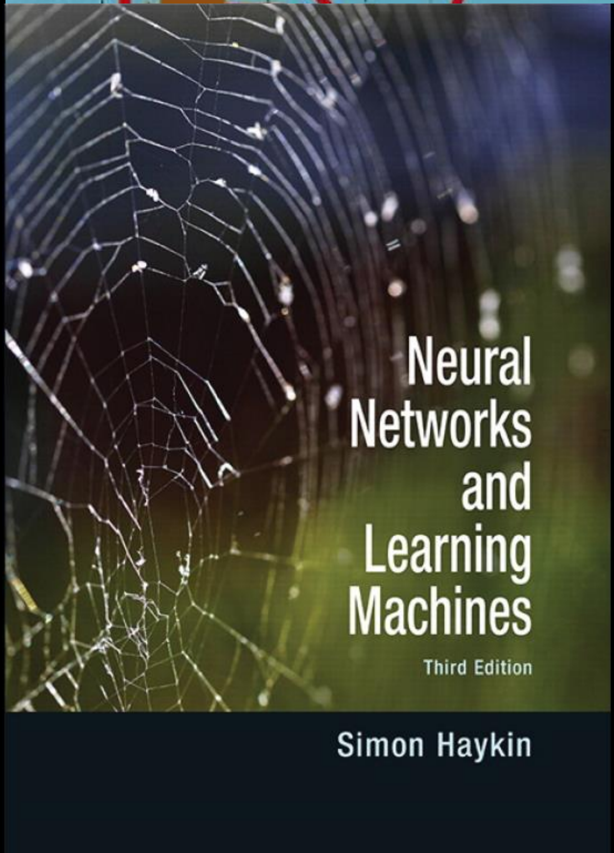
Sections 4.18

4 Multilayer Perceptrons

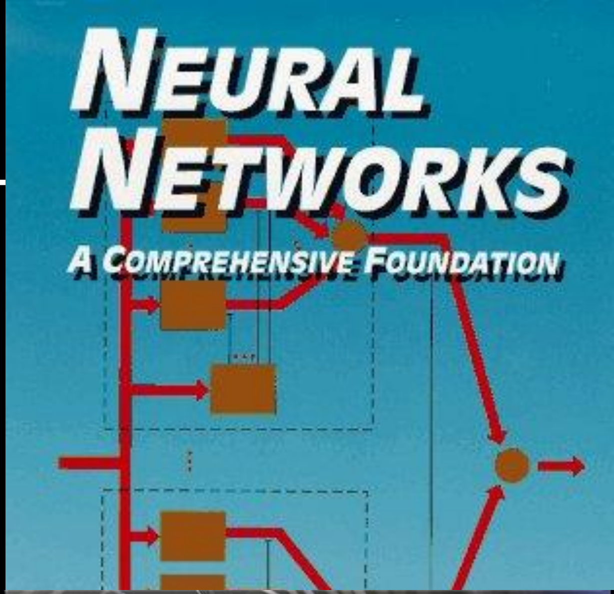
Prof. Dr. Hasan AMCA
Electrical and Electronic Engineering Department
(ee.emu.edu.tr)

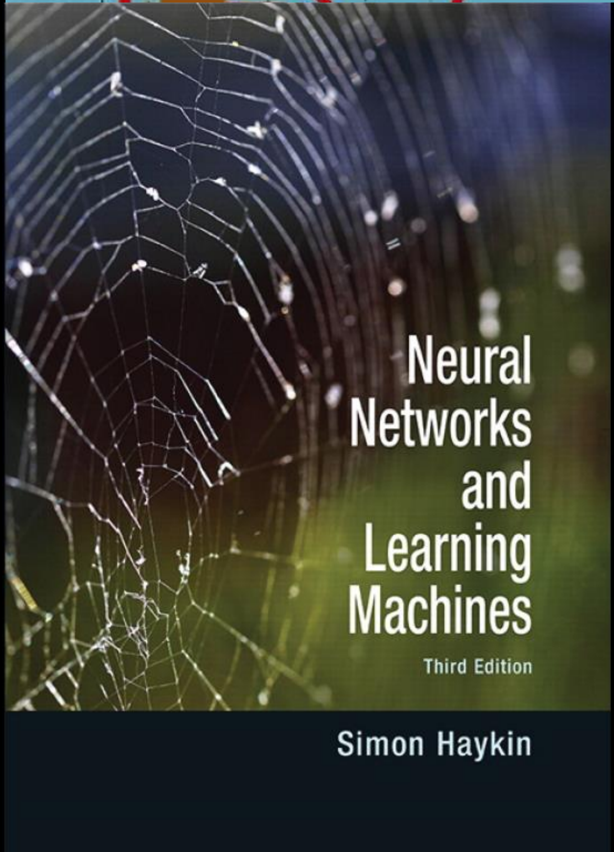
Eastern Mediterranean University
(emu.edu.tr)





4 Multi Layer Perceptrons





Prepared by Prof. Dr. Hasan AMCA

4.18 Nonlinear Filtering

- The typical use of a static neural network is in structural pattern recognition.
- In contrast, in nonlinear filtering, the requirement is to process patterns that evolve over time, with the response at a particular instant of time depending not only on the present value of the input signal but also on past values.
- For a neural network to be dynamic, it must be given short-term memory through the use of time delays, which can be implemented at the synaptic level inside the network or externally at the input layer of the network.

