USC–SIPI REPORT #334

Adaptive Stochastic Resonance

by

Sanya Mitaim

August 1999

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 400
Los Angeles, CA 90089-2564 U.S.A.

To my parents Sawat and Cattleya Mitaim

To my wife Worralux and my daughters Napat and Samon

Their undying love and confidence in me were a true

inspiration and a constant source of strength throughout my research endeavor.

Acknowledgements

I would like to thank my advisor Dr. Bart Kosko for his help and insights. He provided invaluable guidance and honest criticism throughout the years that we worked together. It has been an honor to work with a scholar of such consistent creativity and enthusiasm.

I wish to thank Dr. Robert E. Kalaba and Dr. Armand R. Tanguay, Jr. for sharing their ideas and passions in stochastic resonance. Thanks to Dr. Antonio Ortega and Dr. Michael G. Safonov for serving on my committee and for their technical insights. I also wish to thank Dr. George P. Papavassilopoulos and Dr. Rod Taber for their advice and encouragement.

Many other people deserve my gratitude for helping me to reach this point. My friends and colleagues Dr. Junavit Chalidabhongse, Sukumal Imudom, Margarida Karahalios, Ian Y. Lee, and Sean T. Morrow, often gave sound technical advice and emotional support. Dr. Julie A. Dickerson and Dr. Hyun Mun Kim always gave invaluable advice especially during the early years of my Ph.D. program.

Thanks to Wuttipong Kumwilaisak, Krisda Lengwehasatit, Poonsuk Lohsoonthorn, Dhawat Pansatiankul, Sunil Bharitkar, Raghavendra Singh, Dr. Moe Z. Win, and Dr. Youngjun Yoo for their company and conversations during the past years. I wish to thank Linda Varilla of the SIPI administrative office for her support and encouragement.

I also wish to thank teachers and friends at the USC Aikido Club, especially Robert Dziubla, Gary Wyshel, Jed Mortenson, Jeff Hruby, Mark Colopy, Ellen Valles, Michelle Van Noy, Michael Chapman, Ryan Fong, and Angela Trinh. Thank you for sharing your passions in Aikido, your friendships, and entertainment during and after each practice.

Thank you to my family here in Los Angeles. My mother-in-law Chintana Sookkao, my sisters-in-law Pakpisut and Prairin Sookkao, and family friends Saran Phuangchoey and Piyawan Tangseveepan who always provide invaluable support in so many ways.

I also wish to acknowledge and thank the Royal Thai Government for providing me with this opportunity to conduct Ph.D. research at USC.

Contents

List of Figures

Abstract

This dissertation introduces and explores the new property of adaptive stochastic resonance (ASR). Stochastic resonance (SR) occurs when noise enhances an external forcing signal in a nonlinear dynamical system. ASR uses statistical learning techniques that learn the optimal level of noise to add to a nonlinear system in the sense that this level of noise will maximize the system's signal-to-noise ratio or that it will improve or extremize other measures of how well the system performs. This dissertation studies how adaptive systems can achieve ASR based on only samples from the process or based on these samples and minimal estimates of the system dynamics.

The fundamental result of this research is that stochastic gradient learning can achieve ASR. A statistical learning system can learn the SR effect if it performs a stochastic gradient ascent on a system performance measure such as the system's spectral signal-to-noise ratio. But the gradient becomes so impulsive near optimality that it can destabilize the learning process. A Cauchy noise suppressor solves this problem and lets the stochastic-gradient learning laws train on noisy input-output samples to achieve stochastic resonance.

This research led to new stochastic learning laws for different types of systems and signals and for different types of performance measures. Stochastic gradient ascent on the signal-to-noise ratio led to ASR for narrowband signals. But broadband forcing signals required a correlation performance measure and often required some estimate of the Jacobian structure of the dynamical system.

We discovered stochastic resonance in nonlinear systems with impulsive noise that has infinite variance. An exponential law relates the SR effect or the optimal noise dispersion to the impulsiveness. We also showed that "smart" or black-box function approximators such as adaptive fuzzy systems can learn to induce the SR effect in many nonlinear systems. We developed new fuzzy learning laws for systems that take as input both numerical vectors and entire fuzzy

sets. The appendices present these fuzzy results and apply them to the multimedia problem of teaching an intelligent agent to learn a user's preferences and to search databases on the user's behalf.

This research revealed many problems with ASR. The adaptive system required a large number of input-output samples or it required at least some knowledge of the system dynamics and signals. We found no theorems to guarantee that the stochastic learning algorithms converge. This reflects a fundamental problem of research in SR and ASR. Even simple system nonlinearity can complicate or preclude a closed-form analysis and do so even if we have exact knowledge of the nonlinear signal systems. Future research may lead to new learning laws or to new ways to approximate the dynamics of nonlinear systems that stochastically resonate. This will help answer the key question that underlies ASR: Which noisy dynamical systems show what SR effects for which forcing signals and for which performance measures?

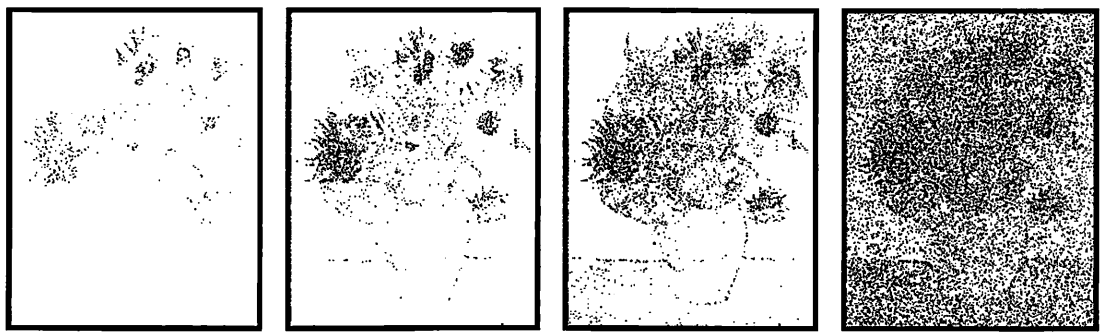
Chapter 1

Introduction

1.1 Stochastic Resonance

Noise can sometimes enhance a signal as well as hurt it. The fact that "noise helps" may seem at odds with almost a century of effort in signal processing to filter noise or to mask or cancel it. But noise is itself a signal and a free source of energy. Noise can amplify a faint signal in some feedback nonlinear systems even though too much noise can swamp the signal. This implies that a system's optimal noise level need not be zero noise. It also suggests that nonlinear signal systems with nonzero-noise optima may be the rule rather than the exception.

Figure 1.1 shows how uniform pixel noise can improve our subjective perception of an image. Here the system quantizes the original gray-scale "12 Sunflowers" image into a binary image of black and white pixels. It gives an output as a white pixel if the input gray-scale pixel equals or exceeds a threshold and a black pixel if the input gray-scale pixel is below the threshold. This quantizer system is biased since it does not set the threshold at the midpoint of the gray scale. So the quantized version of the original image contains almost no information. A small level of noise sharpens the image contours and helps fill in features when it adds to the original image



(a) (b) (c) (d)

Figure 1.1: Uniform pixel noise can improve the subjective response of our nonlinear perceptual system. The noise gives a nonmonotonic response. A small level of noise sharpens the image features while too much noise degrades them. These noisy images result when we apply a pixel threshold to Van Gogh's popular "12 Sunflowers" painting: $y = g((x+n)-\Theta)$ where $g(x) = 1$ if $x \geq 0$ and $g(x) = 0$ if $x < 0$ for an input pixel value $x \in [0, 1]$ and output pixel value $y \in \{0, 1\}$. The input image's gray-scale pixels vary from 0 (black) to 1 (white). The threshold is $\Theta = 0.1$. We threshold the original "12 Sunflowers" image to give the faint image in (a). The uniform noise n has zero mean $m_n = 0$. The noise variance σ_n^2 grows from (b)-(d): $\sigma_n^2 = 2.67 \times 10^{-3}$ in (b), $\sigma_n^2 = 6.51 \times 10^{-2}$ in (c), and $\sigma_n^2 = 1.67 \times 10^{-1}$ in (d).

before the system applies the threshold. Too much noise swamps the image and degrades its contours.

Consider an example where noise can enhance an output of a dynamical system. Figure 1.2 shows the system diagram $\dot{x} = f(x) + s + n$ with input forcing signal s and noise n. The system output $y(t)$ depends on the system state $x(t)$ through $y = g(x)$. Then we take the discrete Fourier transform of the output samples y_t to compute the spectral signal-to-noise ratio (SNR) with (3.15) in Chapter 3. Figure 1.3 shows an example where noise can enhance the output of the popular quartic bistable dynamical system $\dot{x} = x - x^3 + s + n$ [13, 27, 217] discussed in Chapter 2. A weak sinusoidal input signal s and additive white Gaussian noise n force the system. The system output $y(t) = x(t)$ barely respond to the sinusoidal input signal s when the noise n is absent (when $\sigma = 0$). A small amount of noise ($\sigma = 0.5$) improves the system output so that it resembles the sinusoidal input signal while too much noise ($\sigma = 1$) destroys the periodic pattern of the system output. Other signal systems might use a cross-correlation measure as in Figure 2.3.

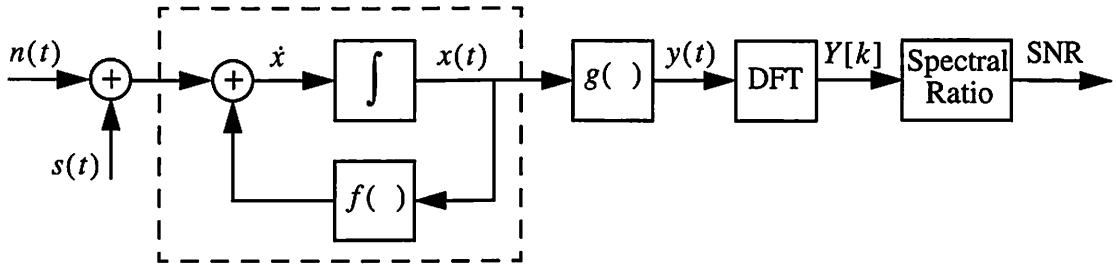


Figure 1.2: System diagram for stochastic resonance within a signal-to-noise framework. A dynamical system $\dot{x} = f(x) + s + n$ has input forcing signal s and noise n. The system output y depends on the system state x through $y = g(x)$ and $Y[k]$ is the discrete Fourier transform (DFT) of the time signal $y(t)$. Wavelets or other transforms can replace the DFT block. The figure shows how the input noise process $n(t)$ drives the spectral signal-to-noise ratio (SNR) of the system. This suggests that stochastic gradient ascent on the SNR can learn the optimal noise level that achieves stochastic resonance.

Stochastic resonance (SR) [11, 12, 13, 22, 24, 27, 77, 93, 102, 148, 188, 189, 219, 223, 224, 234, 313] occurs when noise enhances an external forcing signal in a nonlinear dynamical system. SR occurs in a signal system if and only if the system has a nonzero noise optimum [212]. The classic SR signature is a signal-to-noise ratio (SNR) of the system output that is not monotone. Figure 1.4 shows the SR effect for the quartic bistable dynamical system. The sinusoidal input signal has the form $s(t) = \varepsilon \sin \omega_0 t$ and the additive white Gaussian noise n has zero mean and variance σ^2. Equation (3.15) in Chapter 3 defines the SNR of the system output y at the frequency ω_0. The SNR rises to a maximum and then falls as the variance of the additive white noise grows. Note that the SNR is not infinite for zero input noise because it measures the spectral energy ratio of the system output y. The SNR becomes large when the output y resembles the sinusoidal input signal s so that most of its spectral energy engery lies in the frequency bin ω_0. More complex systems may have multimodal SNRs and so show stochastic "multiresonance" [92, 308].

Stochastic resonance holds promise for the design of engineering systems in a wide range of applications. Engineers may want to shape the noise background of a fixed signal pattern to exploit the SR effect. Or they may want to adapt their signals to exploit a fixed noise background. Engineers now add noise to some systems to improve how humans perceive signals.

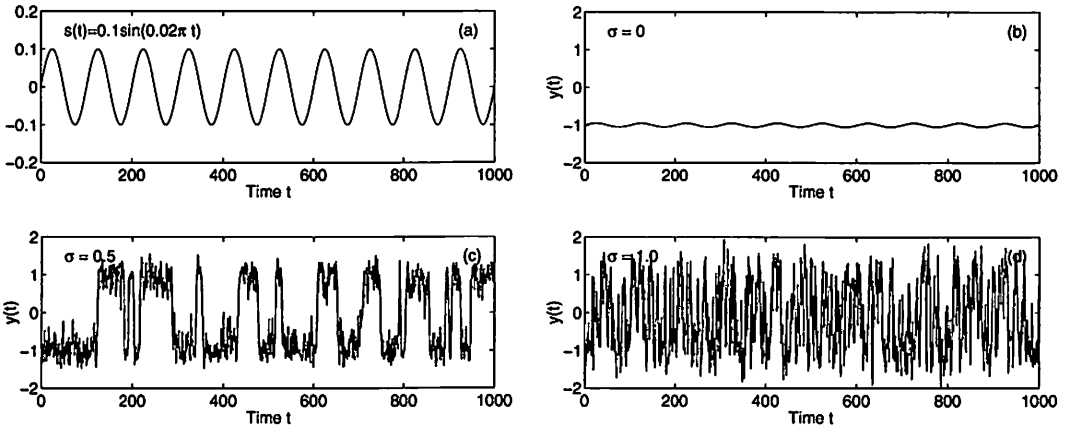


Figure 1.3: Gaussian noise can improve the output signal-to-noise ratio of a quartic bistable dynamical system $\dot{x} = x - x^3 + s + n$. The plot (a) shows the sinusoidal input signal $s(t) = 0.1\sin(2\pi(0.01)t)$. The zero-mean Gaussian noise n has variance σ^2. The graphs (b)-(d) show the system output $y(t) = x(t)$ for different noise levels: $\sigma = 0$, 0.5, and 1. The system barely responds to the input signal s when there is no noise ($\sigma = 0$) in (b). The plot (c) shows that the right amount of noise ($\sigma = 0.5$) maximally enhances the performance of the system so that its output resembles the sinusoidal signal. Too much noise ($\sigma = 1$) destroys the periodic pattern in the output as in (d).

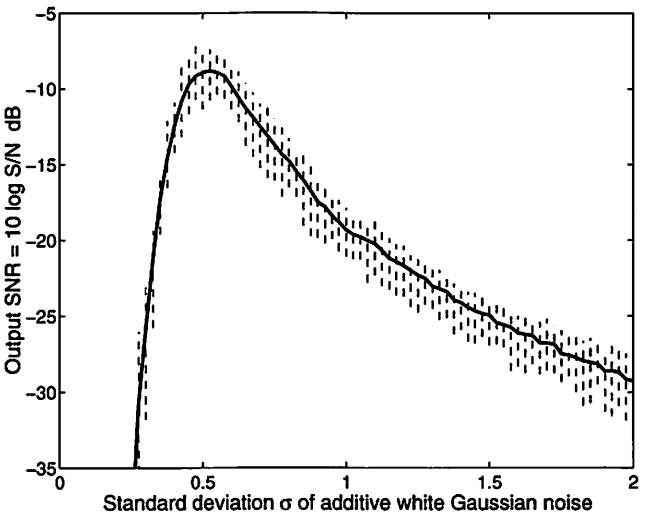


Figure 1.4: The non-monotonic signature of stochastic resonance. The graph shows the smoothed output signal-to-noise ratio of a quartic bistable system as a function of the standard deviation of additive white Gaussian noise. The vertical dashed lines show the absolute deviation between the smallest and largest outliers in each sample average of 20 outcomes. The system has a nonzero noise optimum and thus shows the SR effect. The noisy signal-forced quartic bistable dynamical system has the form $\dot{x} = f(x) + s(t) + n(t) = x - x^3 + \varepsilon\sin\omega_0 t + n(t)$ with binary output $y(t) = \text{sgn}(x(t))$. The Gaussian noise $n(t)$ adds to the external forcing narrowband signal $s(t) = \varepsilon\sin\omega_0 t$. Equation (3.15) in Chapter 3 defines the SNR measure. Other systems can use multiplicative noise [9, 28, 76, 85, 90, 96] or use non-Gaussian noise [41, 43, 44, 92, 251].

These systems include audio compact discs [177], analog-to-digital devices [10], video images [285], schemes for visual perception [268, 269, 291], and cochlear implants [79, 216, 220].

Some control and quantization schemes add a noise-like dither to improve system performance [10, 174, 177, 239, 285]. Additive noise can sometimes stabilize chaotic attractors [16, 89, 201]. Noise can also improve human tactile response [54, 271], muscle contraction [48], and coordination [57]. This suggests that SR designs may improve how robots grasp objects [59] or balance themselves. SR designs might also improve how virtual or augmented reality systems [33, 125] can create or enhance the sensations of touch and balance.

Stochastic resonance designs might lead to better schemes to filter or multiplex the faint signals found in spread spectrum communication systems [82, 290]. These systems transmit and detect faint signals in noisy backgrounds across wide bands of frequencies. SR designs might also exploit the signal-based crosstalk noise found in cellular systems [169, 292], Ethernet packet flows [170], or Internet congestion [132].

The study of SR has emerged largely from physics and biology. The awkward term "stochastic resonance" stems from a 1981 article in which physicists observed "the cooperative effect between internal mechanism and the external periodic forcing" in some nonlinear dynamical systems [13]. (A more apt term might be "noise resonance.") Scientists soon explored SR in climate models [236] to explain how noise could induce periodic ice ages [11, 12, 234, 235]. They conjectured that global or other noise sources could amplify small periodic variations in the Earth's orbit. This might explain the observed 100,000 year primary cycle of the Earth's ice ages. This SR conjecture remains the subject of debate [84, 235, 315]. Physicists have since found stronger evidence of SR in ring lasers [203, 304], threshold hysteretic Schmitt triggers [80, 204], Chua's electrical circuit [4, 5], bistable magnetic systems [111], electron paramagnetic resonance [94, 97, 270], magnetoelastic ribbons [294], superconducting quantum interference devices (SQUIDs) [119, 136, 275], Ising systems [20, 227, 289], coupled diode resonators [178], tunnel diodes [198, 199], Josephson junctions [23, 120], optical systems

[9, 70, 140], chemical systems [71, 83, 114, 124, 149, 172, 218], and quantum-mechanical systems [107, 108, 109, 110, 180, 195, 250, 266, 302].

Some biological systems may have evolved to exploit the SR effect. Most SR studies have searched for the SR effect in the sensory processing of prey and predators. Noisy or turbulent water can help the mechanoreceptor hair cells of the crayfish *Procambarus clarkii* detect faint periodic signals of predators such as a bass's fin motion [67, 68, 224, 247, 254, 258, 313]. Noise helps the mechanosensors of the cricket *Acheta domestica* detect small-amplitude low-frequency air signals from predators [173, 206, 207]. Dogfish sharks use noise in their mouth sensors when they detect periodic signals from prey [17]. The SR effect appears in the mechanoreceptors in a rat's skin [53] and in the neurons in a rat's hippocampus [104]. The SR effect also occurs in a wide range of models of neurons [26, 28, 50, 51, 52, 118, 253, 297] and neural networks [25, 26, 28, 30, 31, 47, 50, 51, 52, 133, 134, 176, 181, 182, 183, 184, 185, 186, 221, 228, 251].

1.2 Current State of Research on Stochastic Resonance

Research in SR has grown from the study of external periodic signals in simple dynamical systems to the study of external aperiodic and broadband signals in more complex dynamical systems [38, 41, 47, 50, 51, 52, 53, 118, 127, 173, 255, 297]. The analyses in current SR research literature focus on systems with small forcing signals and Gaussian noise or other finite-variance noise. These assumptions give rise to simple closed-form approximations of the SR effect. But they do not represent all systems in real-world situations. The breadth of SR systems discussed in the previous section suggests that the SR effect may occur in still more complex dynamical systems for still more complex signals and noise types. These signal systems may prove too complex to obtain closed-form math models. And the math models may still prove too complex to approximate the SR effect with simple closed-form techniques. This suggests in turn that we might use "intelligent" or adaptive model-free techniques to learn or approximate the SR effects.

Most SR research to date deals with Gaussian noise. A few models work with other finite-variance noise such as uniform noise, Laplace noise, or (bounded) chaotic noise [41, 43, 44, 92, 212, 251]. This dissertation explores adaptive stochastic resonance with noise from these densities. It also shows for the first time the SR effect that arises from infinite-variance noise such as alpha-stable noise. Here a *dispersion* measure replaces the more common variance measure.

1.3 Adaptive Stochastic Resonance: Dissertation Objective and Results

Studies of "stochastic resonance" (SR) show that sometimes noise can help a system as well as hurt it. Noise can enhance signals that force a nonlinear system. This raises many new research questions: What is the best level of noise for a given feedback or feedforward system? What is the best type of noise? Which signals best "resonate" with which noise types?

This dissertation explores these questions with what we call *adaptive* stochastic resonance (ASR). ASR uses statistical learning techniques to learn the optimal level of noise to add to a nonlinear system to maximize its signal-to-noise ratio or to improve other measures of how well the system performs. ASR can also tune the forcing signals or tune the nonlinear system itself to improve its performance.

This dissertation studies ASR for two broad categories of signal systems. The first category consists of systems with narrowband or sinusoidal forcing signals. The second broad category consists of systems with any periodic signals and broadband or non-periodic signals. ASR techniques are simpler in the narrowband case. The learning schemes tend to involve less computation and yield more accurate results in the narrowband case than they do in the broadband case.

The fundamental result of this research is that stochastic gradient learning can achieve ASR. A statistical learning system can learn the SR effect (can converge to or near the optimal level of additive noise) if it performs a stochastic gradient ascent on a system performance measure such

as the system's spectral signal-to-noise ratio. But statistical tests confirm that simple gradient-ascent learning laws fail to converge to the optimal noise values. The learning process becomes so impulsive (it approximates a thick-tailed Cauchy probability density) near optimality that it destabilizes the learning process. So the noise optimum acts more as a dynamical repellor than as an attractor. We overcame this problem by including a Cauchy noise suppressor from the theory of robust statistics. This Cauchy-suppressor structure allows many types of stochastic-gradient learning laws to "blindly" train on noisy input-output samples and achieve the SR effect for a wide range of nonlinear dynamical systems, performance measures, forcing signals, and noise types.

This research led to many other results. Different types of systems and signals required different types of performance measures and thus different stochastic learning laws (but all need Cauchy impulse suppression). A system's spectral signal-to-noise ratio worked well for narrowband signals but not for the more general case of broadband forcing signals. Broadband forcing signals required a correlation performance measure and thus a correlation-based learning law. Broadband models often required some estimate of the Jacobian structure of the forced dynamical system.

This research also discovered the first instance of stochastic resonance in nonlinear systems where the forcing noise is so impulsive that it has infinite variance. All published SR models to date assume that the forcing additive noise has finite variance. Almost all SR models have assumed that the noise is Gaussian. Indeed the standard graph of the SR effect plots the signal-to-noise ratio on one axis against this finite-variance term on the other axis. This formulation precludes even the definition of the SR effect in the general case when the noise process has infinite variance. The implicit assumption in such models is that variance equals dispersion. But a mathematical variance is just one of many ways to measure the dispersion in a random variable. We showed that if we parametrize infinite-variance noise densities with the α of alpha-stable probability densities then not only does the SR effect still occur (when we plot a performance

measure against a bell-curve's dispersion parameter γ) but in many cases it obeys an exponential law: $\gamma_{opt}(\alpha) = BA^{\alpha}$. The SR effect tends to lessen as the impulsiveness increases (as α falls).

The research further showed that "smart" or black-box function approximators can learn to induce the SR effect in many nonlinear systems. We showed this for adaptive fuzzy systems that slowly learn from input-output samples of the process how to tune their if-then rules to achieve the SR effect. We developed along the way new fuzzy learning laws for systems that take as input both numerical vectors and entire fuzzy sets. Appendices A and B present this background research on how a large class of feedforward fuzzy systems can learn from sample data and "blindly" approximate functions and how we can apply these new tools to the multimedia problem of teaching an intelligent agent to learn a user's preferences and search databases on the user's behalf.

This research also revealed many problems with the ASR results. Even the best applications involved heavy computation and required a large number of input-output samples that real systems may not provide. The only evidence that the learning algorithms converged came from extensive computer simulations of all known SR models. The research did not produce a formal theorem that proved convergence. This reflects the fundamental problem of research in SR and ASR. We seldom have exact knowledge of the sampled nonlinear signal systems or of the forcing broadband signals. Even the few known SR systems are sufficiently nonlinear to greatly complicate a closed-form analysis. A key challenge is to estimate at least the Jacobian structure of the sampled systems with unknown dynamics. Such research faces a tradeoff. ASR learning schemes will tend to improve as we know more of the dynamical structure of the system or of the spectral structure of its forcing broadband signals. But the more such knowledge we require the more we restrict the application of ASR techniques. Future research may lessen this tradeoff if it leads to new SR performance measures or to new learning laws or to new ways to approximate the dynamics of nonlinear systems that stochastically resonate.

1.4 Dissertation Outline

This dissertation consists of six Chapters. Chapter 1 introduces the notion of stochastic resonance (SR) and states the problems in adaptive stochastic resonance. Chapter 2 briefly reviews nonlinear systems with external forcing signal and noise terms and defines the performance measures for different kinds of input signals and systems. Chapter 3 describes a discrete algorithm that simulates nonlinear dynamical systems with additive forcing signals and noise and defines the performance measures from sample data. It also tests many SR systems with alpha-stable noise. Chapter 4 derives the learning laws that stochastically determine the optimal noise in a nonlinear system and presents the adaptive fuzzy function approximator as a universal approximator. Chapter 5 shows simulation results of adaptive stochastic resonance for many dynamical systems and for many noise types Chapter 6 concludes this dissertation and discusses future research in adaptive stochastic resonance. Appendix A discusses adaptive fuzzy systems as universal function approximators that learn from sample data. Appendix B shows how we apply this universal approximator to learn a user's preferences and search a database on the user's behalf.

Chapter 2

Stochastic Resonance and Optimal Noise

Not all dynamical systems show the SR effect. And the SR effect does not occur for all input forcing signals. This chapter reviews several nonlinear dynamical systems and their forcing input signals and noise that show SR. The sinusoidal or broadband signals and additive noise force the dynamical systems. The output of a dynamical system depends on the system states. The system performance depends on the input signal and output measurements. Below we list many signal systems in the literature that exhibit the SR effect. Their optimal noise inputs are nonzero. Then we review performance measures that determine the SR effect. Then we explore the concept of optimal noise.

2.1 Stochastic Dynamical Systems

Nonlinearity and feedback can lead to stochastic resonance. This section reviews the main known dynamical systems that show the SR effect. These models involve only simple nonlinearities. And they simply add a random noise term to a differential equation rather than use a formal

Ito stochastic differential [49, 69, 99]. There are so far no theorems or formal taxonomies that tell which dynamical systems show SR and which do not.

Consider a dynamical system that relates its input-output response through a differential equation of the form

$$\dot{x} = f(x) + u(x,t) \tag{2.1}$$

$$y(t) = g(x(t)) \tag{2.2}$$

Here a smooth nonlinear function $f : R^n \rightarrow R^n$ governs the system's dynamics. The input u may depend on both time t and on the system's state x. The system is unforced or autonomous when $u(x,t) = 0$ for all x and t. The system output or measurement $y \in R^p$ depends on the state x through $y = g(x)$ for a map $g : R^n \rightarrow R^p$. The output of a simple scalar model neuron may be a signum function: $y = \text{sgn}(x)$.

The input function u can be a deterministic function or a random process or a combination. For example u can have the form

$$u(x,t) = s(t) + n(t) \tag{2.3}$$

when u does not depend on x and the input forcing signal s and noise n are additive. This gives (2.1)-(2.2) a form of

$$\dot{x} = f(x) + s(t) + n(t) \tag{2.4}$$

$$y(t) = g(x(t)). \tag{2.5}$$

This dissertation works with dynamical systems of the form (2.4)-(2.5). Below we list some popular systems in the SR literature.

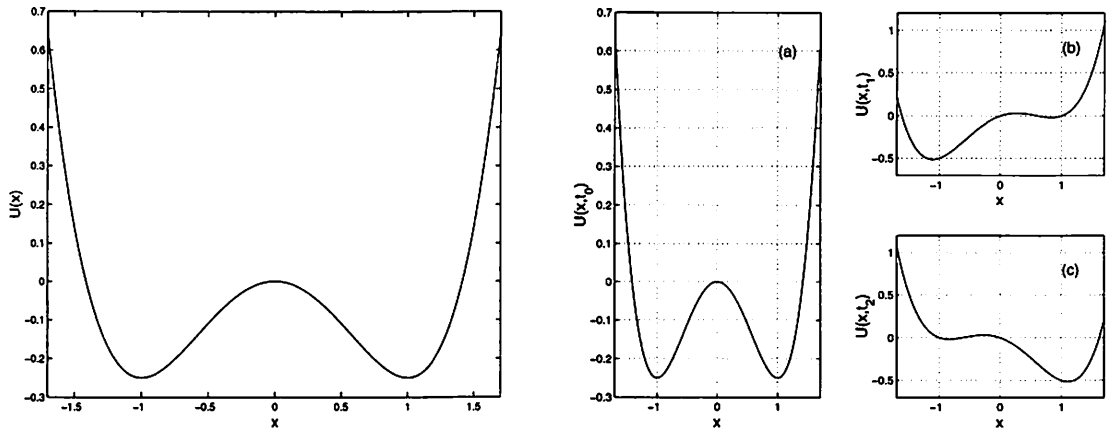


Figure 2.1: (I) Unforced quartic potential: $U(x,t) = -\frac{1}{2}x^2 + \frac{1}{4}x^4$. (II) Forced evolution of the noise-free quartic potential system: $U(x,t) = -\frac{1}{2}x^2 + \frac{1}{4}x^4 + \frac{1}{4}x\sin 2\pi t$. (a) Unforced potential surface at $t = 0$ when the sinusoidal forcing term is zero. (b) Surface $U(x,t)$ at time $t = \frac{1}{4}$. (c) Surface $U(x,t)$ at time $t = \frac{3}{4}$.

• **Quartic Bistable System** [13, 63, 86, 95, 128, 141, 217, 321]. The quartic bistable system is the most studied model that shows the SR effect. It has the form

$$\dot{x} = -\frac{\partial}{\partial x}U(x,t) + s(t) + n(t) \tag{2.6}$$

$$= ax - bx^3 + s(t) + n(t) \tag{2.7}$$

for a quartic potential $U(x,t) = -\frac{a}{2}x^2 + \frac{b}{4}x^4$ with $a > 0$, $b > 0$, input signal s, and white noise n with zero mean and variance σ^2: $E[n] = 0$ and $Var(n) = \sigma^2 < \infty$. The output of this system can have a linear form $y(t) = x(t)$ or a binary form $y(t) = \text{sgn}(x(t))$. Researchers sometimes include the forcing signal s and noise n in the potential function: $U(x,t) = -\frac{a}{2}x^2 + \frac{b}{4}x^4 + x(t)[s(t) + n(t)]$.

The unforced version of (2.7) has the form $\dot{x} = ax - bx^3$. It has two stable fixed points at $x = \pm c = \pm\sqrt{a/b}$ and one metastable fixed point at $x = 0$. These fixed points are the minima and the local maximum of the potential $U(x,t) = -\frac{a}{2}x^2 + \frac{b}{4}x^4$. Figure 2.1(I) shows the quartic potential for $a = b = 1$. The two minima are at $x = \pm 1$. The figure shows the potential at rest and hence with no input force. Figure 2.1(II) shows the potential $U(x,t)$ when the external sinusoidal input modulates it at each time instant t.

• **Threshold Systems** [41, 92, 101, 105, 142, 143, 246]. Threshold systems are among the simplest SR systems. They show the SR effect for many of the performance measures in the next section. A simple threshold system can take the form

$$y(t) \;=\; \text{sgn}(x(t)) \;=\; \begin{cases} -1 & \text{if } x(t) < \Theta \\ 1 & \text{if } x(t) \geq \Theta \end{cases} \tag{2.8}$$

for the signal $x(t) = s(t) + n(t)$ and a threshold $\Theta \in R$.

Thresholds quantize signals. So we state the general forms of uniform infinite quantizers with gain $G > 0$. A uniform mid-tread quantizer with step size Δ has the form

$$y(t) \;=\; Q(x(t)) \;=\; G\Delta \left\lfloor \frac{x(t)}{\Delta} + \frac{1}{2} \right\rfloor. \tag{2.9}$$

A mid-riser quantizer has the form

$$y(t) \;=\; Q(x(t)) \;=\; G\Delta \left\lfloor \frac{x(t)}{\Delta} \right\rfloor + \frac{G\Delta}{2}. \tag{2.10}$$

The floor operator $\lfloor \rfloor$ gives the greatest integer less than or equal to its argument. Researchers have studied the SR effect in M-level quantizers that approximate some dynamical systems [248].

• **Bistable Potential Neuron Model** [28]. This neuron model is a bistable system of the form

$$\dot{x} = -x + (\eta_0 + n_m(t)) \tanh x + n_a(t) + s(t). \tag{2.11}$$

The multiplicative and additive noises n_m and n_a are zero mean and uncorrelated. The term η_0 is a constant. The output of the system has a binary form $y(t) = \text{sgn}(x(t))$.

• **FitzHugh-Nagumo (FHN) Neuron Model** [38, 50, 51, 52, 118, 181, 182, 221, 253, 314]. The FHN neuron model is a two-dimensional limit cycle oscillator that has the form

$$\epsilon \dot{x} = x(x-a)(1-x) - w + A + s(t) + n(t) \tag{2.12}$$

$$\dot{z} = x - z - b \tag{2.13}$$

Here x is a fast (voltage) variable, z is a slow (recovery) variable, A is a constant (tonic) activation signal, s is an input signal, and n is noise. The system output can take the linear form $y(t) = x(t)$ or we can consider the average firing rate where $y(t)$ is a low-pass filtered version of x. Sample constants for the SR effect are $\epsilon = 0.005$, $a = 0.5$, $A = -5/12\sqrt{3} = -0.24056$, and $b = 0.15$ [52].

• **Integrate-Fire Neuron Model** [26, 32, 42, 44, 50, 85, 259, 297]. This neuron model has linear activation dynamics:

$$\dot{x} = \lambda(u_r - x) + \mu + s(t) + n(t) \tag{2.14}$$

where x is cell membrane voltage, μ is a positive drift, λ is a decay constant rate, and u_r is a resting level. A threshold function governs the neuron's output pulse firing and gives the nonlinear system that shows the SR effect.

• **Hodgkin-Huxley Neuron Model** [50, 183, 255]. The Hodgin-Huxley model is among the most studied models in the neural literature:

$$C\dot{x} = -g_{Na}m^3h(x - x_{Na}) - g_K p^4(x - x_K) - g_L(x - x_L) + I + s(t) + n(t) \tag{2.15}$$

$$\dot{m} = \alpha_m(x)(1-m) - \beta_m(x)m \tag{2.16}$$

$$\dot{h} = \alpha_h(x)(1-h) - \beta_h(x)h \tag{2.17}$$

$$\dot{p} = \alpha_p(x)(1-p) - \beta_p(x)p. \tag{2.18}$$

Here x is the membrane potential or activation and m is the sodium activation. The term h is the sodium inactivation, p is the potassium activation, C is the membrane capacitance, x_L is the leakage reversal potential, g_L is the leakage conductance, x_K is the potassium reversal potential, $\bar{g}_K$ is the maximal potassium conductance, ρ_K is the potassium ion-channel density, x_{Na} is the sodium reversal potential, $\bar{g}_{Na}$ is the maximal sodium conductance, ρ_{Na} is the sodium ion-channel density, I is an input current, and s is a subthreshold aperiodic input signal. The membrane potential x determines the output firing of the neuron. SR occurs when a low level of noise n brings the input signal above the neuron's firing threshold.

• **Monostable Systems** [72, 73, 75, 112, 298, 306]. These systems have no potential barriers as do bistable and multistable systems. They have only one stable fixed point. A special case is the single-well Duffing oscillator:

$$\ddot{x} + 2\Gamma\dot{x} + \omega_1^2 x + \gamma x^3 \;=\; s(t) + n(t) \;=\; \varepsilon \cos\omega_0 t + n(t) \tag{2.19}$$

where $\Gamma, |\delta\omega| \ll \omega_0$ and $\gamma\delta\omega > 0$ for $\delta\omega = \omega_0 - \omega_1$. These systems show the SR effect in the small signal limit with an approximate linear response.

• **Array and Coupled Systems** [25, 28, 30, 31, 47, 50, 51, 52, 106, 129, 133, 134, 144, 176, 181, 182, 183, 184, 185, 186, 215, 221, 227, 228, 229, 251, 263, 268, 269]. These systems combine many units of the above systems. They include neural networks and other coupled systems. A special case is the Cohen-Grossberg [158] feedback neural network:

$$C_i\dot{x}_i = -\frac{x_i}{R_i} + \sum_{j=1}^{N} m_{ij} \tanh x_j + s(t) + n(t) \qquad \text{for } i = 1, \ldots, N \tag{2.20}$$

for neural activation potential x_i, synaptic efficacy m_{ij}, and hyperbolic neural firing function $S_j(x_j) = \tanh x_j$. Simulations show that the SR profile grows more peaked as the number N of neurons grows [134]. One study [134] found that the SR effect goes away for $N \geq 10$.

• **Chaotic Systems** [3, 5, 13, 36, 60, 187, 237, 309, 310, 323]. Some chaotic systems show the SR effect. These models include Chua's electric circuit, the Henon map, the Lorenz system, and the following forced Duffing oscillator:

$$\ddot{x} \;=\; -\delta\dot{x} + x + x^3 + \varepsilon\sin(\omega_0 t) + n(t) \tag{2.21}$$

with linear output $y(t) = x(t)$. At least one researcher [100] has argued that noise-induced chaos-order transitions need not be SR.

• **Random Systems** [15, 19, 29, 77, 171, 324]. These systems include many classical random processes such as random walks and Poisson processes. They also include the pulse system [15] whose response is a random train of pulses with a pulse probability r that depends on an input signal V through

$$r(V(t)) \;=\; r(0)\exp(V(t)). \tag{2.22}$$

The input V is the signal plus noise: $V(t) = s(t) + n(t)$. This model includes many kT-driven physio-chemical systems [15].

Other systems in the literature show the SR effect [7, 11, 14, 20, 62, 126, 147, 200, 202, 227, 234, 264, 288, 307, 316, 320]. Special issues of physics journals [24, 219] also present other systems that show SR. Most use the SR measures in the next section.

2.2 Performance Measures

SR depends on how the signal system "performs" as an input noise intensity varies. These performance measures can depend on the forcing signal and noise and can vary from system to system. They also depend on the application of the signal system. A system shows the SR effect when nonzero noise maximizes a given performance measure. Different performance criterion can give different performance curves and thus different "optimal" noise levels for a dynamical system. There is no consensus in the SR literature on how to measure the SR effect.

Some researchers study a stochastic dynamical system in terms of the Fokker-Planck (or forward Kolmogorov) equation [65, 145, 222, 272]:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(a(x,t)\,p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x,t)\,p) \tag{2.23}$$

for the drift term $a(x,t)$ and the diffusion term $b(x,t)$. This partial differential equation stems from a Taylor series and shows how a probability density function p of a Markov system's states evolves in time. But SR dynamical systems in general need not be Markov processes [90, 231]. System nonlinearities often preclude closed-form solutions. Approximations and assumptions such as small noise and small signal effects can give closed-form solutions in some cases. These solutions motivate some of the performance measures below.

2.2.1 Signal-to-Noise Ratio

The most common SR measure is some form of a signal-to-noise ratio (SNR) [80, 86, 98, 130, 202, 321]. This seems the most intuitive measure even though there are many ways to define a SNR in a nonlinear system.

Suppose the input signal is the sine wave $s(t) = \varepsilon \sin \omega_0 t$. Then the SNR measures how much the system output $y = g(x)$ contains the input signal frequency ω_0:

$$\text{SNR} = 10 \log \frac{S}{N} \tag{2.24}$$

$$= 10 \log \frac{S(\omega_0)}{N(\omega_0)} \quad \text{dB}. \tag{2.25}$$

The signal power $S = |Y(\omega_0)|^2$ is the magnitude of the output power spectrum $Y(\omega)$ at the input frequency ω_0. The background noise spectrum $N(\omega_0)$ at input frequency ω_0 is some average of $|Y(\omega)|^2$ at nearby frequencies [135, 202, 321]. The discrete Fourier transform (DFT) $Y[k]$ for $k = 0, \ldots, L-1$ is an exponentially weighted sum of elements of a discrete-time sequence

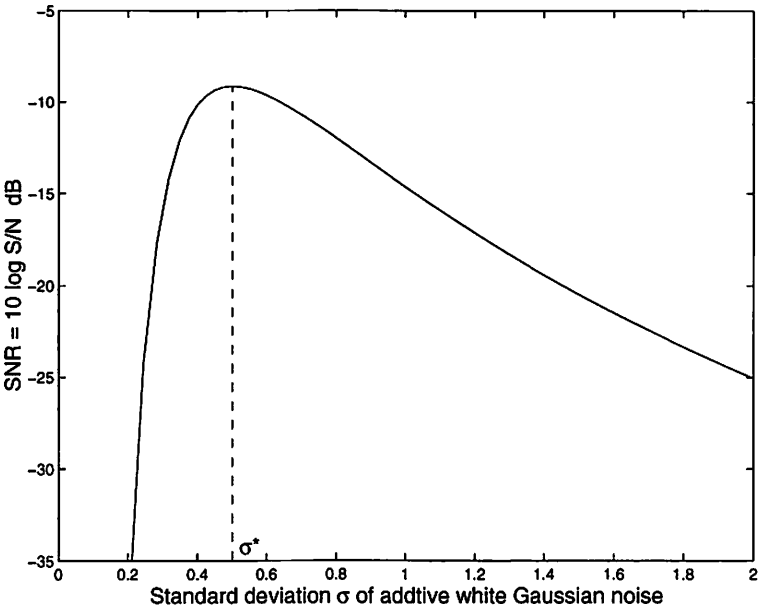


Figure 2.2: The closed form solution of signal-to-noise ratio for a quartic bistable system $\dot{x} = ax - bx^3 + s(t) + n(t)$ as in (2.29). The sinusoidal input $s(t) = \varepsilon \sin 2\pi f_0 t$ has small amplitude $\varepsilon = 0.2$ and low frequency $f_0 = 0.01$. The system's parameters are $a = b = 1$ and so the SNR closed form solution (2.29) gives an estimate of optimal Gaussian noise n at $\sigma^* = \sqrt{U_0} = \sqrt{a^2/4b} = 0.5$.

$\{y_0, y_1, \ldots, y_{L-1}\}$ of output signal samples

$$Y[k] = \sum_{t=0}^{L-1} y_t \exp\{-\frac{2\pi kt}{L}\}. \tag{2.26}$$

The signal frequency ω_0 corresponds to bin k_0 in the DFT for integer $k_0 = L\Delta T f_0$ and for $\omega_0 = 2\pi f_0$. This gives the output signal in terms of a DFT as $S = |Y[k_0]|^2$. The noise power $N = N[k_0]$ is the average power in the adjacent bins $k_0 - M, \ldots, k_0 - 1, k_0 + 1, \ldots, k_0 + M$ for some integer M [6, 321]:

$$N = \frac{1}{2M} \sum_{j=1}^{M} (|Y[k_0 - j]|^2 + |Y[k_0 + j]|^2). \tag{2.27}$$

We expand this noise term to include all energy not due to the signal as discussed in Chapter 5.

An adiabatic approximation [202] can give an explicit signal-to-noise ratio R for the quartic bistable system in (2.7) with sinusoidal input $s(t) = \varepsilon \sin \omega_0 t$:

$$R = \frac{S}{N} = \left[\frac{\sqrt{2}a\varepsilon^2 c^2}{(\sigma^2)^2} \exp\{-\frac{2U_0}{\sigma^2}\}\right] \left[1 - \frac{\frac{4a^2\varepsilon^2c^2}{\pi^2(\sigma^2)^2} \exp\{-\frac{4U_0}{\sigma^2}\}}{\frac{2a^2}{\pi^2} \exp\{-\frac{4U_0}{\sigma^2}\} + \omega_0^2}\right] \tag{2.28}$$

$$\approx \frac{\sqrt{2}a\varepsilon^2 c^2}{(\sigma^2)^2} \exp\{-2U_0/\sigma^2\}. \tag{2.29}$$

Here $U_0 = a^2/4b$ is the barrier height when $\varepsilon = 0$, $x = \pm c = \pm\sqrt{a/b}$ defines the potential minima, and σ^2 is the variance of the additive white Gaussian noise n. This result stems from Kramers rate [164] if the signal amplitude ε is small and if its frequency is smaller than the characteristic rate or curvature at the minimum $U''(\pm c)$ [202]. The SNR approximation (2.29) is zero for zero noise $\sigma^2 = 0$. It grows from zero as σ^2 grows and reaches a maximum at $\sigma^2 = U_0$ before it decays. So the optimum noise intensity is $\sigma^2 = U_0 = a^2/4b$. Figure 2.2 shows the SR profile of the closed form SNR in (2.29) for the quartic bistable system with sinusoidal input. The parameters are $a = b = 1$ and $\epsilon = 0.2$. Then the optimal noise level is at $\sigma^2 = U_0 = a^2/4b = 0.25$.