4.18 Nonlinear Filtering

- Time may be built into the operation of a neural network in two basic ways:
  - ~ **Implicit representation**. In digital implementation of the neural network, the input signal is uniformly sampled and sequence of synaptic weights of each neuron connected to the input layer of the network is convolved with a different sequence of input samples. In so doing, temporal structure of the input signal is embedded in the spatial structure of the network.
  - ~ **Explicit representation**. The echolocation system of a bat operates by emitting a short frequency-modulated (FM) signal, so that the same intensity level is maintained for each frequency channel restricted to a very short period within the FM sweep.

    When an echo is received from the target with an unknown delay, a neuron (in the auditory system) with a matching delay line responds, thereby providing an estimate of the distance to the target.

4.18 Nonlinear Filtering

- Fig. 4.24 shows the block diagram of a nonlinear filter consisting of the cascade connection of two subsystems:
  - short-term memory and
  - static neural network (e.g., multilayer perceptron).
- The static network accounts for *nonlinearity* and the memory accounts for *time*.
- To be specific, suppose we are given a multilayer perceptron with an input layer of size *m*.
- Then, the memory is a single-input, multiple-output (SIMO) structure providing *m* differently delayed versions of the input signal for stimulating the neural network.

4.18 Nonlinear Filtering

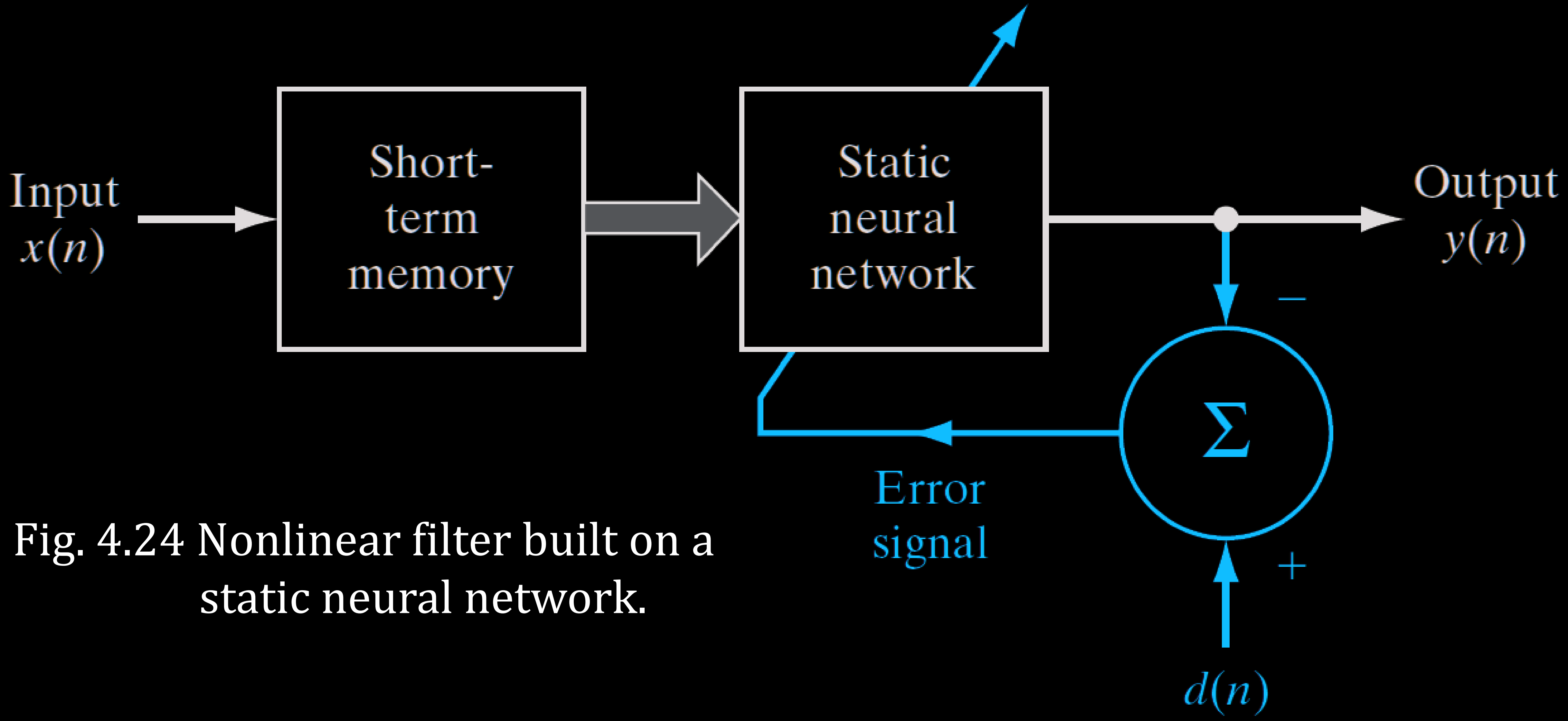


Fig. 4.24 Nonlinear filter built on a static neural network.

4.18 Nonlinear Filtering
Short-Term Memory Structures

- Fig. 4.25 shows the block diagram of a discrete-time memory structure consisting of p identical sections connected in cascade.
- Each section is characterized by an impulse response, denoted by $h(n)$, where n denotes discrete time.
- The number of sections, p, is called the order of the memory.
- Correspondingly, the number of output terminals (i.e., taps) provided by the memory is $p + 1$, which includes the direct connection from the input to the output.
- Thus, with m denoting the size of the input layer of the static neural network, we may set $m = p + 1$.

4.18 Nonlinear Filtering
Short-Term Memory Structures

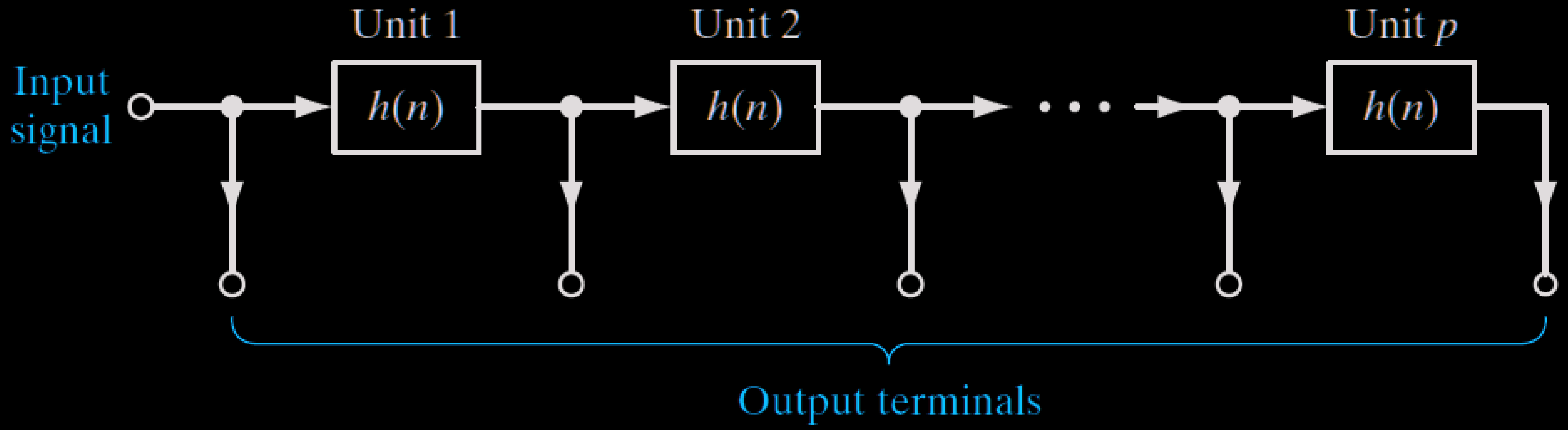


Fig. 4.25 Generalized tapped-delay-line memory of order p.

4.18 Nonlinear Filtering
Short-Term Memory Structures

- The impulse response of each delay section of the memory satisfies two properties:
  - *causality*, which means that $h(n)$ is zero for $n < 0$;
  - *normalization*, which means that $\sum_{n=0}^{\infty}|h(n)| = 1$.
- where $h(n)$ is the generating kernel of the discrete time memory.
- The *memory depth* D is defined as *the first-time moment* of $h_{overall}(n)$, namely

$$D = \sum_{n=0}^{\infty} n h_{overall}(n) \qquad (4.153)$$

- A memory of low depth D holds information content for a relatively short time interval, whereas a high-depth memory holds it much further into past.
- *Memory resolution* R is defined as number of taps in the memory structure per unit of time. For a fixed memory order p, the product of memory depth D and memory resolution R is a constant that turns out to be equal to p.

4.18 Nonlinear Filtering
Short-Term Memory Structures

- Different choices of generating kernel $h(n)$ result in different values for depth D and memory resolution R, as illustrated by the following two memory structures:

  1. *Tapped-delay-line memory*, for which the generating kernel is defined by the unit impulse $\delta(n)$; that is,

$$h(n) = \delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \tag{4.154}$$

  And the overall impulse response is

$$h_{\text{overall}}(n) = \delta(n - p) = \begin{cases} 1, & n = p \\ 0, & n \neq p \end{cases} \tag{4.155}$$

4.18 Nonlinear Filtering
Short-Term Memory Structures

2. *Gamma memory*, for which the generating kernel is defined by,

$$h(n) = \mu(1-\mu)^{n-1}, \qquad n \geq 1 \tag{4.156}$$

where μ is an adjustable parameter.