There is no standard definition of system-level signal and noise in nonlinear systems. This dissertation works with a SNR that is easy to compute and that depends on standard spectral power measures in signal processing as defined in Chapter 3. We apply the SNR measure only to nonlinear systems with sinusoidal input signals. Figure 1.2 shows the diagram of a dynamical system and how we obtain the SNR measure of the system output y.

2.2.2 Cross-Correlation Measures

These "shape matchers" can measure SR when inputs are not periodic signals. Researchers coined the term "aperiodic stochastic resonance" (ASR) [47, 50, 51, 118] for such cases. They defined cross-correlation measures for the input signal s and the system response in terms of

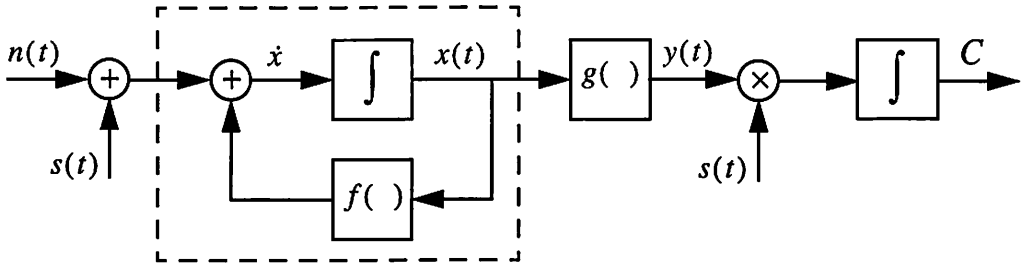


Figure 2.3: System diagram for stochastic resonance within a cross-correlation framework. A dynamical system $\dot{x} = f(x) + s + n$ has input forcing signal s and noise n. The system output y depends on the system state x through $y = g(x)$. The input noise process $n(t)$ drives the cross-correlation measure C. Note that C does not depend on the spectral structure of the input signal or noise processes.

the mean transition rate r in the FHN model in (2.12)-(2.13):

$$C_0 = \max_{\tau}\{\overline{s(t)r(t+\tau)}\} \tag{2.30}$$

and the normalized measure which has the form of sample correlation coefficients

$$C_1 = \frac{C_0}{[\overline{s^2(t)}]^{1/2}\{\overline{[r(t)-\overline{r(t)}]^2}\}^{1/2}}. \tag{2.31}$$

Here $\overline{x}$ is the time average: $\overline{x} = \frac{1}{T}\int_0^T x(t)dt$. Note that the cross-correlation measure C_0 can take any real value $C_0 \in R$ and the correlation coefficient C_1 lies between -1 and 1. Figure 2.3 shows the diagram of a dynamical system and how we obtain the cross-correlation measure C_0 of the system output y. This dissertation considers the cross-correlation measure C_0 for nonlinear systems with both sinusoidal and broadband input signals.

2.2.3 Information and Probability of Detection

Tools from information theory can also measure SR. The information rate of a threshold system shows the SR effect for subthreshold inputs [32, 41, 42, 297]. The FitzHugh-Nagumo (FHN) neuron model (2.12)-(2.13) shows SR for aperiodic input waveforms when we measure the cross-correlation between input and output or the information rate [50, 52, 118]. Noise can also

sometimes maximize the mutual information [58]:

$$I(X;Y) \;=\; H(X) - H(X|Y) \;=\; \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{2.32}$$

The mutual (Kullback) information $I(X;Y)$ and Fisher information [58] can measure SR in some neuron models [32, 230, 297]. Probability of correct detection and other statistics can also measure SR [127, 135, 297].

2.2.4 Probability of Residence Time and Escape Rate

This approach looks at the probability $P(T)$ of the time T that a dynamical system spends in a stable state between consecutive switches between the stable states [62, 77, 95, 141, 322]. So $P(T)$ depends on the input noise intensity. Data can give a histogram of this $P(T)$ to estimate the actual probability for each input noise intensity σ_n^2. The probability of residence time relates to the first passage time density function (FPTDF) or the interspike interval histogram (ISIH) found in the neurophysiological literature [19, 26, 29, 37, 87, 184, 185, 181, 182, 183, 194]. The symmetric bistable system (2.7) with input $s(t) = \varepsilon \sin \omega_0 t$ gives a system that tends to stay at or wander about one stable state for $T = T_0/2 = 2\pi/\omega_0$ seconds and then hops to a new stable state as it tracks the input.

2.2.5 Complexity and Other Performance Measures

Researchers have suggested other ways to measure SR. These include Lyapunov exponents, Shannon entropy, fluctuation complexity that measures the net information gain, and ϵ-complexity for first-order Markov stochastic automata [187, 317].

Other forms of SR measures also occur in the SR literature. They include the other signal-to-noise ratios [73, 143, 157, 175, 179], the amplification characteristic of a system like those found in electronic devices [9, 45, 108, 109, 146], susceptibility [74, 75, 215, 299], "crisis" measure in chaos [36], and prediction error of spike rates [38]. The number of SR performance measures will

likely grow as researchers explore how noise and signals drive other systems in the vast function space of nonlinear dynamical systems.

2.3 Optimal Noise

Stochastic resonance occurs when noise enhances a signal in the output of a system. So SR occurs if and only if the system has nonzero noise optimum. The performance of the SR system improves as noise grows from zero noise to a maximum and then falls as more noise swamps the signal system. This rise and fall of a system performance measure as noise grows implies a local maximum of the performance at some noise level. Some systems may have multiple local maxima or stochastic "multiresonance" [92, 308]. This dissertation consider signal systems with only one maximum in their SR profiles. The noise levels that maximize given performance measures define the *best* noise or the *optimal* noise for such signal systems.

We can also change the shape of the probability density function that underlies the additive noise that induces SR. This gives a more general definition of optimal noise drawn from all possible probability distributions. An even broader scope would consider that an "optimal noise" is some "signal." So any deterministic or random signals that add to a system and maximize its performance measure also qualify as optimal "noise." Some systems may favor a combination of different sinusoidal signals to enhance a particular input sine wave. This form of crosstalk of many signals that enhance each other may prove useful in some nonlinear dynamical systems. ASR for these systems may require knowledge and analysis of the signal systems rather than just a few of their input-output samples.

This dissertation assumes that the form of input noise density does not change for a given signal system. The ASR learning scheme only varies a variance or a dispersion measure of the probability density function. Then the ASR system learns the optimal variance (through a standard deviation) or dispersion (through a noise scale) of this density to achieve the stochastic resonance effect.

Chapter 3

Stochastic Resonance in Computer Simulation

Discrete simulations can model the known continuous-time nonlinear dynamical systems in Chapter 2. A stochastic numerical scheme approximates the system dynamics and its signal and noise response. We use a simple stochastic version of the Euler scheme to model a nonlinear system with input forcing signal and noise. We measure how the system performs based only the system's input-output samples.

3.1 Nonlinear System Simulations

Consider the forced dynamical system (2.4)-(2.5) with additive forcing input signal s and "white" noise n

$$\dot{x} = f(x) + s(t) + n(t) \tag{3.1}$$

$$y(t) = g(x(t)). \tag{3.2}$$

These models simply add noise term to a differential equation rather than use a formal Ito or Stratonovich stochastic differentials [49, 69, 99]. Throughout this dissertation the "whiteness" of a random variable n implies that n is white over some large but finite frequency bandwidth interval $[-B, B]$ for some large $B > 0$. Random numbers from the algorithms in [261, 39, 303] act as noise from various distributions in our simulations. The sections below show how we discretize the continuous-time systems to the discrete-time systems for computer simulation.

3.1.1 Nonlinear Systems with White Gaussian Noise

Consider the dynamical system (2.4) with initial condition $x(t_0) = x_0$. Here the white Gaussian noise w has zero mean and unit variance so $n = \sigma w$ has zero mean and variance σ^2. This system corresponds to the stochastic initial value problem [99]

$$dX = \tilde{f}(t, X) + \sigma(t, X)dW \tag{3.3}$$

for initial condition $X(t_0) = X_0$. Here $\tilde{f}(t, X) = f(X) + s(t)$, $\sigma(t, X) = \sigma$ and W is the standard Wiener process [99]. We obtain its discrete form for computer simulation from the stochastic version of Euler's method (the Euler-Maruyama scheme) [61, 99, 134]:

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \sigma \sqrt{\Delta T} w_t \tag{3.4}$$

$$y_t = g(x_t) \tag{3.5}$$

for $t = 0, 1, 2, \ldots$ and initial condition x_0. The input sample s_t has the value of the signal $s(t\Delta T)$ at time step t. The zero-mean white Gaussian noise sequence $\{w_t\}$ has unit variance $\sigma_w^2 = 1$. The term $\sqrt{\Delta T}$ scales w_t so that $\sqrt{\Delta T} w_t$ conforms with the Wiener increment [99, 134, 222]. The output sample y_t is some transformation g of the system's state x_t.

This simple algorithm gives fairly accurate results for moderate nonlinear systems [99, 134, 183, 222]. There exist other algorithms that give more accurate numerical solutions of the

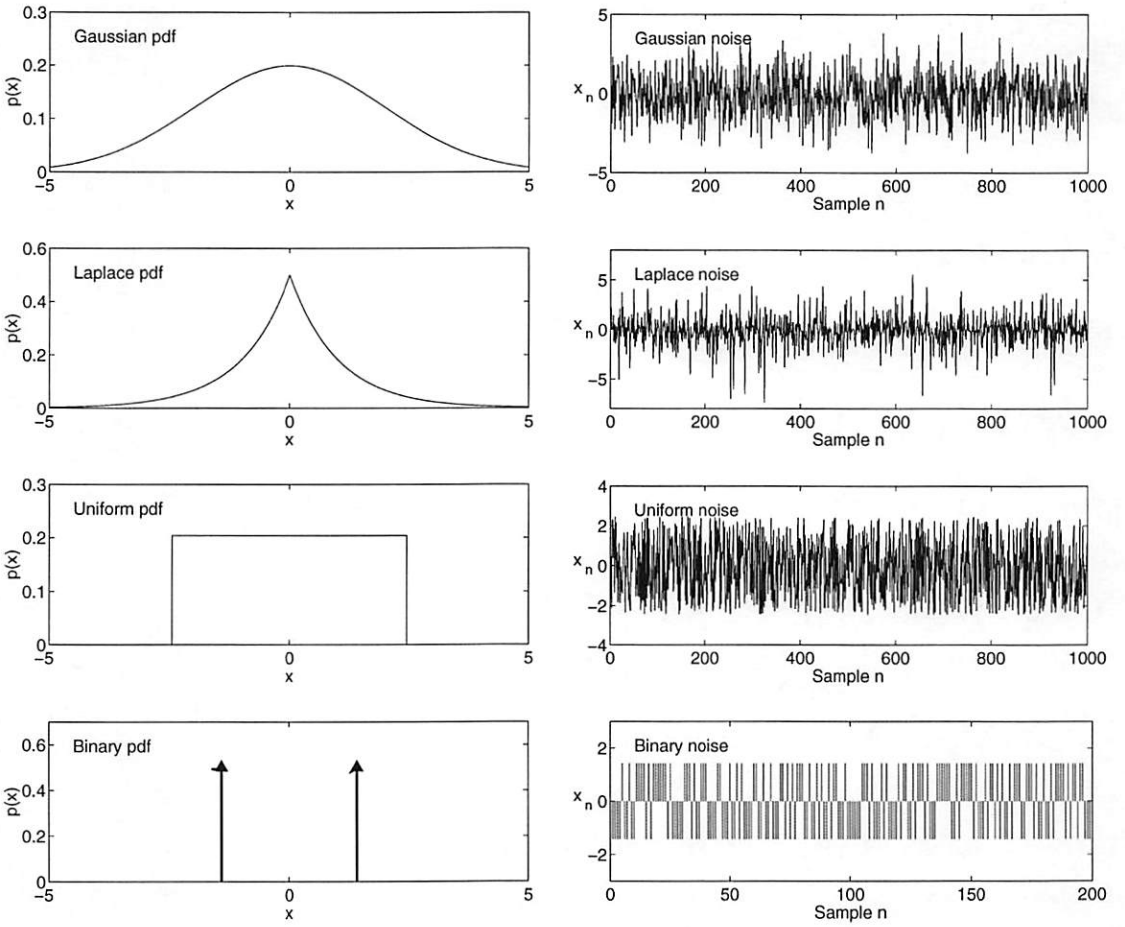


Figure 3.1: Probability density functions and random realizations. The figure shows Gaussian, Laplace, uniform, and binary random variables w with zero mean and variance of two: $E[x] = 0$ and $E[x^2] = \sigma^2 = 2$. The pseudo-random number generators in [261] act as noise sources with these probability densities.

corresponding stochastic differential equations to more complicated system dynamics [99, 197]. All of the simulations in this dissertation will apply the Euler's scheme in (3.4)-(3.5).

The numerical algorithm in [261] generates a sequence of pseudo-random numbers from a Gaussian density with zero mean and unit variance for $\{w_t\}$ in (3.4). The Gaussian probability density function (pdf) has the form

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \tag{3.6}$$

for mean $E[x] = m$ and variance $Var(x) = E[(x-m)^2] = \sigma^2$. Figure 3.1 shows the Gaussian and other pdf's that have zero mean and a variance of two.

3.1.2 Nonlinear Systems with Other Finite-Variance Noise

Now we consider a system (2.4) with any finite-variance noise n with other distributions. Suppose the noise n has variance σ^2. We again apply the above Euler's method for this system

$$x_{t+1} = x_t + \Delta T \Big(f(x_t) + s_t \Big) + \sigma \sqrt{\Delta T} w_t \tag{3.7}$$

$$y_{t+1} = g(x_{t+1}). \tag{3.8}$$

Here the random sequence $\{w_t\}$ has distribution $P(w)$ with zero mean and unit variance. The numerical algorithms in [261] generate sequences of random variables with Laplace, uniform, and binary density functions. A zero-mean Laplace random variable x has a density function of the form [249]

$$p(x) = \frac{c}{2} \exp(-c|x|) \tag{3.9}$$

where x has variance $Var(x) = E[x^2] = \frac{2}{c^2}$. A density function for a zero-mean uniform random variable x has the form

$$p(x) = \begin{cases} \frac{1}{2a} & \text{if } |x| \leq a \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

where x has variance $E[x^2] = \frac{a^2}{3}$. A binary (bipolar) random variable x with zero mean and variance σ^2 has the density function

$$p(x) = \frac{1}{2}\Big(\delta(x-\sigma) + \delta(x+\sigma)\Big). \tag{3.11}$$

Here δ is the Dirac delta function. Figure 3.1 plots the above probability density functions and their realizations with mean zero and variance of two: $E[x] = 0$ and $E[x^2] = 2$. The ASR

simulations in Chapter 5 also test an SR system with input noise from an output of a chaotic time series (the logistic map).

3.1.3 Nonlinear Systems with Alpha-Stable Noise

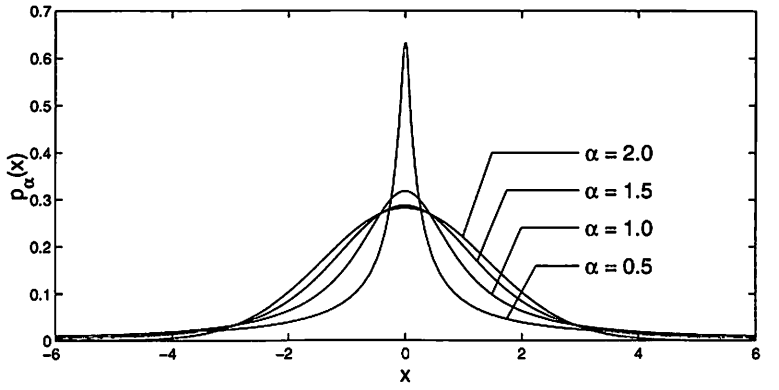
Here we consider a class of symmetric alpha-stable bell curves with parameter α in the characteristic function $\exp\{-|\omega|^\alpha\}$ [18, 81, 286, 287]. The parameter α lies in $0 < \alpha \leq 2$ and gives the Gaussian random variable when $\alpha = 2$ or when $\phi(\omega) = \exp\{-\omega^2\}$. So the standard Gaussian random variable has zero mean and variance $\sigma^2 = 2$. The parameter α gives the thicker-tailed Cauchy bell curve when $\alpha = 1$ or $\phi(\omega) = \exp\{-|\omega|\}$ for a zero *location* and unit *dispersion* Cauchy random variable. The moments of stable distributions with $\alpha < 2$ are finite only up to the order k for $k < \alpha$. The Gaussian density alone has finite variance and higher moments. Alpha-stable random variables characterize the class of normalized sums that converge in distribution to a random variable [18] as in the famous Gaussian version of the central limit theorem. Figure 3.2 shows realizations of some symmetric alpha-stable random variables.

Again we assume that the Euler's method above applies to this class of random variables with inifinite variance. Let w be a standard alpha-stable random variable with parameter α and zero location and unit dispersion: $a = 0$ and $\gamma = 1$. Let $\kappa = \gamma^{1/\alpha}$ denote a "scale" factor of a random variable. Then $n = \kappa w$ has zero location and dispersion $\gamma = \kappa^\alpha$. This leads us to the Euler's numerical solution

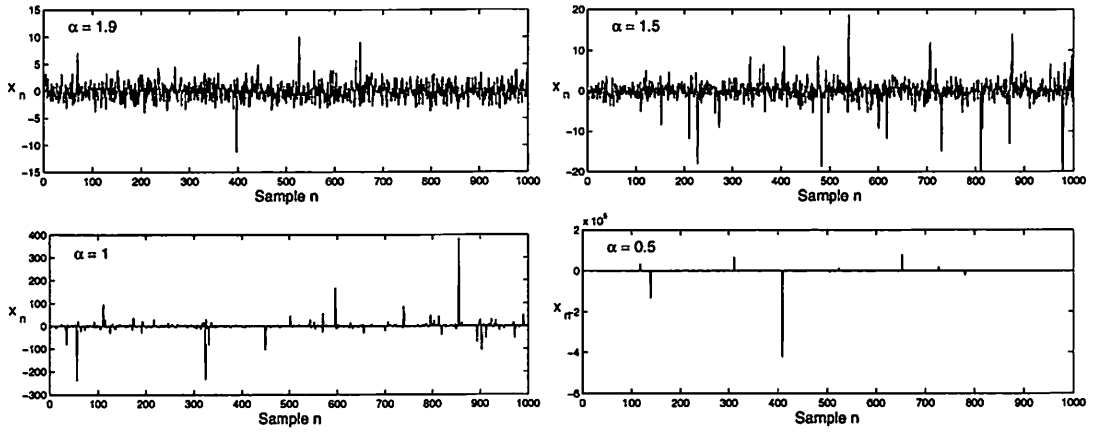
$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \kappa \sqrt{\Delta T} w_t \tag{3.12}$$

$$y_t = g(x_t). \tag{3.13}$$

Here the factor $\sqrt{\Delta T}$ may not give the same interpretation as in the case of Gaussian random variable from the Wiener increment viewpoint. The algorithm in [39, 303] generates a standard alpha-stable random variable w.



(a)



(b)

Figure 3.2: Samples of standard symmetric alpha-stable densities and their realizations. (a) Density functions with zero location ($a = 0$) and unit dispersion ($\gamma = 1$) for $\alpha = 2$, 1.5, 1, and 0.5. The densities are bell curves with thicker tails as α decreases. The case $\alpha = 2$ gives a Gaussian density with variance of two (or unit dispersion). The parameter $\alpha = 1$ gives the Cauchy density. (b) Samples of alpha-stable random variables with zero location and unit dispersion. The plots show realizations when $\alpha = 1.9$, 1.5, 1, and 0.5. Figure 3.1 shows the case of $\alpha = 2$ (Gaussian). Note the scale differences on the y-axes. The alpha-stable variable x becomes more impulsive as the parameter α falls. The algorithm in [39, 303] generates these realizations.

3.2 Performance Measures

The choice of performance measure depends on an application of the system. Engineers often rely on signal-to-noise ratio (SNR) to measure the quality of output signals and how their filters perform. Other popular measures include mean-squared error and bit-error probability in many signal processing and communication applications. These measures may not give the same "optimal" noise level. This dissertation works with the two most popular performance measures in the SR literature: signal-to-noise ratio and cross-correlation. These performance measures depend on the input-output samples of dynamical systems.

3.2.1 Signal-to-Noise Ratio in Nonlinear Systems with Sinusoidal Input

We start with a sinusoidal input and view the output state $y(t) = g(x(t))$ of the dynamical system as a mixture of signal and noise. We arrange the DFT computation so that the energy of the sine term lies in frequency bin k_0. The squared magnitude of this energy spectrum $Y[k_0]$ acts as the system-level signal: $S = 2|Y[k_0]|^2$. We view all else in the spectrum as noise: $N = P - S = P - 2|Y[k_0]|^2$ where the total energy is $P = \sum_{k=0}^{L-1} |Y[k]|^2$. We ignore the factor L that scales S and P since the ratio S/N cancels its effect.

Suppose a nonlinear dynamical system has a sinusoidal forcing function $s(t)$ of known frequency f_0 Hertz. We search the sinusoidal part $r(t)$ of the output $y(t)$ for the known frequency f_0 but unknown amplitude and phase in the system output response $y(t)$. The "noisy signal" $y(t)$ has the form of "signal" plus "noise":

$$y_t = r_t + n_t. \tag{3.14}$$

The signal-to-noise ratio (SNR) at the output is the spectral ratio of the energy of $\{r_t\}$ to the energy of $\{n_t\}$. We assume that the signal $s(t)$ is always present. This ignores the important problem of signal detection but lets us focus on learning the SR effect.

We define the SNR measure as

$$\text{SNR} = \frac{S}{N} = \frac{S}{P-S}. \tag{3.15}$$

Here $S = 2\,|Y[k_0]|^2$, $P = \sum_{k=0}^{L-1} |Y[k]|^2$, and $Y[k]$ is the L-point discrete Fourier transform (DFT) of y_n:

$$Y[k] = \sum_{t=0}^{L-1} y_t \exp\{-i\frac{2\pi k}{L}t\}. \tag{3.16}$$

We assume that the discrete frequency $k_0 = f_0 L T_s > 0$ is an integer for sampling rate $1/T_s$ an d $\omega_0 = 2\pi f_0$. We also assume that there is no aliasing due to sampling. Then we can show that for large L the SNR measure in (3.15) tends to the standard definition of SNR as a ratio of variances:

Theorem:

$$\text{SNR} = \frac{2\,|Y[k_0]|^2}{\sum_{k=0}^{L-1} |Y[k]|^2 - 2\,|Y[k_0]|^2} \longrightarrow \frac{\sigma_r^2}{\sigma_n^2} = \frac{A^2/2}{\sigma_n^2} = \frac{\frac{1}{L}\sum_{t=0}^{L-1} r_t^2}{\frac{1}{L}\sum_{t=0}^{L-1} n_t^2}. \tag{3.17}$$

Here $\sigma_r^2 = \frac{1}{T}\int_0^T (A\sin\omega_0 t)^2 dt = A^2/2$ and $\sigma_n^2 = Var(n) = E[n^2]$. We need further assumptions to derive (3.17). First consider the "energy" in each frequency bin k of the transform $Y[k]$:

$$|Y[k]|^2 = Y[k]Y^*[k] \tag{3.18}$$

$$= (R[k] + N[k])\,(R[k] + N[k])^* \tag{3.19}$$

$$= KJd + A\ ,m,>_N R[k]R^*[k] + R[k]N^*[k] + R^*[k]N[k] + N[k]N^*[k] \tag{3.20}$$

$$= |R[k]|^2 + |N[k]|^2 + 2Re\{R[k]N^*[k]\} \tag{3.21}$$

where $R[k]$ and $N[k]$ are the DFTs of r_t and n_t in (3.14). Suppose the sinusoidal term has the

form

$$r_t = A\cos(2\pi f_0 T_s t + \phi) \tag{3.22}$$

for $t = 0, \ldots, L-1$. Its DFT has the form [240]

$$\begin{aligned} R[k] &= \sum_{t=0}^{L-1} r_t \exp\{-i\frac{2\pi k}{L}t\} && (3.23)\\ &= \sum_{t=0}^{L-1} A\cos(2\pi f_0 T_s t + \phi)\exp\{-i\frac{2\pi k}{L}t\} && (3.24)\\ &= \exp\{ik\Omega(\phi - a)\}\frac{A}{T_s}\Big[\frac{\sin a\Omega(k-k_0)}{\Omega(k-k_0)} + \frac{\sin a\Omega(k+k_0)}{\Omega(k+k_0)}\Big] && (3.25)\\ &= \exp\{ik\Omega(\phi - a)\}\frac{Aa}{T_s}\Big(\delta[k-k_0] + \delta[k-(L-k_0)]\Big) && (3.26) \end{aligned}$$

where $k_0 = f_0 L T_s > 0$ is an integer, $\Omega = \dfrac{2\pi}{LT_s}$ is a frequency band, $a = \dfrac{LT_s}{2}$, and δ is the Kronecker delta function. So $R[k]$ vanishes when both $k \neq k_0$ and $k \neq L - k_0$. This gives

$$\sum_{k=0}^{L-1} |R[k]|^2 = |R[k_0]|^2 + |R[L-k_0]|^2 = 2\,|R[k_0]|^2 = 2\Big(\frac{aA}{T_s}\Big)^2 = \frac{L^2A^2}{2}. \tag{3.27}$$

So $R[k_0]$ and $R[L-k_0]$ contain all the energy of the sinusoidal signal r_t. We define the noise power as $\sigma_n^2 = E\{n_t^2\}$ and assume that n_t is stationary and ergodic with zero mean. Then Parseval's theorem gives

$$\begin{aligned} \sum_{k=0}^{L-1} |N[k]|^2 &= L\sum_{t=0}^{L-1} |n_t|^2 && (3.28)\\ &\approx L\,(L\sigma_n^2) && (3.29)\\ &= L^2\sigma_n^2. && (3.30) \end{aligned}$$

The ergodicity of n_t gives (3.29). Now consider the total output spectrum P:

$$P = \sum_{k=0}^{L-1} |Y[k]|^2 \tag{3.31}$$

$$= |Y[k_0]|^2 + |Y[L-k_0]|^2 + \sum_{k=0,k\neq k_0,L-k_0}^{L-1} |Y[k]|^2 \tag{3.32}$$

$$= 2\,|Y[k_0]|^2 + \sum_{k=0,k\neq k_0,L-k_0}^{L-1} |N[k]|^2 \tag{3.33}$$

$$= 2\,|R[k_0]|^2 + 2\,|N[k_0]|^2 + 4\,Re\{R[k_0]N^*[k_0]\} + \sum_{k=0,k\neq k_0,L-k_0}^{L-1} |N[k]|^2 \tag{3.34}$$

$$= 2\,|R[k_0]|^2 + \Big(\sum_{k=0}^{L-1} |N[k]|^2\Big) + 4Re\{R[k_0]N^*[k_0]\}. \tag{3.35}$$

Then (3.35) and (3.21) give

$$P - 2\,|Y[k_0]|^2 = \sum_{k=0}^{L-1} |N[k]|^2 - 2\,|N[k_0]|^2. \tag{3.36}$$

Then the SNR structure in (3.15) follows:

$$\text{SNR} = \frac{S}{N} = \frac{S}{P-S} \tag{3.37}$$

$$= \frac{2|R[k_0]|^2 + 2|N[k_0]|^2 + 4Re\{R[k_0]N^*[k_0]\}}{\Big(\sum_{k=0}^{L-1} |N[k]|^2\Big) - 2|N[k_0]|^2} \tag{3.38}$$

$$\approx \frac{2|R[k_0]|^2}{\sum_{k=0}^{L-1} |N[k]|^2} = \frac{L^2\,A^2/2}{L^2\sigma_n^2} = \frac{A^2/2}{\sigma_n^2} \tag{3.39}$$

for large L and for small (or null) $|N[k_0]|$ and $|N[L-k_0]|$. Note that $|N[k_0]| = |N[L-k_0]|$ for $k_0 \neq 0$ due to the symmetry of the DFT.

The result (3.39) also holds if the zero-mean noise sequence n_t is not correlated in time and does not correlate with r_t. Then we can take expectations of $2|Y[k_0]|^2$ and $P - 2|Y[k_0]|^2$ to get

$$E\Big[\,|Y[k_0]|^2\,\Big] = E\Big[\,|R[k_0]|^2 + |N[k_0]|^2 + 2Re\{R[k_0]N^*[k_0]\}\,\Big] \tag{3.40}$$

$$= E[\,|R[k_0]|^2\,] + E[\,|N[k_0]|^2\,] + 2Re\{E[\,R[k_0]\,]\,E[\,N^*[k_0]\,]\}\Big] \tag{3.41}$$

$$= E[\,|R[k_0]|^2\,] + E\Big[\Big(\sum_{t=0}^{L-1} n_t\,\exp\{-i\frac{2\pi k}{L}t\}\Big)\Big(\sum_{\tau=0}^{L-1} n_\tau\,\exp\{-i\frac{2\pi k}{L}\tau\}\Big)^*\Big]$$
$$+ 2Re\{R[k_0]\,E\Big[\sum_{\tau=0}^{L-1} n_\tau\,\exp\{-i\frac{2\pi k}{L}\tau\}\Big]^*\} \tag{3.42}$$

$$
\begin{aligned}
&= |R[k_0]|^2 + \sum_{t=0}^{L-1}\sum_{\tau=0}^{L-1} E[\,n_t\,n_\tau\,]\,\exp\{-i\frac{2\pi k}{L}(t-\tau)\} \\
&\quad + 2Re\{E[\,R[k_0]\,]\,\Big(\sum_{\tau=0}^{L-1} E[\,n_\tau\,]\,\exp\{i\frac{2\pi k}{L}\tau\}\Big)\} \qquad (3.43)\\
&= |R[k_0]|^2 + \sum_{t=0}^{L-1}\sum_{\tau=0}^{L-1} \sigma_n^2\,\delta[t-\tau]\,\exp\{-i\frac{2\pi k}{L}(t-\tau)\} + 0 \qquad (3.44)\\
&= |R[k_0]|^2 + L\sigma_n^2 \qquad (3.45)\\
&= L^2\frac{A^2}{4} + L\sigma_n^2 \qquad (3.46)
\end{aligned}
$$

and

$$
\begin{aligned}
E[\,P - 2\,|Y[k_0]|^2\,] &= E\Big[\sum_{k=0}^{L-1} |N[k]|^2 \; - 2|N[k_0]|^2\Big] \qquad (3.47)\\
&= \sum_{k=0}^{L-1} L\sigma_n^2 \; - 2L\sigma_n^2 \qquad (3.48)\\
&= L^2\,\sigma_n^2 - 2L\sigma_n^2. \qquad (3.49)
\end{aligned}
$$

Putting (3.46) and (3.49) into (3.15) gives

$$
\begin{aligned}
\text{SNR} &= \frac{2\,E[\,|Y[k_0]|^2\,]}{E[\,P - 2|Y[k_0]|^2\,]} \qquad (3.50)\\
&= \frac{2(\,L^2\frac{A^2}{4} + L\sigma_n^2\,)}{L^2\sigma_n^2 - 2L\sigma_n^2}. \qquad (3.51)
\end{aligned}
$$

Then SNR $\to \dfrac{A^2/2}{\sigma_n^2}$ as $L \to \infty$.

Figures 1.4 and 3.3 shows the SR profile with this SNR measure for the quartic bistable system with forcing sinusoidal input signal s and Gaussian, Laplace, uniform, and binary noise n. Figure 3.4 shows the SR profiles for a quartic bistable system and the FitzHugh Nagumo neuron model with with forcing sinusoidal input signal and alpha-stable noise. Figure 5.7 shows the SR profiles of the quartic bistable system for other impulsive noise with infinite variance. Figure 5.6 shows the SR profile of the quartic bistable system for chaotic noise from a logistic dynamical system.

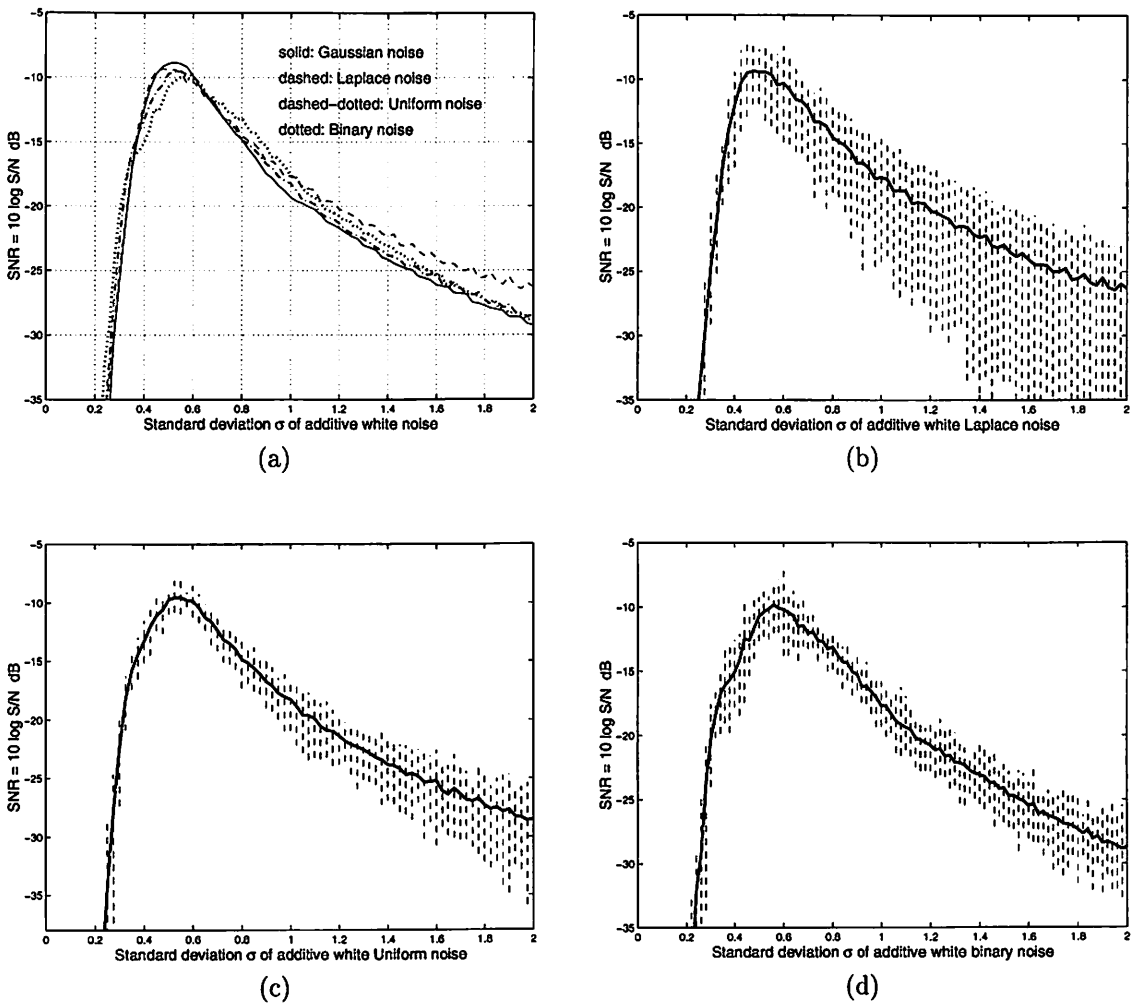
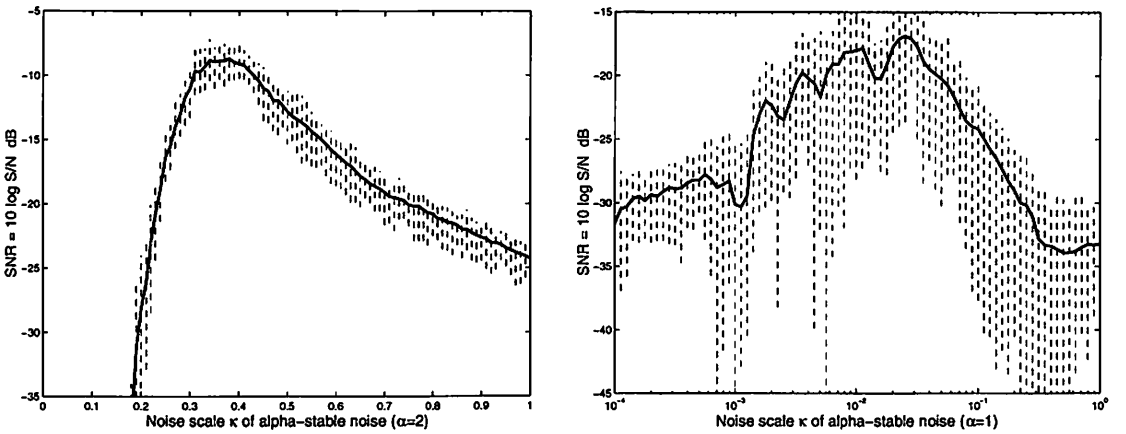
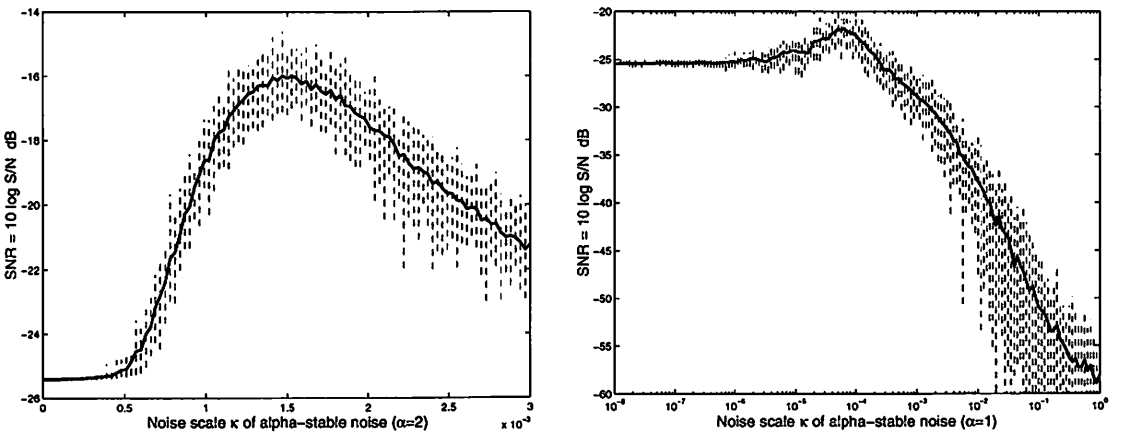


Figure 3.3: SNR measure of the quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ with output $y(t) = \text{sgn}(x(t))$. The sinusoidal input signal s is $s(t) = \varepsilon \sin 2\pi f_0 t$ where $\varepsilon = 0.1$ and $f_0 = 0.01$ Hz. (a) SNR-noise profiles of zero-mean white noise from Gaussian, Laplace, uniform, and binary probability densities. The simulation ran over 20 distinct noise seeds over 10,000 seconds with time step $\Delta T = 10000/1000000 = 0.01$ seconds in the forward Euler formula of numerical analysis. (b) Average SNR-noise profile and its spread for Laplace noise. (c) Average SNR-noise profile and its spread for uniform noise. (d) Average SNR-noise profile and its spread for binary noise. Figure 1.4 shows a like SR profile for Gaussian noise. Figure 5.6 shows the SR profile for the quartic bistable system when chaotic noise drives the system. The plots show distinct spreads of SNR for each kind of noise.



(a) Quartic bistable dynamical system



(b) FitzHugh-Nagumo neuron model

Figure 3.4: SNR-noise profiles of a quartic bistable system and a FitzHugh-Nagumo (FHN) neuron model. The plots show SNR-noise profiles for alpha-stable noise with $\alpha = 2$ (Gaussian density) and $\alpha = 1$ (Cauchy density). The densities have zero location $a = 0$ and dispersion γ that depends on a noise scale κ through $\gamma = \kappa^\alpha$. The scale κ equals $\sqrt{2}\sigma$ for a Gaussian density (when $\alpha = 2$) with variance σ^2. (a) The quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ with binary output $y(t) = \text{sgn}(x(t))$. We limit the magnitude of the system state x in (3.57) so that $|x| < 10$ as in Section 3.3.1. The sinusoidal input signal $s(t) = \varepsilon \sin 2\pi f_0 t$ has amplitude $\varepsilon = 0.1$ and frequency $f_0 = 0.01$. (b) The FHN model has the form $\epsilon\dot{x} = -x(x^2 - \frac{1}{4}) - z + A + s(t) + n(t)$ and $\dot{z} = x - z$ for $\epsilon = 0.005$ and $A = -(5/12\sqrt{3} + 0.07) = -.31056$ with a sinusoidal input $s(t) = \varepsilon \sin 2\pi f_0 t$ where $\varepsilon = 0.01$ and $f_0 = 0.5$. We limit the magnitude of the FHN model to $|x| < 2$. Figure 3.5 shows the SR profiles of these systems for a cross-correlation C performance measure.

3.2.2 Cross-Correlation Measure

The cross-correlation measure is simple and straightforward to compute. Here we follow the definition defined in [47, 50, 51, 118] as in (2.30) except that we omit the maximum argument. We let $\tau = 1$ replace the maximum argument over the time τ. Then the cross-correlation measure C for discrete simulations in this dissertation has the form

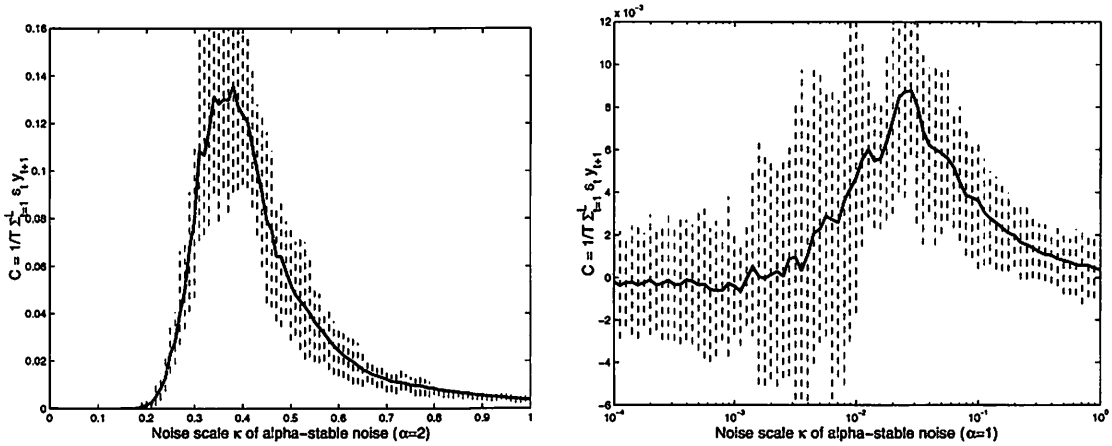
$$C = \frac{1}{L}\sum_{t=0}^{L-1} s_t r_{t+1}. \tag{3.52}$$

Figure 3.5 shows the SR profiles with the cross-correlation measure C for the quartic bistable system and the FitzHugh-Nagumo neuron model with sinusoidal inputs and noise with Gaussian and Cauchy densities.

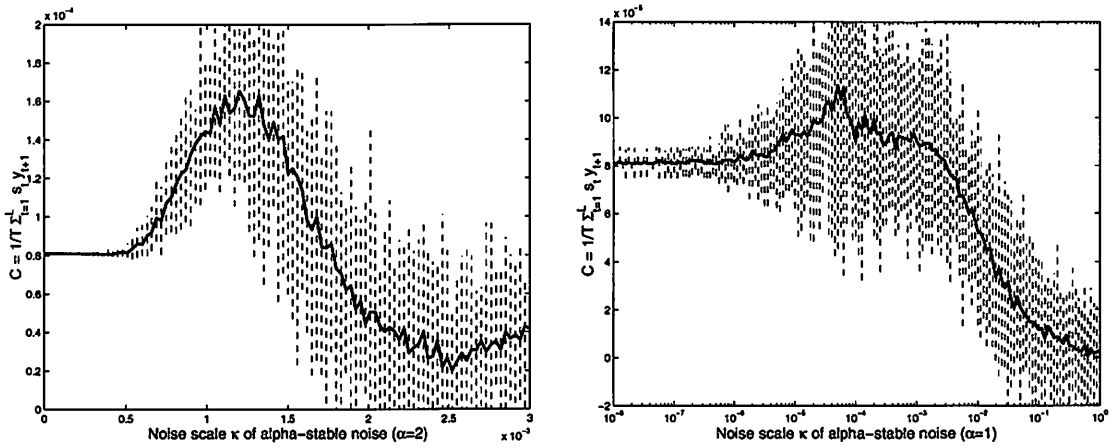
3.3 Stochastic Resonance and Alpha-Stable Noise

Most SR studies deal with Gaussian noise. A few systems work with other finite-variance noise such as uniform noise, Laplace noise, or (bounded) chaotic noise [41, 43, 44, 92, 212, 251]. This raises new questions: Will the signal system show the SR effect if the input noise has infinite variance? How does impulsiveness reshape the SR profile of a signal system? Does an increase in impulsiveness affect all SR systems in the same way? How does the increase in impulsiveness change the SR effect? What happens when the input noise grows more impulsive?