For $h(n)$ to be convergent (i.e., for the short-term memory to be stable), we require that, $0 < \mu < 2$ and the overall impulse response of the gamma memory is

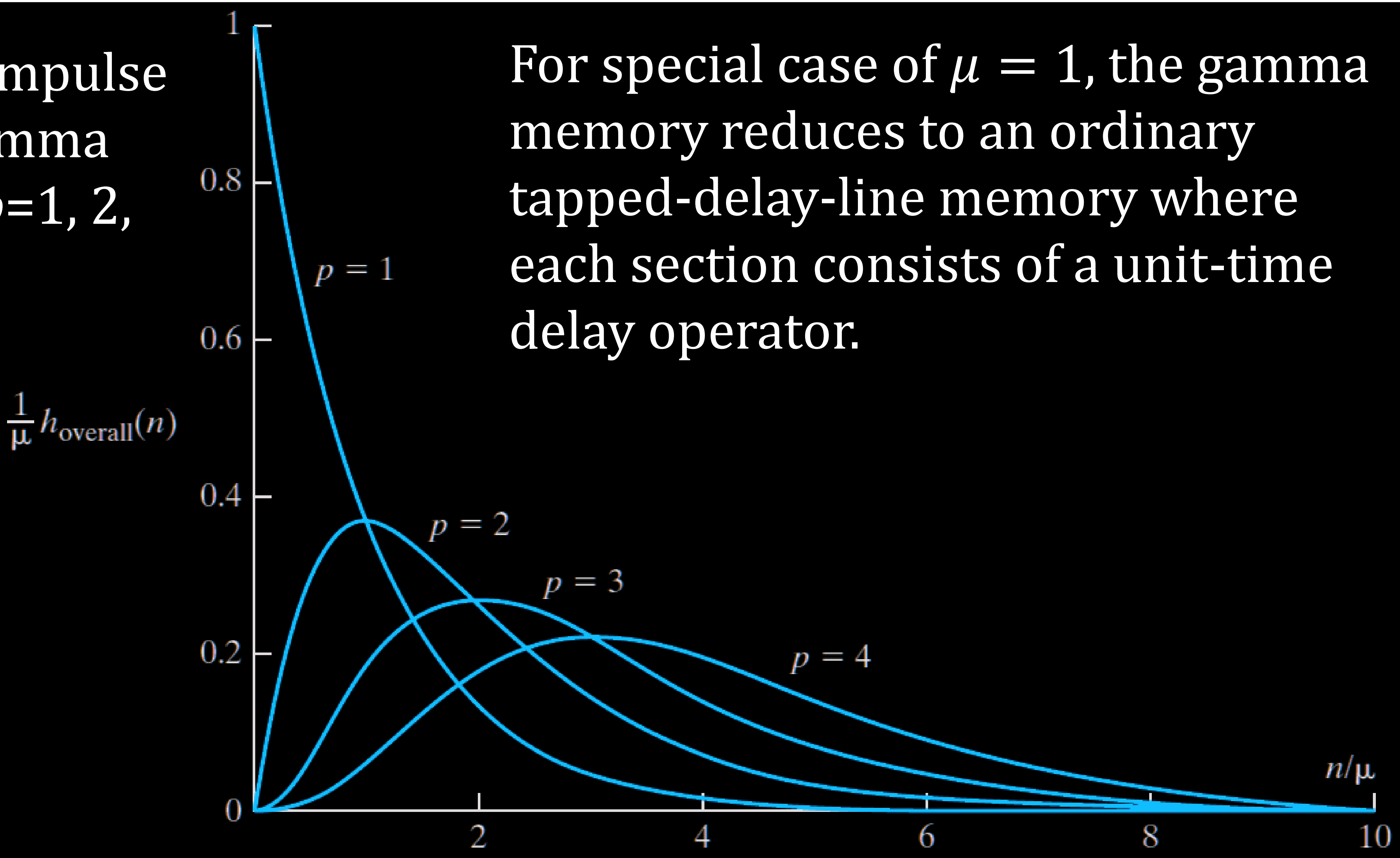
$$h_{\text{overall}}(n) = \binom{n-1}{p-1}\mu^{p}(1-\mu)^{n-p}, \qquad n \geq p \tag{4.157}$$

Where $\binom{\cdot}{\cdot}$ is a binomial coefficient.

- Figure 4.26 plots $h_{\text{overall}}(n)$, normalized with respect to μ, for varying memory order p=1, 2, 3, 4 and $\mu = 0.7$.

4.18 Nonlinear Filtering
Short-Term Memory Structures

Fig. 4.26 Family of impulse responses of the gamma memory for order *p*=1, 2, 3, 4 and $\mu = 0.7$



For special case of $\mu = 1$, the gamma memory reduces to an ordinary tapped-delay-line memory where each section consists of a unit-time delay operator.

4.18 Nonlinear Filtering
Universal Myopic Mapping Theorem

- The nonlinear filter of Fig. 4.24 may be generalized to that shown in Fig. 4.27
- The block labeled $\{h_j\}_{j=1}^{L}$ represents multiple convolutions in the time domain, that is, a bank of parallel linear filters.
- N represents a static (i.e., memoryless) nonlinear feedforward network such as the multilayer perceptron.
- For any single-variable, shift-invariant $[x(n-n_0) \rightarrow y(n-n_0)]$, causal $[x(n)=0, \forall n<0]$, uniformly fading memory map, there is a gamma memory and static neural network, the combination of which approximates the map uniformly and arbitrarily well.
- The *universal myopic mapping theorem* can be stated as

Any shift-invariant myopic dynamic map can be uniformly approximated arbitrarily well by a structure consisting of two functional blocks: a bank of linear filters feeding a static neural network.

4.18 Nonlinear Filtering
Universal Myopic Mapping Theorem

The h_j are (set of real-valued kernels) representing the impulse response of linear filter.