The dissertation empirically explores the relationships of the parameter α in alpha-stable distributions that determines impulsiveness and the optimal levels of noise in many stochastic resonance systems. This section shows simulation results of nonlinear systems with alpha-stable noise for different α's. We seek the relationship between α and the optimal dispersion γ_{opt}.



(a) Quartic bistable dynamical system



(b) FitzHugh-Nagumo neuron model

Figure 3.5: Cross-correlation C versus noise profiles of a quartic bistable system and a FitzHugh-Nagumo (FHN) neuron model. The plots show the cross-correlation C profiles for alpha-stable noise with $\alpha = 2$ (Gaussian density) and $\alpha = 1$ (Cauchy density). A noise scale κ relates a dispersion γ through $\gamma = \kappa^\alpha$. The scale κ equals $\sqrt{2}\sigma$ for a Gaussian density ($\alpha = 2$) with variance σ^2. (a) The quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ with binary output $y(t) = \text{sgn}(x(t))$ with the modification that $|x| < 10$. The sinusoidal input signal $s(t) = \varepsilon \sin 2\pi f_0 t$ has amplitude $\varepsilon = 0.1$ and frequency $f_0 = 0.01$. (b) The FHN model has the form $\epsilon\dot{x} = -x(x^2 - \frac{1}{4}) - z + A + s(t) + n(t)$ and $\dot{z} = x - z$ with $|x| < 2$ and $\epsilon = 0.005$ and $A = -(5/12\sqrt{3} + 0.07) = -.31056$. The sinusoidal input $s(t) = \varepsilon \sin 2\pi f_0 t$ has parameters $\varepsilon = 0.01$ and $f_0 = 0.5$.

3.3.1 SR Systems and Simulation Models

The computer simulation uses a discrete version in (3.12)-(3.13).

$$x_{t+1} = x_t + \Delta T \left(f(x_t) + s_t \right) + \sqrt{\Delta T} \kappa w_t \tag{3.53}$$

$$y_{t+1} = g(x_{t+1}) \tag{3.54}$$

with initial condition x_0 and output y_t. We assume that this discrete model applies to systems with alpha-stable noise. The zero-location white alpha-stable random sequence $\{w_t\}$ has unit dispersion $\gamma_w = 1$. So $n_t = \kappa w_t$ has dispersion $\gamma = \kappa^\alpha$. Note that a unit dispersion for Gaussian density (when $\alpha = 2$) equals a variance of two. We tested the following models:

- **Quartic bistable system (modified)**: The forced quartic bistable system has the form

$$\dot{x} = x - x^3 + s(t) + n(t) \tag{3.55}$$

$$y(t) = \text{sgn}(x(t)) \tag{3.56}$$

for binary output $y(t)$. We test the quartic bistable system model with the sinusoid input $s(t) = \varepsilon \sin 2\pi f_0 t$ for $\varepsilon = 0.1$ and $f_0 = 0.01$. The discrete version of the quartic bistable follows from (3.4)-(3.5) as

$$x_{t+1} = x_t + \Delta T(x_t - x_t^3 + s_t) + \sqrt{\Delta T} \kappa w_t \tag{3.57}$$

$$y_{t+1} = \text{sgn}(x_{t+1}). \tag{3.58}$$

We limit the magnitude of the system state x_t to 10 in the simulation model (3.57) because the impulsiveness of the alpha-stable noise could take x_t to $\pm\infty$ in computer simulations. We let $x_{t+1} = 10$ when $x_{t+1} > 10$ and let $x_{t+1} = -10$ when $x_{t+1} < -10$ in the discrete dynamic (3.57). This gives a modified version of the quartic bistable system. The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha) = \kappa^\alpha$ for the noise scale κ in (3.57).

• **FHN model (modified)**: The forced FHN model has the form

$$\begin{aligned} \epsilon\dot{x} &= -x(x^2 - \frac{1}{4}) - z + A' + s'(t) + n'(t) && (3.59) \\ \dot{z} &= x - z && (3.60) \\ y(t) &= x(t) && (3.61) \end{aligned}$$

for $\epsilon = 0.005$ and $A' = -(5/12\sqrt{3} + 0.07) = -.31056$ as in [118] and linear output $y(t)$. We use a sinusoidal input $s'(t) = \varepsilon \sin 2\pi f_0 t$ where $\varepsilon = 0.01$ and $f_0 = 0.5$. We can rewrite (3.59)-(3.61) as

$$\begin{aligned} \dot{x} &= -\frac{1}{\epsilon}x(x^2 - \frac{1}{4}) - \frac{1}{\epsilon}z + \frac{1}{\epsilon}A + \frac{1}{\epsilon}s'(t) + \frac{1}{\epsilon}n'(t) \\ &= -\frac{1}{\epsilon}x(x^2 - \frac{1}{4}) - \frac{1}{\epsilon}z + A + s(t) + n(t) && (3.62) \\ \dot{z} &= x - z && (3.63) \\ y(t) &= x(t) && (3.64) \end{aligned}$$

for $A = A'/\epsilon$. Then (3.4)-(3.5) give the discrete version to simulate the FHN model as

$$\begin{aligned} x_{t+1} &= x_t + \Delta T \left(-\frac{1}{\epsilon}x(x^2 - \frac{1}{4}) - \frac{1}{\epsilon}z + A + s_t \right) + \sqrt{\Delta T}\kappa w_t && (3.65) \\ z_{t+1} &= z_t + \Delta T(x_t - z_t) && (3.66) \\ y_{t+1} &= x_{t+1}. && (3.67) \end{aligned}$$

We also modify the recursive relation (3.65) so that the magnitude of x_{t+1} does not exceed 2. The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha) = \kappa^\alpha$ for the noise scale κ in (3.65).

• **Bistable neuron model**: The bistable potential neuron model with Gaussian white noise has the form

$$\dot{x} = -x + 2\tanh x + s(t) + n(t) \quad (3.68)$$

$$y(t) = \text{sgn}(x(t)). \tag{3.69}$$

The sinusoid input is $s(t) = \varepsilon \sin 2\pi f_0 t$ for $\varepsilon = 0.1$ and $f_0 = 0.01$. The discrete version has the form

$$x_{t+1} = x_t + \Delta T(-x_t + 2\tanh x_t + s_t) + \sqrt{\Delta T}\kappa w_t \tag{3.70}$$

$$y_{t+1} = \text{sgn}(x_{t+1}). \tag{3.71}$$

We test this neuron model with sinusoidal input $s(t) = \varepsilon \sin 2\pi f_0 t$ where $\varepsilon = 0.1$ and $f_0 = 0.01$.

- **Duffing oscillator (modified)**: The forced duffing oscillator has the form

$$\ddot{x} = -0.15\dot{x} + x - x^3 + \varepsilon \sin(\omega_0 t) + n(t) \tag{3.72}$$

$$y(t) = x(t). \tag{3.73}$$

We test the duffing oscillator with sinusoidal input $s(t) = \varepsilon \sin 2\pi f_0 t$ for $\varepsilon = 0.3$ and $f_0 = 0.01$. The discrete version of the duffing oscillator has the form

$$x_{t+1} = x_t + \Delta T z_t \tag{3.74}$$

$$z_{t+1} = z_t + \Delta T(-\delta z_t + x_t - x_t^3 + s_t) + \sqrt{\Delta T}\kappa w_t \tag{3.75}$$

$$y_{t+1} = \text{sgn}(x_{t+1}). \tag{3.76}$$

- **Threshold system**: The output y of a simple feedforward threshold system has the form

$$y_t = \text{sgn}(s_t + n_t - \Theta) = \text{sgn}(s_t + \kappa w_t - \Theta) \tag{3.77}$$

The optimal dispersion γ_{opt} has the form $\gamma_{opt}(\alpha) = \kappa^\alpha$ for κ in (3.77).

• **Pulse system:** This doubly Poisson process generates a pulse train with probability r then depends on the input $V(t) = s(t) + n(t)$

$$r(V(t)) \;=\; r(0)\exp(V(t)). \tag{3.78}$$

Here we let $r(0) = 1$. The sinusoid input is $s(t) = \varepsilon \sin 2\pi f_0 t$ for $\varepsilon = 0.5$ and $f_0 = 0.05$. The system generates an output $y(t)$ as a unit pulse with a rate $r(t)$.

3.3.2 Exponential Law with Linear Least-Square Fit of Log Data

Consider the SR-noise profiles for the quartic bistable system and the FitzHugh-Nagumo neuron model with alpha-stable noise. Figures 3.4-3.5 shows the profiles of these systems and alpha-stable noise for $\alpha = 2$ and $\alpha = 1$. Note how the optimal noise scale κ_{opt} falls as α falls. Since the dispersion depends on κ through $\gamma = \kappa^\alpha$ we then hypothesize that the optimal dispersion $\gamma_{opt}(\alpha)$ of the system obeys the exponential law

$$\gamma_{opt}(\alpha) = BA^\alpha \tag{3.79}$$

for some factors $B, A \in R$. Then

$$\log \gamma_{opt}(\alpha) \;=\; \log B + \alpha \log A \;\equiv\; a\alpha + b \tag{3.80}$$

for $a = \log A$ and $b = \log B$. The least square method gives the a and b values as

$$a = \frac{\sum_{i=1}^{N} (\alpha_i - \overline{\alpha})\, w_i}{\sum_{i=1}^{N} \alpha_i^2 - N(\overline{\alpha})^2} \qquad \text{and} \qquad b = \overline{w} - a\overline{\alpha} \tag{3.81}$$

for N data pairs (α_i, w_i) where $w_i = \log \gamma_{opt}(\alpha_i)$ at the experiment i with the parameter α_i. This method is the same as the minimum variance method for arbitrary random variables and the maximum likelihood method for normal random variables [249].

The correlation coefficient r^2 (coefficient of determination) indicates how good the linear model fits the data

$$r^2 = \frac{\sum(\hat{w}_i - \overline{w})^2}{\sum(w_i - \overline{w})^2} = \frac{\left(\sum(\alpha_i - \overline{\alpha})(w_i - \overline{w})\right)^2}{\sum(\alpha_i - \overline{\alpha})^2 \sum(w_i - \overline{w})^2} \tag{3.82}$$

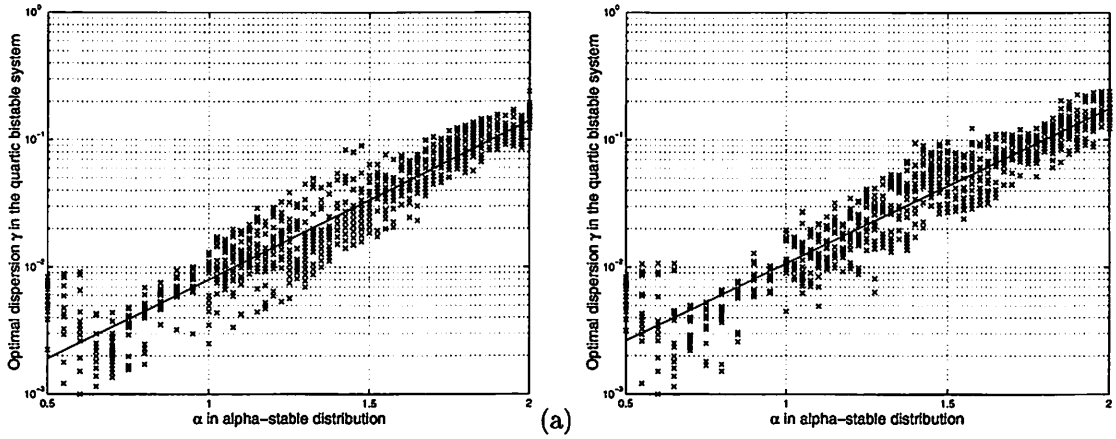
where $0 \leq |r| \leq 1$ and $|r| = 1$ iff $w_i = \hat{w}_i = a\alpha_i = b$ for every i. The positive and negative signs reflect the positive and negative slopes.

3.3.3 Test Results

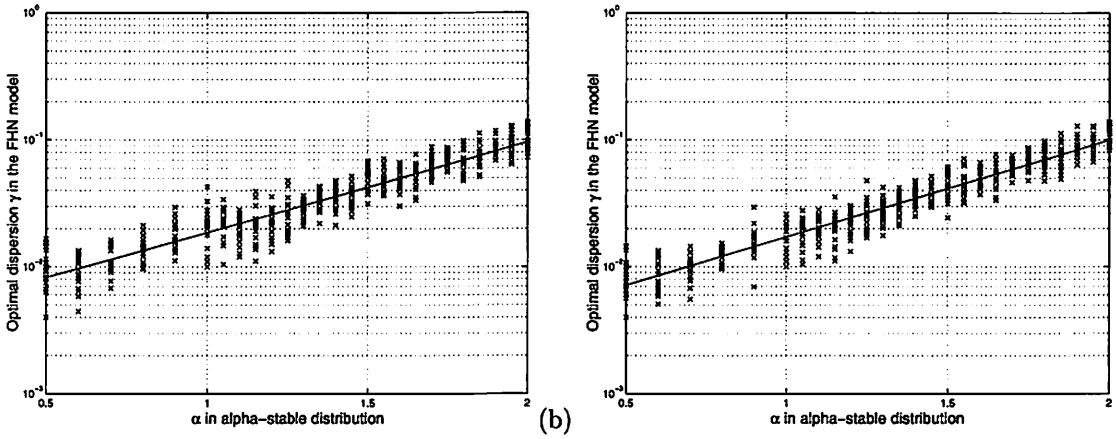
Table 3.1 below shows the parameters a and b of the linear least square fit of logarithm of the optimal dispersion γ_{opt} and the parameter α. The correlation coefficients r^2 measure how well the regression $a\alpha + b$ fit the data. Figure 3.6) shows the optimal dispersion $\gamma_{opt}(\alpha)$ and the parameters α. The plots in Figures 3.6(a)-(d) for feedback systems seem to agree with the exponential law. Figures 3.6(e)-(f) shows the plots for the feedforward systems which do not seem to follow the exponential law. Their correlation coefficients r^2 are small.

	SNR		C	
	parameters	r^2	parameters	r^2
Quartic bistable	$a = 1.2855, b = -3.4053$	0.8524	$a = 1.2075, b = -3.1682$	0.8442
FHN	$a = 0.7602, b = -2.5183$	0.8320	$a = 0.8061, b = -2.5968$	0.8690
Bistable neuron	$a = 1.8823, b = -3.9775$	0.9646	$a = 1.9175, b = -3.9628$	0.9605
Duffing oscillator	$a = 0.7320, b = -3.3057$	0.7444	$a = 0.8912, b = -3.3204$	0.8175
Threshold	$a = -0.6157, b = 0.3346$	0.8952	$a = -0.6142, b = 0.3312$	0.8949
Pulse	$a = 0.0692, b = 0.2267$	0.0406	$a = 0.2478, b = 0.2516$	0.3361

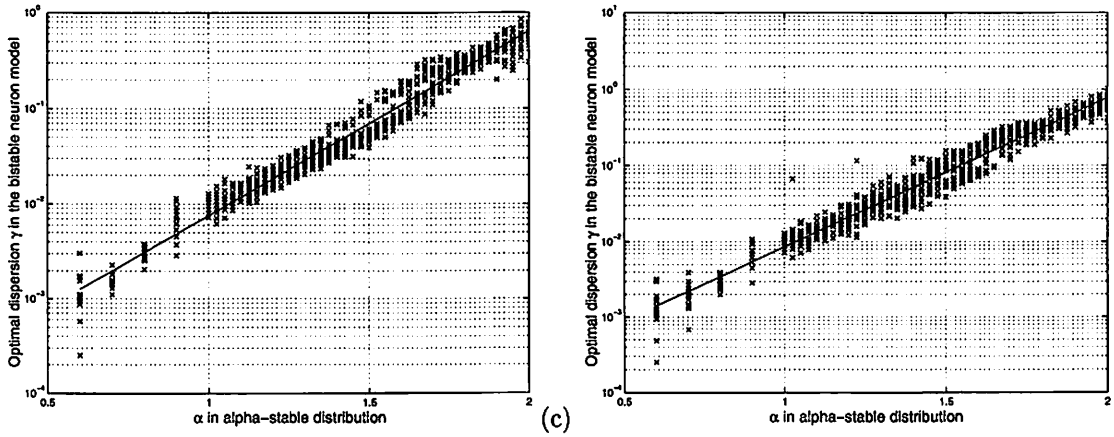
Table 3.1: Linear least square fit of the log of optimal dispersion γ and the parameter α in alpha-stable distribution. The parameters a and b relate $\log\gamma$ and α through a straight line: $\log\gamma(\alpha) = a\alpha + b$.


Optimal dispersion γ in the quartic bistable system
α in alpha-stable distribution
Optimal dispersion γ in the quartic bistable system
α in alpha-stable distribution

(a)


Optimal dispersion γ in the FHN model
α in alpha-stable distribution
Optimal dispersion γ in the FHN model
α in alpha-stable distribution

(b)


Optimal dispersion γ in the bistable neuron model
α in alpha-stable distribution
Optimal dispersion γ in the bistable neuron model
α in alpha-stable distribution

(c)

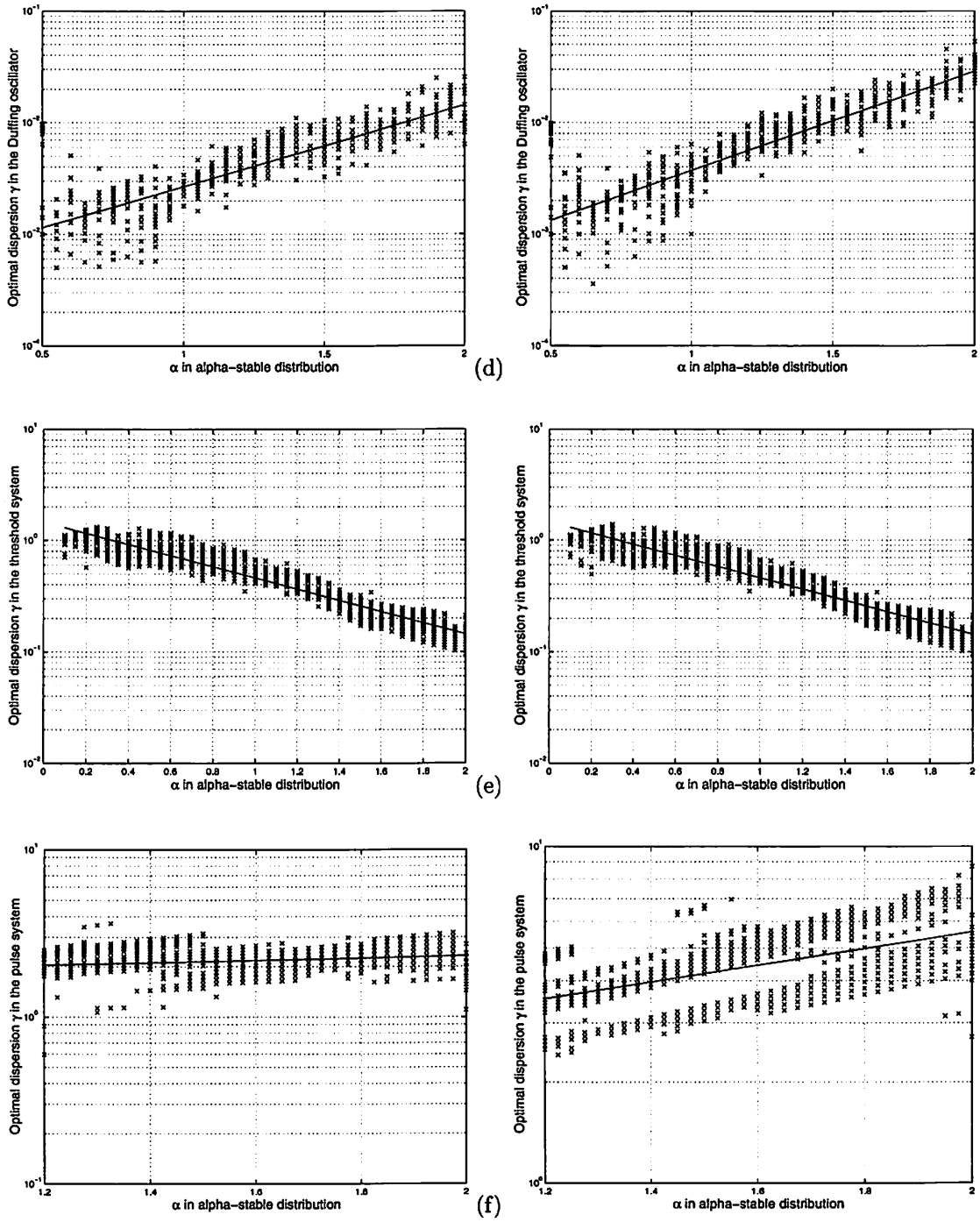


Figure 3.6: Optimal dispersion γ_{opt} versus α in alpha-stable distribution for (a) the quartic bistable system, (b) the FHN model, (c) the bistable potential neuron model (Cohen-Grossberg one-neuron neural network), (d) the duffing oscillator, (e) the threshold system, and (f) the pulse system with sinusoidal inputs. The plots on the left-handed side use the SNR performance measure and the plots on the right-handed side use the cross-correlation measure C. The mark x shows the "optimal" dispersion for each α-stable noise seed. Least-square regression defines the straight lines.

Chapter 4

Adaptive Stochastic Resonance with Gradient Learning

How can an adaptive system learn the SR effect? This chapter explores that question. The adaptive technique should depend on only the input-output samples of the system and not on a math model of the system dynamics. Then we can in theory apply the adaptive technique to any nonlinear dynamical system and for any performance measure.

This raises some key questions: How can the adaptive technique capture the structure of the performance measure and input noise from system input-output data so it can add the right amount of noise? How can the technique extract crucial information from the input-output data so it can quickly adapt the additive noise when the underlying statistics of the signal and system have changed? And how can the technique process the input-output data so it can use fewer data samples and fewer computations and can still maintain its performance to track the best noise?

Here we assume full control of a noise source with a specific random distribution. Such device is a pseudo-random number generator. Numerical algorithms let us generate pseudo-

random numbers for many forms of distributions that we use as noise in discrete simulations of dynamical systems [39, 261, 303].

We seek to maximize the performance measure P of the dynamical system $\dot{x} = f(x) + s + n$ with output $y = g(x)$. The performance measure $P : R^L \rightarrow R$ is a function of L output measurements $\{y_t\}$. Throughout this dissertation we apply the *stochastic* gradient algorithm on the performance measure P:

$$m_j(t+1) = m_j(t) + \mu_t \frac{\partial P}{\partial m_j} \tag{4.1}$$

where m_j is the jth parameter of an adaptive system F and where μ_t is a learning rate at iteration t. Formal stochastic approximation [273] further requires that the learning rate μ_n must decrease slowly but not too slowly:

$$\sum_{n=1}^{\infty} \mu_n^2 < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \mu_n = \infty. \tag{4.2}$$

Linear decay terms $\mu_n = 1/n$ obey (4.2). We used small but constant learning rates in most simulations.

This learning law requires the gradient $\frac{\partial P}{\partial m_j}$. The question is how can we estimate this gradient from the input-output data? The adaptive system needs to extract crucial information from input-output samples. Below we show how we can obtain a noisy measurement of the gradient from the system math model and also a noisy estimate from the input-output data.

4.1 Stochastic Gradient Learning on the Signal-to-Noise Ratio

An adaptive system can learn a SR noise pattern that maximizes a dynamical system's SNR. The learning law updates a parameter m_j of the adaptive system F at iteration n with the

deterministic law

$$m_j(n+1) \;=\; m_j(n) + \mu_n \frac{\partial E[\text{SNR}]}{\partial m_j}. \tag{4.3}$$

for learning coefficients $\{\mu_n\}$. This is gradient *ascent* learning. We assume that the first-order moment of the SNR exists. We seldom know the probability structure or the expectation of the SNR. So we estimate this expectation with its random realization at each time step: $E[\text{SNR}] \approx \text{SNR}$. This gives the *stochastic* gradient learning law

$$m_j(n+1) \;=\; m_j(n) + \mu_n \frac{\partial \text{SNR}}{\partial m_j} \tag{4.4}$$

or simple random hill climbing. We assume the chain rule holds (at least approximately) to give

$$\frac{\partial \text{SNR}}{\partial m_j} \;=\; \frac{\partial \text{SNR}}{\partial \sigma} \frac{\partial \sigma}{\partial m_j}. \tag{4.5}$$

Here σ is the noise level or standard deviation of the forcing noise term $n(t)$. We want the adaptive system F such as a fuzzy system in Section 4.4 to approximate the optimum noise level $\hat{\sigma}$ for any input signal or initial condition of the dynamical system: $F \approx \hat{\sigma}$. We then use σ and F interchangeably:

$$\frac{\partial \text{SNR}}{\partial m_j} \;=\; \frac{\partial \text{SNR}}{\partial \sigma} \frac{\partial F}{\partial m_j}. \tag{4.6}$$

The term $\frac{\partial F}{\partial m_j}$ shows how any adaptive system F depends on its jth parameter m_j. We again assume that the chain rule holds to get

$$\frac{\partial \text{SNR}}{\partial \sigma} \;=\; \frac{\partial \text{SNR}}{\partial S} \frac{\partial S}{\partial \sigma} + \frac{\partial \text{SNR}}{\partial N} \frac{\partial N}{\partial \sigma}. \tag{4.7}$$

Then $\text{SNR} = S/N$ implies that

$$\frac{\partial \text{SNR}}{\partial S} \;=\; \frac{\partial}{\partial S}\frac{S}{N} \;=\; \frac{1}{N} \tag{4.8}$$

$$\frac{\partial \text{SNR}}{\partial N} \;=\; \frac{\partial}{\partial N}\frac{S}{N} \;=\; -\frac{S}{N^2} \;=\; -\frac{\text{SNR}}{N}. \tag{4.9}$$

Like results hold for the decibel definition SNR = $10 \log S/N$ dB for the base-10 logarithm:

$$\frac{\partial \text{SNR}}{\partial S} = \frac{\partial}{\partial S} 10 \log \frac{S}{N} = (10 \log e) \frac{1}{S} \tag{4.10}$$

$$\frac{\partial \text{SNR}}{\partial N} = \frac{\partial}{\partial N} 10 \log \frac{S}{N} = -(10 \log e) \frac{1}{N}. \tag{4.11}$$

We next put (4.8)-(4.11) into (4.7) to get the log term that drives SR learning:

$$\frac{\partial \text{SNR}}{\partial \sigma} = \begin{cases} \frac{1}{N} \frac{\partial S}{\partial \sigma} - \frac{\text{SNR}}{N} \frac{\partial N}{\partial \sigma} & \text{if SNR} = \frac{S}{N} \\ (10 \log e) \left(\frac{1}{S} \frac{\partial S}{\partial \sigma} - \frac{1}{N} \frac{\partial N}{\partial \sigma} \right) & \text{if SNR} = 10 \log \frac{S}{N}. \end{cases} \tag{4.12}$$

4.1.1 Learning Law from the System Math Model

We now derive the SR learning laws in terms of DFTs. The math model in (2.1)-(2.2) gives the exact learning laws. Recall that the L-point DFT [240] for a sequence of states $\{y_t\}$ has the form

$$Y_n[k] = \sum_{l=0}^{L-1} y_{l+(n+1-L)} \exp\{-i \frac{2\pi k}{L} l\} \qquad k = 0, \ldots, L-1. \tag{4.13}$$

The time index n denotes the current time $t = nT_s$ for the sampling period T_s. Let $\frac{\partial S_n}{\partial y_j}$ denote the partial derivative of the signal energy S at iteration n with respect to the output y evaluated at time step j: $\frac{\partial S_n}{\partial y_j} = \frac{\partial S_n}{\partial y}[j]$. We likewise put $\frac{\partial N_n}{\partial y_j} = \frac{\partial N_n}{\partial y}[j]$ and $\frac{\partial y_j}{\partial \sigma} = \frac{\partial y}{\partial \sigma}[j]$. We assume some form of the chain rule holds to give

$$\frac{\partial S_n}{\partial \sigma} = \sum_{j=n+1-L}^{n} \frac{\partial S_n}{\partial y_j} \frac{\partial y_j}{\partial \sigma} \quad \text{and} \quad \frac{\partial N_n}{\partial \sigma} = \sum_{j=n+1-L}^{n} \frac{\partial N_n}{\partial y_j} \frac{\partial y_j}{\partial \sigma}. \tag{4.14}$$

We first derive $\frac{\partial S_n}{\partial y_j}$ and $\frac{\partial N_n}{\partial y_j}$ in (4.14). Consider the partial derivative of $|Y_n[k]|^2$ with respect to y at time step j:

$$\frac{\partial}{\partial y_j} \left| Y_n[k] \right|^2 = \frac{\partial}{\partial y_j} Y_n[k] Y_n^*[k] \tag{4.15}$$

$$= Y_n[k]\frac{\partial}{\partial y_j}Y_n^*[k] + Y_n^*[k]\frac{\partial}{\partial y_j}Y_n[k] \tag{4.16}$$

$$= Y_n[k]\exp\{i\frac{2\pi k}{L}(j-(n+1-L))\} + Y_n^*[k]\exp\{-i\frac{2\pi k}{L}(j-(n+1-L))\} \tag{4.17}$$

$$= 2\mathrm{Re}\{Y_n[k]\exp\{i\frac{2\pi k}{L}(j-(n+1-L))\}\} \tag{4.18}$$

$$= 2\,\mathrm{Re}\{Y_n[k]\}\cos\Big(\frac{2\pi k}{L}(j-(n+1-L))\Big) - 2\,\mathrm{Im}\{Y_n[k]\}\sin\Big(\frac{2\pi k}{L}(j-(n+1-L))\Big) \tag{4.19}$$

So the partial derivative of the signal spectrum $S_n = 2\,|Y_n[k_0]|^2$ is

$$\frac{\partial S_n}{\partial y_j} = 4\,\mathrm{Re}\{Y_n[k_0]\}\cos\Big(\frac{2\pi k_0}{L}(j-(n+1-L))\Big) - 4\,\mathrm{Im}\{Y_n[k_0]\}\sin\Big(\frac{2\pi k_0}{L}(j-(n+1-L))\Big). \tag{4.20}$$

The partial derivative $\frac{\partial N_n}{\partial y_j}$ follows in like manner:

$$\frac{\partial N_n}{\partial y_j} = \frac{\partial}{\partial y_j}(P_n - S_n) \tag{4.21}$$

$$= \frac{\partial}{\partial y_j}\sum_{k=0}^{L-1}\Big|Y_n[k]\Big|^2 - \frac{\partial S_n}{\partial y_j} \tag{4.22}$$

$$= \frac{\partial}{\partial y_j}L\sum_{t=0}^{L-1}y_t^2 - \frac{\partial S_n}{\partial y_j} \quad \text{from Parseval's relation} \tag{4.23}$$

$$= 2\,L\,y_j - \frac{\partial S_n}{\partial y_j}. \tag{4.24}$$

We can consider the term $\frac{\partial y_j}{\partial \sigma}$ in (4.14) as a sample of $\frac{\partial y}{\partial \sigma}$ at the time step j. We need to estimate $\frac{\partial y}{\partial \sigma}$ from sample data or from the system math model.

Recall the math model of the dynamical system (2.1)-(2.2) and let $G(x,u,t) = f(x)+u(x,t)$. Assume that this is a scalar system and that $u(x,t) = s(t)+n(t)$ where $n(t) = \sigma w(t)$ for the

zero-mean white noise process $w(t)$ with unit variance $E[w^2] = 1$. So the model becomes

$$\dot{x} = G(x, s, \sigma, w, t) = f(x) + s(t) + \sigma w(t) \tag{4.25}$$

$$y(t) = g(x(t)). \tag{4.26}$$

The chain rule gives

$$\frac{\partial y}{\partial \sigma} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial \sigma}. \tag{4.27}$$

Let $\eta(t)$ denote $\frac{\partial x}{\partial \sigma}$. Assume that G is sufficiently differentiable. Then differentiate η with respect to time [8] to get

$$\frac{d\eta}{dt} = \frac{d}{dt}\left(\frac{\partial x}{\partial \sigma}\right) = \frac{\partial \frac{dx}{dt}}{\partial \sigma} = \frac{\partial G(x, s, \sigma, w)}{\partial \sigma} \tag{4.28}$$

$$= \frac{\partial G}{\partial x}\frac{\partial x}{\partial \sigma} + \frac{\partial G}{\partial \sigma} = \frac{\partial G}{\partial x}\eta(t) + \frac{\partial G}{\partial \sigma}. \tag{4.29}$$

The last derivative $\frac{\partial G}{\partial \sigma}$ results from G's explicit dependence on σ. So the derivatives $\frac{\partial G}{\partial x}$ and $\frac{\partial G}{\partial \sigma}$ for the additive case $G(x, s, \sigma, w) = f(x) + s(t) + \sigma w(t)$ follow

$$\frac{\partial G}{\partial x} = \frac{\partial f}{\partial x} \tag{4.30}$$

$$\frac{\partial G}{\partial \sigma} = \frac{\partial}{\partial \sigma}[f(x) + s(t) + \sigma w(t)] = w(t). \tag{4.31}$$

The extension to multi-variable system is straightforward. Suppose m differential equations represent a dynamical system with m forcing signals and noise. Then the states x, signal s, and n are m-D vectors. We can express the dynamics in the form

$$\dot{x}_1 = f_1(x) + s_1(t) + n_1(t) = G_1(x, s_1, n_1) \tag{4.32}$$

$$\vdots$$

$$\dot{x}_m = f_m(x) + s_m(t) + n_m(t) = G_m(x, s_m, n_m) \tag{4.33}$$

For simplicity we consider scalar system output $y \in R$. We again let $n_j(t) = \sigma_j w_j(t)$ for zero-mean white noise processes $w_j(t)$. The chain rule gives

$$\frac{\partial y}{\partial \sigma_j} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \frac{\partial x_i}{\partial \sigma_j}. \tag{4.34}$$

Let $\eta_{ij}(t)$ denote $\frac{\partial x_i}{\partial \sigma_j}$. It follows that

$$\begin{aligned}
\dot{\eta}_{ij} &= \frac{\partial \dot{x}_i}{\partial \sigma_j} = \frac{\partial G_i}{\partial \sigma_j} && (4.35)\\
&= \sum_{k=1}^{n} \left(\frac{\partial G_i}{\partial x_k} \frac{\partial x_k}{\partial \sigma_j} + \frac{\partial G_i}{\partial s_k} \frac{\partial s_k}{\partial \sigma_j} + \frac{\partial G_i}{\partial n_k} \frac{\partial n_k}{\partial \sigma_j} \right) && (4.36)\\
&= \sum_{k=1}^{n} \left(\frac{\partial G_i}{\partial x_k} \frac{\partial x_k}{\partial \sigma_j} \right) + \delta_{ij} w_i(t) && (4.37)
\end{aligned}$$

where we let $\frac{\partial s_k}{\partial \sigma_j} = 0$ for all j, k (because the input signal does not depend on the noise level), $\frac{\partial G_i}{\partial n_k} = \delta_{ik}$, and $\frac{\partial n_k}{\partial \sigma_j} = \frac{\partial \sigma_k w_k}{\partial \sigma_j} = \delta_{jk} w_j$. If the system has only one input forcing signal and noise (both of which may enter the system at different states) then

$$\begin{aligned}
\dot{\eta}_i &= \frac{\partial \dot{x}_i}{\partial \sigma} = \frac{\partial G_i}{\partial \sigma} && (4.38)\\
&= \sum_{k=1}^{n} \frac{\partial G_i}{\partial x_k} \frac{\partial x_k}{\partial \sigma} + \frac{\partial G_i}{\partial s} \frac{\partial s}{\partial \sigma} + \frac{\partial G_i}{\partial n} \frac{\partial n}{\partial \sigma} && (4.39)\\
&= \begin{cases} \sum_k \left(\frac{\partial G_i}{\partial x_k} \frac{\partial x_k}{\partial \sigma_j} \right) + w(t) & \text{if the noise } n \text{ enters the system at } x_i \\ \sum_k \left(\frac{\partial G_i}{\partial x_k} \frac{\partial x_k}{\partial \sigma_j} \right) & \text{otherwise} \end{cases} && (4.40)
\end{aligned}$$

or we can rewrite in matrix form (here we assume that $n(t) = \sigma w(t)$ enters the systems at x_i)

$$\dot{\eta} = J(x)\eta + e_i w(t) \tag{4.41}$$

where $\eta = [\eta_1 \ \cdots \eta_m]^T$ and the unit vector $e_i = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T$ has all zero elements but unity at the ith slot. Here $J(x) = \left[\frac{\partial f_i}{\partial x_j} \right]$ is the Jacobian matrix of the system evaluated at x.

We need to simulate the evolution (4.29) for $\frac{\partial x}{\partial \sigma}$ and obtain $\frac{\partial y}{\partial \sigma}$ from (4.27). Then we put (4.20), (4.24), and $\frac{\partial y}{\partial \sigma}$ into (4.14) to get the stochastic gradient learning law:

$$\begin{aligned}
\sigma(n+1) &= \sigma(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} && (4.42)\\
&= \sigma(n) + \mu_n \Big(\frac{\partial \mathrm{SNR}_n}{\partial S_n} \frac{\partial S_n}{\partial \sigma} + \frac{\partial \mathrm{SNR}_n}{\partial N_n} \frac{\partial N_n}{\partial \sigma} \Big) && (4.43)\\
&= \sigma(n) + \mu_n \Big(\frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} \Big). && (4.44)
\end{aligned}$$

Here we omit the constant factor $10 \log e$ from (4.8)-(4.11) or view it as part of the learning rate μ_n in (4.44). The learning law for the parameters m_j of a function approximator F that approximates the surface of optimal noise levels follows in like manner. Here F replaces the parameter σ so the learning law becomes

$$\begin{aligned}
m_j(n+1) &= m_j(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial m_j} && (4.45)\\
&= m_j(n) + \mu_n \Big(\frac{\partial \mathrm{SNR}_n}{\partial S_n} \frac{\partial S_n}{\partial m_j} + \frac{\partial \mathrm{SNR}_n}{\partial N_n} \frac{\partial N_n}{\partial m_j} \Big) && (4.46)\\
&= m_j(n) + \mu_n \Big(\frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial F} \frac{\partial F}{\partial m_j} \\
&\qquad - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial F} \frac{\partial F}{\partial m_j} \Big). && (4.47)
\end{aligned}$$

We get (4.44) if σ replaces F and m_j. Appendix A derives the last partial derivative $\frac{\partial F}{\partial m_j}$ in the chain-rule expansion (4.6) for all SAM fuzzy parameters m_j. This is again the step where users can insert other adaptive function approximators F and derive learning laws for their parameters m_j by expanding $\frac{\partial F}{\partial m_j}$.

4.1.2 Approximation of the Learning Term $\frac{\partial \mathrm{SNR}}{\partial \sigma}$

The above learning law requires a complete knowledge of the math model that describes the dynamical system. It also needs accurate estimation of the evolution of (4.29). This may not be practical in many cases since we do not know the mathematical models (2.1)-(2.2). Suppose

we have access to the system input $\{s_t\}$ and output $\{y_t\}$. These input and output time series let us compute the performance of the system. SNR requires the knowledge of input spectrum while cross-correlation computes the inner product of the input and output sequences.

Suppose we have access to the output samples of the SR system with sinusoid input. Then we can compute the DFTs of the L samples to get the signal energy S and noise energy N in the SNR. Then we can approximate $\frac{\partial S}{\partial \sigma}$ and $\frac{\partial N}{\partial \sigma}$ with a ratio of time differences at each iteration n:

$$\frac{\partial S_n}{\partial \sigma_n} \approx \frac{\Delta S_n}{\Delta \sigma_n} = \frac{S_n - S_{n-1}}{\sigma_n - \sigma_{n-1}} \tag{4.48}$$

$$\frac{\partial N_n}{\partial \sigma_n} \approx \frac{\Delta N_n}{\Delta \sigma_n} = \frac{N_n - N_{n-1}}{\sigma_n - \sigma_{n-1}}. \tag{4.49}$$

So the learning law becomes

$$\sigma_{n+1} = \sigma_n + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} \tag{4.50}$$

$$= \sigma_n + \mu_n \left(\frac{1}{S_n} \frac{\partial S_n}{\partial \sigma} - \frac{1}{N_n} \frac{\partial N_n}{\partial \sigma} \right) \tag{4.51}$$

$$\approx \sigma_n + \mu_n \left(\frac{1}{S_n} \frac{S_n - S_{n-1}}{\sigma_n - \sigma_{n-1}} - \frac{1}{N_n} \frac{N_n - N_{n-1}}{\sigma_n - \sigma_{n-1}} \right). \tag{4.52}$$

We also replace the difference $\sigma_n - \sigma_{n-1}$ with its sign $\mathrm{sgn}(\sigma_n - \sigma_{n-1})$ to avoid numerical instability. The gradient becomes

$$\frac{\partial \mathrm{SNR}_n}{\partial \sigma} \approx \left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n} \right) \mathrm{sgn}(\sigma_n - \sigma_{n-1}). \tag{4.53}$$

This approximation gives the SR learning law

$$\sigma_{n+1} = \sigma_n + \mu_n \left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n} \right) \mathrm{sgn}(\sigma_n - \sigma_{n-1}). \tag{4.54}$$

This learning law does not require that we know the dynamical model. It depends only on samples from the system dynamics and from the input signal $s(t)$. Stochastic approximation

and perturbation give similar form of learning when we can sample the system performance measures [122, 277, 293].

4.1.3 SR Optimality for the SNR Measure

Recall the general form of the learning law (4.12)

$$\frac{\partial \text{SNR}}{\partial \sigma} = \begin{cases} \frac{1}{N}\frac{\partial S}{\partial \sigma} - \frac{\text{SNR}}{N}\frac{\partial N}{\partial \sigma} & \text{if SNR} = \frac{S}{N} \\ (10 \log e)\left(\frac{1}{S}\frac{\partial S}{\partial \sigma} - \frac{1}{N}\frac{\partial N}{\partial \sigma}\right) & \text{if SNR} = 10 \log \frac{S}{N}. \end{cases} \tag{4.55}$$

The right side of (4.55) leads to the first-order condition for an SNR extremum:

$$\frac{1}{S}\frac{\partial S}{\partial \sigma} - \frac{1}{N}\frac{\partial N}{\partial \sigma} = 0 \tag{4.56}$$

or simply

$$\frac{S}{N} = \frac{S'}{N'}. \tag{4.57}$$

We can rewrite this optimality condition as

$$\left.\frac{S}{N}\right|_{\sigma_{opt}} = \left.\frac{\partial S/\partial \sigma}{\partial N/\partial \sigma}\right|_{\sigma_{opt}} \tag{4.58}$$

when the partial derivatives of S and N with respect to σ are not zero at $\sigma = \sigma_{opt}$. Equations (4.56) and (4.58) give a necessary condition for the SR maximum. The result (4.58) says that at SR the ratio of the rate of changes of S and N must equal the ratio of S and N. This has the same form as the result in microeconomics [165] that the marginal rates of substitution of two goods must at optimality equal the partial derivatives of the utility function with respect to each good. But (4.57) and (4.58) hold only in a stochastic sense for sufficiently well-behaved random processes.

We find the second-order condition for an SR maximum when SNR $= 10 \log S/N$ from

$$
\begin{aligned}
0 > \frac{\partial^2 \mathrm{SNR}}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma} \frac{\partial \mathrm{SNR}}{\partial \sigma} && (4.59)\\
&= \frac{\partial}{\partial \sigma} (10 \log e) \Big[\frac{1}{S} \frac{\partial S}{\partial \sigma} - \frac{1}{N} \frac{\partial N}{\partial \sigma} \Big] && (4.60)\\
&= (10 \log e) \Big[\Big(\frac{1}{S} \frac{\partial^2 S}{\partial \sigma^2} + \frac{\partial S}{\partial \sigma} \Big(- \frac{1}{S^2} \frac{\partial S}{\partial \sigma} \Big) \Big) - \Big(\frac{1}{N} \frac{\partial^2 N}{\partial \sigma^2} + \frac{\partial N}{\partial \sigma} \Big(- \frac{1}{N^2} \frac{\partial N}{\partial \sigma} \Big) \Big) \Big] && (4.61)\\
&= (10 \log e) \Big[\frac{1}{S} \frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{S^2} \Big(\frac{\partial S}{\partial \sigma} \Big)^2 - \frac{1}{N} \frac{\partial^2 N}{\partial \sigma^2} + \frac{1}{N^2} \Big(\frac{\partial N}{\partial \sigma} \Big)^2 \Big] && (4.62)\\
&= (10 \log e) \Big[\frac{1}{S} \frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{N} \frac{\partial^2 N}{\partial \sigma^2} \Big] && (4.63)
\end{aligned}
$$

or $\frac{S''}{S} < \frac{N''}{N}$. The last equality follows from the first-order condition $\frac{1}{S} \frac{\partial S}{\partial \sigma} - \frac{1}{N} \frac{\partial N}{\partial \sigma} = 0$ or $\frac{S'}{S} = \frac{N'}{N}$ since then $\frac{(S')^2}{S^2} = \frac{(N')^2}{N^2}$. A like result holds for SNR $= S/N$. We still get the second-order condition

$$
\frac{1}{S} \frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{N} \frac{\partial^2 N}{\partial \sigma^2} < 0. \tag{4.64}
$$

These first- and second-order conditions show how the signal power S and noise power N relate to each other and to their derivatives at the SR maximum.

4.2 Stochastic Gradient Learning on the Cross-Correlation Measure

This section derives a learning law for the cross correlation measure C in (2.30) for the system

$$
\begin{aligned}
\dot{x} &= f(x) + s(t) + n(t) && (4.65)\\
y(t) &= g(x(t)). && (4.66)
\end{aligned}
$$

The cross-correlation measures the performance of the system from its input-output samples with the form

$$
C_n = \sum_{t=n-L+1}^{n} s_t y_{t+\tau} \tag{4.67}
$$

where we assume that the dynamical system has time lag $\tau = 1$ in (2.30). Our goal is to find the scale σ that maximize C. Stochastic gradient ascent can learn this σ through the learning law:

$$\sigma_{n+1} = \sigma_n + \mu \frac{\partial C_n}{\partial \sigma}. \tag{4.68}$$

Now we derive the learning term

$$\frac{\partial C_n}{\partial \sigma} = \frac{\partial}{\partial \sigma} \sum_{t=n-L+1}^{n} s_t y_{t+1} \tag{4.69}$$

$$= \sum_{t=n-L+1}^{n} \frac{\partial}{\partial \sigma} s_t y_{t+1} \tag{4.70}$$

$$= \sum_{t=n-L+1}^{n} \frac{\partial s_t}{\partial \sigma} y_{t+1} + s_t \frac{\partial y_{t+1}}{\partial \sigma} \tag{4.71}$$

$$= \sum_{t=n-L+1}^{n} s_t \frac{\partial y_{t+1}}{\partial \sigma} \tag{4.72}$$

where we let $\frac{\partial s_t}{\partial \sigma} = 0$ since the input signal s does not depend on noise level σ. Again the difficulty here is to obtain the partial derivative $\frac{\partial y_{t+1}}{\partial \sigma}$. We use the algorithm (4.41) in the previous section to simulate this term where we use the math model of the system. Future research needs to find a way to approximate it to relax the assumption of the math model.

4.3 Robust SR Learning

The performance gradients $\frac{\partial \text{SNR}}{\partial \sigma}$ and $\frac{\partial C}{\partial \sigma}$ give crucial information for adaptive systems to achieve stochastic resonance. But their estimates may be too noisy to prove useful. This section examines the statistics of the noisy gradients $\frac{\partial \text{SNR}}{\partial \sigma}$ and $\frac{\partial C}{\partial \sigma}$ and how they affect adaptive stochastic resonance.

4.3.1 Impulsiveness of the Learning Term

Much of the noisiness and complexity of the random learning law (4.4) stems from the probability structure that underlies the learning term $\frac{\partial \text{SNR}}{\partial \sigma}$. The probability density of $\frac{\partial \text{SNR}}{\partial \sigma}$ depends on the statistics of the input noise, the differential equation that defines the dynamical system, and how we define the signal and noise terms S and N. Below we test statistics of the random processes $\frac{\partial \text{SNR}}{\partial \sigma}$ and $\frac{\partial C}{\partial \sigma}$ for the quartic bistable system and the FitzHugh-Nagumo neuron model with sinusoidal inputs.

Impulsiveness of the Learning Term $\frac{\partial \text{SNR}}{\partial \sigma}$ from Math Model

The results in Figure 4.1 suggest that in some cases the density of $\frac{\partial \text{SNR}}{\partial \sigma}$ is Cauchy or otherwise belongs to the "impulsive" or thick-tailed family of symmetric alpha-stable bell curves with parameter α in the characteristic function $\exp\{-|\omega|^\alpha\}$ [18, 81, 286, 287]. The parameter α lies in $0 < \alpha \leq 2$ and gives the Gaussian random variable when $\alpha = 2$ or $\phi(\omega) = \exp\{-\omega^2\}$. It gives the thicker-tailed Cauchy bell curve when $\alpha = 1$ or $\phi(\omega) = \exp\{-|\omega|\}$. The moments of stable distributions with $\alpha < 2$ are finite only up to the order k for $k < \alpha$. The Gaussian density alone has finite variance and higher moments. Alpha-stable random variables characterize the class of normalized sums that converge in distribution to a random variable [18] as in the famous Gaussian version of the central limit theorem. The noisiness or impulsiveness of the $\frac{\partial \text{SNR}}{\partial \sigma}$-based learning grows as α falls. Note also that the ratio X/Y is a Cauchy random variable if X and Y are Gaussian [81, 249] or if they obey certain more general statistical conditions [163, 167]. The simulations found that the impulsiveness of $\frac{\partial \text{SNR}}{\partial \sigma}$ and hence of the learning process stemmed at least in part from the step size of the successive DFTs in (4.13).

We simulate the random gradient $\frac{\partial \text{SNR}_n}{\partial \sigma}$ with the partial derivatives from (4.20), (4.24), and $\frac{\partial y}{\partial \sigma}$ from (4.29):

$$\frac{\partial \text{SNR}_n}{\partial \sigma} = \frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma}. \tag{4.73}$$

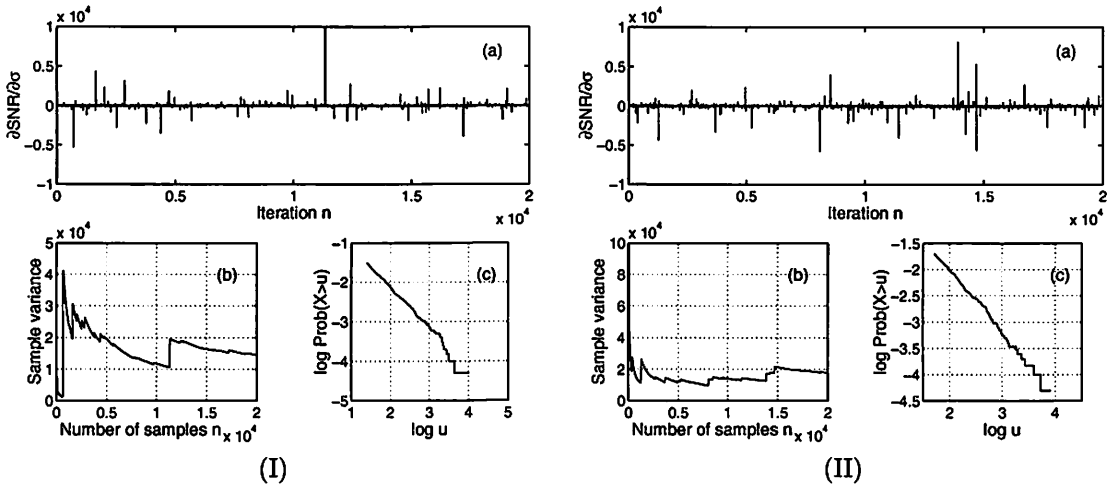


Figure 4.1: Visual display of $\frac{\partial \text{SNR}_n}{\partial \sigma} = \frac{1}{S_n}\frac{\partial S_n}{\partial \sigma} - \frac{1}{N_n}\frac{\partial N_n}{\partial \sigma}$ in (4.73) with simulation of $\frac{\partial y}{\partial \sigma}$ from math model as in (4.29) for the quartic bistable system with sinusoidal input s and Gaussian noise $n(t)$: $\dot{x} = x - x^3 + \varepsilon \sin 2\pi f + n(t)$ where $\varepsilon = 0.1$ and $f = 0.01$ Hz. The system has linear output $y(t) = x(t)$ in (I) and binary output $y(t) = \text{sgn}(x(t))$ in (II). The noise variances are the constants $\sigma_n^2 = 0.25$. (a) Cauchy-like samples of $\frac{\partial \text{SNR}_n}{\partial \sigma}$ at each iteration n. (b) Converging variance test as test of infinite variance. The sequence of sample variances converges to a finite value if the underlying probability density has finite variance. Else it has infinite variance. (c) Log-tail test of the parameter α for an alpha-stable bell curve. The test looks at the plot of log Prob$(X > u)$ versus $\log u$ for large u. If the underlying density is alpha-stable with $\alpha < 2$ then the slope of this plot is approximately $-\alpha$. This test found that $\alpha \approx 1$ and so the density was approximately Cauchy.

The simulations confirm that the random gradient $\frac{\partial \text{SNR}_n}{\partial \sigma}$ is often impulsive and can destabilize the learning process (4.44) at or near the optimal noise level. The impulsiveness of $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in Figure 4.1 suggests that $\frac{\partial \text{SNR}_n}{\partial \sigma}$ may have an alpha-stable probability density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\frac{\partial \text{SNR}_n}{\partial \sigma}$ again has an approximate Cauchy distribution.

Impulsiveness of the Learning Term $\frac{\partial C}{\partial \sigma}$ from Math Model

Consider the learning term $\frac{\partial C}{\partial \sigma}$ from (4.72)

$$\frac{\partial C_n}{\partial \sigma} = \sum_{t=n-L+1}^{n} s_t \frac{\partial y_{t+1}}{\partial \sigma} \tag{4.74}$$

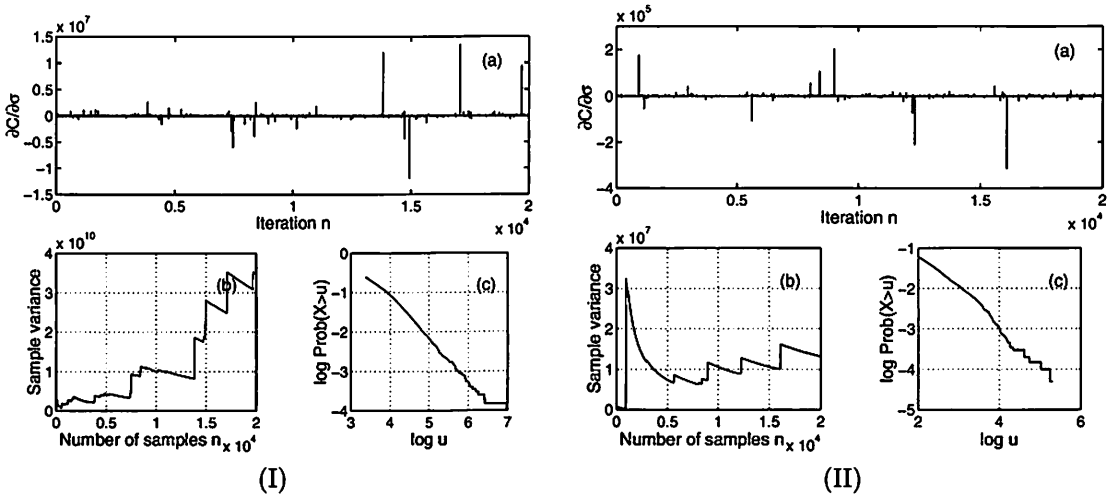


Figure 4.2: Visual display of $\frac{\partial C_n}{\partial \sigma} = \sum_{t=n-L+1}^{n} s_t \frac{\partial y_{t+1}}{\partial \sigma}$ with simulation of $\frac{\partial y}{\partial \sigma}$ in (4.29) from the system's Jacobian. (I) The quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ where $s(t) = \varepsilon \sin 2\pi f$ with $\varepsilon = 0.1$ and $f = 0.01$ Hz and linear output $y(t) = x(t)$. (II) The FHN model $\epsilon \dot{x} = -x(x^2 - \frac{1}{4}) - z + A + s(t) + n(t)$ and $\dot{z} = x - z$ with sinusoidal input signal with $\varepsilon = 0.01$ and $f_0 = 0.5$, output $y(t) = x(t)$, and parameters $\epsilon = 0.005$ and $A = -(5/12\sqrt{3} + 0.07) = -.31056$. The noise variances are the constants $\sigma_n^2 = 0.25$ for the quartic bistable system and $\sigma_n^2 = 4 \times 10^{-6}$ for the FHN model. (a) Cauchy-like samples of $\frac{\partial C_n}{\partial \sigma}$ at each iteration n. (b) Converging variance test as test of infinite variance. The sequence of sample variances converges to a finite value if the underlying probability density has finite variance. Else it has infinite variance. (c) Log-tail test of the parameter α for an alpha-stable bell curve. The test looks at the plot of log Prob$(X > u)$ versus $\log u$ for large u. If the underlying density is alpha-stable with $\alpha < 2$ then the slope of this plot is approximately $-\alpha$. This test found that $\alpha \approx 1$ and so the density was approximately Cauchy for both cases.

where we simulate the term $\frac{\partial y}{\partial \sigma}$ from (4.29). The simulations confirm that the random gradient $\frac{\partial C}{\partial \sigma}$ is impulsive and may have an alpha-stable density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\frac{\partial C}{\partial \sigma}$ in (4.72) also has approximate Cauchy distribution. Figure 4.2 shows the tests for the Quartic bistable system and the FHN model.

Impulsiveness of the Learning Term $\frac{\partial \text{SNR}}{\partial \sigma}$ from Approximation

Consider the learning term $\frac{\partial \text{SNR}}{\partial \sigma}$ from the approximation (4.53)

$$\frac{\partial \text{SNR}_n}{\partial \sigma} \approx \left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n} \right) \text{sgn}(\sigma_n - \sigma_{n-1}). \tag{4.75}$$

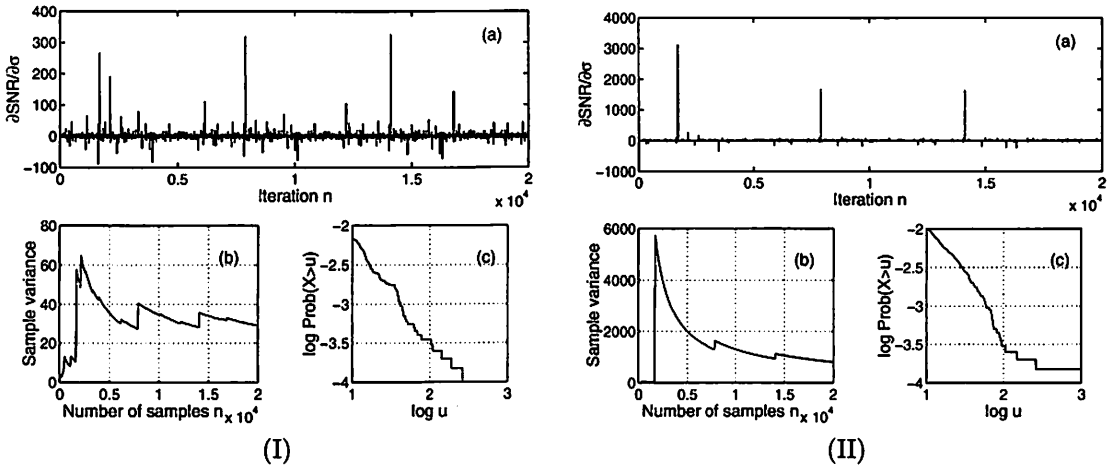


Figure 4.3: Visual display of sample statistics of approximated $\frac{\partial \text{SNR}_n}{\partial \sigma}$ for the quartic bistable system $\dot{x} = x - x^3 + s + n$ with sinusoidal input $s(t) = 0.1 \sin 2\pi(0.01)t$ and Gaussian noise $n(t)$. The system has linear output $y(t) = x(t)$ in (I) and binary output in (II). (a) Cauchy-like samples of $\frac{\partial \text{SNR}_n}{\partial \sigma}$ at each iteration n We compute $\frac{\partial \text{SNR}_n}{\partial \sigma}$ at each iteration from $\frac{\partial \text{SNR}_n}{\partial \sigma} \approx (\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n})\text{sgn}(\sigma_n - \sigma_{n-1})$ in (4.53). We vary the noise level σ_n between $\sigma_n = 0.50$ and $\sigma_n = 0.51$ so that $\text{sgn}(\sigma_n - \sigma_{n-1})$ changes values between 1 and -1. The plot shows impulsiveness of the random variable $\frac{\partial \text{SNR}_n}{\partial \sigma}$. (b) Converging variance test as test of infinite variance. The sequence of sample variances converges to a finite value if the underlying probability density has finite variance. Else it has infinite variance. (c) Log-tail test of the parameter α in for an alpha-stable bell curve. The test looks at the plot of $\log \text{Prob}(X > u)$ versus $\log u$ for large u. If the underlying density is alpha-stable with $\alpha < 2$ then the slope of this plot is approximately $-\alpha$. This test found that $\alpha \approx 1$ and so the density was approximately Cauchy. The result is that we need to apply the Cauchy noise suppressor (4.77) to the approximate SR gradient $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.53) as well as to the exact SR gradient in (4.73).

Simulations also show that this approximation of $\frac{\partial \text{SNR}}{\partial \sigma}$ is often impulsive and can destabilize the learning process (4.54) in Figure 5.3. The impulsiveness of $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in Figure 4.3 suggests that $\frac{\partial \text{SNR}_n}{\partial \sigma}$ may have an alpha-stable probability density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.53) also has an approximate Cauchy distribution.

Other Statistics at Optimality

As an aside we also form the random optimality "error" process $\mathcal{E}$:

$$\mathcal{E} = \frac{S}{N} - \frac{\partial S / \partial \sigma}{\partial N / \partial \sigma} \tag{4.76}$$

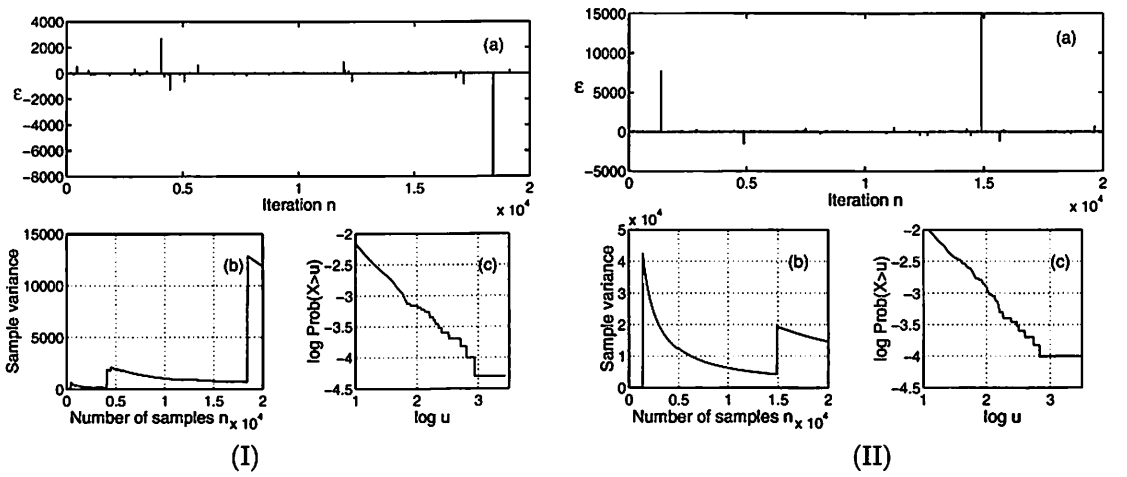
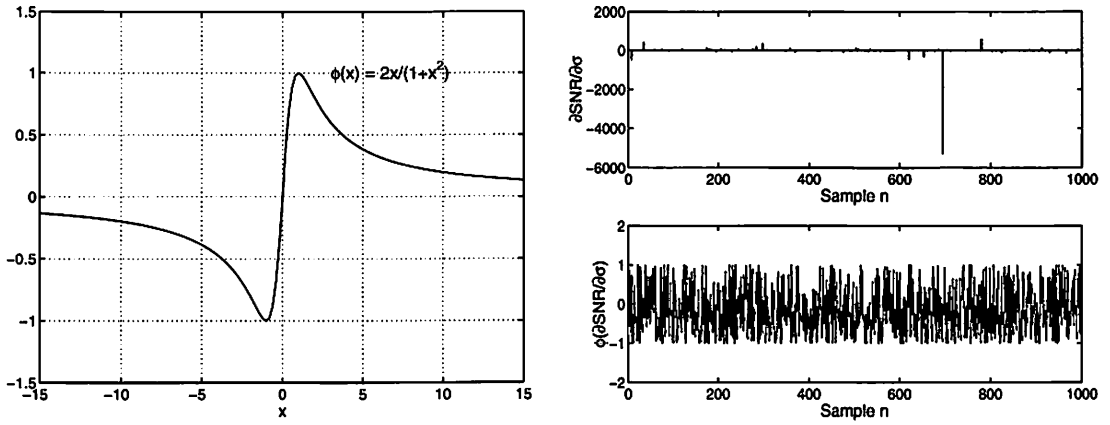


Figure 4.4: Visual display of samples from the equilibrium term $\mathcal{E}_n = \frac{S_n}{N_n} - \left(\frac{\partial S_n}{\partial \sigma} \Big/ \frac{\partial N_n}{\partial \sigma}\right)$. (a) Cauchy-like impulsive samples of $\mathcal{E}_n$ at each iteration n for the discretized version of the quartic bistable system $\dot{x} = x - x^3 + \varepsilon \sin 2\pi f_0 t + n(t)$ where $\varepsilon = 0.1$ and $f_0 = 0.01$ Hz. The system outputs are (I) $y_t = x_t$ and (II) $y_t = \text{sgn}(x_t)$. The noise intensity is the constant $\sigma_n^2 = 0.25$ that lies near the optimal level. (b) Converging variance test as a test for infinite variance. The sequence of sample variances will converge to a finite value if the underlying probability density has finite variance and diverges if it has infinite variance. (c) Log-tail test of the parameter α in an alpha-stable probability density. The test plots $\log \text{Prob}(X > u)$ versus $\log u$ for large u. If the density is alpha-stable with $\alpha < 2$ then the slope of this plot is approximately $-\alpha$. The test found $\alpha \approx 1$. So the probability density of $\mathcal{E}_n$ was approximately Cauchy.

near the optimum noise $\sigma = \sigma_{opt}$. The statistics of $\mathcal{E}_n$ change with the noise level σ^2 and with the sinewaves values ε and f_0. The empirical histogram of $\mathcal{E}_n$ is a bell curve. A key question is how thick are its tails. Figure 4.4 shows $\mathcal{E}_n$ samples from the quartic bistable system (5.3)-(5.4) with Gaussian noise $n(t) = \sigma w(t)$. The convergence of variance test [286] confirms that $\mathcal{E}_n$ had infinite variance in our simulations. The log-tail test [286] of parameter α in the family of alpha-stable probability densities leads to the estimate $\alpha \approx 1.0$. So the $\mathcal{E}_n$ density is approximately Cauchy. Recall also that $Z = X/Y$ is a Cauchy random variable if X and Y are Gaussian [81, 249] or if they obey certain more general statistical conditions [163, 167]. This suggests that much of the impulsive nature of $\mathcal{E}_n$ may stem from the ratio of derivatives in (4.76).



(a) (b)

Figure 4.5: Cauchy suppressor. The graph in (a) shows the Cauchy suppressor as a function $\phi(x) = \frac{2x}{1+x^2}$. The plots in (b) show samples of the impulsive gradients $\frac{\partial \text{SNR}_n}{\partial \sigma}$ and their Cauchy suppressed samples.

4.3.2 Cauchy Suppressor

This section introduces a Cauchy noise suppressor from the theory of robust statistics [131]. The previous section and the simulations in Chapter 5 show that the learning term $\frac{\partial \text{SNR}_n}{\partial \sigma}$ is impulsive at the optimality. This impulsiveness can destabilize the learning process of σ_n to find the optimal noise level. The theory of robust statistics [131] suggests one way to reduce the impulsiveness of $\frac{\partial \text{SNR}_n}{\partial \sigma}$. We can replace the noisy random sample z_n with a Cauchy-like noise suppressor $\phi(z_n)$ [131]:

$$\phi(z_n) = \frac{2z_n}{1+z_n^2}. \tag{4.77}$$

Figure 4.5 shows the Cauchy suppressor function and when it applies to the samples $\frac{\partial \text{SNR}_n}{\partial \sigma}$. We let $\phi(\frac{\partial \text{SNR}_n}{\partial \sigma})$ replace the noise gradient $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.73). This gives the robust SR learning law

$$\sigma(n+1) = \sigma(n) + \mu_n \phi\Big(\frac{\partial \text{SNR}_n}{\partial \sigma}\Big). \tag{4.78}$$

Simulations in chapter 5 show that the Cauchy noise suppressor can reduce the impulsiveness of the learning term and stabilize the learning processes.

4.4 Additive Fuzzy Systems and Function Approximation

We can replace an adaptive system F in (4.1) with an adaptive fuzzy function approximators [158, 159, 160, 161, 162]. The learning laws update the parameters m_j of a SAM fuzzy system. Part of this dissertation explores how to learn the SR effect with this adaptive fuzzy function approximator for the quartic bistable and other dynamical systems with sinusoidal inputs.

Adaptive fuzzy systems approximate functions with if-then rules that relate tunable fuzzy subsets of input and outputs. Each rule defines a fuzzy patch or subset of the input-output state space. The fuzzy system approximates a function as its rule patches cover the graph of the function. These systems resemble the radial-basis function networks found in neural networks [116, 214, 162]. Neural-like learning laws tune and move the fuzzy rule patches as they tune the shape of the fuzzy sets that make up the rule patches. The learning laws in the appendix use input-output data from the sampled noisy dynamical system. The rule patches move quickly to cover optimal or near-optimal regions of the function (such as its extrema). Experts can also state verbal if-then rules in some cases and add them to the fuzzy patch covering. These rules offer a simple way to endow a fuzzy approximator with prior knowledge or "hints" [1, 2] that can improve how well a fuzzy system approximates a function or how well it generalizes from training samples [238]. Fuzzy systems achieve their patch-covering approximation at the high cost of rule explosion [161, 162]. The number of rules grows exponentially with the state-space dimension of the fuzzy system. We stress that our SR learning laws can also tune non-fuzzy adaptive systems.

Adaptive fuzzy systems offer a balance between the structured and symbolic rulebased expert systems found in artificial intelligence [276] and the unstructured but numeric approximators found in modern neural networks [116, 117, 158]. These or other adaptive model-free approximators might better model the SR effect in some dynamical systems. Our first goal was to show that adaptive systems can learn to shape the input noise and perhaps shape other terms to achieve SR in the main closed-form dynamical systems that scientists have shown produce

to produce the SR effect. Our second goal was to suggest through these simulation experiments that adaptive fuzzy systems or other model-free approximators might achieve SR in more complex dynamical systems that defy easy math modeling or measurement.

The scalar SAM fuzzy system $F : R^n \rightarrow R$ can learn the SR pattern of optimum noise of an unknown dynamical system if it uses enough rules and if it samples enough data from a dynamical system that stochastically resonates. Below we derive a gradient-based learning law that tunes the SAM parameters to achieve SR from samples of system dynamics. It can also tune the parameters in other adaptive systems. We first define a practical SNR measure in terms of discrete Fourier transforms. Other SR measures can give other learning laws.

A fuzzy system $F : R^n \rightarrow R^p$ stores m rules of the word form "If $X = A_j$ Then $Y = B_j$" or the patch form $A_j \times B_j \subset X \times Y = R^n \times R^p$. The if-part fuzzy sets $A_j \subset R^n$ and then-part fuzzy sets $B_j \subset R^p$ have set functions $a_j : R^n \rightarrow [0, 1]$ and $b_j : R^p \rightarrow [0, 1]$. Generalized fuzzy sets map to intervals other than [0,1]. The scalar sinc set functions in Figure 5.12 map real inputs to "membership degrees" in the bipolar range [-0.217,1]. The system design must take care when these negative set values enter the SAM ratio in (4.80). The system can use the joint set function a_j or some factored form such as $a_j(x) = a_j^1(x_1) \cdots a_j^n(x_n)$ or $a_j(x) = \min(a_j^1(x_1), \ldots, a_j^n(x_n))$ or any other conjunctive form for input vector $x = (x_1, \ldots, x_n) \in R^n$ [158].

An additive fuzzy system [158, 159] sums the "fired" then-part sets B_j' :

$$B(x) \quad = \quad \sum_{j=1}^{m} w_j B_j' \quad = \quad \sum_{j=1}^{m} w_j a_j(x) B_j. \tag{4.79}$$

Figure 4.6(a) shows the parallel fire-and-sum structure of the standard additive model (SAM). These nonlinear systems can uniformly approximate any continuous (or bounded measurable) function f on a compact domain [159, 162]. Engineers often apply fuzzy systems to problems of control [138] but fuzzy systems can also apply to problems of communication [242] and signal processing [154] and other fields such as multimedia software agent [211, 213] in Appendix B.

Figure 4.6(b) shows how three rule patches can cover part of the graph of a scalar function

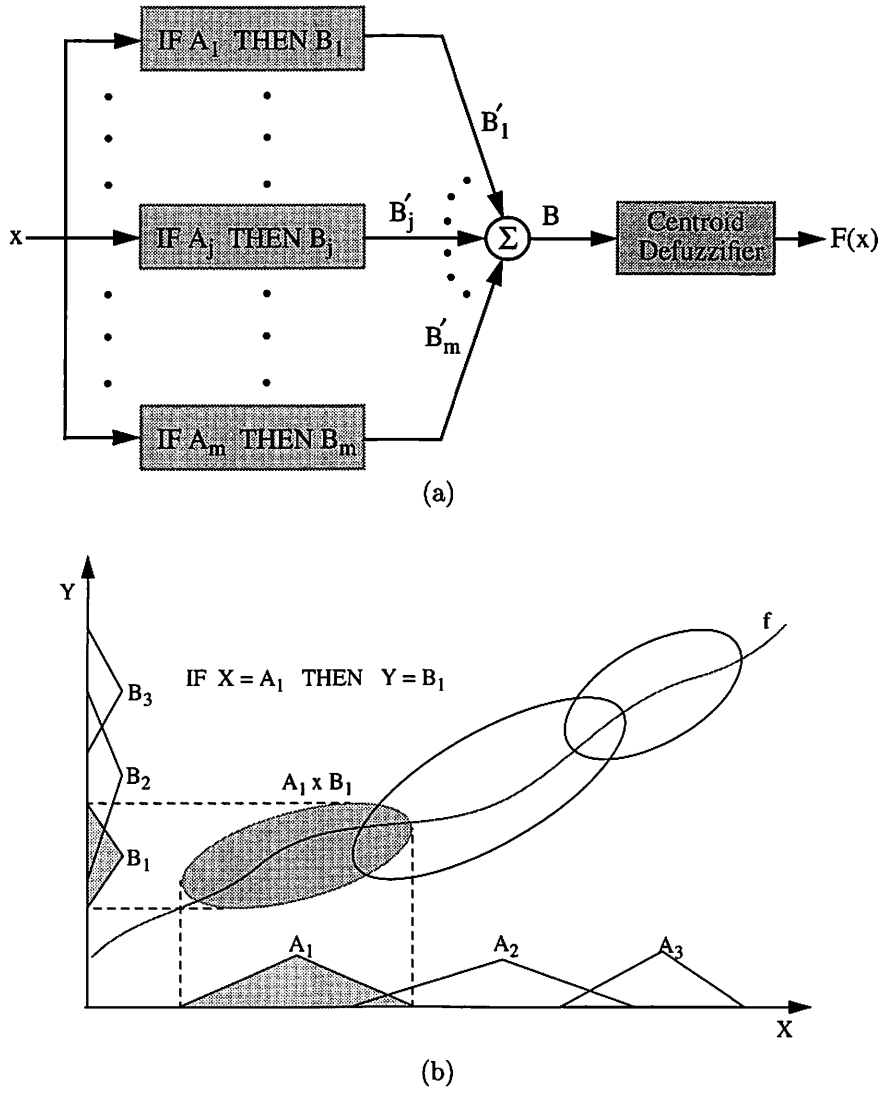


Figure 4.6: Feedforward fuzzy function approximator. (a) The parallel associative structure of the additive fuzzy system $F : R^n \to R^p$ with m rules. Each input $x_0 \in R^n$ enters the system F as a numerical vector. At the set level x_0 acts as a delta pulse $\delta(x - x_0)$ that combs the if-part fuzzy sets A_j and gives the m set values $a_j(x_0) = \int_{R^n} \delta(x - x_0) a_j(x) dx$. The set values "fire" or scale the then-part fuzzy sets B_j to give B'_j. A standard additive model (SAM) scales each B_j with $a_j(x)$. Then the system sums the B'_j sets to give the output "set" B. The system output $F(x_0)$ is the centroid of B. (b) Fuzzy rules define Cartesian rule patches $A_j \times B_j$ in the input-output space and cover the graph of the approximand f. This leads to exponential rule explosion in high dimensions. Optimal lone rules cover the extrema of the approximand as in Figure 4.7.

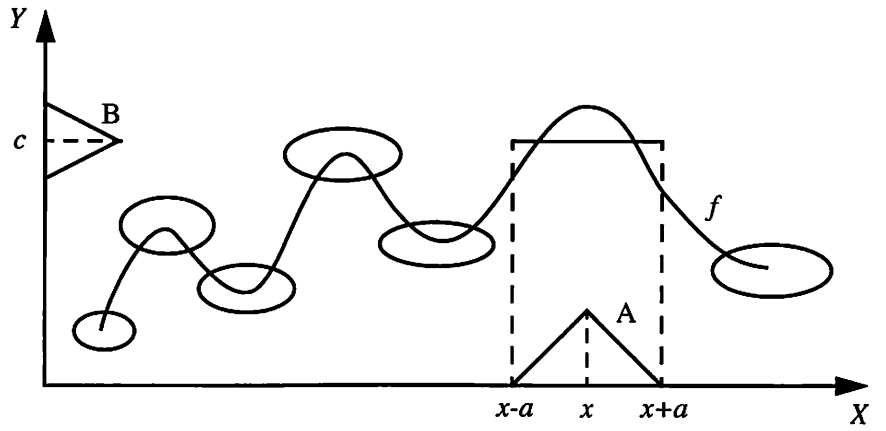


Figure 4.7: Lone optimal fuzzy rule patches cover the extrema of approximand f. A lone rule defines a flat line segment that cuts the graph of the local extremum in at least two places. The mean value theorem implies that the extremum lies between these points. This can reduce much of fuzzy function approximation to the search for zeroes $\hat{x}$ of the derivative map $f' : f'(\hat{x}) = 0$.

$f : R \to R$. The patch-cover structure implies that fuzzy systems $F : R^n \to R^p$ suffer from *rule explosion* in high dimensions. A fuzzy system F needs on the order of k^{n+p-1} rules to cover the graph and thus to approximate a vector function $f : R^n \to R^p$. Optimal rules can help deal with the exponential rule explosion. Lone or local mean-squared optimal rule patches cover the extrema of the approximand f [161, 162]. They "patch the bumps" as in Figure 4.7. Better learning schemes move rule patches to or near extrema and then fill in between extrema with extra rule patches if the rule budget allows.

The scaling choice $B'_j = a_j(x)B_j$ gives a *standard additive model* or SAM. Appendix A.1 shows that taking the centroid of $B(x)$ in (4.79) gives the following SAM ratio [158, 159, 160, 161]

$$F(x) = \frac{\sum_{j=1}^{m} w_j a_j(x) V_j c_j}{\sum_{j=1}^{m} w_j a_j(x) V_j} = \sum_{j=1}^{m} p_j(x) c_j. \tag{4.80}$$

Here V_j is the finite positive volume or area of then-part set B_j and c_j is the centroid of B_j or its center of mass. The convex weights $p_1(x), \ldots, p_m(x)$ have the form $p_j(x) = \frac{w_j a_j(x) V_j}{\sum_{i=1}^{m} w_i a_i(x) V_i}$. The convex coefficients $p_j(x)$ change with each input vector x.

Now we give a simple *local* description of optimal lone fuzzy rules [161, 162]. We move a fuzzy rule patch so that it most reduces an error. We look (locally) at a minimal fuzzy system $F : R \to R$ of just one rule. So the fuzzy system is constant in that region: $F = c$. Suppose that $f(x) \neq c$ for $x \in [a, b]$ and define the error

$$e(x) = (f(x) - F(x))^2 = (f(x) - c)^2. \tag{4.81}$$

We want to find the best place $\hat{x}$. So the first-order condition gives $\nabla e = \mathbf{0}$ or

$$0 = \frac{\partial e(x)}{\partial x} = 2(f(x) - c)\frac{\partial f(x)}{\partial x}. \tag{4.82}$$

Then $f(x) \neq c$ implies that

$$\frac{\partial e(x)}{\partial x} = 0 \iff \frac{\partial f(x)}{\partial x} = 0 \tag{4.83}$$

at $x = \hat{x}$. So the extrema of e and f coincide in this case. Figure 4.7 shows how fuzzy rule patches can "patch the bumps" and so help minimize the error of approximation.

Supervised gradient ascent changes the SAM parameters with performance data such as the SNR. Appendix A.2 derives the supervised SAM learning algorithms. The learning laws for the centroid and volume have the form

$$c_j(t+1) = c_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} p_j(x) \tag{4.84}$$

$$\text{and} \quad V_j(t+1) = V_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} \frac{p_j(x)}{V_j} [c_j - F(x)]. \tag{4.85}$$

Learning laws for set parameters depend on how we define the set functions. Figure 5.12 shows the sinc set functions [162, 209, 210] that we use in the SR simulations. The scalar sinc set

function has the form

$$a_j(x) = \frac{\sin\left(\frac{x - m_j}{d_j}\right)}{\left(\frac{x - m_j}{d_j}\right)}. \tag{4.86}$$

So to the learning laws for the parameters m_j and d_j have the form

$$m_j(t+1) = m_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \left(a_j(x) - \cos\left(\frac{x - m_j}{d_j}\right)\right) \frac{1}{x - m_j} \tag{4.87}$$

$$d_j(t+1) = d_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \left(a_j(x) - \cos\left(\frac{x - m_j}{d_j}\right)\right) \frac{1}{d_j}. \tag{4.88}$$

Like results hold for the learning laws of product n-D set functions as discussed in Appendix A.2. The learning laws update each SAM parameter to maximize the performance measure P of the SR dynamical system. This process repeats as needed for a large number of sample data pairs (x_t, y_t). Figure 4.8 shows how supervised learning moves and shapes the fuzzy rule patches to give a finer approximation as the system samples more input-output data. Figure 4.8(f) displays the absolute error of the sinc-based fuzzy function approximation.

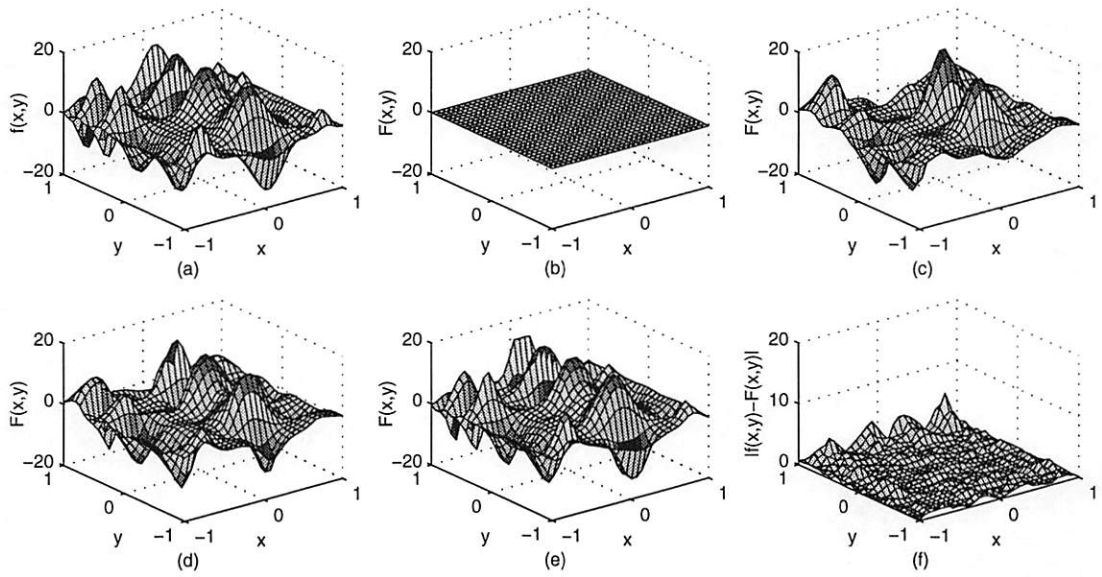


Figure 4.8: Fuzzy function approximation. 2-D Sinc standard additive model (SAM) function approximation with 100 fuzzy if-then rules and supervised gradient descent learning. (a) Desired function or approximand f. (b) SAM initial phase as a flat sheet or constant approximator F. (c) SAM approximator F after it initializes its centroids to the samples: $c_j = f(m_j)$. (d) SAM approximator F after 100 epochs of learning. (e) SAM approximator F after 6000 epochs of learning. (f) Absolute error of the fuzzy function approximation ($|f - F|$).

Chapter 5

Simulation Results

This chapter shows how the stochastic SR learning laws in Chapter 4 tend to find the optimal noise levels in many dynamical systems and for many noise distributions. This holds for both the SNR and cross-correlation measures for systems with sinusoidal inputs. The learning process updates the noise parameter σ_n at each iteration n. The learning process is noisy and may not be stable due to the impulsiveness of the random gradient $\frac{\partial \text{SNR}_n}{\partial \sigma_n}$ for SNR measure and $\frac{\partial C_n}{\partial \sigma_n}$ for cross-correlation measure C. A Cauchy noise suppressor from the theory of robust statistics [131] can help stabilize the learning process. Then sample paths of σ_n converge and wander about the optimal values if the initial values were close to the optimum.

The response of a system depends on its dynamics and on the nature of its input signals. We apply the SNR measure to the quartic bistable and other dynamical systems with sinusoidal inputs. Then we test the cross-correlation measure with sinusoidal and other broadband input signals.

Figure 5.13(a) shows how the optimum noise level varies for each sinusoidal input signal in the quartic bistable system. The learning process samples the system's input-output response as it learns the optimum noise for sinusoidal cases. An adaptive fuzzy system can encode this pattern of optimum noise in its if-then rules when gradient learning tunes its parameters. The fuzzy

system learns this optimum noise level as it varies the output of a random noise generator. More complex fuzzy systems can themselves act as adaptive random number generators [162, 242].

5.1 Adaptive Stochastic Resonance: Signal-to-Noise Ratio

This section presents simulation results for adaptive stochastic resonance with signal-to-noise ratio measure. This measure applies to narrowband sinusoidal signals. We set up a discrete computer simulation for each system as described in Chapter 3. The learning process itself does not use the system model in any calculation. It needs access only to the system's input-output responses. The learning process's sampling period T_s differs from the time step ΔT of the dynamical system's simulator in (3.4)-(3.5). The subsampling rate for the quartic bistable system is 1:50. We ignored all aliasing effects.

5.1.1 SR Test Case: The Quartic Bistable System

We test the quartic bistable system (2.7) in detail because of its wide use in the SR literature as a benchmark SR dynamical system. The quartic bistable system for $a = b = 1$ with binary output has the form [223]

$$\dot{x} = x - x^3 + s(t) + n(t) \tag{5.1}$$

$$y(t) = \text{sgn}(x(t)) \tag{5.2}$$

or $y(t) = x(t)$ in the linear-output case. The sinusoidal input forcing term is $s(t) = \varepsilon \sin \omega_0 t$. The term $n(t) = \sigma w(t)$ is a zero-mean additive white Gaussian noise with variance σ_n^2 and where $E[w] = 0$ and $E[w^2] = 1$. The discrete version has the form (3.4)-(3.5):

$$x_{t+1} = x_t + \Delta T \left(x_t - x_t^3 + \varepsilon \sin 2\pi f_0 \Delta T t \right) + \sigma \sqrt{\Delta T} w_t \tag{5.3}$$

$$y_t = \text{sgn}(x_t) \quad \text{or} \quad y_t = x_t \tag{5.4}$$

with initial condition x_0. The time step is $\Delta T = 0.0195$. The sampling period is $T_s = 0.976$ with 1:50 subsampling.