$\{h_j\}_{j=1}^{L}$ represents a bank of parallel linear filters.

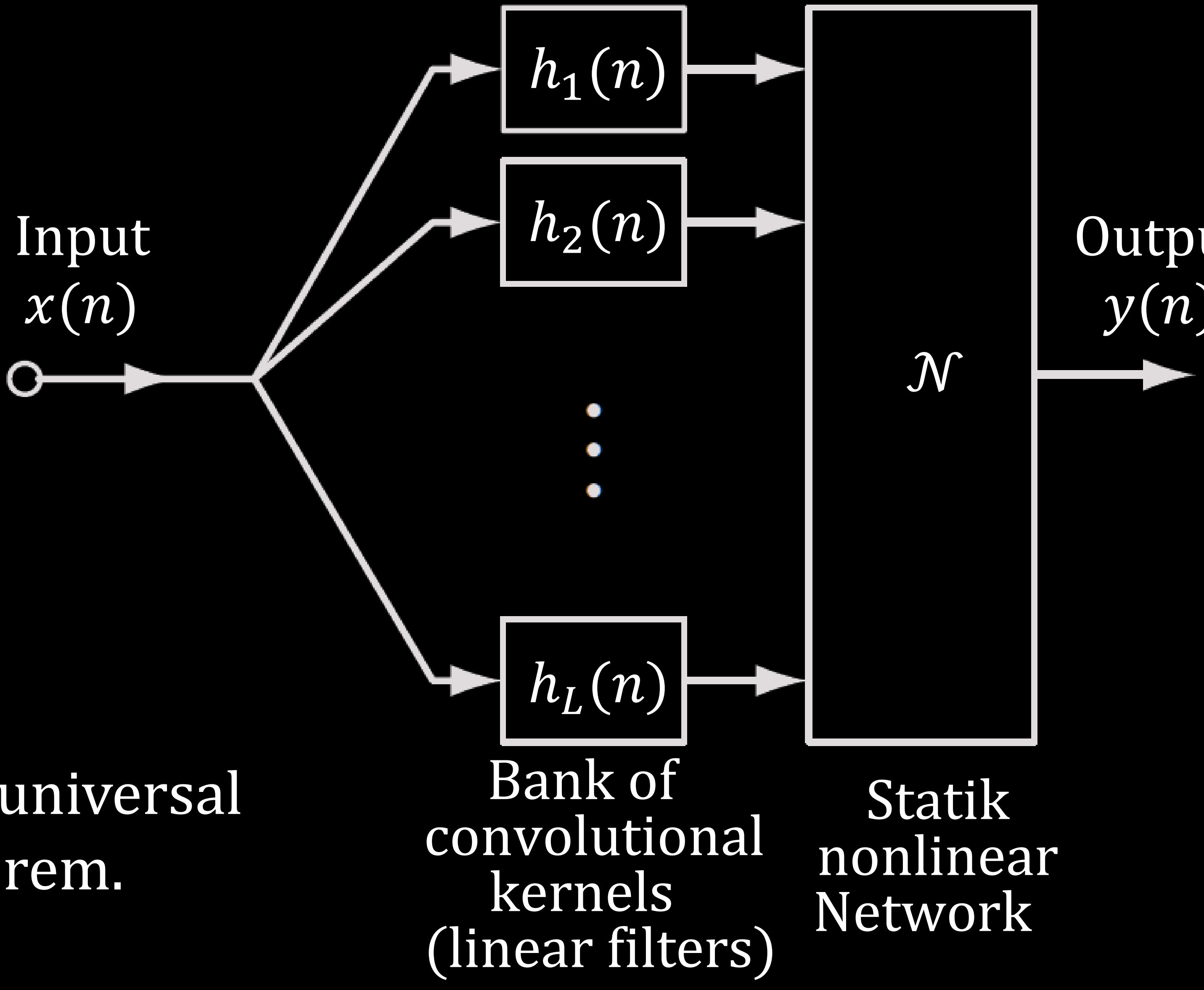


Fig. 4.27 Generic structure for universal myopic mapping theorem.

4.18 Nonlinear Filtering

Practical Implications of the Theorem

- The *universal myopic mapping theorem* has profound practical implications:

1. The theorem provides justification for NETtalk, which is a massively parallel distributed multilayer perceptrons network that converts English speech to basic linguistic unit (phonemes) as shown in Fig. 4.28.

   the input layer consists of 203 sensory (source) nodes,

   the hidden layer of 80 neurons, and

   the output layer of 26 neurons.

   a sigmoid (logistic) activation functions.

   the synaptic connections in the network were specified by a total of 18,629 weights, including a variable threshold for each neuron; threshold is the negative of bias.

   The standard back-propagation algorithm was used to train the network.