We can freely choose the time length between the iteration step n and the step $n+1$. Longer time lengths can better show how the noise intensity σ at iteration n affects S_n, N_n, and SNR_n. We chose the time length $T_{n+1} - T_n = 2000$ seconds for the simulations of the quartic bistable system. The sampling period was $T_s = 0.976$ seconds. This yields 2048 samples per iteration. This long period of time allows for low frequency signals such as $f_0 = 0.001$ Hz.

The simulations use Gaussian noise, Laplace noise, uniform noise, and impulsive alpha-stable noise. We also test the quartic bistable system with the chaotic noise from the logistic map. Figures 1.4, 3.3, and 5.6 show the output SNR for input signal $s(t) = 0.1 \sin 2\pi(0.01)t$ for Gaussian noise, Laplace noise, uniform noise, and chaotic noise from the logistic map.

SR Learning from the Math Model

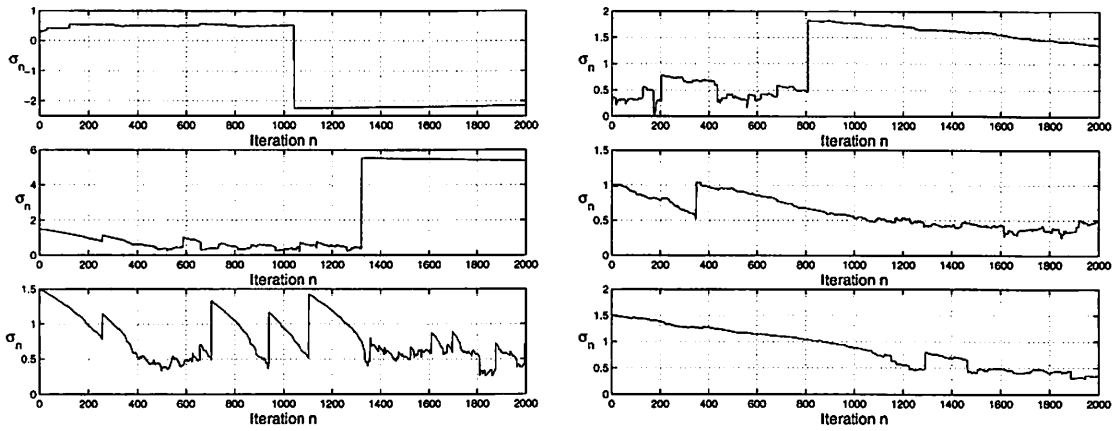
In this section tests the exact learning law (4.44) on the quartic bistable system. The Jacobian of the quartic bistable system has the form

$$\frac{\partial G}{\partial x} = \frac{\partial}{\partial x}[x - x^3 + s(t) + \sigma w(t)] \tag{5.5}$$

$$= 1 - 3x^2. \tag{5.6}$$

Then the partial derivative $\frac{\partial G}{\partial \sigma} = w(t)$ from (4.31) gives the evolution of $\eta(t) = \frac{\partial x}{\partial \sigma}$ for the quartic bistable system

$$\dot{\eta} = (1 - 3x^2)\,\eta(t) + w(t). \tag{5.7}$$



(a) (b)

Figure 5.1: Learning paths for the quartic bistable system $\dot{x} = x - x^3 + s + n$ with sinusoidal input s and Gaussian noise n. The sinusoidal input $s(t) = \varepsilon \sin 2\pi f t$ has parameters $\varepsilon = 0.1$ and $f = 0.01$. The system has linear output $y(t) = x(t)$ in (a) and binary output $y(t) = \text{sgn}(x(t))$ in (b). The learning law takes the form (4.44). The optimal noise level is $\sigma \approx 0.5$ for both cases. The impulsiveness of the learning term $\frac{\partial \text{SNR}}{\partial \sigma}$ destabilizes the learning process near the optimal noise level.

Its discrete version has the form

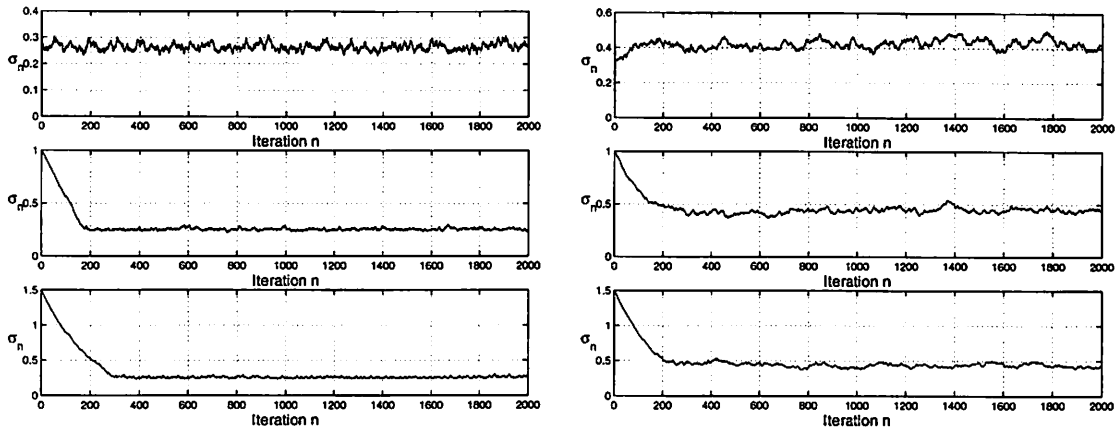
$$\eta(t+1) \quad = \quad \eta(t) + \Delta T \left(1 - 3x_t^2\right) \eta(t) + \sqrt{\Delta T} w_t. \tag{5.8}$$

We used the initial condition $\frac{\partial x_0}{\partial \sigma} = 0$ in simulations. Then we get $\frac{\partial y}{\partial \sigma}$ from (4.27) for use in the learning law (4.44). The linear output $y = g(x) = x$ has $\frac{\partial g}{\partial x} = 1$. We can approximate a binary output as $g(x) = \text{sgn}(x) \approx \tanh(cx)$ for a large positive $c \gg 0$. Then $\frac{\partial g}{\partial x} = c(1 - \tanh^2(cx))$.

We test the learning law (4.44)

$$\sigma(n+1) \quad = \quad \sigma(n) + \mu_n \Big(\frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} \Big). \tag{5.9}$$

Figure 5.1 shows the simulation results. It displays the unstability in the learning due to the impulsiveness of the random gradient $\frac{\partial \text{SNR}}{\partial \sigma}$ as shown in Figure 4.1. So we replace the noisy random sample $\frac{\partial \text{SNR}_n}{\partial \sigma}$ with a Cauchy-like noise suppressor $\phi(\frac{\partial \text{SNR}_n}{\partial \sigma})$ [131] as described in



(a) (b)

Figure 5.2: Learning paths for the quartic bistable system $\dot{x} = x - x^3 + s + n$ with sinusoidal input $s(t) = \varepsilon \sin 2\pi f t$ and Gaussian noise n. The system has linear output $y(t) = x(t)$. The learning law has the form (5.10). The parameters of input sine waves are (a) $\varepsilon = 0.1$ and $f = 0.001$ and (b) (a) $\varepsilon = 0.1$ and $f = 0.01$. Optimal noise levels are (a) $\sigma \approx 0.35$ and (b) $\sigma \approx 0.5$. The learning paths converge close to the optimal levels.

Section 4.3.2:

$$\sigma(n+1) \;=\; \sigma(n) + \mu_n \, \phi\Big(\frac{\partial \mathrm{SNR}_n}{\partial \sigma}\Big). \tag{5.10}$$

Figure 5.2 shows the results of the SR learning law (5.10) with the gradient in (4.73). The σ_n learning paths in (4.44) converge near the optimal noise level.

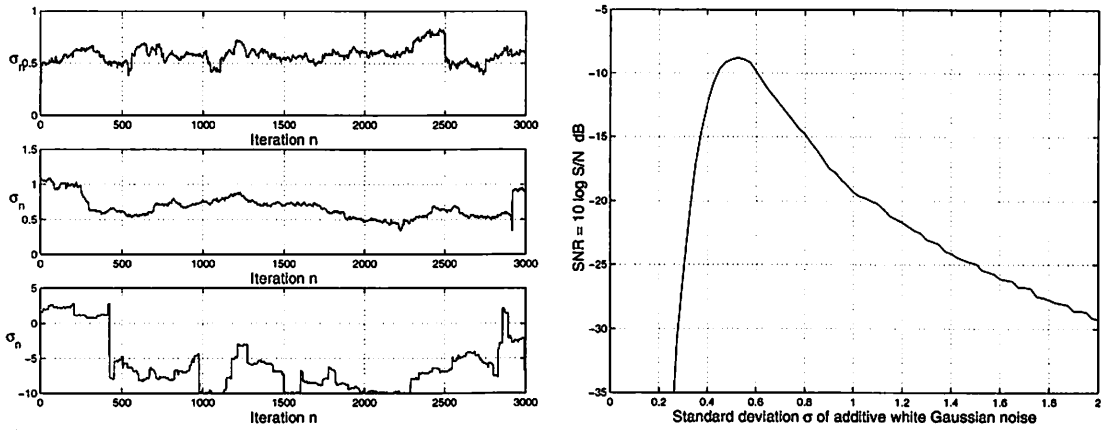
Learning Law with Approximation of $\frac{\partial \mathrm{SNR}_n}{\partial \sigma}$

This section tests the learning law with approximation of the gradient term as in section 4.1.2:

$$\sigma_{n+1} \;=\; \sigma_n + \mu_n \Big(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n} \Big) \operatorname{sgn}(\sigma_n - \sigma_{n-1}). \tag{5.11}$$

This learning law does not use the math model of the system. It depends only on samples from the system dynamics and from the input signal $s(t)$.

Figure 5.3(a) shows sample learning paths of σ_n for the quartic bistable system and approximation (4.53). Figure 5.3(b) shows the noise-SNR profile of the dynamical system. The σ_n learning paths converge to the optimum noise values only in some cases. The chance of path



(a) (b)

Figure 5.3: Impulsive effects on learning paths of noise intensity σ_n. The quartic bistable system has the form $\dot{x} = x - x^3 + s(t) + n(t)$ with binary output $y(t) = \text{sgn}(x(t))$ and initial condition $x(0) = -1$. The input sinusoid signal function is $s(t) = 0.1 \sin 2\pi(0.01)t$. (a) The sequence σ_n with different initial values that differ from the optimum noise intensity. (b) Noise-SNR profile of the quartic bistable system. The graph shows that the optimum noise intensity lies near $\sigma = 0.5$. The paths of σ_n do not converge to the optimum noise. This stems from the impulsiveness of the derivative term $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in the approximate SR learning law (5.11).

convergence is higher for larger amplitudes of sinusoidal inputs. The paths do not converge as often for small amplitudes. The simulations confirm that the random gradient $\frac{\partial \text{SNR}_n}{\partial \sigma_n}$ in (4.53) is often impulsive and can destabilize the learning process (5.11) as in Figure 5.3. The impulsiveness of $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in Figure 4.3 suggests that $\frac{\partial \text{SNR}_n}{\partial \sigma}$ may have an alpha-stable probability density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.53) also has an approximate Cauchy distribution.

We again apply the Cauchy-like noise suppressor from robust statistics [131] to reduce the impulsiveness of the approximated term $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.53). So $\phi(\frac{\partial \text{SNR}_n}{\partial \sigma})$ replaces the approximation of the noise gradient $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in (4.53) to give the robust SR learning law

$$\sigma_{n+1} = \sigma_n + \mu_n \phi\left(\left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n}\right) \text{sgn}(\sigma_n - \sigma_{n-1})\right). \tag{5.12}$$

Figure 5.4 shows the results of the SR learning law (5.12). The σ_n learning paths converge to

the optimum noise level if the initial value lies close enough to it. Then σ_n wanders in a small Brownian-like motion about the optimum noise level.

Like results hold for other noise densities with finite variance such as Laplace and uniform noise. Figure 5.5 shows σ_n learning paths for the quartic bistable system (5.3)-(5.4) with Laplace noise and uniform noise. We also use a chaotic time series as the forcing noise n_t in the quartic bistable dynamical system [137]. The simple and popular logistic map creates the noise sequence $\{z_t\}$:

$$z_{t+1} = 4\,z_t\,(1 - z_t) \tag{5.13}$$

from the initial value $z_0 = .123456789$ [137]. The positive sequence $\{z_t\}$ stays bounded within the unit interval: $z_t \in (0, 1)$. The chaotic noise n_t comes from

$$n_t = A\,(z_t - \frac{1}{2}). \tag{5.14}$$

The factor $A > 0$ acts as the scaled power or standard deviation if the term $(z_t - \frac{1}{2})$ is a zero-mean random variable with unit variance. Learning tunes A so that the dynamical system shows the SR effect. Figure 5.6 shows a sample chaotic noise sequence and shows two A learning paths on their way to stochastic convergence.

We also test the quartic bistable system with alpha-stable noise. Figure 5.7 shows the paths of the optimal dispersion γ_n for $\alpha = 1.9$, 1.8, and 1. The learning degrades as α falls and the alpha-stable bell curves have thicker tails.

5.1.2 Other SR Systems

The SR learning schemes also work for other SR models. We here show only the results for zero-mean white Gaussian noise.

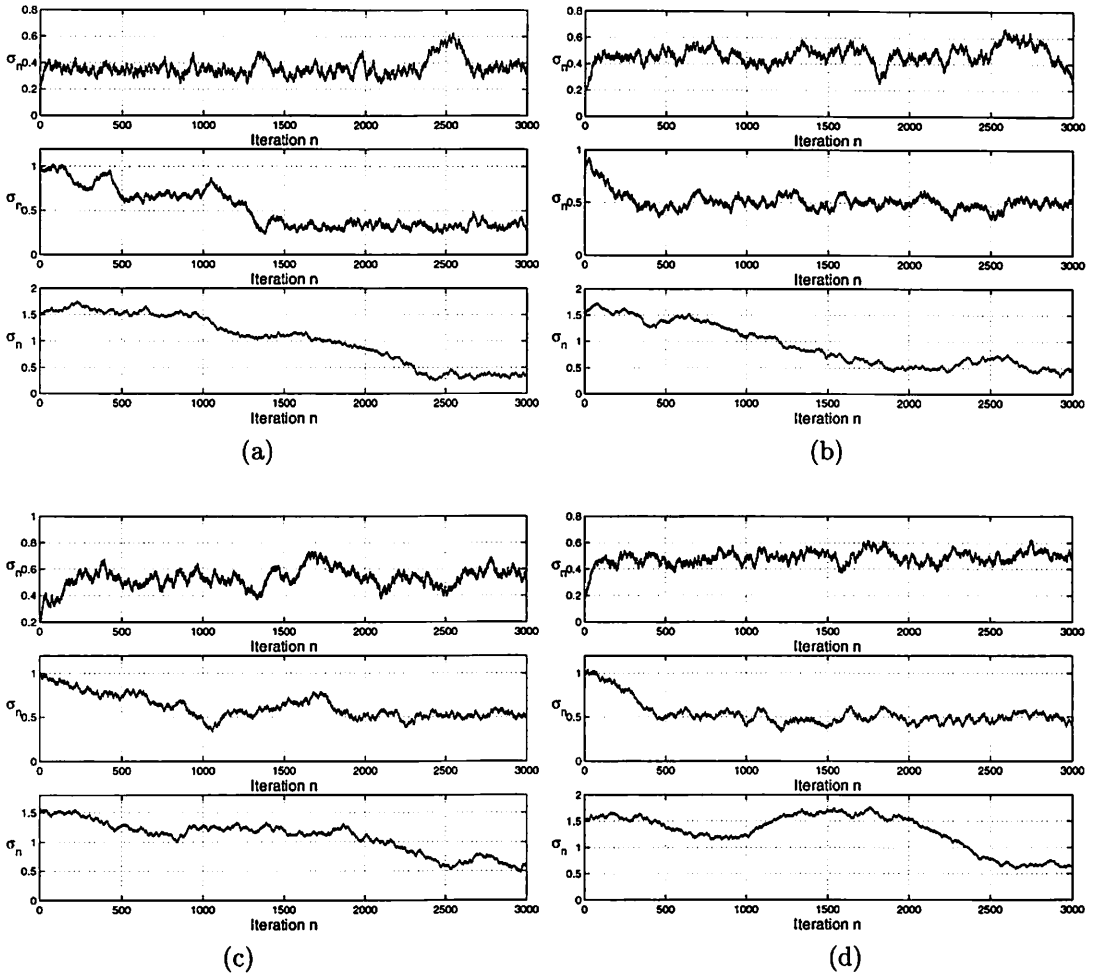


Figure 5.4: Learning paths of σ_n with the Cauchy noise suppressor $\phi(z) = 2z/(1+z^2)$ for the quartic bistable system with binary threshold output $y_t = \text{sgn}(x_t)$. The term $\phi(\frac{\partial \text{SNR}_n}{\partial \sigma})$ replaces $\frac{\partial \text{SNR}_n}{\partial \sigma}$ in the SR learning law (4.50). The paths of σ_n wander in a Brownian-like motion around the optimum noise. The suppressor function ϕ makes the learning algorithm more robust against impulsive shocks. The input signals are (a) $s(t) = 0.1 \sin 2\pi(0.001)t$, (b) $s(t) = 0.1 \sin 2\pi(0.005)t$, (c) $s(t) = 0.1 \sin 2\pi(0.01)t$, and (d) $s(t) = 0.2 \sin 2\pi(0.01)t$.

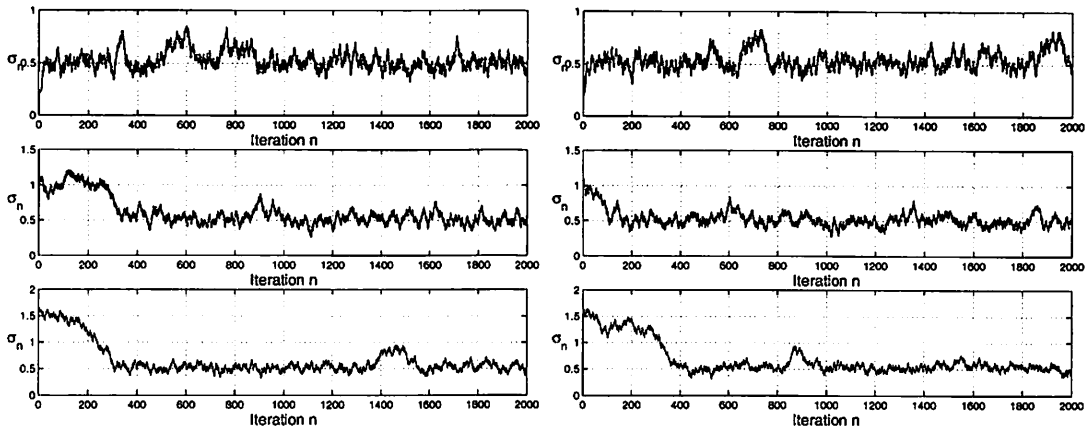


Figure 5.5: Learning paths of σ_n for other noise densities in the quartic bistable system with binary output $y_t = \text{sgn}(x_t)$. The input signal is $s(t) = 0.1 \sin 2\pi(0.01)t$. The optimal noise lies near $\sigma = 0.5$ for both cases of (a) Laplace noise and (b) uniform noise.

Threshold neuron model

We first test the discrete-time threshold neuron model

$$y_t = \begin{cases} 1 & \text{if } s_t + n_t \geq \Theta \\ -1 & \text{if } s_t + n_t < \Theta \end{cases} \tag{5.15}$$

for $t = 0, 1, 2, \ldots$ The threshold Θ sets the output of the neuron. The sinusoidal input has the form $s_t = \varepsilon \sin 2\pi f_0 \Delta T t$. The Gaussian noise n_t has variance σ_n^2. The threshold system is not a dynamical system but it does show SR. Figure 5.8 shows the result of learning when $f_0 = 0.001$, $\varepsilon = 0.1$, and $\Theta = 0.5$ and when $f_0 = 0.001$, $\varepsilon = 0.5$, and $\Theta = 1$. The sampling period is $T_s = \Delta T = 1$.

Bistable potential neuron model

We next test the bistable potential neuron model with Gaussian white noise [28]

$$\dot{x} = -x + 2 \tanh x + s(t) + n(t) \tag{5.16}$$

$$y(t) = \text{sgn}(x(t)). \tag{5.17}$$

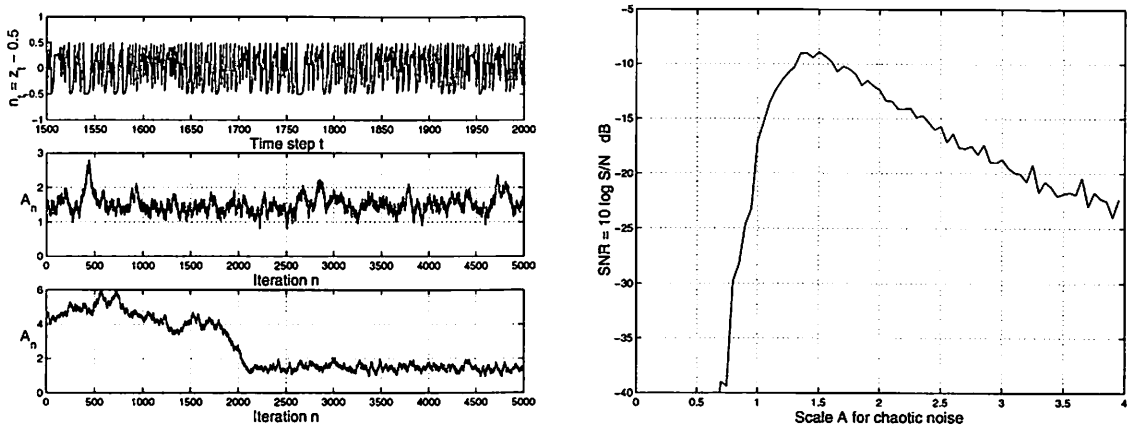


Figure 5.6: Learning paths of the scaling factor A_n in chaotic noise $n_t = A_n(z_t - \frac{1}{2})$ from the logistic dynamical system $z_{t+1} = 4z_t(1 - z_t)$. The dynamical system is the quartic bistable system with binary output $y_t = \text{sgn}(x_t)$. The input signal is $s(t) = \varepsilon \sin 2\pi f_0 t$ where $f_0 = 0.01$ Hz and $\varepsilon = 0.1$. The top figure shows a sample noise path n_t from the chaotic logistic map when $A_n = 1$.

We ignore the multiplicative noise in (2.11). Figure 5.9 shows the SR learning paths of σ_n. The sinusoidal input is $s(t) = \varepsilon \sin 2\pi f_0 t$ where $f_0 = 0.01$ Hz and $\varepsilon = 0.1$ and the $\varepsilon = 0.3$. The time step in the discrete simulation is $\Delta T = 0.0195$. The sampling period is $T_s = 0.975$ or 50 times the time step ΔT.

FitzHugh-Nagumo neuron model

We next test the forced FitzHugh-Nagumo neuron model [221]. We rewrite (2.12)-(2.13) with $a = 0.5$ and with the changes of variables $x \to x + 0.5$, $z \to z - b + 0.5$, and $A \to A - b + 0.5$ [52]:

$$\epsilon \dot{x} = -x(x^2 - \frac{1}{4}) - z + A + s(t) + n(t) \tag{5.18}$$

$$\dot{z} = x - z \tag{5.19}$$

$$y(t) = x(t). \tag{5.20}$$

The constants are $\epsilon = 0.005$, $a = 0.5$, and $A = -(5/12\sqrt{3} + 0.07) = -.31056$ as in [118]. The sinusoidal input is $s(t) = \varepsilon \sin 2\pi f_0 t$ with $\varepsilon = 0.01$, $f_0 = 0.1$ and 0.5 Hertz. The sampling period

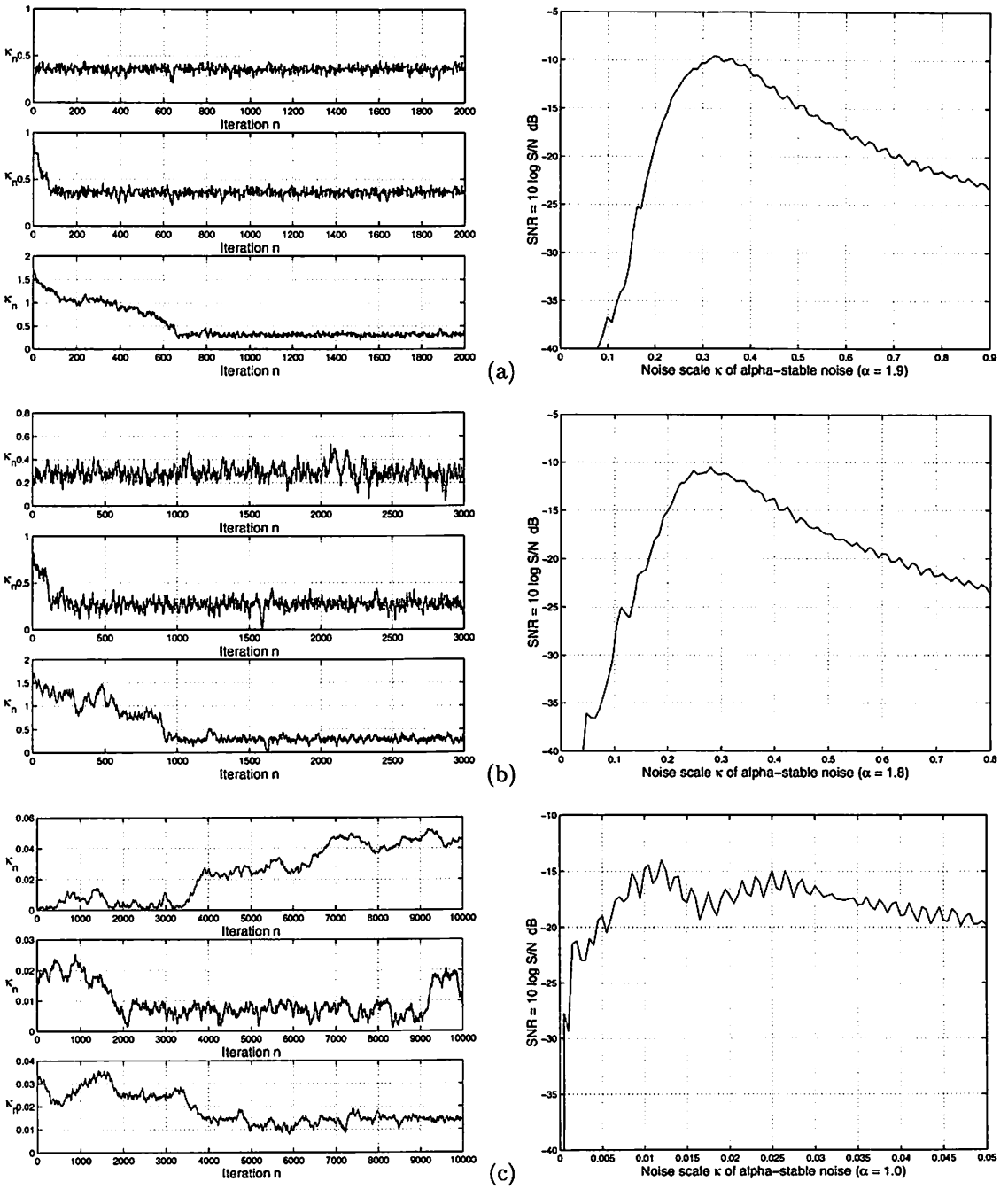


Figure 5.7: Learning paths of κ_n for alpha-stable noise in the quartic bistable system with binary output $y_t = \text{sgn}(x_t)$. The input signal is $s(t) = 0.1 \sin 2\pi(0.01)t$. (a) $\alpha = 1.9$. (b) $\alpha = 1.8$. (c) $\alpha = 1$. The noise scale κ acts like a standard deviation and controls the width of the alpha-stable bell curve through the dispersion $\gamma = \kappa^\alpha$. Learning becomes more difficult as α falls and the bell curves have thicker tails. The impulsiveness is so severe in the Cauchy case (c) that κ_n often fails to converge. Note the noisy multimodal nature of the SNR profiles.

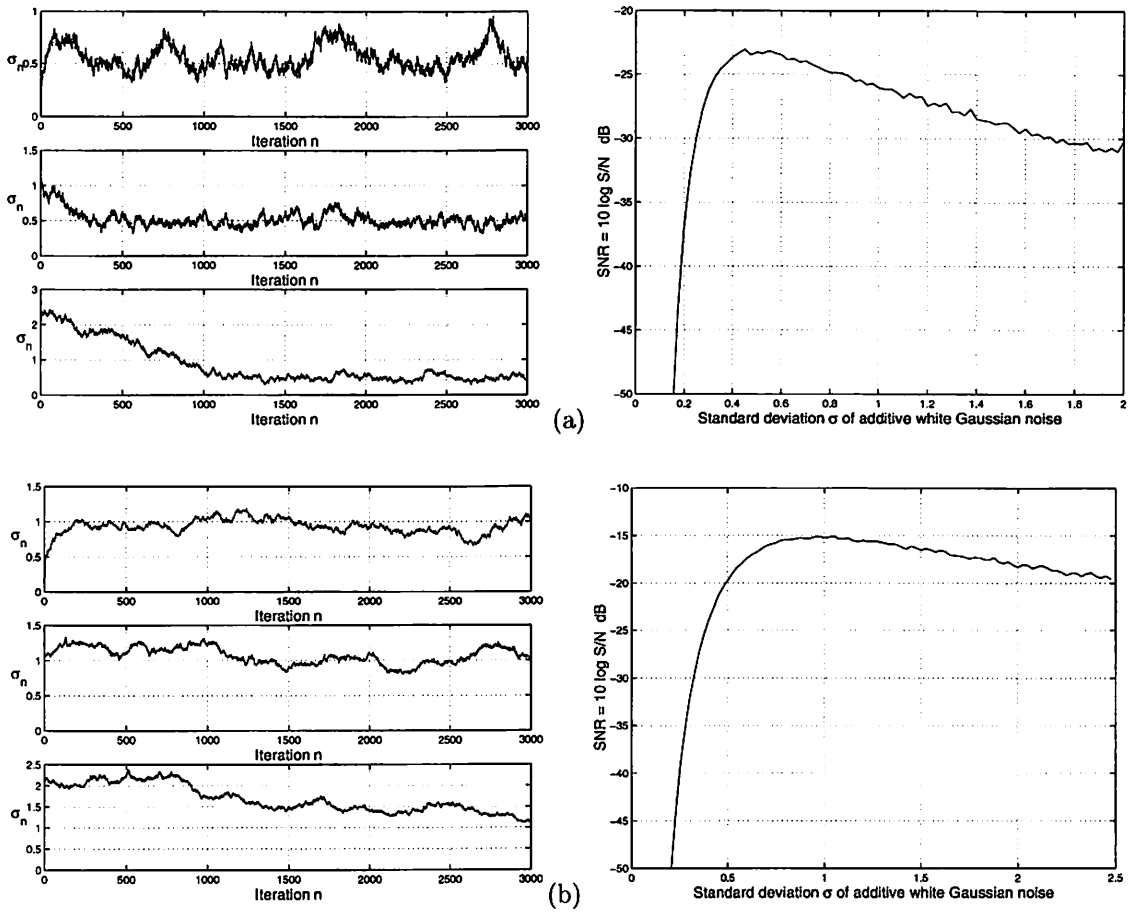


Figure 5.8: SR learning paths of σ_n for the threshold system $y_t = \text{sgn}(s_t + n_t - \Theta)$ where $\text{sgn}(x) = 1$ if $x \geq 0$ and $\text{sgn}(x) = -1$ if $x < 0$. The sinusoidal input is $s_t = \varepsilon \sin 2\pi f_0 t$ with additive white Gaussian noise sequence n_t. The parameters are (a) $f_0 = 0.001$, $\varepsilon = 0.1$, and $\Theta = 0.5$ and (b) $f_0 = 0.001$, $\varepsilon = 0.5$, and $\Theta = 1$.

is $T_s = 0.01$ with $\Delta T = 0.001$.

The Jacobian of the FHN model (5.18)-(5.20) has the form

$$J(x,z) = \begin{bmatrix} -\frac{3}{\epsilon}x^2 + \frac{1}{4\epsilon} & -\frac{1}{\epsilon} \\ 1 & -1 \end{bmatrix} \tag{5.21}$$

Figure 5.10 shows the learning paths of the standard deviation σ_n of the Gaussian noise n.

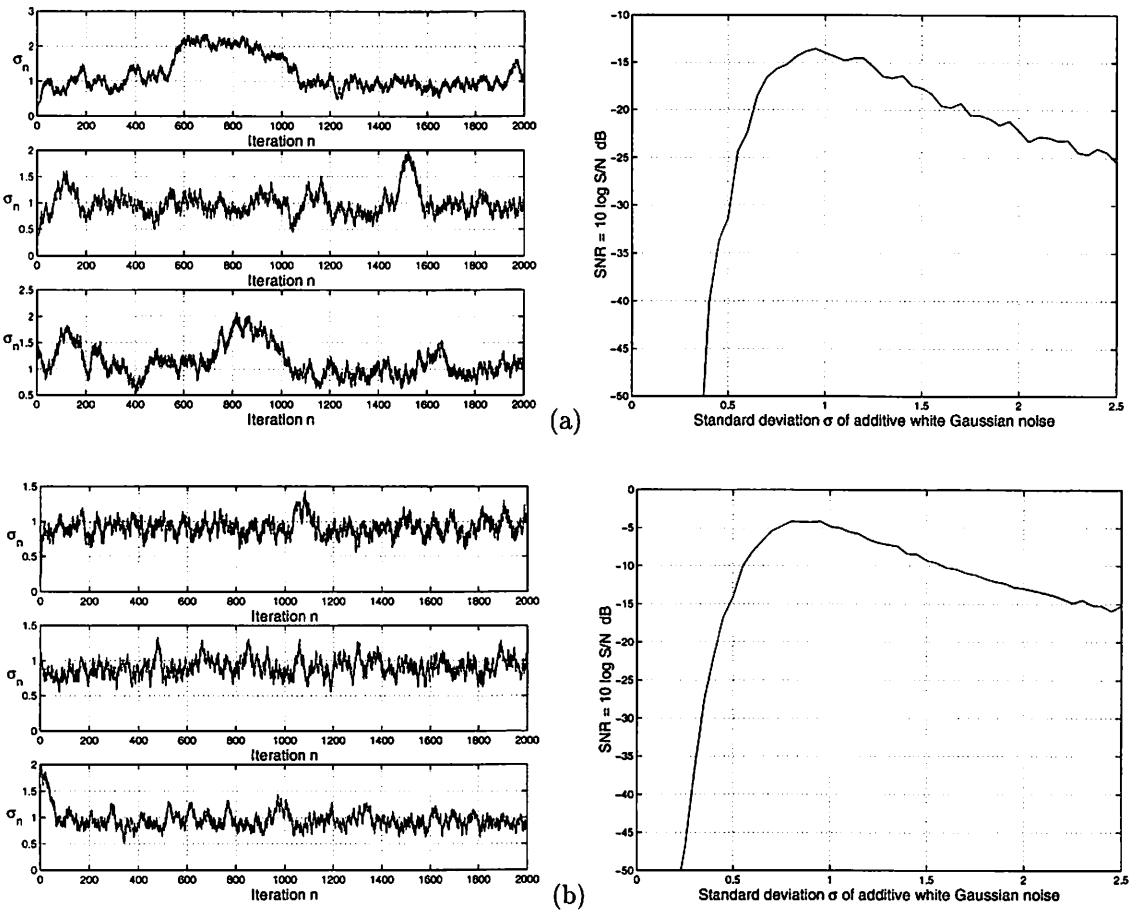


Figure 5.9: SR learning paths of σ_n for the forced bistable neuron model $\dot{x} = -x + 2\tanh x + \varepsilon \sin 2\pi f_0 t + n(t)$ with binary output $y(t) = \text{sgn}(x(t))$. The parameters of the sinusoidal inputs are $f_0 = 0.01$ Hz and (a) $\varepsilon = 0.1$ and (b) $\varepsilon = 0.3$.

Duffing oscillator

We also show SR learning in the forced Duffing oscillator with Gaussian white noise n [237]:

$$\ddot{x} = -0.15\dot{x} + x - x^3 + \varepsilon \sin(\omega_0 t) + n(t) \tag{5.22}$$

$$y(t) = x(t). \tag{5.23}$$

Figure 5.11 shows the learning paths of σ_n for sinusoidal input s with frequency $f_0 = 0.01$ Hz and with amplitudes $\varepsilon = 0.1$ and $\varepsilon = 0.3$. The sampling period is $T_s = 0.02$ with $\Delta T = 0.005$.

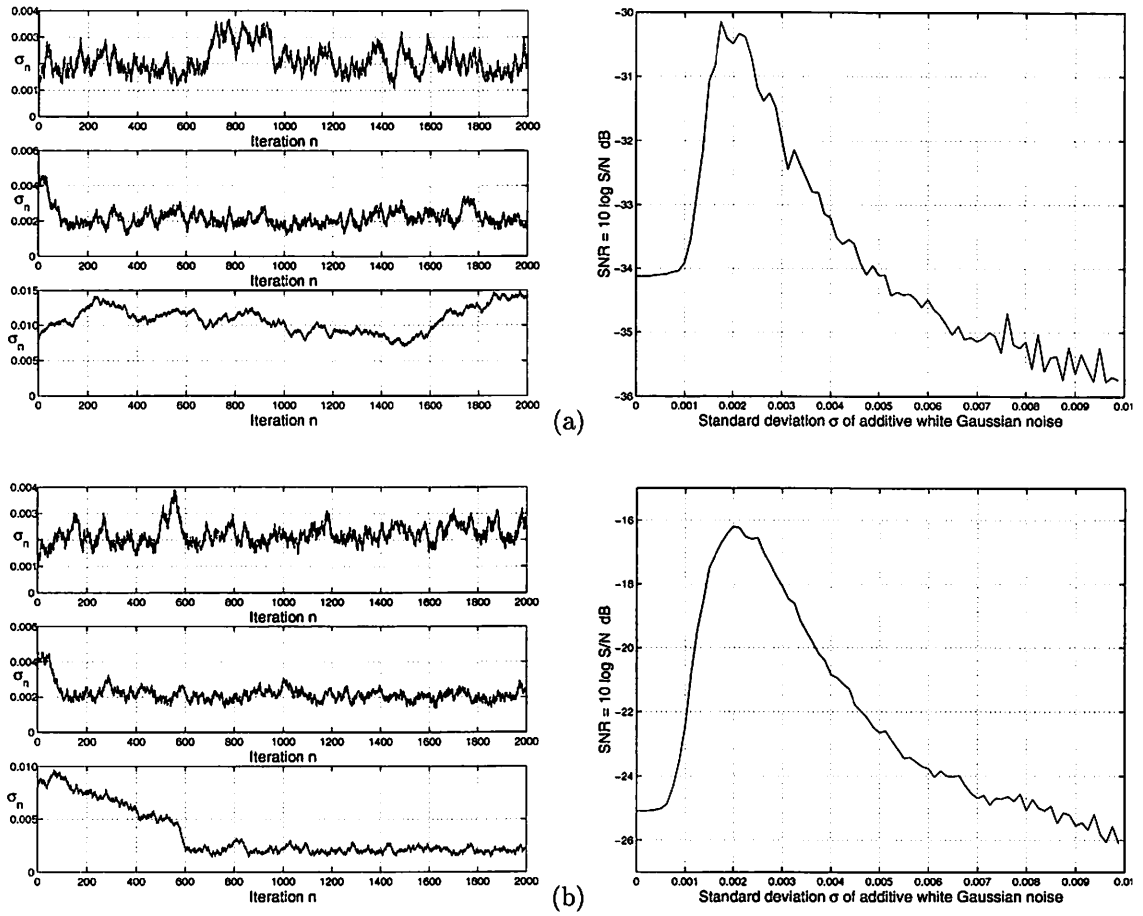


Figure 5.10: SR learning paths of σ_n for the FitzHugh-Nagumo neuron model $\epsilon\dot{x} = -x(x^2 - \frac{1}{4}) - z + A + s(t) + n(t)$ and $\dot{z} = x - z$ with output $y(t) = x(t)$. The parameters are $\epsilon = 0.005$ and $A = -(5/12\sqrt{3} + 0.07) = -.31056$. The sinusoidal input signal is $s(t) = \varepsilon \sin 2\pi f_0 t$ where (a) $\varepsilon = 0.01$ and $f_0 = 0.1$ Hz and (b) $\varepsilon = 0.01$ and $f_0 = 0.5$ Hz. Figures (a) and (b) show how SR learning convergence can depend on initial conditions. The distant starting point $\sigma_0 > 7.5 \times 10^{-3}$ leads to divergence in the third learning sample in (a) but leads to convergence in the third learning sample in (b).

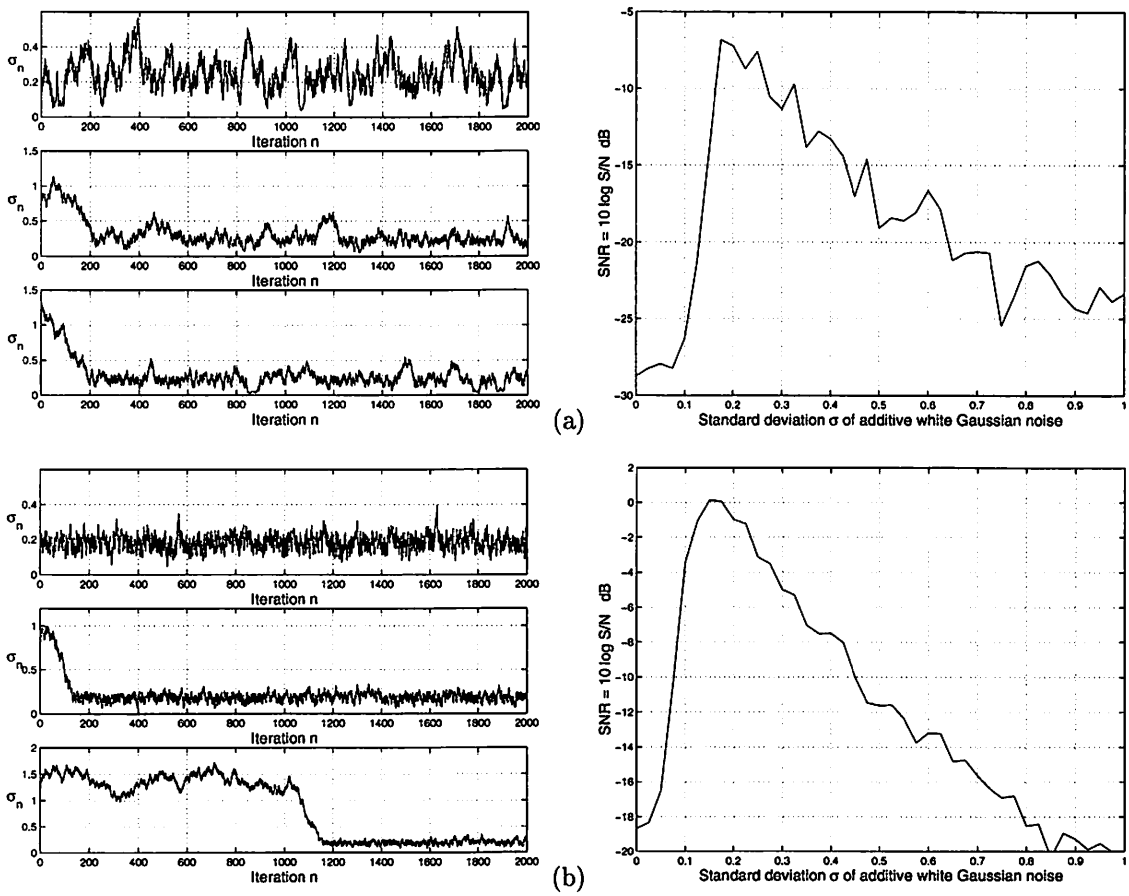


Figure 5.11: SR learning paths of σ_n for the forced Duffing oscillator $\ddot{x} = -\delta\dot{x} + x - x^3 + \varepsilon \sin 2\pi f_0 t + n(t)$ with output $y(t) = x(t)$ and $\delta = 0.15$. The parameters of the sinusoidal inputs are $f_0 = 0.01$ Hz and (a) $\varepsilon = 0.1$ and (b) $\varepsilon = 0.3$.

5.1.3 Fuzzy SR Learning: The Quartic Bistable System

We use a fuzzy function approximator $F : R^n \rightarrow R$ to learn and store the entire surface of optimal noise values for the quartic bistable system with sinusoidal inputs. The fuzzy system had as its input the 2-D vector of sine wave amplitude ε and frequency f_0. We test the system with the fixed input initial value $x(0) = -1$. The fuzzy system itself defines a vector function $F : R^2 \rightarrow R$ and uses 200 rules. The chain rule extends the learning laws in the previous sections to tune the fuzzy system's parameters m_j as in (4.4):

$$m_j(n+1) = m_j(n) + \mu_n \frac{\partial \text{SNR}_n}{\partial m_j} \tag{5.24}$$

$$= m_j(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} \frac{\partial F}{\partial m_j}. \tag{5.25}$$

Appendix A.2 derives the partial derivative $\frac{\partial F}{\partial m_j}$ for the sinc SAM fuzzy system that we use. The Cauchy noise suppressor gives the learning law as

$$m_j(n+1) = m_j(n) + \mu_n \phi\Big(\frac{\partial \mathrm{SNR}_n}{\partial \sigma}\Big) \frac{\partial F}{\partial m_j}. \tag{5.26}$$

Figure 5.12 shows how we form a first set of rules on the product space of the two variables ϵ and f_0. It also shows how the learning laws move and shape the width of the if-part sinc set. Figure 5.13 shows the results of SAM learning of the optimal noise pattern for the quartic bistable system. The sinc SAM uses 200 rules. Fewer rules give a coarser approximation.

5.2 Adaptive Stochastic Resonance: Cross-Correlation Measure

This section explores adaptive stochastic resonance for a quartic bistable system and a FitzHugh-Nagumo neuron model with the cross-correlation measure C in (3.52). The cross-correlation measure can apply to both narrowband and broadband signals. Signal-to-noise ratio works well with the sinusoidal signal because it computes the spectral ratio of output energy at a given frequency ω_0. This concept might not extend to broadband signals in nonlinear systems since their energy can spread over a broad range of frequencies. Cross-correlation measure and correlation coefficients define how well the input and output matches in shapes.

This section explores ASR for systems with broadband signals. Most real-world systems deal with broadband signals rather than sinusoidal signals because broadband signals contain more information [50]. Examples of broadband SR systems include crickets [173], rat [53], and human [48, 54, 57, 271]. Figure 5.14 shows samples of broadband signals. We create broadband signals by convolving a sequence of Gaussian noise with windows in signal processing such as

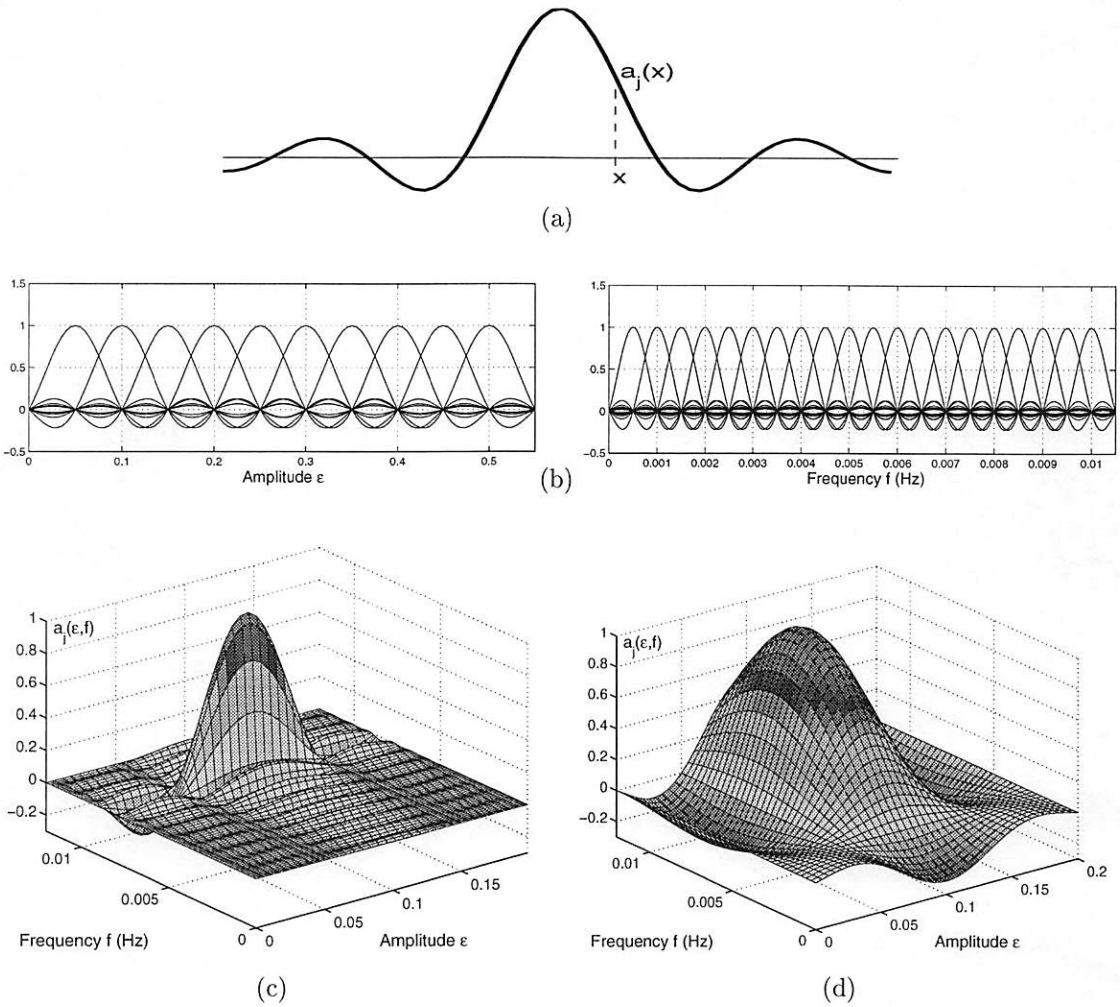
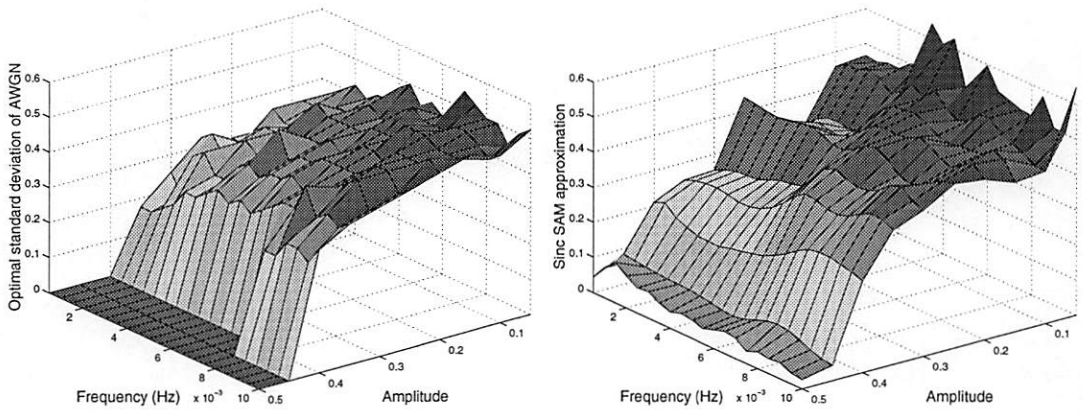


Figure 5.12: If-part sinc fuzzy sets. (a) Scalar sinc set function $a_j(x) = \sin x/x$. Sinc sets are generalized fuzzy sets with "membership values" in [-.217,1]. Element x belongs to set A_j to degree $a_j(x)$: Degree($x \in A_j$) $= a_j(x)$. (b) Initial subsets for sine wave amplitudes and frequencies. There are 10 fuzzy sets for amplitude ε and 20 fuzzy sets for frequency f_0. The product of two 1-D sets gives the 2-D joint sets: $a_j(x) = a_j(\varepsilon, f_0) = a_j^1(\varepsilon)\, a_j^2(f_0)$. So the product space gives $10 \times 20 = 200$ if-part sets in the if-then rules. (c) One of the 2-D if-part sinc sets in the 200 rules at the initial location. (d) Learning laws tune the location and width of the same set in (c) after 30 epochs of learning.



(a) (b)

Figure 5.13: Optimal noise levels in terms of the signal-to-noise ratio for the quartic bistable system with binary output. (a) The optimum noise pattern when inputs are sine waves with distinct amplitudes and frequencies. (b) SAM fuzzy approximation of the optimum noise after 30 epochs. The sinc SAM uses 200 rules. One epoch uses 20 iterations that trains on 200 input amplitudes and frequencies. The quartic bistable system has the form $\dot{x} = x - x^3 + s(t) + n(t)$ with initial condition $x(0) = -1$. The initialized SAM gave the output value 0.2 as its first estimate of the optimal noise level.

Hanning window [240]. This limits the range of the signals's spectrum. We also include non-sinusoid periodic signals in this category because they can have broad spectrum. We test the ASR learning schemes for both types of input forcing signals and Gaussian noise.

The learning law for optimal noise level σ for the cross-correlation measure has the form (4.68)

$$\sigma_{n+1} = \sigma_n + \mu \frac{\partial C_n}{\partial \sigma} \tag{5.27}$$

with the learning term (4.72)

$$\frac{\partial C_n}{\partial \sigma} = \sum_{t=n-L+1}^{n} s_t \frac{\partial y_{t+1}}{\partial \sigma}. \tag{5.28}$$

The learning scheme uses L input-output samples to compute the performance gradient $\frac{\partial C_n}{\partial \sigma}$ and update σ_n at each iteration n. This follows the procedure in the previous section for the

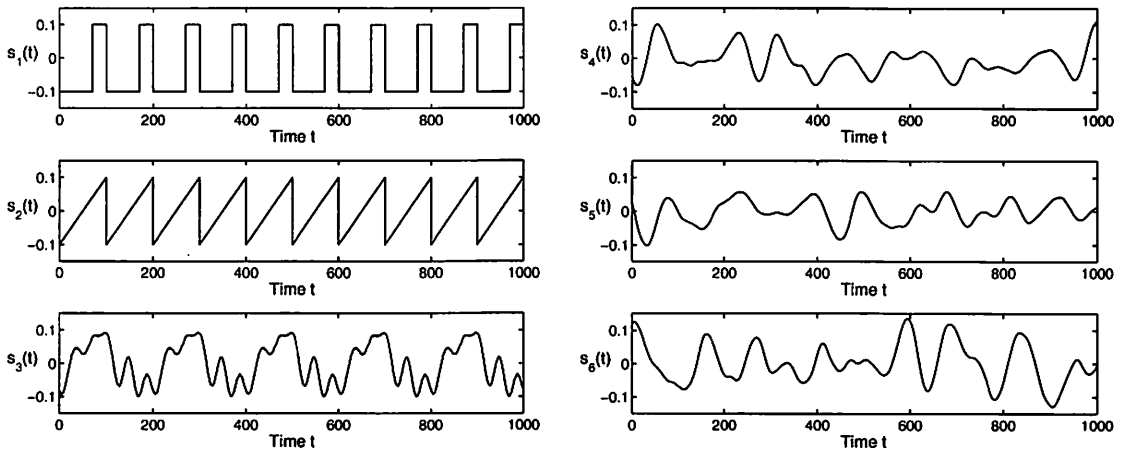


Figure 5.14: Samples of broadband signals. The signals s_1, s_2, and s_3 are periodic signals with broad spectrum while the other three signals s_4, s_5, and s_6 are aperiodic broadband signals. We generate broadband signals by convolving a sequence of Gaussian noise with Hanning windows [240].

SNR measure. In this section we will also test a gradient learning scheme that updates σ_n at every input-output sample (s_t, y_{t+1}).

Quartic bistable system

First we consider the popular quartic bistable system with sinusoidal input signal and cross-correlation measure. We test the exact learning law that requires the Jacobian matrix of the system. The Jacobian for the quartic bistable system $\dot{x} = x - x^3 + s + n$ has the form

$$\frac{\partial G}{\partial x} = \frac{\partial}{\partial x}[f(x) + s + \sigma w] = \frac{df}{dx} = \frac{d}{dx}(x - x^3) = 1 - 3x^2. \tag{5.29}$$

So the evolution of $\eta = \frac{\partial y}{\partial \sigma}$ for the quartic bistable system with linear output $y(t) = x(t)$ has the form

$$\dot{\eta} = (1 - 3x^2)\eta + w(t). \tag{5.30}$$

The derivative $\frac{\partial G}{\partial x}$ is the Jacobian matrix for multi-dimensional systems. We simulate $\frac{\partial y}{\partial \sigma}$ in (5.30) with its discretized version

$$\eta(t+1) = \eta(t) + \Delta T(1 - 3x_t^2)\eta(t) + \sqrt{\Delta T} w_t \tag{5.31}$$

for a zero-mean unit variance random sequence w_t from a random number generator that adds noise $n_t = \sigma w_t$ to the system. The learning scheme uses this partial derivative $\frac{\partial y}{\partial \sigma}$ to compute the performance gradient $\frac{\partial C}{\partial \sigma}$ in (5.28). Figure 5.15(a) shows the SR learning paths of σ_n for the forced quartic bistable system. The Cauchy suppressor ϕ replaces the performance gradient $\frac{\partial C}{\partial \sigma}$ with $\phi(\frac{\partial C}{\partial \sigma})$ in the learning law (5.27). We can approximate the performance gradient $\frac{\partial C}{\partial \sigma}$ with sample differences

$$\frac{\partial C}{\partial \sigma} \approx \frac{C(\sigma + \Delta\sigma) - C(\sigma - \Delta\sigma)}{2\Delta\sigma} \tag{5.32}$$

where $\Delta\sigma$ and μ_n satisfy certain technical conditions of stochastic approximation [152]. We use constant μ and $\Delta\sigma$. Figure 5.15(b) shows that SR learning paths of σ_n slowly converge and wander about the optimal noise level. The learning law that use Jacobian gives better learning paths.

Next we update the noise level σ_n in (5.27) at each input-output sample (s_t, y_{t+1}). Figure 5.16(a) shows the impulsiveness of the term $\frac{\partial y}{\partial \sigma}$ from the above evolution of the quartic bistable system with sinusoidal input and constant noise levels. This might also cause of impulsiveness in the learning of SNR. Figure 5.16(b) shows the learning paths of the parameter σ at each time step t. The learning algorithm updates the parameters σ_t at every sample y_t of the dynamical system. The impulsiveness of $\frac{\partial y}{\partial \sigma}$ pushes the trajectory away from the optimal noise levels.

Since we update the noise level σ at every time step t it is logical to replace $\frac{\partial y}{\partial \sigma}$ with the Cauchy-suppressed version $\phi(\frac{\partial y}{\partial \sigma})$ rather than the total learning term $\phi(\frac{\partial C}{\partial \sigma})$. Figure 5.17(a) shows the learning paths for $\phi(\frac{\partial C}{\partial \sigma})$ (as in our results in SNR case) comparing to the results

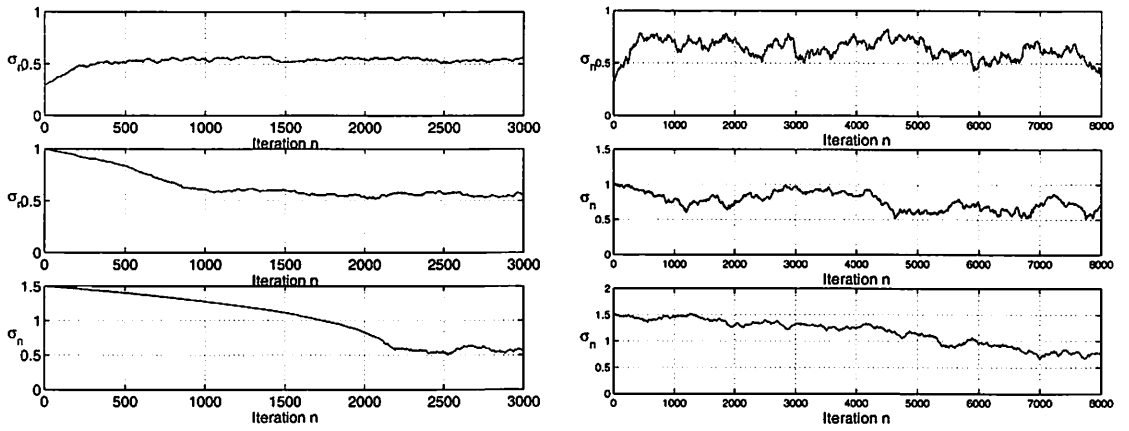


Figure 5.15: Learning paths of σ_n for the quartic bistable system $\dot{x} = x - x^3 + s + n$ with linear output $y(t) = x(t)$ and cross-correlation measure C. The sinusoidal input signal is $s = \varepsilon \sin 2\pi f_0 t$ for $\varepsilon = 0.1$ and $f_0 = 0.01$ and n is Gaussian noise. The optimal noise level is $\sigma = 0.6$. (a) The learning process uses Jacobian of the system to simulate $\frac{\partial y}{\partial \sigma}$ for $\frac{\partial C}{\partial \sigma}$ in (5.28). A Cauchy suppressor applies to the term $\frac{\partial C}{\partial \sigma}$. (b) The learning scheme approximates the performance gradient $\frac{\partial C}{\partial \sigma}$ with the differences $\frac{\partial C}{\partial \sigma} \approx \frac{C(\sigma+\Delta\sigma)-C(\sigma-\Delta\sigma)}{2\Delta\sigma}$ as in stochastic approximation framework [152]. The paths converge and wander near the optimal noise level.

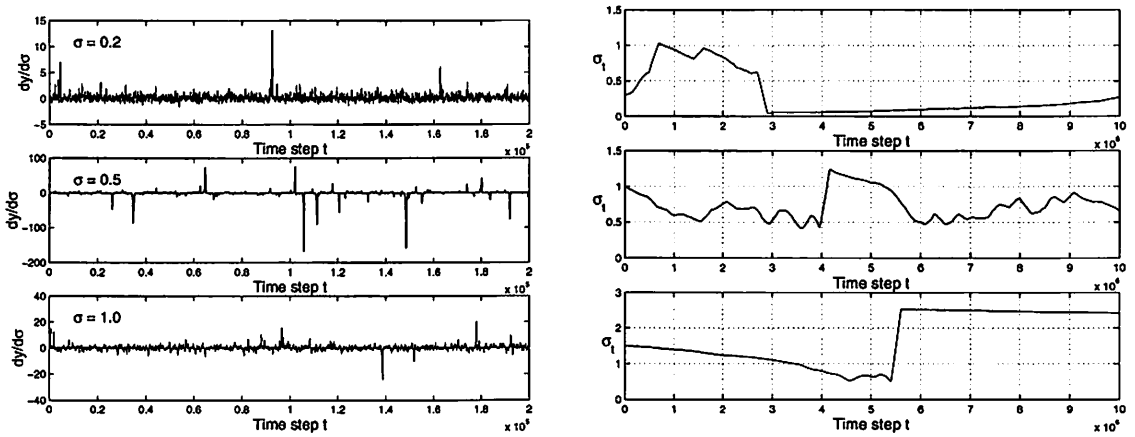
when the learning law uses $\frac{\partial y}{\partial \sigma}$ in Figure 5.17(b). Figure 5.18(a) shows the results when the algorithm updates the noise level at a subsampling rate of 100:1. The paths look smoother.

We also test the Cauchy suppressor to the term $s_t \frac{\partial y_{t+1}}{\partial \sigma}$ in (5.28). This derives from the conjecture that the term $\frac{\partial s_t y_{t+1}}{\partial \sigma}$ is impulsive instead of just the term $\frac{\partial y_{t+1}}{\partial \sigma}$. Figure 5.18b shows the learning paths.

Then we test the quartic bistable system with broadband signals. Figure 5.19 shows the learning paths. The learning scheme uses the Cauchy-suppressed term $\phi(\frac{\partial y}{\partial \sigma})$ in (5.28). The paths converge close to the optimal noise level.

FitzHugh-Nagumo neuron model

Now we consider the FitzHugh-Nagumo (FHN) neuron model (5.18)-(5.20) with sinusoidal input signal. We test the learning law (5.27) that updates after L samples. We apply a Cauchy noise suppressor to the term $\frac{\partial C}{\partial \sigma}$. Then we test the learning that updates every sample y_t. The Cauchy



(a) (b)

Figure 5.16: The partial derivative $\frac{\partial y}{\partial \sigma}$ and the learning paths for the quartic bistable system with output $y(t) = x(t)$. The optimal noise level is $\sigma = 0.6$. The learning process update σ_n at every sample time t. (a) The adaptive system simulate $\frac{\partial y}{\partial \sigma}$ from the system Jacobian (5.31). The term $\frac{\partial y}{\partial \sigma}$ shows an impulsive nature at the optimal noise level and thus can destabilize the learning process in (b).

noise suppressor applies to the term $\frac{\partial y}{\partial \sigma}$ and update the noise level every time we sample the dynamics.

The results for the FHN model with cross-correlation measure were not satisfactory for both cases of sinusoidal and broadband signals and for both ways of updating the optimal noise level. The paths did not converge to the optimal noise level even when we used the math model (the Jacobian) to simulate the term $\frac{\partial y}{\partial \sigma}$ in the learning law. This suggests that there might be other structure effects of the learning process besides the impulsiveness of $\frac{\partial y}{\partial \sigma}$ and $\frac{\partial C}{\partial \sigma}$ that we are not aware of. Future research would examine this effect in more detail or derive a new learning law that might overcome this effect.

5.3 Conclusion

Stochastic gradient ascent can learn to find the SR mode of at least some simple dynamical systems. This learning scheme may fail to scale up for more complex nonlinear dynamical systems of higher dimension or may get stuck in the local maxima of multimodal SNR profiles.

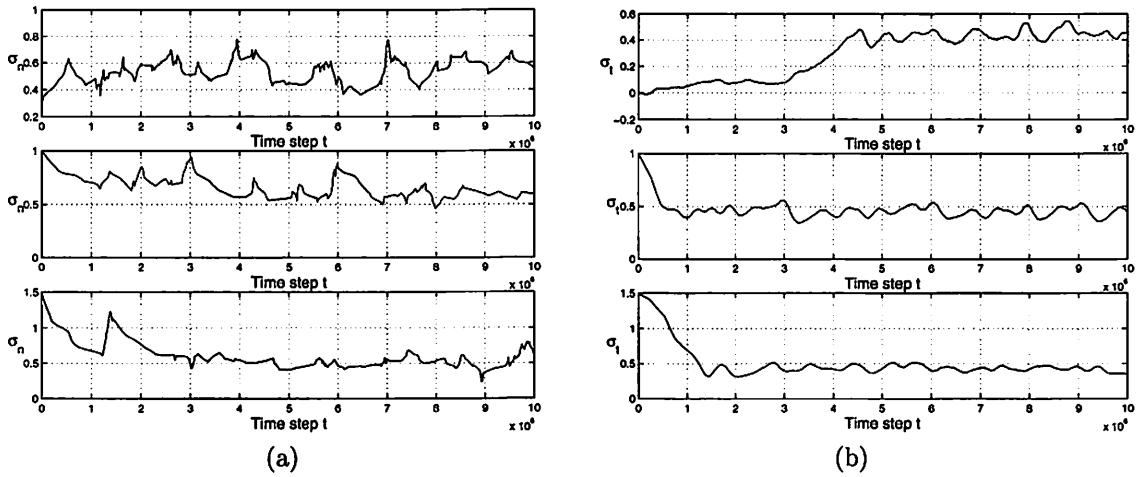


Figure 5.17: Learning paths for the quartic bistable system $\dot{x} = x - x^3 + s + n$ with output $y(t) = x(t)$. The input signal is $s(t) = 0.1 \sin 2\pi(0.01)t$ and n is additive Gaussian noise. The optimal noise level is $\sigma = 0.6$ for a cross-correlation measure C. (a) The learning law replaces $\frac{\partial C}{\partial \sigma}$ with $\phi(\frac{\partial C}{\partial \sigma})$. (b) The learning law replaces $\frac{\partial y}{\partial \sigma}$ with $\phi(\frac{\partial y}{\partial \sigma})$. The results show that $\phi(\frac{\partial y}{\partial \sigma})$ gives better learning.

Simulations showed that impulsive noise can destabilize the SR learning process even though the learning process does not minimize a mean-squared error. Simulations showed that the key learning term itself can give rise to strong impulsive shocks in the learning process. These shocks often approached Cauchy noise in intensity. A Cauchy noise suppressor gave a working SR learning scheme for the DFT-based SNR measure. Other SNR measures or other process statistics may favor other types of robust noise suppressors or may favor still other techniques to lessen the impulsiveness.

The learning scheme updates the standard deviation σ or the noise scale κ of zero-mean noise w with unit variance (or unit dispersion). A sign of this scale κ has no effect because the noise has zero mean. This also implies that there is a local extremum of the SR profile at the zero noise level. The expected value of a performance gradient equals zero at zero noise level: $E[\frac{\partial P}{\partial \sigma}] = 0$ at $\sigma = 0$ (if the expectation exists). The gradient-based learning can therefore wander about zero noise as well as nonzero noise since the adaptive system uses only noisy measurements of the performance gradients.

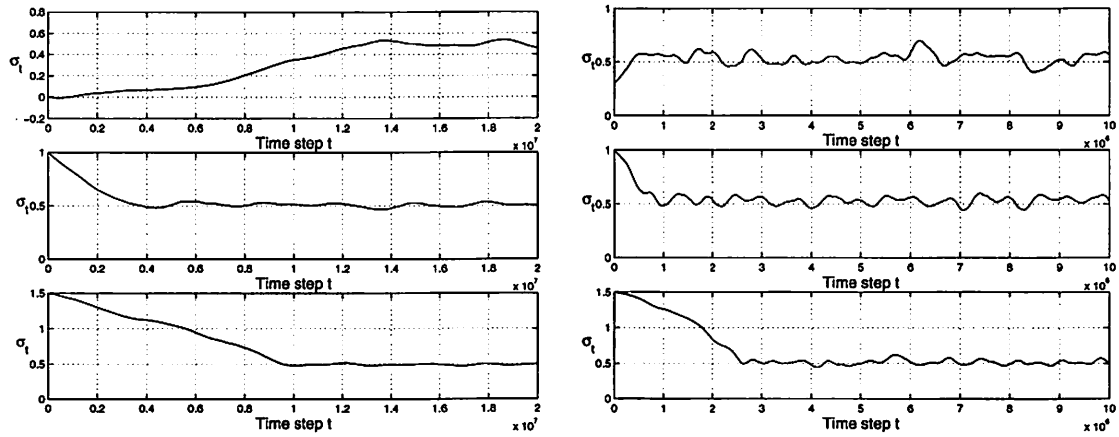


Figure 5.18: Learning paths for the quartic bistable system with sinusoidal input and linear output $y(t) = x(t)$. The optimal noise level is $\sigma = 0.6$. (a) A Cauchy suppressor replaces $\frac{\partial y}{\partial \sigma}$ with $\phi(\frac{\partial y}{\partial \sigma})$ in the learning law (5.27) and (5.28). Here the update has a subsampling rate at 100:1 while the sampling rate in Figure 5.17 is 1:1 with the dynamical system. (b) Here the learning law applies the Cauchy suppressor to the term $s_t \frac{\partial y_t}{\partial \sigma}$ to obtain $\phi(s_t \frac{\partial y_t}{\partial \sigma})$ in the performance gradient (5.28). The learning paths σ_t converge close to and wander about the optimal noise level.

Better learning schemes should escape from this trap of local minima. An adaptive system might perform some random search such as simulated annealing [155] but the system should spend most of its time at or near the optimal noise. An ideal system might encode the information in the signal and noise structures it has learned. Then it would spend less time searching for the optimal noise. This system would map a space of input signal to optimal noise levels. The system might use a finite-dimension space of features of input signals such as a few frequencies of interest as inputs. The system might also combine the knowledge it has learned from other signals to help search for the optimal noise of a new input signal. An adaptive fuzzy system is a candidate for such a system but it suffers exponential rule explosion when the number of features is large.

Zero-mean noise might not be the best noise for all systems. An adaptive system might also adapt the mean or bias of the noise that it adds to adaptively achieve stochastic resonance. A more complicated learning scheme might learn the best distribution of input noise. A learning system of this type would combine a random search on the space of all possible probability

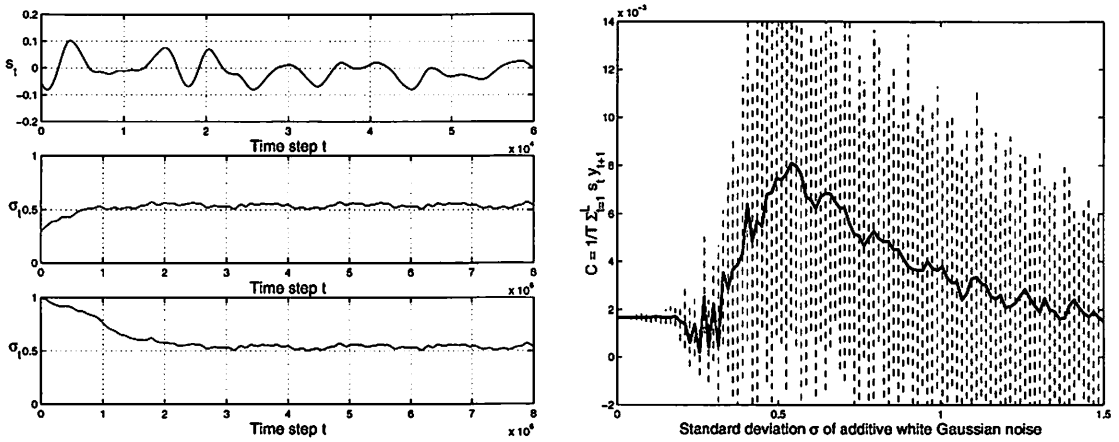


Figure 5.19: Learning paths of σ_t for the quartic bistable system with linear output $y(t) = x(t)$ and cross-correlation measure C. The top figure shows part of the broadband input signal $s(t)$ that forces the quartic bistable system $\dot{x} = x - x^3 + s + n$. The optimal noise level is $\sigma \approx 0.6$ as in the SR profile of a cross-correlation measure C. The learning law replaces $\frac{\partial y}{\partial \sigma}$ with $\phi(\frac{\partial y}{\partial \sigma})$.

densities with some analysis of the past input-output data to find the best noise. Better systems might learn the waveform or sequence of added "noise" that maximizes the performance measures.

Chapter 6

Future Research

This dissertation studies how we can adaptively achieve stochastic resonance. The SR learning schemes converge to and wander about the optimal noise level for sinusoidal input forcing signals. This follows from the stationarity statistics of the SR effect because the power spectrum of a sinusoidal signal is constant over periods of time. So the best noise is fixed at a constant level if the nonlinear system is time-invariant. Then the adaptive system can update the noise level after it samples enough data points to better estimate the performance gradients. It can also apply other schemes such as gradient averaging [277] that give more reliable estimates of the performance gradients $\frac{\partial P}{\partial \sigma}$.

Broadband signals complicate the learning process. The power spectrum of a broadband signal changes over time when we consider its Fourier spectrum over small windows of time. If the qualitative response of a nonlinear system depends on the spectrum of the input forcing signal then the statistics of the SR effect can also change over time. So the optimal noise can vary as the input signal spectrum changes. The adaptive system needs to track these changes so it can still add the right amount of noise to achieve the SR effect. The learning scheme may need to sample the dynamical system more frequently and may need to update the noise level more frequently. The dissertation derives such a learning law that uses the Jacobian of a nonlinear

dynamical system. The learning scheme assumes that we know the system dynamics and that we have access to all of the system states. These assumptions do not hold in many cases.

Future systems may relax these assumptions as they approximate the dynamics from input-output samples. A straightforward extension is to apply the "embedding theorem" that states that there exists an F that maps the past samples of the time series $\{y(t - j\tau)\}$ to the current sample $y(t)$:

$$y(t) = F(y(t-\tau), \ldots, y(t - d_E\tau)) \tag{6.1}$$

for a sufficiently large *embedding dimension* d_E and time delay τ [196, 243, 284, 295, 301]. So we can construct or estimate F from sample data $\{y(t - j\tau)\}$ and derive the learning law with the technique discussed in Chapter 4 for a dynamical system (6.1). Such learning laws will use the Jacobian matrix $J = \frac{\partial F}{\partial y}$ from the estimated map F. Local or global dynamic modeling can adaptively approximate the map F [262]. The method of least squares can give an estimate of the Jacobian J from data points within local region in the reconstruction space [282]. Another gradient system such as the LMS algorithm [262, 312] can give an estimate of the Jacobian matrix J from the input-output sample data:

$$J_{t+1} = J_t - \mu_n \nabla_J E. \tag{6.2}$$

Here E is the error of a linear modeling $E = \frac{1}{2}\|y(t+1) - \tilde{y}(t+1)\|^2$ where $\tilde{y}(t+1) = J_t y(t) + b_t$. Then the adaptive system simulates $\frac{\partial y}{\partial \sigma}$ from the Jacobian estimate as in Chapter 4.

Do we need to estimate the entire Jacobian matrix? Accurate Jacobian estimation requires a large amount of data and computation [262, 282] that complicates the ASR system. This dissertation derived a gradient-based ASR for the SNR and cross-correlation measures that relies on the gradients $\frac{\partial y}{\partial \sigma}$. One way to obtain this gradient is to simulate it from the Jacobian of the system. This raises the question whether we need the Jacobian to obtain $\frac{\partial y}{\partial \sigma}$. The

idea is to obtain or estimate the gradient $\frac{\partial y}{\partial \sigma}$ directly from input-output data. Applications of perturbation analysis [35, 103, 123, 293] may give a way to estimate the gradient $\frac{\partial y}{\partial \sigma}$ directly from input-output data. Ideal adaptive stochastic resonance would derive a learning law that truly depends on only data. It might depend on some form of rough approximation of the Jacobian matrix J or on the gradients $\frac{\partial y}{\partial \sigma}$ or on both or neither. But it needs to quickly capture the structure of F (from the data $\{y(t - j\tau)\}$) so it can track the nonstationary statistics of the SR effect.

Future research will also derive learning laws for ASR with other performance measures such as mutual information, probability of correct detection, probability of residence time, and Lyapunov exponents. These measures may require different techniques to extract SR information. A gradient-based ASR for the mutual information $I(s, y)$ as in (2.32) may require an estimate of the joint probability density of the input signal s and the output y: $\frac{\partial p_{s,y}(s,y)}{\partial \sigma}$. This may not be practical and we may need to find other ways to derive learning laws for mutual information measure.

ASR simulations also show that the gradient $\frac{\partial y}{\partial \sigma}$ is impulsive even when the learning system uses the math model (the system Jacobian), when we have has access to all of the system states, and when we use Gaussian or other finite-variance noise. We lack a formal explanation of this observed effect. The simulations also show that straightforward application of gradient learning does not always give convergence to the optimal noise level. A Cauchy noise suppressor partially overcomes and helps stabilize the learning processes. But the ASR simulations for the cross-correlation measure fail to converge to the optimal noise levels for the FHN neuron model. This suggests other structural effects in the learning processes. Better ASR systems should identify these problems and find solutions to improve convergence.

This dissertation shows in sum that gradient-ascent learning can find the SR mode of the main known dynamical models that show the SR effect and can do so in the presence of a wide range of noise types and signal types. This suggests that SR may occur in many multivariable dynamical systems in science and engineering and that simple learning schemes can sometimes

measure or approximate this behavior. We lack formal results that describe when and how such SR learning algorithms will converge for which types of SR systems. This reflects the general lack of a formal taxonomy in this promising new field: Which noisy dynamical systems show what SR effects for which forcing signals and for which performance measures?

Bibliography

[1] Y. S. Abu-Mostafa, "Learning from Hints in Neural Networks," *Journal of Complexity*, vol. 6, pp. 192–198, 1990.

[2] Y. S. Abu-Mostafa, "Hints," *Neural Computation*, vol. 7, pp. 639–671, 1995.

[3] V. S. Anishchenko, A. B. Neiman, and M. A. Safonova, "Stochastic Resonance in Chaotic Systems," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 183–196, 1993.

[4] V. S. Anishchenko, M. A. Safonova, and L. O. Chua, "Stochastic Resonance in Chua's Circuit," *International Journal of Bifurcation and Chaos*, vol. 2, no. 2, pp. 397–401, 1992.

[5] V. S. Anishchenko, M. A. Safonova, and L. O. Chua, "Stochastic Resonance in Chua's Circuit Driven by Amplitude or Frequency Modulated Signals," *International Journal of Bifurcation and Chaos*, vol. 4, no. 2, pp. 441–446, 1994.

[6] A. S. Asdi and A. H. Tewfik, "Detection of Weak Signals Using Adaptive Stochastic Resonance," in *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-95)*, May 1995, vol. 2, pp. 1332–1335.

[7] P. Babinec, "Stochastic Resonance in the Weidlich Model of Public Opinion Formation," *Physics Letters A*, vol. 225, pp. 179–181, January 1997.

[8] P. Baldi, "Gradient Descent Learning Algorithm Overview: A General Dynamical Systems Perspective," *IEEE Transactions on Neural Networks*, vol. 6, pp. 182–195, January 1995.

[9] R. Bartussek, P. Hänggi, and P. Jung, "Stochastic Resonance in Optical Bistable Systems," *Physical Review E*, vol. 49, no. 5, pp. 3930–3939, May 1994.

[10] M. Bartz, "Large-Scale Dithering Enhances ADC Dynamic Range," *Microwave & RF*, vol. 32, pp. 192–194+, May 1993.

[11] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic Resonance in Climatic Change," *Tellus*, vol. 34, pp. 10–16, 1982.

[12] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "A Theory of Stochastic Resonance in Climatic Change," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 565–578, June 1983.

[13] R. Benzi, A. Sutera, and A. Vulpiani, "The Mechanism of Stochastic Resonance," *Journal of Physics A: Mathematical and General*, vol. 14, pp. L453–L457, 1981.

[14] R. Benzi, A. Sutera, and A. Vulpiani, "Stochastic Resonance in the Landau-Ginzberg Equation," *Journal of Physics A: Mathematical and General*, vol. 18, pp. 2239–2245, 1985.

[15] S. M. Bezrukov and I. Vodyanoy, "Stochastic Resonance in Non-dynamical Systems without Response Thresholds," *Nature*, vol. 385, pp. 319–321, January 1997.

[16] Y. Braiman, J. F. Lindner, and W. L. Ditto, "Taming Spatiotemporal Chaos with Disorder," *Nature*, vol. 378, pp. 465–467, November 1995.

[17] H. A. Braun, H. Wissing, K. Schäfer, and M. C. Hirsch, "Oscillation and Noise Determine Signal Transduction in Shark Multimodal Sensory Cells," *Nature*, vol. 367, pp. 270–273, January 1994.

[18] L. Breiman, *Probability*, Addison-Wesley, 1968.

[19] J. J. Brey, J. Casado-Pascul, and B. Sánchez, "Resonant Behavior of a Poisson Process Driven by a Periodic Signal," *Physical Review E*, vol. 52, no. 6, pp. 6071–6081, December 1995.

[20] J. J. Brey and A. Prados, "Stochastic Resonance in a One-Dimension Ising Model," *Physics Letters A*, vol. 216, pp. 240–246, June 1996.

[21] R. Brooks, "Intelligence Without Reason," in *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*, L. Steels and R. Brooks, Eds., chapter 2, pp. 25–81. Lawrence Erlbaum Associates, Inc., 1995.

[22] K. S. Brown, "Noises On," *New Scientist*, vol. 150, pp. 28–31, June 1996.

[23] P. Bryant, K. Wiesenfeld, and B. McNamara, "The Nonlinear Effects of Noise on Parametric Amplification: An Analysis of Noise Rise in Josephson Junctions and Other Systems," *Journal of Applied Physics*, vol. 62, no. 7, pp. 2898–2913, October 1987.

[24] A. Bulsara, S. Chillemi, L. Kiss, P. V. E. McClintock, R. Mannella, F. Marchesoni, K. Nicolis, and K. Wiesenfeld, Eds., *Il Nuovo Cimento, Special Issue on Fluctuation in Physics and Biology: Stochastic Resonance, Signal Processing, and Related Phenomena*, vol. 17 D, no. 7-8, Luglio-Agosto 1995.

[25] A. R. Bulsara, R. D. Boss, and E. W. Jacobs, "Noise Effects in an Electronic Model of a Single Neuron," *Biological Cybernetics*, vol. 61, pp. 211–222, 1989.

[26] A. R. Bulsara, T. C. Elston, C. R. Doering, S. B. Lowen, and K. Lindenberg, "Cooperative Behavior in Periodically Driven Noisy Integrate-Fire Models of Neuronal Dynamics," *Physical Review E*, vol. 53, no. 4, pp. 3958–3969, April 1996.

[27] A. R. Bulsara and L. Gammaitoni, "Tuning in to Noise," *Physics Today*, pp. 39–45, March 1996.

[28] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, "Stochastic Resonance in a Single Neuron Model: Theory and Analog Simulation ," *Journal of Theoretical Biology*, vol. 152, pp. 531–555, 1991.

[29] A. R. Bulsara, S. B. Lowen, and C. D. Rees, "Cooperative Behavior in the Periodically Modulated Wiener Process: Noise-Induced Complexity in a Model Neuron," *Physical Review E*, vol. 49, no. 6, pp. 4989–5000, June 1994.

[30] A. R. Bulsara, A. J. Maren, and G. Schmera, "Single Effective Neuron: Dendritic Coupling Effects and Stochastic Resonance," *Biological Cybernetics*, vol. 70, pp. 145–156, 1993.

[31] A. R. Bulsara and G. Schmera, "Stochastic Resonance in Globally Coupled Nonlinear Oscillators," *Physical Review E*, vol. 47, no. 5, pp. 3734–3737, May 1993.

[32] A. R. Bulsara and A. Zador, "Threshold Detection of Wideband Signals: A Noise-Induced Maximum in the Mutual Information," *Physical Review E*, vol. 54, no. 3, pp. R2185–R2188, September 1996.

[33] G. C. Burdea, *Force and Touch Feedback for Virtual Reality*, John Wiley & Sons, 1996.

[34] T. Caelli and D. Reye, "On the Classification of Image Regions by Colour, Texture and Shape," *Pattern Recognition*, vol. 26, no. 4, pp. 461–470, 1993.

[35] X.-R. Cao and Y.-C. Ho, "Sensitivity Analysis and Optimization of Throughput in a Production Line with Blocking," *IEEE Transactions on Automatic Control*, vol. AC-32, no. 11, pp. 959–967, November 1987.

[36] T. L. Carroll and L. M. Pecora, "Stochastic Resonance and Crises," *Physical Review Letters*, vol. 70, no. 5, pp. 576–579, February 1993.

[37] J. M. Casado and M. Morillo, "Distribution of Escape Times in a Driven Stochastic Model," *Physical Review E*, vol. 49, no. 2, pp. 1136–1139, February 1994.

[38] R. Castro and T. Sauer, "Chaotic Stochastic Resonance: Noise-Enhanced Reconstruction of Attractors," *Physical Review Letters*, vol. 79, no. 6, pp. 1030–1033, August 1997.

[39] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, vol. 71, no. 354, pp. 340–344, 1976.

[40] S. F. Chang and J. R. Smith, "Extracting Multi-Dimensional Signal Features for Content-Based Visual Query," in *SPIE Symposium on Visual Communications and Signal Processing*, May 1995.

[41] F. Chapeau-Blondeau, "Noise-Enhanced Capacity via Stochastic Resonance in an Asymmetric Binary Channel," *Physical Review E*, vol. 55, no. 2, pp. 2016–2019, February 1997.

[42] F. Chapeau-Blondeau and X. Godivier, "Stochastic Resonance in Nonlinear Transmission of Spike Signals: An Exact Model and an Application to the Neuron," *International Journal of Bifurcation and Chaos*, vol. 6, no. 11, pp. 2069–2076, 1996.

[43] F. Chapeau-Blondeau and X. Godivier, "Theory of Stochastic Resonance in Signal Transmission by Static Nonlinear System ," *Physical Review E*, vol. 55, no. 2, pp. 1478–1495, February 1997.

[44] F. Chapeau-Blondeau, X. Godivier, and N. Chambet, "Stochastic Resonance in a Neuron Model that Transmits Spike Trains," *Physical Review E*, vol. 53, no. 1, pp. 1273–1275, January 1996.

[45] A. K. Chattah, C. B. Briozzo, O. Osenda, and M. O. Cáceres, "Signal-to-Noise Ratio in Stochastic Resonance," *Modern Physics Letters B*, vol. 10, no. 22, pp. 1085–1094, 1996.

[46] A. Chavez and P. Maes, "Kasbah: An Agent Marketplace for Buying and Selling Goods," in *Proceedings of the Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, April 1996.

[47] D. R. Chialvo, A. Longtin, and J. Müller-Gerking, "Stochastic Resonance in Models of Neuronal Ensembles," *Physical Review E*, vol. 55, no. 2, pp. 1798–1808, February 1997.