4.18 Nonlinear Filtering
Practical Implications of the Theorem

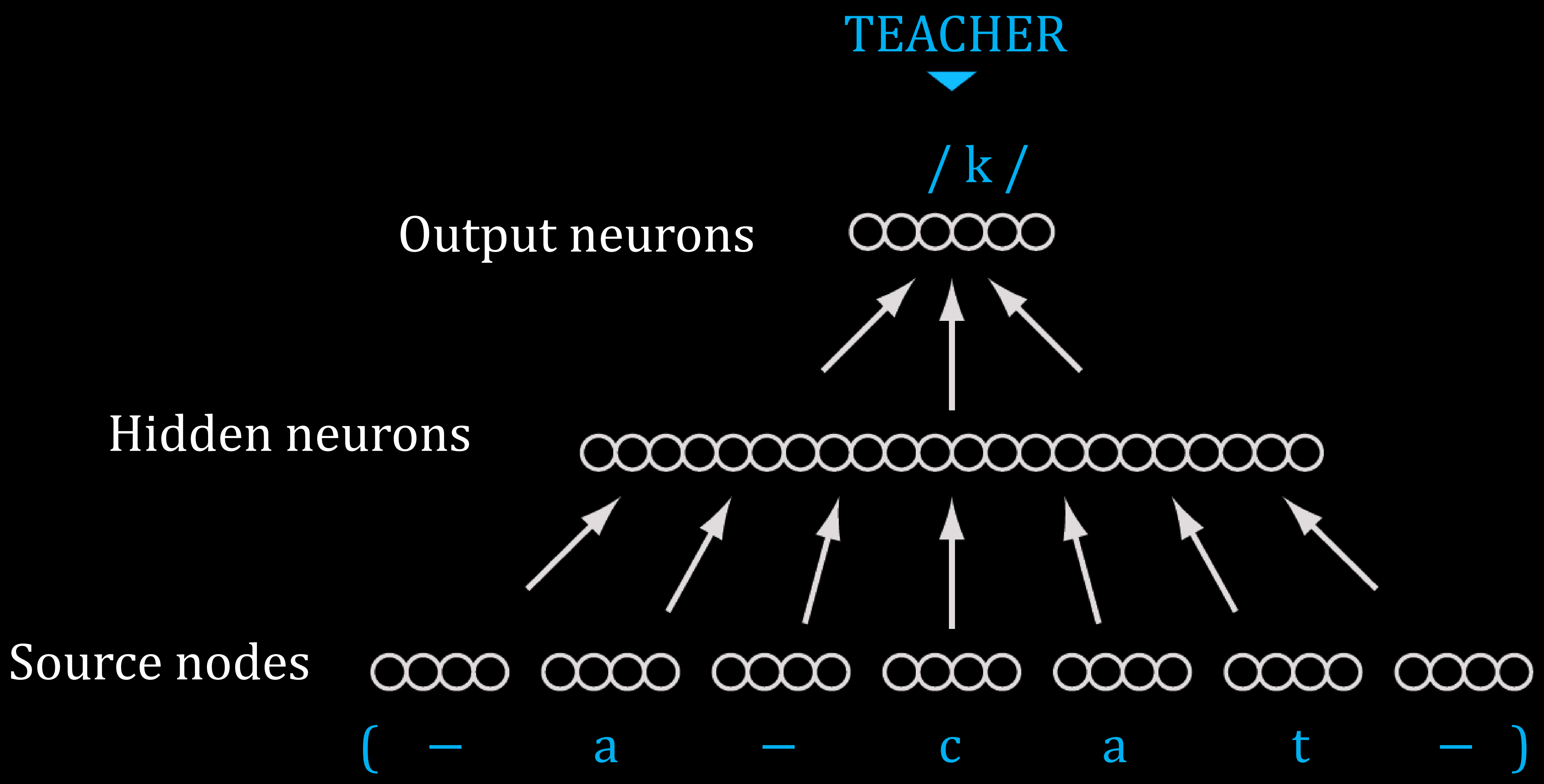


Fig. 4.28 Schematic diagram of the NETtalk network architecture.

4.18 Nonlinear Filtering
Practical Implications of the Theorem

The network had seven groups of nodes in the input layer, for encoding letters of the input text.

Strings of seven letters were thus presented to input layer at any one time.

The desired response for the training process was specified as the correct phoneme associated with the center (i.e., fourth) letter in the seven-letter window.

The other six letters (three on either side of the center letter) provided a partial context for each decision made by the network.

The text was stepped through the window on a letter-by-letter basis.

At each step in the process, the network computed a phoneme, and after each word the synaptic weights of the network were adjusted according to how closely the computed pronunciation matched the correct one.

4.18 Nonlinear Filtering
Practical Implications of the Theorem

2. The universal myopic theorem lays down the framework for the design of more elaborate models of nonlinear systems.

   The multiple convolutions at the front end of the structure in Fig. 4.27 may be implemented using linear filters with a finite duration impulse response (FIR) or infinite-duration impulse response (IIR).