[48] F. Y. Chiou-Tan, K. N. Magee, S. M. Tuel, L. R. Robinson, T. A. Krouskop, M. R. Netson, and F. Moss, "Augmented Sensory Nerve Action Potentials During Distant Muscle Contraction," *American Journal of Physical Medicine and Rehabitation*, vol. 76, no. 1, pp. 14–18, January/February 1997.

[49] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, Birkhäuser, second edition, 1990.

[50] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "Aperiodic Stochastic Resonance," *Physical Review E*, vol. 54, no. 5, pp. 5575–5584, November 1996.

[51] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Aperiodic Stochastic Resonance in Excitable Systems," *Physical Review E*, vol. 52, no. 4, pp. R3321–R3324, October 1995.

[52] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, pp. 236–238, July 1995.

[53] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-Enhanced Information Transmission in Rat SA1 Cutaneous Mechanoreceptors via Aperiodic Stochastic Resonance," *Journal of Neurophysiology*, vol. 76, no. 1, pp. 642–645, July 1996.

[54] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-Enhanced Tactile Sensation," *Nature*, vol. 383, pp. 770, October 1996.

[55] M. Colombetti and M. Dorigo, "Training Agents to Perform Sequential Behavior," *Adaptive Behavior*, vol. 2, no. 3, pp. 247–275, 1994.

[56] W. D. Cook and M. Kress, *Ordinal Information and Preference Structures: Decision Models and Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[57] P. Cordo, J. T. Inglis, S. Vershueren, J. J. Collins, D. M. Merfeld, S. Rosenblum, S. Buckley, and F. Moss, "Noise in Human Muscle Spindles," *Nature*, vol. 383, pp. 769–770, October 1996.

[58] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[59] J. J. Craig, *Introduction to Robotics: Mechanics and Control*, Addison-Wesley, second edition, 1989.

[60] A. Crisanti, M. Falcioni, G. Paladin, and A. Vulpiani, "Stochastic Resonance in Deterministic Chaotic Systems," *Journal of Physics A: Mathematical and General*, vol. 27, pp. L597–L603, 1994.

[61] G. Dahlquist and Å. Björck, *Numerical Methods*, Prentice Hall, 1974.

[62] G. Debnath, F. Moss, Th. Leiber, H. Risken, and F. Marchesoni, "Holes in the Two-Dimensional Probability Density of Bistable Systems Driven by Strongly Colored Noise," *Physical Review A*, vol. 42, no. 2, pp. 703–710, July 1990.

[63] G. Debnath, T. Zhou, and F. Moss, "Remarks on Stochastic Resonance," *Physical Review A*, vol. 39, no. 8, pp. 4323–4326, April 1989.

[64] G. Debreu, "Representation of a Preference Ordering by a Numerical Function," in *Mathematical Economics: Twenty Papers of Gerard Debreu*, chapter 6, pp. 105–110. Cambridge University Press, 1983.

[65] J. L. Doob, *Stochastic Processes*, John Wiley & Sons, 1953.

[66] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant System: Optimization by a Colony of Cooperating Agents," *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, vol. 26, no. 1, pp. 29–41, February 1996.

[67] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance," *Nature*, vol. 365, pp. 337–340, September 1993.

[68] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Stochastic Resonance in Crayfish Hydrodynamic Receptors Stimulated with External Noise," in *AIP Conference Proceedings 285: Noise in Physical Systems and 1/f Fluctuations*, P. H. Hanel and A. L. Chung, Eds., 1993, pp. 712–715.

[69] R. Durrett, *Stochastic Calculus: A Practical Introduction*, CRC Press, 1996.

[70] M. I. Dykman, G. P. Golubev, I. Kh. Kaufman, D. G. Luchinsky, P. V. E. McClintock, and E. A. Zhukov, "Noise-Enhanced Optical Heterodyning in an All-Optical Bistable System," *Applied Physics Letters*, vol. 67, no. 3, pp. 308–310, July 1995.

[71] M. I. Dykman, T. Horita, and J. Ross, "Statistical Distribution and Stochastic Resonance in a Periodically Driven Chemical System," *Journal of Chemical Physics*, vol. 103, no. 3, pp. 966–972, July 1995.

[72] M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, S. M. Soskin, and N. D. Stein, "Resonant Subharmonic Absorption and Second-Harmonic Generation by a Fluctuating Nonlinear Oscillator," *Physical Review E*, vol. 54, no. 3, pp. 2366–2377, September 1996.

[73] M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Nonconventional Stochastic Resonance," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 479–499, January 1993.

[74] M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Stochastic Resonance: Linear Response and Giant Nonlinearity," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 463–478, January 1993.

[75] M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Supernarrow Spectral Peaks and High-Frequency Stochastic Resonance in Systems with Coexisting Periodic Attractors," *Physical Review E*, vol. 49, no. 2, pp. 1198–1215, February 1994.

[76] M. I. Dykman, D. G. Luchinsky, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Stochastic Resonance for Periodically Modulated Noise Intensity," *Physical Review A*, vol. 46, no. 4, pp. R1713–R1716, August 1992.

[77] J.-P. Eckmann and L. E. Thomas, "Remarks on Stochastic Resonances," *Journal of Physics A: Mathematical and General*, vol. 15, pp. L261–L266, 1982.

[78] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

[79] K. Ehrenberger, D. Felix, and K. Svozil, "Stochastic Resonance in Cochlear Signal Transduction," *Acta Otolaryngol (Stockh)*, vol. 119, pp. 166–170, 1999.

[80] S. Fauve and F. Heslot, "Stochastic Resonance in a Bistable System," *Physics Letters A*, vol. 97, no. 1,2, pp. 5–7, August 1983.

[81] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, 1966.

[82] P. G. Flikkema, "Spread-Spectrum Techniques for Wireless Communication," *Signal Processing Magazine*, vol. 14, no. 3, pp. 26–36, 1997.

[83] A. Förster, M. Merget, and F. W. Schneider, "Stochastic Resonance in Chemistry. 2. The Peroxidase-Oxidase Reaction," *Journal of Physical Chemistry*, vol. 100, pp. 4442–4447, 1996.

[84] A. M. Forte and J. X. Mitrovica, "A Resonance in the Earth's Obliquity and Precession over the Past 20 Myr Driven by Mantle Convection," *Nature*, vol. 390, pp. 676–680, December 1997.

[85] J. Foss, F. Moss, and J. Milton, "Noise, Multistability, and Delayed Recurrent Loops," *Physical Review E*, vol. 55, no. 4, pp. 4536–4543, April 1997.

[86] R. F. Fox, "Stochastic Resonance in a Double Well," *Physical Review A*, vol. 39, no. 8, pp. 4148–4153, April 1989.

[87] R. F. Fox and Y. N. Lu, "Analytic and Numerical Study of Stochastic Resonance," *Physical Review E*, vol. 48, no. 5, pp. 3390–3398, November 1993.

[88] J. N. Franklin, *Matrix Theory*, Prentice Hall, 1968.

[89] W. J. Freeman, H.-J. Chang, B. C. Burke, P. A. Rose, and J. Badler, "Taming chaos: Stabilization of aperiodic attractors by noise," *IEEE Transactions on Circuits and Systems–I: Fundamental Theory and Applications*, vol. 44, no. 10, pp. 989–996, October 1997.

[90] A. Fuliński, "Relaxation, Noise-Induced Transitions, and Stochastic Resonance Driven by Non-Markovian Dichotomic Noise," *Physical Review E*, vol. 52, no. 4, pp. 4523–4526, October 1995.

[91] B. V. Funt, "Color Constant Color Indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 522–529, May 1995.

[92] L. Gammaitoni, "Stochastic Resonance in Multi-Threshold Systems," *Physics Letters A*, vol. 208, pp. 315–322, December 1995.

[93] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic Resonance," *Reviews of Modern Physics*, vol. 70, no. 1, pp. 223–287, January 1998.

[94] L. Gammaitoni, F. Marchesoni, M. Martinelli, L. Pardi, and S. Santucci, "Phase Shifts in Bistable EPR Systems at Stochastic Resonance," *Physics Letters A*, vol. 158, no. 9, pp. 449–452, September 1991.

[95] L. Gammaitoni, F. Marchesoni, E. Menichella-Saetta, and S. Santucci, "Stochastic Resonance in Bistable Systems," *Physical Review Letters*, vol. 62, no. 4, pp. 349–352, January 1989.

[96] L. Gammaitoni, F. Marchesoni, E. Menichella-Saetta, and S. Santucci, "Multiplicative Stochastic Resonance," *Physical Review E*, vol. 49, no. 6, pp. 4878–4881, June 1994.

[97] L. Gammaitoni, M. Martinelli, L. Pardi, and S. Santucci, "Observation of Stochastic Resonance in Bistable Electron-Paramagnetic-Resonance Systems," *Physical Review Letters*, vol. 67, no. 13, pp. 1799–1802, September 1991.

[98] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, F. Marchesoni, and C. Pressilla, "Periodically Time-Modulated Bistable Systems: Stochastic Resonance," *Physical Review A*, vol. 40, no. 4, pp. 2114–2119, August 1989.

[99] T. C. Gard, *Introduction to Stochastic Differential Equations*, Marcel Dekker, Inc., 1988.

[100] F. Gassmann, "Noise-Induced Chaos-Order Transitions," *Physical Review E*, vol. 55, no. 3, pp. 2215–2221, March 1997.

[101] Z. Gingl, L. B. Kiss, and F. Moss, "Non-Dynamical Stochastic Resonance: Theory and Experiments with White and Arbitrarily Coloured Noise," *Europhysics Letters*, vol. 29, no. 3, pp. 191–196, January 1995.

[102] J. Glanz, "Mastering the Nonlinear Brain," *Science*, vol. 277, pp. 1758–1760, September 1997.

[103] P. Glasserman, *Gradient Estimation via Perturbation Analysis*, Kruwer Academic Publishers, 1991.

[104] B. J. Gluckman, T. I. Netoff, E. J. Neel, W. L. Ditto, M. L. Spano, and S. J. Schiff, "Stochastic Resonance in a Neuronal Network from Mammalian Brain," *Physical Review Letters*, vol. 77, no. 19, pp. 4098–4101, November 1996.

[105] X. Godivier and F. Chapeau-Blondeau, "Noise-Assisted Signal Transmission in a Nonlinear Electronic Comparator: Experiment and Theory," *Signal Processing*, vol. 56, pp. 293–303, 1997.

[106] D. C. Gong, G. Hu, X. D. Wen, C. Y. Yang, G. R. Qin, R. Li, and D. F. Ding, "Experimental Study of the Signal-to-Noise Ratio of Stochastic Resonance Systems," *Physical Review A*, vol. 46, no. 6, pp. 3243–3249, September 1992.

[107] M. Grifoni, "Dynamics of the Dissipative Two-State System under AC Modulation of Bias and Coupling Energy," *Physical Review E*, vol. 54, no. 4, pp. R3086–R3089, October 1996.

[108] M. Grifoni and P. Hänggi, "Nonlinear Quantum Stochastic Resonance," *Physical Review E*, vol. 54, no. 2, pp. 1390–1401, August 1996.

[109] M. Grifoni, L. Hartmann, S. Berchtold, and P. Hänggi, "Quantum Tunneling and Stochastic Resonance," *Physical Review E*, vol. 53, no. 6, pp. 5890–5898, June 1996.

[110] M. Grifoni, M. Sassetti, P. Hänggi, and U. Weiss, "Cooperative Effects in the Nonlinearly Driven Spin-Boson System," *Physical Review E*, vol. 52, no. 4, pp. 3596–3607, October 1995.

[111] A. N. Grigorenko and P. I. Nikitin, "Stochastic Resonance in a Bistable Magnetic System," *IEEE Transactions on Magnetics*, vol. 31, no. 5, pp. 2491–2493, September 1995.

[112] A. N. Grigorenko, S. I. Nikitin, and G. V. Roschepkin, "Stochastic Resonance at Higher Harmonics in Monostable Systems," *Physical Review E*, vol. 56, no. 5, pp. R4907–4910, November 1997.

[113] W. I. Grosky, "Multimedia Information Systems," *IEEE Multimedia*, vol. 1, no. 1, pp. 12–24, Spring 1994.

[114] A. Guderian, G. Dechert, K.-P. Zeyer, and F. W. Schneider, "Stochastic Resonance in Chemistry. 1. The Belousov-Zhabitinsku Reaction," *Journal of Physical Chemistry*, vol. 100, pp. 4437–4441, 1996.

[115] P. T. Harker, "Incomplete Pairwise Comparison in the Analytic Hierarchy Process," *Mathematical Modelling*, vol. 9, no. 11, pp. 837–848, 1987.

[116] S. Haykin, *Neural Networks: A Comprehensive Foundation*, McMillan, 1994.

[117] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, Reading, MA, 1990.

[118] C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich, "Information Measures Quantifying Aperiodic Stochastic Resonance," *Physical Review E*, vol. 54, no. 3, pp. R2228–R2231, 1996.

[119] A. Hibbs, E. W. Jacobs, J. Bekkedahl, A. R. Bulsara, and F. Moss, "Signal Enhancement in a r.f. SQUID Using Stochastic Resonance," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 811–817, Luglio-Agosto 1995.

[120] A. D. Hibbs, A. L. Singsaas, E. W. Jacobs, A. R. Bulsara, J. J. Bekkedahl, and F. Moss, "Stochastic Resonance in a Superconducting Loop with Josephson Junction," *Journal of Applied Physics*, vol. 77, no. 6, pp. 2582–2590, March 1995.

[121] W. Hildenbrand and A. P. Kirman, *Introduction to Equilibrium Analysis*, North Holland, 1976.

[122] Y.-C. Ho, "Performance Evaluation and Perturbation Analysis of Discrete Event Dynamical Systems," *IEEE Transactions on Automatic Control*, vol. AC-23, no. 7, pp. 563–572, July 1987.

[123] Y.-C. Ho and X.-R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems*, Kruwer Academic Publishers, 1991.

[124] W. Hohmann, J. Müller, and F. W. Schneider, "Stochastic Resonance in Chemistry. 3. The Minimal-Bromate Reaction," *Journal of Physical Chemistry*, vol. 100, pp. 5388–4492, 1996.

[125] R. Hollands, *The Virtual Reality Home Brewer's Handbook*, John Wiley & Sons, 1996.

[126] G. Hu, T. Ditzinger, C. Z. Ning, and H. Haken, "Stochastic Resonance without External Periodic Force," *Physical Review Letters*, vol. 71, no. 6, pp. 807–810, August 1993.

[127] G. Hu, D. C. Gong, X. D. Wen, C. Y. Yang, G. R. Qing, and R. Li, "Stochastic Resonance in a Nonlinear System Driven by an Aperiodic Force," *Physical Review A*, vol. 46, no. 6, pp. 3250–3254, September 1992.

[128] G. Hu, H. Haken, and C. Z. Ning, "A Study of Stochastic Resonance without Adiabatic Approximation," *Physics Letters A*, vol. 172, no. 1,2, pp. 21–28, December 1992.

[129] G. Hu, H. Haken, and F. Xie, "Stochastic Resonance with Sensitive Frequency Dependence in Globally Coupled Continuous Systems," *Physical Review Letters*, vol. 77, no. 10, pp. 1925–1928, September 1996.

[130] G. Hu, G. Nicolis, and N. Nicolis, "Periodically Forced Fokker-Planck Equation and Stochastic Resonance," *Physical Review A*, vol. 42, no. 4, pp. 2030–2041, August 1990.

[131] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.

[132] B. A. Huberman and R. M. Lukose, "Social Dilemmas and Internet Congestion," *Science*, vol. 277, pp. 535–537, July 1997.

[133] M. E. Inchiosa and A. R. Bulsara, "Coupling Enhances Stochastic Resonance in Nonlinear Dynamic Elements Driven by a Sinusoidal Plus Noise," *Physics Letters A*, vol. 200, pp. 283–288, April 1995.

[134] M. E. Inchiosa and A. R. Bulsara, "Nonlinear Dynamic Elements with Noisy Sinusoidal Forcing: Enhancing Response via Nonlinear Coupling," *Physical Review E*, vol. 52, no. 1, pp. 327–339, July 1995.

[135] M. E. Inchiosa and A. R. Bulsara, "Signal Detection Statistics of Stochastic Resonators," *Physical Review E*, vol. 53, no. 3, pp. R2021–R2024, March 1996.

[136] M. E. Inchiosa, A. R. Bulsara, and L. Gammaitoni, "Higher-Order Resonant Behavior in Asymmetric Nonlinear Stochastic Systems," *Physical Review E*, vol. 55, no. 4, pp. 4049–4056, April 1997.

[137] E. Ippen, J. Lindner, and W. L. Ditto, "Chaotic Resonance: A Simulation," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 437–450, January 1993.

[138] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neurofuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, 1996.

[139] N. R. Jennings and M. Wooldridge, "Software Agents," *IEE Review*, vol. 42, no. 1, pp. 17–20, January 1996.

[140] B. M. Jost and B. E. A. Saleh, "Signal-to-Noise Ratio Improvement by Stochastic Resonance in a Unidirectional Photorefractive Ring Resonator," *Optics Letters*, vol. 21, no. 4, pp. 287–289, February 1996.

[141] P. Jung, "Thermal Activation in Bistable Systems under External Periodic Forces," *Zeitschrift für Physik B*, vol. 76, pp. 521–535, 1989.

[142] P. Jung, "Threshold Devices: Fractal Noise and Neural Talk," *Physical Review E*, vol. 50, no. 4, pp. 2513–2522, October 1994.

[143] P. Jung, "Stochastic Resonance and Optimal Design of Threshold Detectors," *Physics Letters A*, vol. 207, pp. 93–104, October 1995.

[144] P. Jung, U. Behn, E. Pantazelou, and F. Moss, "Collective Response in Globally Coupled Bistable Systems," *Physical Review A*, vol. 46, no. 4, pp. R1709–R1712, August 1992.

[145] P. Jung and P. Hänggi, "Resonantly Driven Brownian Motion: Basic Concepts and Exact Results," *Physical Review A*, vol. 41, no. 6, pp. 2977–2988, March 1990.

[146] P. Jung and P. Hänggi, "Amplification of Small Signals via Stochastic Resonance," *Physical Review A*, vol. 44, no. 12, pp. 8032–8042, December 1991.

[147] P. Jung and P. Talkner, "Suppression of Higher Harmonics at Noise Induced Resonances," *Physical Review E*, vol. 51, no. 3, pp. 2640–2643, March 1995.

[148] P. Jung and K. Wiesenfeld, "Too Quiet to Hear a Whisper," *Nature*, vol. 385, pp. 291, January 1997.

[149] S. Kádár, J. Wang, and K. Showalter, "Noise-supported travelling waves in sub-excitable media," *Nature*, vol. 391, pp. 770–772, February 1998.

[150] R. L. Keeney and H. Raiffa, *Decision with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Sons, New York, 1976.

[151] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, Edward Arnold, A division of Hodder & Stoughton, London, fifth edition, 1990.

[152] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *Annals of Mathematical Statistics*, vol. 23, pp. 462–466, 1952.

[153] F. A. Kilpatrick, G. H. Gunsch, and E. Santos Jr., "Induction and State-Space Search for an Intelligent Training System," in *Proceedings of the Midwest Artificial Intelligence and Cognitive Science Conference*, 1996.

[154] H. M. Kim and B. Kosko, "Neural Fuzzy Motion Estimation and Compensation," *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2515–2532, October 1997.

[155] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, pp. 671–680, 1983.

[156] J. Kirman, A. Nicholson, M. Lejter, T. Dean, and E. Santos Jr., "Using Goals to Find Plans with High Expected Utility," in *Proceedings of the Second European Workshop on Planning*, 1993, pp. 158–170.

[157] L. B. Kiss, "Possible Breakthrough: Significant Improvement of Signal to Noise Ratio by Stochastic Resonance," in *AIP Conference Proceedings 375: Chaotic, Fractal, and Nonlinear Signal Processing, 1995*, R. A. Katz, Ed., 1996, pp. 382–396.

[158] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1991.

[159] B. Kosko, "Fuzzy Systems as Universal Approximators," *IEEE Transactions on Computers*, vol. 43, no. 11, pp. 1329–1333, November 1994.

[160] B. Kosko, "Combining Fuzzy Systems," in *Proceedings of the IEEE International Conference on Fuzzy Systems (IEEE FUZZ-95)*, March 1995, pp. 1855–1863.

[161] B. Kosko, "Optimal Fuzzy Rules Cover Extrema," *International Journal of Intelligent Systems*, vol. 10, no. 2, pp. 249–255, February 1995.

[162] B. Kosko, *Fuzzy Engineering*, Prentice Hall, 1996.

[163] I. Kotlasi, "On Random Variables Whose Quotient Follows the Cauchy Law," *Colloquium Mathematicum*, vol. VII, pp. 277–284, 1960.

[164] H. A. Kramers, "Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions," *Physica*, vol. VII, no. 4, pp. 284–304, April 1940.

[165] D. M. Kreps, *A Course in Microeconomic Theory*, Princeton University Press, 1990.

[166] R. Krishnapuram and J. M. Keller, "A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, pp. 98–110, May 1993.

[167] R. G. Laha, "On a Class of Distribution Functions Where the Quotients Follows the Cauchy Law," *Transactions of the American Mathematical Society*, vol. 93, pp. 205–215, November 1959.

[168] J. E. Laird, A. Newell, and P. S. Rosenbloom, "SOAR: An Architecture for General Intelligence," *Artificial Intelligence*, vol. 33, pp. 1–64, 1987.

[169] W. C. Y. Lee, *Mobile Cellular Telecommunications: Analog and Digital Systems*, McGraw-Hill, 1995.

[170] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic," *Computer Communication Review: Proceedings of the SIGCOMM-93*, vol. 23, pp. 183–193, September 1993.

[171] D. S. Leonard, "Stochastic Resonance in a Random Walk," *Physical Review A*, vol. 46, no. 10, pp. 6742–6744, November 1992.

[172] D. S. Leonard and L. E. Reichl, "Stochastic Resonance in a Chemical Reaction," *Physical Review E*, vol. 49, no. 2, pp. 1734–1737, February 1994.

[173] J. E. Levin and J. P. Miller, "Broadband Neural Encoding in the Cricket Cercal Sensory System Enhanced by Stochastic Resonance," *Nature*, vol. 380, pp. 165–168, March 1996.

[174] P. H. Lewis and C. Yang, *Basic Control Systems Engineering*, Prentice Hall, 1997.

[175] R. Li, G. Hu, C. Y. Yang, X. D. Wen, G. R. Qing, and H. J. Zhu, "Stochastic Resonance in Bistable Systems Subject to Signal and Quasimonochromatic Noise," *Physical Review E*, vol. 51, no. 5, pp. 3964–3967, May 1995.

[176] J. F. Lindner, B. K. Meadows, W. L. Ditto, M. E. Inchiosa, and A. R. Bulsara, "Scaling Laws for Spatiotemporal Synchronization and Array Enhanced Stochastic Resonance," *Physical Review E*, vol. 53, no. 3, pp. 2081–2086, March 1996.

[177] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and Dither: A Theoretical Survey," *Journal of Audio Engineering Society*, vol. 40, no. 5, pp. 355–374, May 1992.

[178] M. Löcher, G. A. Johnson, and E. R. Hunt, "Spatiotemporal Stochastic Resonance in a System of Coupled Diode Resonators," *Physical Review Letters*, vol. 77, no. 23, pp. 4698–4701, December 1996.

[179] K. Loerincz, Z. Gingl, and L. B. Kiss, "A Stochastic Resonator is Able to Greatly Improve Signal-to-Noise Ratio," *Physics Letters A*, vol. 224, pp. 63–67, December 1996.

[180] R. Löfstedt and S. N. Coppersmith, "Quantum Stochastic Resonance," *Physical Review Letters*, vol. 72, no. 13, pp. 1947–1950, March 1994.

[181] A. Longtin, "Stochastic Resonance in Neuron Models," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 309–327, January 1993.

[182] A. Longtin, "Synchronization of the Stochastic Fitzhugh-Nagumo Equations to Periodic Forcing," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 835–846, Luglio-Agosto 1995.

[183] A. Longtin, "Autonomous Stochastic Resonance in Bursting Neurons," *Physical Review E*, vol. 55, no. 1, pp. 868–876, January 1997.

[184] A. Longtin, A. R. Bulsara, and F. Moss, "Time-Interval Sequences in Bistable Systems and the Noise-Induced Transmission of Information by Sensory Neurons," *Physical Review Letters*, vol. 67, no. 5, pp. 656–659, July 1991.

[185] A. Longtin, A. R. Bulsara, D. Pierson, and F. Moss, "Bistability and the Dynamics of Periodically Forced Sensory Neurons," *Biological Cybernetics*, vol. 70, pp. 569–578, 1994.

[186] A. Longtin and K. Hinzer, "Encoding with Bursting, Subthreshold Oscillations, and Noise in Mammalian Cold Receptors," *Neural Computation*, vol. 8, pp. 215–255, 1996.

[187] V. Loreto, G. Paladin, and A. Vulpiani, "Concept of Complexity in Random Dynamical Systems," *Physical Review E*, vol. 53, no. 3, pp. 2087–2098, March 1996.

[188] J. Maddox, "Towards the Brain-Computer's Code?," *Nature*, vol. 352, pp. 469, August 1991.

[189] J. Maddox, "Bringing More Order out of Noisiness," *Nature*, vol. 369, pp. 271, May 1994.

[190] P. Maes, "Agents that Reduce Work and Information Overload," *Communications of the ACM*, vol. 37, no. 7, pp. 31–40, July 1994.

[191] P. Maes, "Artificial Life Meets Entertainment: Lifelike Autonomous Agents," *Communications of the ACM*, vol. 38, no. 11, pp. 108–114, November 1995.

[192] P. Maes, "Modeling Adaptive Autonomous Agents," in *Artificial Life: An Overview*, C. G. Langton, Ed., pp. 135–162. MIT Press, 1995.

[193] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The ALIVE System: Wireless, Full-body Interaction with Autonomous Agents," *Multimedia Systems*, Spring 1996.

[194] M. C. Mahato and S. R. Shenoy, "Hysteresis Loss and Stochastic Resonance: A Numerical Study of a Double-Well Potential," *Physical Review E*, vol. 50, no. 4, pp. 2503–2512, October 1994.

[195] D. E. Makarov and N. Makri, "Stochastic Resonance and Nonlinear Response in Double-Quantum-Well Structures," *Physical Review B*, vol. 52, no. 4, pp. R2257–R2260, July 1995.

[196] R. Mañé, "On the Dimension of the Compact Invariant Sets of Certain Non-Linear Maps," in *Dynamical Systems and Turbulence*, D. A. Rand and L.-S. Young, Eds., pp. 230–242. Springer-Verlag, 1980, Springer Lecture Notes in Mathematics, Vol. 898.

[197] R. Mannella and V. Palleschi, "Fast and Precise Algorithm for Computer Simulation of Stochastic Differential Equations," *Physical Review A*, vol. 40, no. 6, pp. 3381–3386, September 1989.

[198] R. N. Mantegna and B. Spagnolo, "Stochastic Resonance in a Tunnel Diode," *Physical Review E*, vol. 49, no. 3, pp. R1792–R1795, March 1994.

[199] R. N. Mantegna and B. Spagnolo, "Stochastic Resonance in a Tunnel Diode in the Presence of White or Coloured Noise," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 873–881, Luglio-Agosto 1995.

[200] F. Marchesoni, L. Gammaitoni, and A. R. Bulsara, "Spatiotemporal Stochastic Resonance in a ϕ^4 Model of Kink-Antikink Nucleation," *Physical Review Letters*, vol. 76, no. 15, pp. 2609–2612, April 1996.

[201] K. Matsumoto and I. Tsuda, "Noise-induced order," *Journal of Statistical Physics*, vol. 31, no. 1, pp. 87–106, 1983.

[202] B. McNamara and K. Wiesenfeld, "Theory of Stochastic Resonance," *Physical Review A*, vol. 39, no. 9, pp. 4854–4869, May 1989.

[203] B. McNamara, K. Wiesenfeld, and R. Roy, "Observation of Stochastic Resonance in a Ring Laser," *Physical Review Letters*, vol. 60, no. 25, pp. 2626–2629, June 1988.

[204] V. I. Melnikov, "Schmitt Trigger: A Solvable Model of Stochastic Resonance," *Physical Review E*, vol. 48, no. 4, pp. 2481–2489, October 1993.

[205] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review*, vol. 63, no. 2, pp. 81–97, March 1956.

[206] J. P. Miller, G. A. Jacobs, and F. E. Theunissen, "Representation of Sensory Information in the Cricket Cercal Sensory System. I. Response Properties of the Primary Interneurons," *Journal of Neurophysiology*, vol. 66, no. 5, pp. 1680–1689, November 1991.

[207] J. P. Miller, G. A. Jacobs, and F. E. Theunissen, "Representation of Sensory Information in the Cricket Cercal Sensory System. II. Information Theoretic Calculation of System Accuracy and Optimal Tuning-Curve Widths of Four Primary Interneurons," *Journal of Neurophysiology*, vol. 66, no. 5, pp. 1690–1703, November 1991.

[208] I. Millet and P. T. Harker, "Globally Effective Questioning in the Analytic Hierarchy Process," *European Journal of Operational Research*, vol. 48, pp. 88–97, 1990.

[209] S. Mitaim and B. Kosko, "What is the Best Shape for a Fuzzy Set in Function Approximation?," in *Proceedings of the 5th IEEE International Conference on Fuzzy Systems (FUZZ-96)*, September 1996, vol. 2, pp. 1237–1243.

[210] S. Mitaim and B. Kosko, "Adaptive Joint Fuzzy Sets for Function Approximation," in *Proceedings of the 1997 IEEE International Conference on Neural Networks (ICNN-97)*, June 1997, vol. 1, pp. 537–542.

[211] S. Mitaim and B. Kosko, "Neural Fuzzy Agents for Profile Learning and Object Matching," in *Proceedings of the 1st International Conference on Autonomous Agents (AA '97)*, W. L. Johnson, Ed., Marina del Rey, CA, February 1997, pp. 544–545, ACM Press.

[212] S. Mitaim and B. Kosko, "Adaptive Stochastic Resonance," *Proceedings of the IEEE: Special Issue on Intelligent Signal Processing*, vol. 86, no. 11, pp. 2152–2183, November 1998.

[213] S. Mitaim and B. Kosko, "Neural Fuzzy Agents for Profile Learning and Adaptive Object Matching," *Presence: Special Issue on Autonomous Agents, Adaptive Behavior, and Distributed Simulations*, vol. 7, no. 6, pp. 617–637, December 1998.

[214] J. Moody and C. Darken, "Fast Learning in Networks of Locally-Tuned Processing Unit," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 1989.

[215] M. Morillo, J. Gómez-Ordo nez, and J. M. Casado, "Stochastic Resonance in a Mean-Field Model of Cooperative Behavior," *Physical Review E*, vol. 52, no. 1, pp. 316–320, July 1995.

[216] R. P. Morse and E. F. Evans, "Enhancement of Vowel Coding for Cochlear Implants by Addition of Noise," *Nature Medicine*, vol. 2, no. 8, pp. 928–932, August 1996.

[217] F. Moss, "Stochastic Resonance: From the Ice Ages to the Monkey's Ear," in *Contemporary Problems in Statistical Physics*, G. H. Weiss, Ed., chapter 5, pp. 205–253. SIAM, 1994.

[218] F. Moss, "Noisy waves," *Nature*, vol. 391, pp. 743–744, February 1998.

[219] F. Moss, A. Bulsara, and M. Shlesinger, Eds., *Journal of Statistical Physics, Special Issue on Stochastic Resonance in Physics and Biology (Proceedings of the NATO Advanced Research Workshop)*, vol. 70, no. 1/2, January 1993.

[220] F. Moss, F. Chiou-Tan, and R. Klinke, "Will There be Noise in Their Ears?," *Nature Medicine*, vol. 2, no. 8, pp. 860–862, August 1996.

[221] F. Moss, J. K. Douglass, L. Wilkens, D. Pierson, and E. Pantazelou, "Stochastic Resonance in an Electronic FitzHugh-Nagumo Model," in *Annals of the New York Academy of Sciences Volume 706: Stochastic Processes in Astrophysics*, J. R. Buchler and H. E. Kandrup, Eds., 1993, pp. 26–41.

[222] F. Moss and P. V. E. McClintock, Eds., *Noise in Nonlinear Dynamical Systems*, vol. I-III, Cambridge University Press, 1989.

[223] F. Moss, D. Pierson, and D. O'Gorman, "Stochastic Resonance: Tutorial and Update," *International Journal of Bifurcation and Chaos*, vol. 4, no. 6, pp. 1383–1397, 1994.

[224] F. Moss and K. Wiesenfeld, "The Benefits of Background Noise," *Scientific American*, vol. 273, no. 2, pp. 66–69, August 1995.

[225] A. Moukas, "Amalthaea: Information Discovery and Filtering Using a Multiagent Evolving Ecosystem," in *Proceedings of the Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, April 1996.

[226] T. Mullen and M. P. Wellman, "A Simple Computational Market for Network Information Services," in *Proceedings of the First International Conference on Multi-Agent Systems*, June 1995, pp. 283–289.

[227] Z. Néda, "Stochastic Resonance in Ising Systems," *Physical Review E*, vol. 51, no. 6, pp. 5315–5317, June 1995.

[228] A. Neiman and L. Schimansky-Geier, "Stochastic Resonance in Two Coupled Bistable Systems," *Physics Letters A*, vol. 197, pp. 379–386, February 1995.

[229] A. Neiman, L. Schimansky-Geier, and F. Moss, "Linear Response Theory Applied to Stochastic Resonance in Models of Ensembles of Oscillators," *Physical Review E*, vol. 56, no. 1, pp. R9–R12, July 1997.

[230] A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, "Dynamic Entropies Applied to Stochastic Resonance," *Physical Review Letters*, vol. 76, no. 23, pp. 4299–4302, June 1996.

[231] A. Neiman and W. Sung, "Memory Effects on Stochastic Resonance," *Physics Letters A*, vol. 223, pp. 341–347, December 1996.

[232] A. Newell and P. S. Rosenbloom, "Mechanisms of Skill Acquisition and the Law of Practice," in *Cognitive Skills and Their Acquisition*, J. R. Anderson, Ed., chapter 1, pp. 1–55. Lawrence Erlbaum Associates, Inc., 1981.

[233] W. Niblack, R. Barber, W. Equitz, M. Flikner, E. Glassman, D. Petkovic, P. Yanker, and C. Faloutsos, "The QBIC Project: Querying Images by Content Using Color, Texture, and Shape," Research Report RJ 9203 (81511), IBM, February 1993.

[234] C. Nicolis, "Stochastic Aspects of Climatic Transitions–Response to a Periodic Forcing," *Tellus*, vol. 34, pp. 1–9, 1982.

[235] C. Nicolis, "Long-Term Climatic Transitions and Stochastic Resonance," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 4–14, January 1993.

[236] C. Nicolis and G. Nicolis, "Stochastic Aspects of Climate Transitions–Additive Fluctuations," *Tellus*, vol. 33, pp. 225–234, 1981.

[237] G. Nicolis, C. Nicolis, and D. McKernan, "Stochastic Resonance in Chaotic Dynamics," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 125–139, January 1993.

[238] P. Niyogi and F. Girosi, "On the Relationship between Generalization Errors Hypothesis Complexity and Sample Complexity for Radial Basis Functions," *Neural Computation*, vol. 8, pp. 819–842, 1996.

[239] K. Ogata, *Modern Control Engineering*, Prentice Hall, third edition, 1997.

[240] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1989.

[241] G. Owen, *Game Theory*, Academic Press, third edition, 1995.

[242] P. J. Pacini and B. Kosko, "Adaptive Fuzzy Frequency Hopper," *IEEE Transactions on Communications*, vol. 43, no. 6, pp. 2111–2117, June 1995.

[243] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a Time Series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, September 1980.

[244] N. R. Pal and J. C. Bezdek, "On Cluster Validity for the Fuzzy c-Means Model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, August 1995.

[245] N. R. Pal, J. C. Bezdek, and R. J. Hathaway, "Sequential Competitive Learning and the Fuzzy c-Means Clustering Algorithms," *Neural Networks*, vol. 9, pp. 787–96, July 1996.

[246] A. Restrepo (Palacios), L. F. Zuluaga, and L. E. Pino, "Optimal Noise Levels for Stochastic Resonance," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*, April 1997, vol. III, pp. 2365–2368.

[247] E. Pantazelou, C. Dames, F. Moss, J. Douglass, and L. Wilkens, "Temperature Dependence and the Role of Internal Noise in Signal Transduction Efficiency of Crayfish Mechanoreceptors," *International Journal of Bifurcation and Chaos*, vol. 5, no. 1, pp. 101–108, 1995.

[248] H. C. Papadopoulos and G. W. Wornell, "A Class of Stochastic Resonance Systems for Signal Processing Applications," in *Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*, May 1996, pp. 1617–1620.

[249] A. Papoulis, *Probability and Statistics*, Prentice Hall, 1990.

[250] T. P. Pareek, M. C. Mahato, and A. M. Jayannavar, "Stochastic Resonance and Nonlinear Response in a Dissipative Quantum Two-State System," *Physical Review B*, vol. 55, no. 15, pp. 9318–9321, April 1997.

[251] B. R. Parnas, "Noise and Neuronal Populations Conspire to Encode Simple Waveforms Reliably," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 3, pp. 313–318, March 1996.

[252] S. Parsons and N. R. Jennings, "Negotiation through Argumentation–A Preliminary Report," in *Proceedings of the International Conference on Multi-Agent Systems*, 1996.

[253] X. Pei, K. Bachmann, and F. Moss, "The Detection Threshold, Noise and Stochastic Resonance in the Fitzhugh-Nagumo Neuron Model," *Physics Letters A*, vol. 206, pp. 61–65, October 1995.

[254] X. Pei, L. Wilkens, and F. Moss, "Light Enhances Hydrodynamic Signaling in the Multimodal Caudal Photoreceptor Interneurons of the Crayfish," *Journal of Neurophysiology*, vol. 76, no. 5, pp. 3002–3011, November 1996.

[255] X. Pei, L. Wilkens, and F. Moss, "Noise-Mediated Spike Timing Precision from Aperiodic Stimuli in an Array of Hodgkin-Huxley-Type Neurons," *Physical Review Letters*, vol. 77, no. 22, pp. 4679–4682, November 1996.

[256] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases," in *SPIE: Storage and Retrieval for Image and Video Database II*, February 1994, vol. 2185, pp. 34–47.

[257] R. W. Picard and T. P. Minka, "Vision Texture for Annotation," *Multimedia Systems*, vol. 3, pp. 3–14, 1995.

[258] D. Pierson, J. K. Douglass, E. Pantazelou, and F. Moss, "Using an Electronic FitzHugh-Nagumo Simulator to Mimic Noisy Electrophysiological Data from Stimulated Crayfish Mechanoreceptor Cells," in *AIP Conference Proceedings 285: Noise in Physical Systems and 1/f Fluctuations*, P. H. Hanel and A. L. Chung, Eds., 1993, pp. 731–734.

[259] H. E. Plesser and S. Tanaka, "Stochastic Resonance in a Model Neuron with Reset," *Physics Letters A*, vol. 225, pp. 228–234, February 1997.

[260] W. K. Pratt, *Digital Image Processing*, Wiley Interscience, second edition, 1991.

[261] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, second edition, 1993.

[262] J. C. Principe, L. Wang, and M. A. Motter, "Local Dynamic Modeling with Self-Organizing Maps and Applications to Nonlinear System Identification and Control," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2240–2258, 1998.

[263] W.-J. Rappel and A. Karma, "Noise-Induced Coherence in Neural Networks," *Physical Review Letters*, vol. 77, no. 15, pp. 3256–3259, October 1996.

[264] W.-J. Rappel and S. H. Strogatz, "Stochastic Resonance in an Autonomous System with a Nonuniform Limit Cycle," *Physical Review E*, vol. 50, no. 4, pp. 3249–3250, October 1994.

[265] D. W. Rasmus, "Intelligent Agents: DAI Goes to Work," *PC AI*, pp. 27–32, January/February 1995.

[266] D. O. Reale, A. K. Pattanayak, and W. C. Schieve, "Semiquantal Corrections to Stochastic Resonance," *Physical Review E*, vol. 51, no. 4, pp. 2925–2932, April 1995.

[267] W. S. Reilly and J. Bates, "Natural Negotiation for Believable Agents," Technical Report CMU-CS-95-164, Carnegie Mellon University, Pittsburgh, PA, June 1995.

[268] M. Riani and E. Simonotto, "Stochastic Resonance in the Perceptual Interpretation of Ambiguous Figures: A Neural Network Model," *Physical Review Letters*, vol. 72, no. 19, pp. 3120–3123, May 1994.

[269] M. Riani and E. Simonotto, "Periodic Perturbation of Ambiguous Figure: A Neural Network Model and a Non-Simulated Experiment," *Il Nuovo Cimento*, vol. 17 D, no. 7-8, pp. 903–913, Luglio-Agosto 1995.

[270] T. F. Ricci and C. Scherer, "Linear Response and Stochastic Resonance of Superparamagnets," *Journal of Statistical Physics*, vol. 86, no. 3/4, pp. 803–819, February 1997.

[271] K. A. Richardson, T. T. Imhoff, R. Grigg, and J. J. Collins, "Using Electrical Noise to Enhance the Ability of Humans to Detect Subthreshold Mechanical Cutaneous Stimuli," *Chaos: Focus Issue on the Constructive Role of Noise in Fluctuation Driven Transport and Stochastic Resonance*, vol. 8, no. 3, pp. 599–603, September 1998.

[272] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Application*, Springer-Verlag, 1984.

[273] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, vol. 22, pp. 400–407, 1951.

[274] J. S. Rosenschein and G. Zlotkin, "Consenting Agents: Designing Conventions for Automated Negotiation," *AI Magazine*, vol. 15, no. 3, pp. 29–46, Fall 1994.

[275] R. Rouse, S. Han, and J. E. Lukens, "Flux Amplification Using Stochastic Superconducting Quantum Interference Devices," *Applied Physics Letters*, vol. 66, no. 1, pp. 108–110, January 1995.

[276] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.

[277] A. Ruszczyński and W. Syski, "Stochastic Approximation Method with Gradient Averaging for Unconstrained Problems," *IEEE Transactions on Automatic Control*, vol. AC-28, no. 12, pp. 1097–1105, December 1983.

[278] T. L. Saaty, "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, vol. 15, pp. 234–281, 1977.

[279] T. L. Saaty, "Axiomatic Foundation of the Analytic Hierarchy Process," *Management Science*, vol. 32, no. 7, pp. 841–855, July 1986.

[280] T. L. Saaty, "Highlights and Critical Points in the Theory and Application of the Analytic Hierarchy Process," *European Journal of Operational Research*, vol. 74, pp. 426–447, 1994.

[281] T. Sandholm and V. Lesser, "Issues in Automated Negotiation and Electronic Commerce: Extending the Contract Net Framework," in *Proceedings of the First International Conference on Multi-Agent Systems*, June 1995, pp. 328–335.

[282] M. Sano and Y. Sawada, "Measurement of the Lyapunov Spectrum from a Chaotic Time Series," *Physical Review Letters*, vol. 55, no. 10, pp. 1082–1085, September 1985.

[283] E. Santos Jr. and D. O. Banks, "Acquiring Consistent Knowledge," Technical Report AFIT/EN/TR96-01, Air Force Institute of Technology, January 1996.

[284] T. Sauer, J. A. Yorke, and M. Casdagi, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3/4, pp. 579–616, 1991.

[285] D. W. E. Schobben, R. A. Beuker, and W. Oomen, "Dither and Data Compression," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2097–2101, August 1997.

[286] M. Shao and C. L. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications," *Proceedings of the IEEE*, vol. 81, pp. 984–1010, July 1993.

[287] M. Shao and C. L. Nikias, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, 1995.

[288] V. A. Shneidman, P. Jung, and P. Hänggi, "Power Spectrum of a Driven Bistable System," *Europhysics Letters*, vol. 26, no. 8, pp. 571–576, June 1994.

[289] S. W. Sides, R. A. Ramos, P. A. Rikvold, and M. A. Novotny, "Kinetic Ising System in an Oscillating External Field: Stochastic Resonance and Residence-Time Distributions," *Journal of Applied Physics*, vol. 81, no. 8, pp. 5597–5599, April 1997.

[290] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*, McGraw Hill, 1994.

[291] E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss, "Visual Perception of Stochastic Resonance," *Physical Review Letters*, vol. 78, no. 6, pp. 1186–1189, February 1997.

[292] C. Smith and C. Gervelis, *Cellular System Design and Optimization*, McGraw-Hill, 1996.

[293] J. C. Spall, "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation," *IEEE Transactions on Automatic Control*, vol. 36, no. 3, pp. 332–341, March 1992.