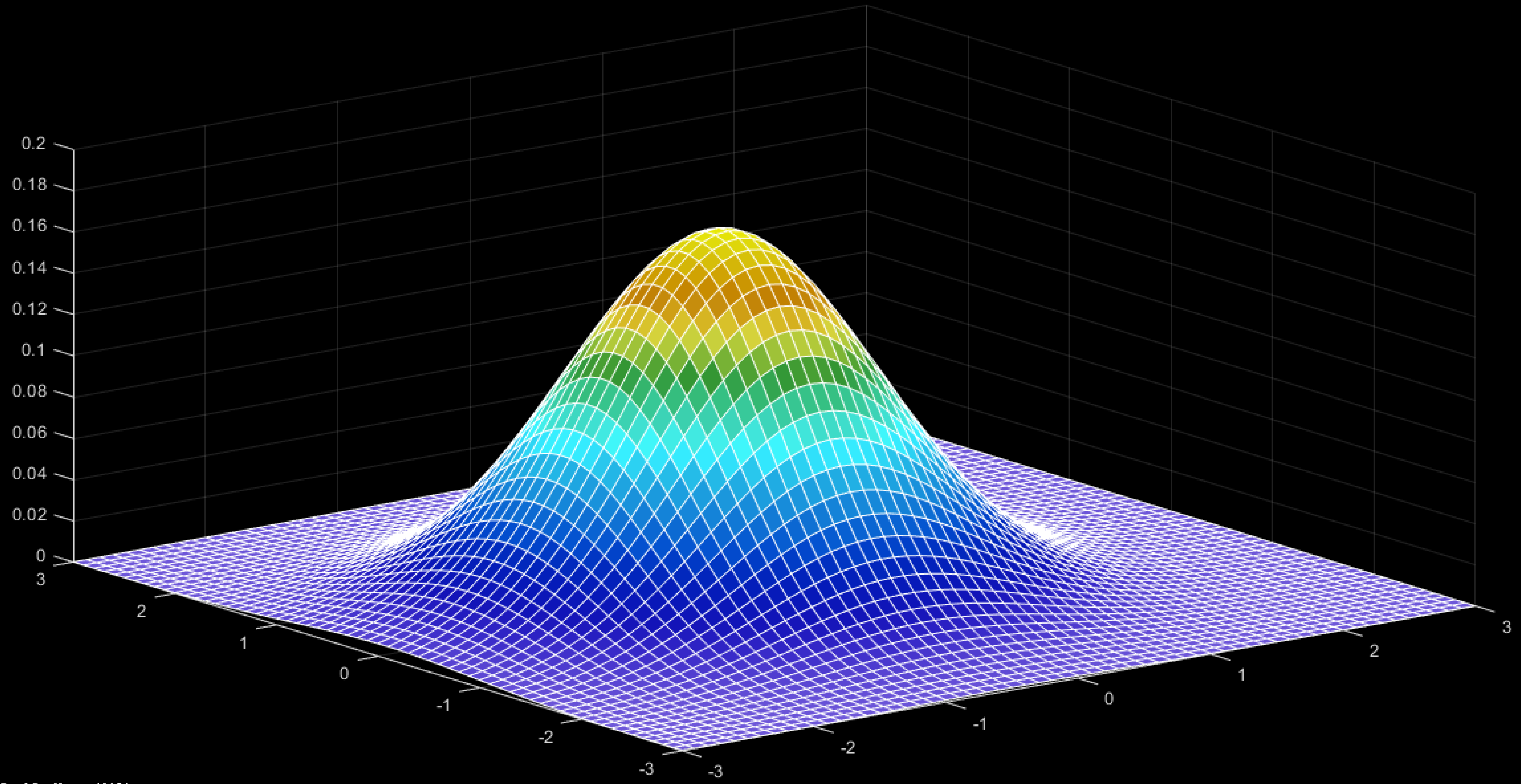
   These filters could take care of of short-term memory and memoryless nonlinearity in building a stable dynamic system.

3. Given a stationary time series $x(1), x(2), \dots, x(n)$, we may use *universal myopic mapping structure* of Fig. 4.27 to build a predictive model of the underlying nonlinear physical laws responsible for generating the time series by setting $y(n) = x(n+1)$, no matter how complex the laws are.

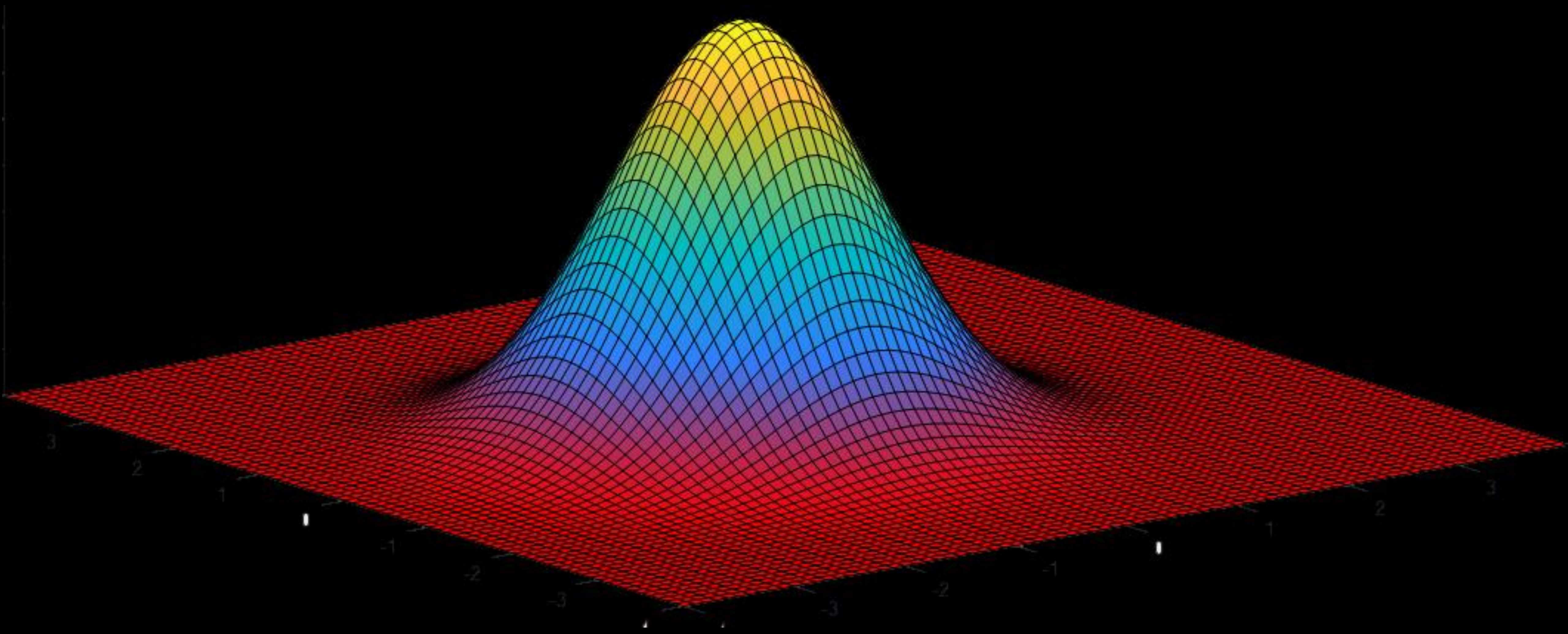
   When a multilayer perceptron is used as the static network in Fig. 4.27 for such an application, it is advisable to provide a linear neuron for the output unit in the network.

End of Section 4.19

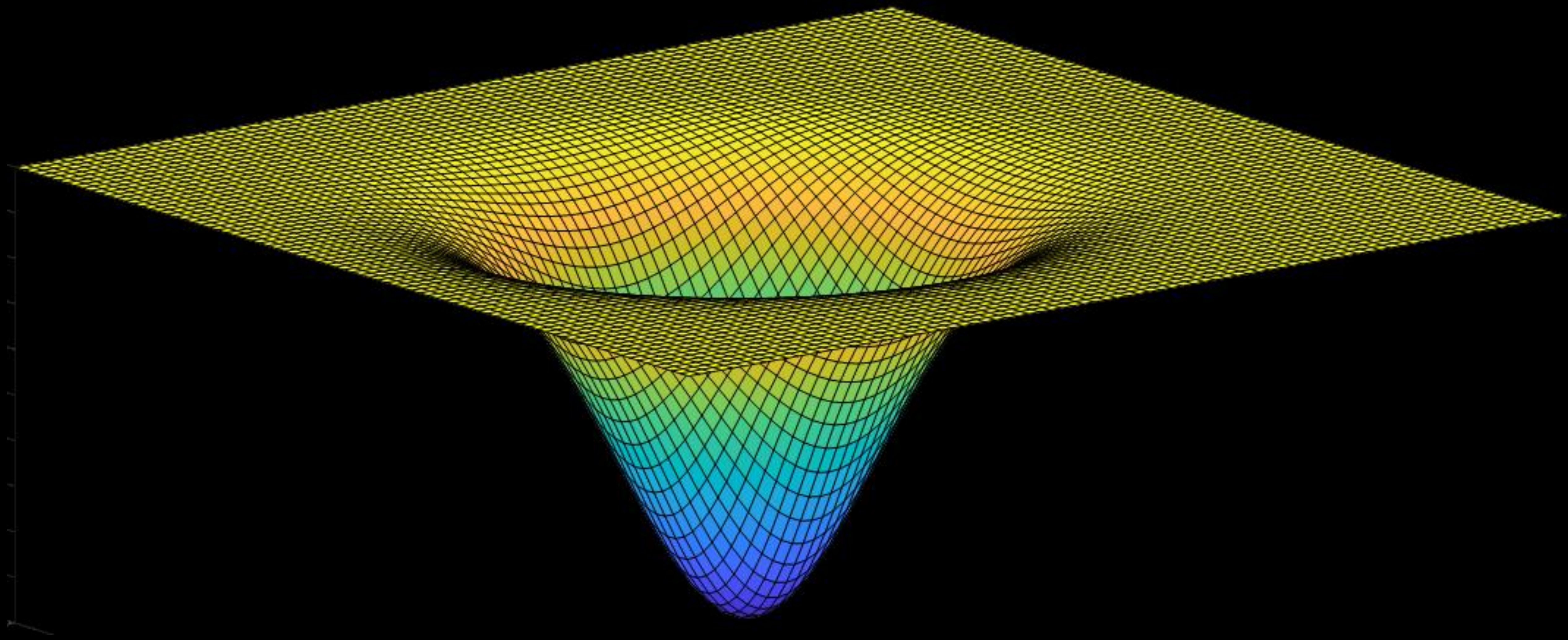
Multivariate Gaussian PDF Plot



Multivariate Gaussian PDF Plot



Multivariate Gaussian PDF Plot



Tips and Tricks:
Getting Started Using Optimization with MATLAB

https://www.mathworks.com/videos/tips-and-tricks-getting-started-using-optimization-with-matlab-81594.html?s_iid=doc_rw_GD_footer