[294] M. L. Spano, M. Wun-Fogle, and W. L. Ditto, "Experimental Observation of Stochastic Resonance in a Magnetoelastic Ribbon," *Physical Review A*, vol. 46, no. 8, pp. 5253–5256, October 1992.

[295] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, "Takens Embedding Theorems for Forced and Stochastic Systems," *Nonlinear Analysis, Theory, Methods & Applications*, vol. 30, no. 8, pp. 5303–5314, 1997.

[296] L. Steels, "The Artificial Life Roots of Artificial Intelligence," in *Artificial Life: An Overview*, C. G. Langton, Ed., pp. 75–110. MIT Press, 1995.

[297] M. Stemmler, "A Single Spike Suffices: The Simplest Form of Stochastic Resonance in Model Neurons," *Network: Computation in Neural Systems*, vol. 7, pp. 687–716, 1996.

[298] N. G. Stocks, N. D. Stein, and P. V. E. McClintock, "Stochastic Resonance in Monostable Systems," *Journal of Physics A: Mathematical and General*, vol. 26, pp. L385–L390, 1993.

[299] N. G. Stocks, N. D. Stein, S. M. Soskin, and P. V. E. McClintock, "Zero-Dispersion Stochastic Resonance," *Journal of Physics A: Mathematical and General*, vol. 25, pp. L1119–L1125, 1992.

[300] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[301] F. Takens, "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence*, D. A. Rand and L.-S. Young, Eds., pp. 365–381. Springer-Verlag, 1980, Springer Lecture Notes in Mathematics, Vol. 898.

[302] M. Thorwart and P. Jung, "Dynamical Hysteresis in Bistable Quantum Systems," *Physical Review Letters*, vol. 78, no. 13, pp. 2503–2506, March 1997.

[303] P. Tsakalides and C. L. Nikias, "The Robust Covariation-Based MUSIC (ROC-MUSIC) Algorithm for Bearing Estimation in Impulsive Noise Environments," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, July 1996.

[304] G. Vemuri and R. Roy, "Stochastic Resonance in a Bistable Ring Laser," *Physical Review A*, vol. 39, no. 9, pp. 4668–4674, May 1989.

[305] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, 1995.

[306] J. M. G. Vilar and J. M. Rubí, "Divergent Signal-to-Noise Ratio and Stochastic Resonance in Monostable Systems," *Physical Review Letters*, vol. 77, no. 14, pp. 2863–2866, September 1996.

[307] J. M. G. Vilar and J. M. Rubí, "Spatiotemporal Stochastic Resonance in the Swift-Hohenberg Equation," *Physical Review Letters*, vol. 78, no. 15, pp. 2886–2889, April 1997.

[308] J. M. G. Vilar and J. M. Rubí, "Stochastic Multiresonance," *Physical Review Letters*, vol. 78, no. 15, pp. 2882–2885, April 1997.

[309] S. T. Vohra and F. Bucholtz, "Observation of Stochastic Resonance near a Subcritical Bifurcation," *Journal of Statistical Physics*, vol. 70, no. 1/2, pp. 413–421, January 1993.

[310] S. T. Vohra and L. Fabiny, "Induced Stochastic Resonance near a Subcritical Bifurcation," *Physical Review E*, vol. 50, no. 4, pp. R2391–2394, October 1994.

[311] M. P. Wellman and J. Doyle, "Preferential Semantics for Goals," in *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, July 1991, pp. 698–703.

[312] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[313] K. Wiesenfeld and F. Moss, "Stochastic Resonance and the Benefits of Noise: From Ice Ages to Crayfish and SQUIDs," *Nature*, vol. 373, pp. 33–36, January 1995.

[314] K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss, "Stochastic Resonance on a Circle," *Physical Review Letters*, vol. 72, pp. 2125–2129, April 1994.

[315] I. J. Winograd, T. B. Coplen, J. M. Landwehr, A. C. Riggs, K. R. Ludwig, B. J. Szabo, P. T. Kolesar, and K. M. Revesz, "Continuous 500,000-Year Climate Record from Vein Calcite in Devils Hole, Nevada," *Science*, vol. 258, pp. 255–260, October 1992.

[316] H. S. Wio, "Stochastic Resonance in a Spatially Extended System," *Physical Review E*, vol. 54, no. 4, pp. R3075–3078, October 1996.

[317] A. Witt, A. Neiman, and J. Kurths, "Characterizing the Dynamics of Stochastic Bistable Systems by Measures of Complexity," *Physical Review E*, vol. 55, no. 5, pp. 5050–5059, May 1997.

[318] J. K. Wu, A. D. Narasimhalu, B. M. Mehtre, and Y. J. Gao, "CORE: A Content-Based Retrieval Engine for Multimedia Information Systems," *Multimedia Systems*, vol. 3, pp. 25–41, 1995.

[319] B. Yamauchi and R. Beer, "Integrating Reactive, Sequential, and Learning Behavior Using Dynamical Neural Networks," in *Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, D. Cliff, P. Husbands, J. A. Meyer, and S. Wilson, Eds. 1994, pp. 382–391, MIT Press.

[320] W. Yang, M. Ding, and G. Hu, "Trajectory (Phase) Selection in Multistable Systems: Stochastic Resonance, Signal Bias, and the Effect of Signal Phase," *Physical Review Letters*, vol. 74, no. 20, pp. 3955–3958, May 1995.

[321] T. Zhou and F. Moss, "Analog Simulations of Stochastic Resonance," *Physical Review A*, vol. 41, no. 8, pp. 4255–4264, April 1990.

[322] T. Zhou, F. Moss, and P. Jung, "Escape-Time Distributions of a Periodically Modulated Bistable System with Noise," *Physical Review A*, vol. 42, no. 6, pp. 3161–3169, September 1990.

[323] S. Zozor and P.-O. Amblard, "Stochastic Resonance in a Discrete Time Nonlinear SE-TAR(1,2,0,0) Model," in *Proceedings of the 1997 IEEE Workshop on Higher-Order-Statistics*, July 1997.

[324] U. Zürcher and C. R. Doering, "Thermally Activated Escape over Fluctuating Barriers," *Physical Review E*, vol. 47, no. 6, pp. 3862–3869, June 1993.

Appendix A

Fuzzy Sets and Fuzzy Function Approximation

A.1 The Standard Additive Model (SAM) Theorem

This appendix derives the basic ratio structure (4.80) of a standard additive fuzzy system.

SAM Theorem. Suppose the fuzzy system $F : R^n \to R^p$ is a standard additive model: $F(x) = \text{Centroid}(B(x)) = \text{Centroid}(\sum_{j=1}^{m} w_j a_j(x) B_j)$ for if-part joint set function $a_j : R^n \to [0, 1]$, rule weights $w_j \geq 0$, and then-part fuzzy set $B_j \subset R^p$. Then $F(x)$ is a convex sum of the m then-part set centroids:

$$F(x) = \frac{\sum_{j=1}^{m} w_j a_j(x) V_j c_j}{\sum_{j=1}^{m} w_j a_j(x) V_j} = \sum_{j=1}^{m} p_j(x) c_j. \tag{A.1}$$

The convex coefficients or discrete probability weights $p_1(x), \ldots, p_m(x)$ depend on the input x through

$$p_j(x) = \frac{w_j a_j(x) V_j}{\sum_{i=1}^{m} w_i a_i(x) V_i}. \tag{A.2}$$

V_j is the finite positive volume (or area if $p = 1$) and c_j is the centroid of then-part set B_j:

$$V_j = \int_{R^p} b_j(y_1, \ldots, y_p)\, dy_1 \cdots dy_p > 0, \tag{A.3}$$

$$c_j = \frac{\int_{R^p} y\, b_j(y_1, \ldots, y_p)\, dy_1 \cdots dy_p}{\int_{R^p} b_j(y_1, \ldots, y_p)\, dy_1 \cdots dy_p}. \tag{A.4}$$

Proof. There is no loss of generality to prove the theorem for the scalar-output case $p = 1$ when $F : R^n \to R^p$. This simplifies the notation. We need but replace the scalar integrals over R with the p-multiple or volume integrals over R^p in the proof to prove the general case. The scalar case $p = 1$ gives (A.3) and (A.4) as

$$V_j = \int_{-\infty}^{\infty} b_j(y)\, dy \tag{A.5}$$

$$c_j = \frac{\int_{-\infty}^{\infty} y\, b_j(y)\, dy}{\int_{-\infty}^{\infty} b_j(y)\, dy}. \tag{A.6}$$

Then the theorem follows if we expand the centroid of B and invoke the SAM assumption $F(x) = \text{Centroid}(B(x)) = \text{Centroid}(\sum_{j=1}^{m} w_j\, a_j(x)\, B_j)$ to rearrange terms:

$$F(x) = \text{Centroid}(B(x)) \tag{A.7}$$

$$= \frac{\int_{-\infty}^{\infty} y\, b(y)\, dy}{\int_{-\infty}^{\infty} b(y)\, dy} \tag{A.8}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^{m} w_j \, b'_j(y) \, dy}{\int_{-\infty}^{\infty} \sum_{j=1}^{m} w_j \, b'_j(y) \, dy} \tag{A.9}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^{m} w_j \, a_j(x) \, b_j(y) \, dy}{\int_{-\infty}^{\infty} \sum_{j=1}^{m} w_j \, a_j(x) \, b_j(y) \, dy} \tag{A.10}$$

$$= \frac{\sum_{j=1}^{m} w_j \, a_j(x) \int_{-\infty}^{\infty} y \, b_j(y) \, dy}{\sum_{j=1}^{m} w_j \, a_j(x) \int_{-\infty}^{\infty} b_j(y) \, dy} \tag{A.11}$$

$$= \frac{\sum_{j=1}^{m} w_j \, a_j(x) \, V_j \, \frac{\int_{-\infty}^{\infty} y \, b_j(y) \, dy}{V_j}}{\sum_{j=1}^{m} w_j \, a_j(x) \, V_j} \tag{A.12}$$

$$= \frac{\sum_{j=1}^{m} w_j \, a_j(x) \, V_j \, c_j}{\sum_{j=1}^{m} w_j \, a_j(x) \, V_j}. \tag{A.13}$$

Generalizing the SAM system leads to the *set* SAM F that maps fuzzy sets A in the input space R^n to vector points y in the output space R^p. So the set SAM $F : F(2^{R^n}) \rightarrow R^p$ has as its domain the fuzzy power set $F(2^{R^n})$ or the set of all fuzzy subsets $A \subset R^n$ with arbitrary set function $a : R^n \rightarrow [0, \infty)$. The point SAM is a special case of the set SAM for a singleton input fuzzy set $A = \{x_0\} \subset R^n$: $a(x) = \delta(x - x_0)$ where δ is a Dirac delta function in the continuous case or a unit bit vector in the discrete case. Correlation computes the "fired" fit value of the jth set $a_j(A)$ as [162]

$$a_j(A) = \int a(x) \, a_j(x) \, dx. \tag{A.14}$$

Then the fired fit value $a_j(x_0)$ of the singleton set $A = \{x_0\}$ follows from the sifting property of

delta pulses:

$$a_j(A) = \int a(x)\, a_j(x)\, dx \tag{A.15}$$

$$= \int \delta(x - x_0)\, a_j(x)\, dx \tag{A.16}$$

$$= a_j(x_0). \tag{A.17}$$

The set SAM equation follows from the SAM additive combiner $B(A) = \sum_{j=1}^{m} w_j\, a_j(A)\, B_j$ [162]:

$$F(A) = \text{Centroid}(B(A)) \tag{A.18}$$

$$= \text{Centroid}\Big(\sum_{j=1}^{m} w_j\, a_j(A)\, B_j\Big) \tag{A.19}$$

$$= \frac{\sum_{j=1}^{m} w_j\, a_j(A)\, V_j\, c_j}{\sum_{j=1}^{m} w_j\, a_j(A)\, V_j} = \sum_{j=1}^{m} p_j(A)\, c_j. \tag{A.20}$$

where the convex coefficients $p_1(A), \ldots, p_m(A)$ depend on the input fuzzy set A through

$$p_j(A) = \frac{w_j\, a_j(A)\, V_j}{\sum_{i=1}^{m} w_i\, a_i(A)\, V_i}. \tag{A.21}$$

A.2 Supervised SAM Learning

Supervised gradient descent can tune all the parameters in the SAM model (4.80) [160, 162]. A gradient ascent learning law for a SAM parameter ξ has the form

$$\xi(t+1) = \xi(t) + \mu_t \frac{\partial P}{\partial \xi} \tag{A.22}$$

where μ_t is a learning rate at iteration t. We seek to maximize the performance measure P of the dynamical system $\dot{q} = h(q, u)$. Here the signal-to-noise ratio (SNR) defines the performance P. Other systems can use different measures such as cross-correlation or mean-squared error.

Let ξ_j^k denote the kth parameter in the set function a_j. Then the chain rule gives the gradient of the SNR with respect to ξ_j^k, with respect to the then-part set centroid c_j, and with respect to the then-part set volume V_j:

$$\frac{\partial \text{SNR}}{\partial \xi_j^k} = \frac{\partial \text{SNR}}{\partial F} \frac{\partial F}{\partial a_j} \frac{\partial a_j}{\partial \xi_j^k}, \quad \frac{\partial \text{SNR}}{\partial c_j} = \frac{\partial \text{SNR}}{\partial F} \frac{\partial F}{\partial c_j}, \quad \text{and} \quad \frac{\partial \text{SNR}}{\partial V_j} = \frac{\partial \text{SNR}}{\partial F} \frac{\partial F}{\partial V_j}. \tag{A.23}$$

We have derived the partial derivative $\frac{\partial \text{SNR}}{\partial F} = \frac{\partial \text{SNR}}{\partial \sigma}$ in Section 4.1. We next derive the partial derivatives for the SAM parameters:

$$\frac{\partial F}{\partial a_j} = \frac{\left(\sum_{i=1}^{m} w_i\, a_i(x)\, V_i\right)(w_j\, V_j\, c_j) - w_j\, V_j \left(\sum_{i=1}^{m} w_i\, a_i(x)\, V_i\, c_i\right)}{\left(\sum_{i=1}^{m} w_i\, a_i(x)\, V_i\right)^2} \tag{A.24}$$

$$= \frac{[c_j - F(x)]\, w_j\, V_j}{\sum_{i=1}^{m} w_i\, a_i(x)\, V_i} = [c_j - F(x)] \frac{p_j(x)}{a_j(x)}. \tag{A.25}$$

The SAM ratio (4.80) gives [160]

$$\frac{\partial F}{\partial c_j} = \frac{w_j\, a_j(x)\, V_j}{\sum_{i=1}^{m} w_i\, a_i(x)\, V_i} = p_j(x) \tag{A.26}$$

$$\text{and} \quad \frac{\partial F}{\partial V_j} = \frac{w_j\, a_j(x)\, [c_j - F(x)]}{\sum_{i=1}^{m} w_i\, a_i(x)\, V_i} = \frac{p_j(x)}{V_j} [c_j - F(x)]. \tag{A.27}$$

Then the learning laws for the centroid and volume have the final form

$$c_j(t+1) = c_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} p_j(x) \tag{A.28}$$

$$\text{and} \quad V_j(t+1) = V_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma} \frac{p_j(x)}{V_j} [c_j - F(x)]. \tag{A.29}$$

Learning laws for set parameters depend on how we define the set functions. The partial derivatives for the scalar sinc set function $a_j(x) = \sin\left(\frac{x-m_j}{d_j}\right)\Big/\left(\frac{x-m_j}{d_j}\right)$ have the form

$$\frac{\partial a_j}{\partial m_j} = \begin{cases} \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{x-m_j} & \text{for } x \neq m_j \\ 0 & \text{for } x = m_j \end{cases} \tag{A.30}$$

$$\frac{\partial a_j}{\partial d_j} = \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{d_j}. \tag{A.31}$$

So this scalar set function leads to the learning laws

$$m_j(t+1) = m_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma}[c_j - F(x)]\frac{p_j(x)}{a_j(x)}\left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{x-m_j} \tag{A.32}$$

$$d_j(t+1) = d_j(t) + \mu_t \frac{\partial \text{SNR}}{\partial \sigma}[c_j - F(x)]\frac{p_j(x)}{a_j(x)}\left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{d_j}. \tag{A.33}$$

The application in Appendix B applies the SAM system to minimize the squared-error $E(x) = \frac{1}{2}(f(x) - F(x))^2$ of the function approximation. So we use a gradient *descent* learning law for a SAM parameter ξ

$$\xi(t+1) = \xi(t) - \mu_t \frac{\partial E}{\partial \xi}. \tag{A.34}$$

Then the chain rule gives the gradient of the error function with respect to ξ_j^k, with respect to the then-part set centroid c_j, and with respect to the then-part set volume V_j as

$$\frac{\partial E}{\partial \xi_j^k} = \frac{\partial E}{\partial F}\frac{\partial F}{\partial a_j}\frac{\partial a_j}{\partial \xi_j^k}, \quad \frac{\partial E}{\partial c_j} = \frac{\partial E}{\partial F}\frac{\partial F}{\partial c_j}, \quad \text{and} \quad \frac{\partial E}{\partial V_j} = \frac{\partial E}{\partial F}\frac{\partial F}{\partial V_j}. \tag{A.35}$$

We have derived the partial derivatives $\frac{\partial F}{\partial a_j}$, $\frac{\partial F}{\partial c_j}$, and $\frac{\partial F}{\partial V_j}$ in (A.25),(A.26), and (A.27). We need only derive the partial derivative $\frac{\partial E}{\partial F}$

$$\frac{\partial E}{\partial F} = -(f(x) - F(x)) = -\varepsilon(x) \tag{A.36}$$

Then the learning laws for the centroid and volume have the form

$$c_j(t+1) = c_j(t) + \mu_t\, \varepsilon(x)\, p_j(x) \tag{A.37}$$

$$\text{and} \quad V_j(t+1) = V_j(t) + \mu_t\, \varepsilon(x)\, \frac{p_j(x)}{V_j}\, [c_j - F(x)]. \tag{A.38}$$

The learning laws for set parameters follow in like manner. For example the scalar sinc set function has the learning laws

$$m_j(t+1) = m_j(t) + \mu_t\, \varepsilon(x)\, [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \left(a_j(x) - \cos\left(\frac{x - m_j}{d_j}\right)\right) \frac{1}{x - m_j} \tag{A.39}$$

$$d_j(t+1) = d_j(t) + \mu_t\, \varepsilon(x)\, [c_j - F(x)] \frac{p_j(x)}{a_j(x)} \left(a_j(x) - \cos\left(\frac{x - m_j}{d_j}\right)\right) \frac{1}{d_j}. \tag{A.40}$$

The partial derivatives of the scalar Laplace set function $a_j(x) = \exp\left\{-\left|\frac{x - m_j}{d_j}\right|\right\}$ with respect to its two parameters m_j and d_j have the form

$$\frac{\partial a_j}{\partial m_j} = \text{sign}(x - m_j) \frac{1}{|d_j|} a_j(x) \tag{A.41}$$

$$\frac{\partial a_j}{\partial m_j} = \text{sign}(d_j) \frac{|x - m_j|}{|d_j^2|} a_j(x) \tag{A.42}$$

where we define the sign function as

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{A.43}$$

Substitute (A.41)–(A.42) in (A.35) and in (A.34) to obtain the learning laws

$$m_j(t+1) = m_j(t) + \mu_t \varepsilon(x)\, [c_j - F(x)]\, p_j(x)\, \text{sign}(x - m_j) \frac{1}{|d_j|} \tag{A.44}$$

$$d_j(t+1) = d_j(t) + \mu_t \varepsilon(x)\, [c_j - F(x)]\, p_j(x)\, \text{sign}(d_j)\, \frac{|x - m_j|}{d_j^2}. \tag{A.45}$$

Like results hold for the learning laws of product n-D set functions. A factored set function $a_j(x) = a_j^1(x_1) \cdots a_j^n(x_n)$ leads to a new form for the error gradient. The gradient with respect to the parameter m_j^k of the jth set function a_j has the form

$$\frac{\partial E}{\partial m_j^k} = \frac{\partial E}{\partial F}\frac{\partial F}{\partial a_j}\frac{\partial a_j}{\partial a_j^k}\frac{\partial a_j^k}{\partial m_j^k} \quad \text{where} \quad \frac{\partial a_j}{\partial a_j^k} = \prod_{i \neq k}^{n} a_j^i(x_i) = \frac{a_j(x)}{a_j^k(x_k)}. \tag{A.46}$$

Products of the scalar sinc set functions defined the if-part fuzzy sets $A_j \subset R^n$ in the SAM approximator for adaptive stochastic resonance simulations in Chapter 5.

The application in Appendix B used product of the scalar sinc set function to define the if-part fuzzy set in the fuzzy profile approximator and used product of the scalar Laplace set function for the fuzzy equality measure. But it used the set SAM system instead of the simple point SAM. The learning laws follow from the structure of the set SAM.

We now derive learning laws for the set SAM. The chain-rule terms in (A.35) become

$$\frac{\partial E}{\partial \xi_j^k}(A) = \frac{\partial E}{\partial F}(A)\frac{\partial F}{\partial a_j}(A)\frac{\partial a_j}{\partial \xi_j^k}(A) \tag{A.47}$$

$$\frac{\partial E}{\partial c_j}(A) = \frac{\partial E}{\partial F}(A)\frac{\partial F}{\partial c_j}(A) \tag{A.48}$$

$$\frac{\partial E}{\partial V_j}(A) = \frac{\partial E}{\partial F}(A)\frac{\partial F}{\partial V_j}(A). \tag{A.49}$$

Then (A.25)–(A.27), and (A.36) give

$$\frac{\partial E}{\partial F}(A) = -(f(A) - F(A)) = -\varepsilon(A) \tag{A.50}$$

$$\frac{\partial F}{\partial a_j}(A) = [c_j - F(A)]\frac{p_j(A)}{a_j(A)} \tag{A.51}$$

$$\frac{\partial F}{\partial c_j}(A) = p_j(A) \tag{A.52}$$

$$\frac{\partial F}{\partial V_j}(A) = \frac{p_j(A)}{V_j}[c_j - F(A)]. \tag{A.53}$$

The learning laws for then-part set centroids c_j and volumes V_j are

$$c_j(t+1) = c_j(t) + \mu_t \varepsilon(A) p_j(A) \tag{A.54}$$

$$V_j(t+1) = V_j(t) + \mu_t \varepsilon(A) \frac{p_j(A)}{V_j} [c_j - F(A)]. \tag{A.55}$$

But the partial derivative of the jth set function with respect to its parameters ξ_j^k has the new form

$$\frac{\partial a_j}{\partial \xi_j^k}(A) = \frac{\partial}{\partial \xi_j^k} \int a_j(x)\, a(x)\, dx \tag{A.56}$$

$$= \begin{cases} \int \frac{\partial a_j}{\partial \xi_j^k}(x)\, a(x)\, dx & \text{continuous case} \\ \sum \frac{\partial a_j}{\partial \xi_j^k}(x)\, a(x) & \text{discrete case} \end{cases} \tag{A.57}$$

Then we substitute these partial derivatives into (A.34) to obtain the set-SAM learning rules.

A.3 Sets as Points: The Geometry of Discrete Fuzzy Sets

This appendix reviews the unit-cube geometry of discrete fuzzy system and derive the new adaptive equality measure. Let X be a set of n elements: $X = \{x_1, \ldots, x_n\}$. Any subset $A \subset X$ defines a point in the n-D unit hypercube $I^n = [0,1]^n$. The set of all fuzzy subsets of X or $F(2^X)$ fill in the cube. So the ordinary power set 2^X or the set of all 2^n subsets of X equals the Boolean n-cube $B^n : 2^X = B^n$. Fuzzy subsets $A \subset X$ define the points *inside* or on the n-D unit hypercube [158, 162] as in Figure A.1. A set $A \subset X$ is fuzzy when the "laws" of noncontradiction and excluded middle do not hold: $A \cap A^c \neq \emptyset$ and $A \cup A^c \neq X$.

Figure A.1 shows an example when $X = \{x_1, x_2\}$. Then there are 4 binary subsets of $X : 2^X = \{\emptyset, \{x_1\}, \{x_2\}, \{x_1, x_2\}\}$. The space $X = \{x_1, x_2\}$ lies at (1, 1). The empty set $\emptyset$ lies at the origin (0, 0) and the other two (standard) subsets $\{x_1\}$ and $\{x_2\}$ are at (1, 0) and (0, 1). A fuzzy subset $A \subset X$ defines the *fuzzy unit* or *fit* vectors $A = (a_1, a_2) \in I^2$ for

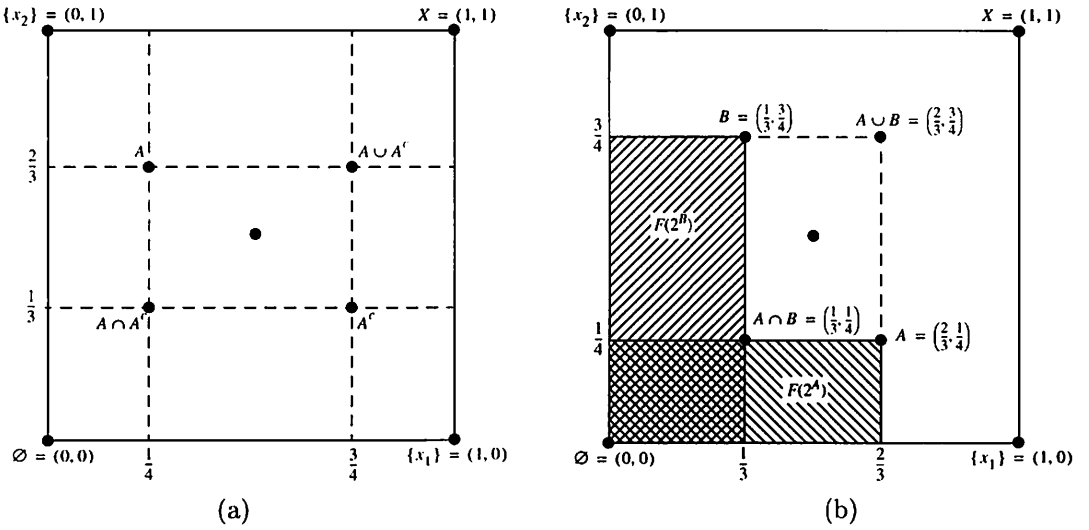


Figure A.1: Geometry of discrete fuzzy sets. Sets as points in a unit hypercube or fuzzy cube. Fuzzy set $A \subset X = \{x_1, \ldots, x_n\}$ defines a point in the fuzzy cube $[0,1]^n$. Here $X = \{x_1, x_2\}$, $A = (\frac{2}{3}, \frac{1}{4})$, and $B = (\frac{1}{3}, \frac{3}{4})$. We define fuzzy-set intersection fitwise with pairwise minimum, union with pairwise maximum, and complementation with order reversal ($a^c(x) = 1 - a(x)$). $F(2^A)$ and $F(2^B)$ define the fuzzy power sets or the sets of of all fuzzy subsets of A and B. Each set $C \subset X$ is a subset of A to some degree and so C belongs to $F(2^A)$ to some degree. C is a 100% subset of A if and only if $c(x) \leq a(x)$ for all $x \in X$. Then $C \in F(2^A)$ and so the set point C lies on or inside the hyper-rectangle $F(2^A)$. Partial subsets lie outside $F(2^A)$.

$a_1, a_2 \in [0,1]$. Figure A.1(a) shows an example of a fuzzy set A. The geometrical view reveals the 2^n-fold symmetry of the set A and its set operation products with respect to the midpoint. The midpoint is the maximal fuzzy set. It alone obeys $A = A^c$. The midpoint alone has spherical symmetry and lies equidistant to all 2^n cube vertices.

Figure A.1(b) shows the 2-D cube with the fuzzy sets $A = (\frac{2}{3}, \frac{1}{4})$ and $B = (\frac{1}{3}, \frac{3}{4})$. We can define fuzzy-set intersection fitwise with pairwise minimum, union with pairwise maximum, and complementation with order reversal:

$$a \cap b(x) = \min(a(x), b(x)) \tag{A.58}$$

$$a \cup b(x) = \max(a(x), b(x)) \tag{A.59}$$

$$a^c(x) = 1 - a(x) \tag{A.60}$$

The subsethood theorem [158] measures the degree to which a set A contains in a set B and

does so in a simple ratio of cardinalities:

$$S(A,B) \;=\; \text{Degree}(A \subset B) \;=\; \frac{c(A \cap B)}{c(A)} \tag{A.61}$$

where c is a *counting* or *cardinality* [158] measure

$$c(A) \;=\; \sum_{x_i \in X} a(x_i) \qquad \text{or} \qquad c(A) \;=\; \int_X a(x)\,dx \tag{A.62}$$

for integrable fuzzy set function $a : X \to [0,1]$. This positive measure stems from the geometric interpretation of the fuzzy power sets $F(2^A)$ and $F(2^B)$ [158, 162]. The subsethood measure extends the histogram intersection in (B.21). The subsethoods need not be symmetric: $S(A,B) \neq S(B,A)$. So we use a new symmetric measure [162] of fuzzy equality as in (B.9) :

$$\mathcal{E}(A,B) \;=\; \text{Degree}(A = B) \;=\; \frac{c(A \cap B)}{c(A \cup B)} \tag{A.63}$$

$$\;=\; \frac{S(A,B)\,S(B,A)}{S(A,B) + S(B,A) - S(A,B)\,S(B,A)}. \tag{A.64}$$

Then we use the identities $\min(a,b) = \frac{1}{2}(a + b - |a - b|)$ and $\max(a,b) = \frac{1}{2}(a + b + |a - b|)$ to derive (A.68):

$$\mathcal{E}(A,B) \;=\; \frac{c(A \bigcap B)}{c(A \bigcup B)} \;=\; \frac{\int \min(a(x), b(x))\,dx}{\int \max(a(x), b(x))\,dx} \tag{A.65}$$

$$\;=\; \frac{\int a(x) + b(x) - |a(x) - b(x)|\,dx}{\int a(x) + b(x) + |a(x) - b(x)|\,dx} \tag{A.66}$$

$$\;=\; \frac{1 - \dfrac{\int |a(x) - b(x)|\,dx}{\int a(x) + b(x)\,dx}}{1 + \dfrac{\int |a(x) - b(x)|\,dx}{\int a(x) + b(x)\,dx}} \tag{A.67}$$

$$\;=\; \frac{1 - \bar{d}(A,B)}{1 + \bar{d}(A,B)} \tag{A.68}$$

where fuzzy set $A \subset R^n$ has set function $a : R^n \to [0,1]$ and $B \subset R^n$ has set function $b : R^n \to [0,1]$ and

$$\|A - B\| = \int |a(x) - b(x)|\, dx \tag{A.69}$$

$$\|A + B\| = \int |a(x) + b(x)|\, dx \tag{A.70}$$

$$\text{and} \quad \bar{d}(A,B) = \frac{\|A - B\|}{\|A + B\|} = \frac{\int |a(x) - b(x)|\, dx}{\int |a(x) + b(x)|\, dx}. \tag{A.71}$$

Sums can replace the integrals in the discrete case.

We next derive a supervised learning law to tune the parameters of the set functions. Square error for a desired matching value D has the form $E = \frac{1}{2}(D - \mathcal{E})^2$. The chain rule gives the derivative of the squared error with respect to the kth parameter of the jth set function m_j^k as

$$\frac{\partial E}{\partial m_j^k} = \frac{\partial E}{\partial \mathcal{E}} \frac{\partial \mathcal{E}}{\partial \bar{d}} \frac{\partial \bar{d}}{\partial m_j^k}. \tag{A.72}$$

The derivatives have the form

$$\frac{\partial E}{\partial \mathcal{E}} = -[D(A,B) - \mathcal{E}(A,B)] \tag{A.73}$$

$$\frac{\partial \mathcal{E}}{\partial \bar{d}} = -\frac{1 + \mathcal{E}(A,B)}{1 + \bar{d}(A,B)} \tag{A.74}$$

$$\frac{\partial \bar{d}}{\partial m_j^k} = \frac{1}{\|A + B\|} \Big(\frac{\partial}{\partial m_j^k} \|A - B\| - \bar{d}(A,B) \frac{\partial}{\partial m_j^k} \|A + B\| \Big). \tag{A.75}$$

We now derive the derivatives of the "norms" $\|A - B\|$ and $\|A + B\|$ for the discrete sets A and B with respect to the parameter m_j^k in our image matching problems. The result follows from equations (B.14)-(B.15) and

$$\bar{d}(A,B) = \frac{\|A - B\|}{\|A + B\|} = \frac{\sum_{i=1}^{N} |A(\bar{x}_i) - B(\bar{x}_i)|}{\sum_{i=1}^{N} |A(\bar{x}_i) + B(\bar{x}_i)|} \tag{A.76}$$

and the assumption that each set has its own independent parameters (so $\frac{\partial a_i}{\partial m_j^k} = 0$ for $i \neq j$) :

$$
\begin{aligned}
&\frac{\partial}{\partial m_j^k}\|A - B\| \\
&= \frac{\partial}{\partial m_j^k}\sum_{i=1}^{N} |A(\bar{x}_i) - B(\bar{x}_i)| && \text{(A.77)} \\
&= \sum_{i=1}^{N} \text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\frac{\partial}{\partial m_j^k}(A(\bar{x}_i) - B(\bar{x}_i)) && \text{(A.78)} \\
&= \sum_{i=1}^{N} \Big[\text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\frac{\partial}{\partial m_j^k}\Big(\sum_{l=1}^{m} A_l\, a_l(\bar{x}_i) - \sum_{l=1}^{m} B_l\, a_l(\bar{x}_i)\Big)\Big] && \text{(A.79)} \\
&= \sum_{i=1}^{N} \Big[\text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\frac{\partial}{\partial m_j^k}\Big(\sum_{l=1}^{m} a_l(\bar{x}_i)\,(a_l(T_A) - a_l(T_B))\Big)\Big] && \text{(A.80)} \\
&= \sum_{i=1}^{N} \Big[\text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\sum_{l=1}^{m}\Big(\frac{\partial}{\partial m_j^k} a_l(\bar{x}_i)\,(a_l(T_A) - a_l(T_B))\Big)\Big] && \text{(A.81)} \\
&= \sum_{i=1}^{N} \text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\Big[[a_j(T_A) - a_j(T_B)]\frac{\partial a_j(\bar{x}_i)}{\partial m_j^k} \\
&\qquad + a_j(\bar{x}_i)\Big(\frac{\partial a_j(T_A)}{\partial m_j^k} - \frac{\partial a_j(T_B)}{\partial m_j^k}\Big)\Big] && \text{(A.82)} \\
&= [a_j(T_A) - a_j(T_B)]\sum_{i=1}^{N}\text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\,\frac{\partial a_j(\bar{x}_i)}{\partial m_j^k} \\
&\quad + \Big(\frac{\partial a_j(T_A)}{\partial m_j^k} - \frac{\partial a_j(T_B)}{\partial m_j^k}\Big)\sum_{i=1}^{N}\text{sign}(A(\bar{x}_i) - B(\bar{x}_i))\,a_j(\bar{x}_i) && \text{(A.83)}
\end{aligned}
$$

The derivation proceeds in like manner for $\frac{\partial}{\partial m_j^k}\|A + B\|$ as

$$\frac{\partial}{\partial m_j^k}\|A + B\| = [a_j(T_A) + a_j(T_B)]\sum_{i=1}^{N}\frac{\partial a_j(\bar{x}_i)}{\partial m_j^k} + \Big(\frac{\partial a_j(T_A)}{\partial m_j^k} + \frac{\partial a_j(T_B)}{\partial m_j^k}\Big)\sum_{i=1}^{N} a_j(\bar{x}_i) \quad \text{(A.84)}$$

since $a(x) \geq 0$ for all $x \in X$. The condition $a_j(T_A) = \sum_{i=1}^{N} t_A^i\, a_j(\bar{x}_i)$ and $a_j(T_B) = \sum_{i=1}^{N} t_B^i\, a_j(\bar{x}_i)$ from (B.8) gives

$$\frac{\partial a_j}{\partial m_j^k}(T_A) = \sum_{i=1}^{N} t_A^i\,\frac{\partial a_j}{\partial m_j^k}(\bar{x}_i) \quad \text{(A.85)}$$

$$\frac{\partial a_j}{\partial m_j^k}(T_B) \;=\; \sum_{i=1}^{N} t_B^i \, \frac{\partial a_j}{\partial m_j^k}(\bar{x}_i). \tag{A.86}$$

Appendix A.2 derives the partial derivatives of the Laplace set function a_j with respect to its two parameters in equations (A.41)-(A.42). Then substitute (A.85)-(A.86) into (A.83)-(A.84) to obtain (A.75) and substitute (A.73)-(A.75) to obtain (A.72) and the learning law for each parameter in the form of (A.34):

$$\begin{aligned}
m_j^k(t+1) \;=\; & m_j^k(t) - \mu_t \Big(D(A,B) - \mathcal{E}(A,B) \Big) \frac{1+\mathcal{E}(A,B)}{1-\bar{d}(A,B)} \frac{1}{\|A+B\|} \times \\
& \Bigg[\Big(a_j(T_A) - a_j(T_B) \Big) \sum_{i=1}^{N} \operatorname{sign}\Big(A(\bar{x}_i) - B(\bar{x}_i) \Big) \operatorname{sign}(\bar{x}_i - m_j^k) \frac{1}{|d_j^k|} a_j(\bar{x}_i) \\
& + \Big(\sum_{i=1}^{N} (t_A^i - t_B^i) \operatorname{sign}(\bar{x}_i - m_j^k) \frac{1}{|d_j^k|} a_j(\bar{x}_i) \Big) \sum_{i=1}^{N} \operatorname{sign}\Big(A(\bar{x}_i) - B(\bar{x}_i) \Big) a_j(\bar{x}_i) \\
& - \bar{d}(A,B) \Bigg(\Big(a_j(T_A) - a_j(T_B) \Big) \sum_{i=1}^{N} \operatorname{sign}(\bar{x}_i - m_j^k) \frac{1}{|d_j^k|} a_j(\bar{x}_i) \\
& \quad + \Big(\sum_{i=1}^{N} (t_A^i - t_B^i) \operatorname{sign}(\bar{x}_i - m_j^k) \frac{1}{|d_j^k|} a_j(\bar{x}_i) \Big) \sum_{i=1}^{N} a_j(\bar{x}_i) \Bigg) \Bigg]
\end{aligned} \tag{A.87}$$

$$\begin{aligned}
d_j^k(t+1) \;=\; & d_j^k(t) - \mu_t \Big(D(A,B) - \mathcal{E}(A,B) \Big) \frac{1+\mathcal{E}(A,B)}{1-\bar{d}(A,B)} \frac{1}{\|A+B\|} \times \\
& \Bigg[\Big(a_j(T_A) - a_j(T_B) \Big) \sum_{i=1}^{N} \operatorname{sign}\Big(A(\bar{x}_i) - B(\bar{x}_i) \Big) \operatorname{sign}(d_j^k) \frac{|\bar{x}_i - m_j^k|}{|d_j^k|^2} a_j(\bar{x}_i) \\
& + \Big(\sum_{i=1}^{N} (t_A^i - t_B^i) \operatorname{sign}(d_j^k) \frac{|\bar{x}_i - m_j^k|}{|d_j^k|^2} a_j(\bar{x}_i) \Big) \sum_{i=1}^{N} \operatorname{sign}\Big(A(\bar{x}_i) - B(\bar{x}_i) \Big) a_j(\bar{x}_i) \\
& - \bar{d}(A,B) \Bigg(\Big(a_j(T_A) - a_j(T_B) \Big) \sum_{i=1}^{N} \operatorname{sign}(d_j^k) \frac{|\bar{x}_i - m_j^k|}{|d_j^k|^2} a_j(\bar{x}_i) \\
& \quad + \Big(\sum_{i=1}^{N} (t_A^i - t_B^i) \operatorname{sign}(d_j^k) \frac{|\bar{x}_i - m_j^k|}{|d_j^k|^2} a_j(\bar{x}_i) \Big) \sum_{i=1}^{N} a_j(\bar{x}_i) \Bigg) \Bigg].
\end{aligned} \tag{A.88}$$

Appendix B

Neural Fuzzy Agents for Profile Learning and Adaptive Object Matching

A neural fuzzy system can learn an agent profile of a user when it samples user question-answer data. A fuzzy system uses if-then rules to store and compress the agent's knowledge of the user's likes and dislikes. A neural system uses training data to form and tune the rules. The profile is a preference map or a bumpy utility surface defined over the space of search objects. Rules define fuzzy patches that cover the surface bumps as learning unfolds and as the fuzzy agent system gives a finer approximation of the profile. The agent system searches for preferred objects with the learned profile and with a new fuzzy measure of similarity. The appendix derives the supervised learning law that tunes this matching measure with fresh sample data. We test the fuzzy agent profile system on object spaces of flowers and sunsets and test the fuzzy agent matching system on an object space of sunset images. Rule explosion and data acquisition impose fundamental limits on the system designs.

B.1 Smart Agents: Profile Learning and Object Matching

How can we teach an agent what we like and dislike? How can an agent search new databases on our behalf? These are core questions for both human agents and intelligent software agents. We explore these questions with the joint tools of fuzzy rule-based systems and neural learning. These tools exploit the filter and set-theoretic structure of agent search.

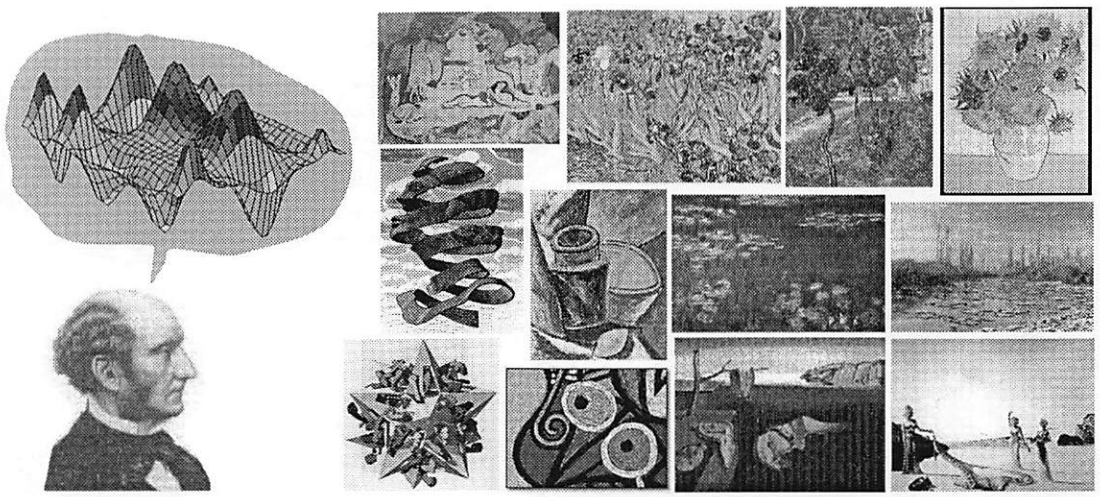


Figure B.1: Profile learning. A neural fuzzy agent learns a user's utility surface as the user samples a database of classic paintings. The 12 bumps or extrema on the preference map show how much the user (or the agent who acts on the user's behalf) likes or dislikes the 12 paintings. Here the evolving utility surface forms in the "mind's eye" of a neural fuzzy agent based on nineteenth-century English philosopher John Stuart Mill.

An intelligent agent can act as a smart database filter [113, 190]. The agent can search a database or search a space of objects on behalf of its user. The agent can find and retrieve objects that the user likes. Or the agent can find and then ignore or delete objects that the user does not like. Or it can perform some mix of both. The agent acts as a filter because it maps a set of objects to one or more of its subsets. The agent is "smart" [21, 191, 296] to the degree that it can quickly and accurately learn the user's tastes or object profile and to the degree that it can use that profile map to search for and to rank preferred objects. Figure B.1 shows how an agent can learn and store user tastes as a bumpy preference surface defined over search objects.

Agent search depends on set structure in a still deeper way. The search system itself may have many parts to its design and may perform many functions in many digital venues [55, 319]. But at some abstract level the agent partitions the object space into two fuzzy or multivalued sets with blurred borders. The agent partitions the space into the fuzzy set of objects that it assumes the users likes and into the complement fuzzy set of objects that it assumes the user does not like. All search objects belong to both of these fuzzy sets to some degree. Then the agent can rank some or all of the objects in the preferred set and can pick some of the extremal objects as its output set.

The agent needs a profile of its user so that it can group objects and rank them. The agent must somehow learn what patterns of objects the user likes or dislikes and to what degree he likes or dislikes them [192, 265]. This profile is some form of the user's implicit preference map. The user may state part of this map in ordinal terms: "I like these red flowers more than I like those blue flowers. I like the large purple flowers about the same as I like the small red-white flowers." The objects may be fuzzy patterns or fuzzy clusters in some feature space [166, 244, 245].

Microeconomic theory ensures that under certain technical conditions these complete ordinal rankings define a numerical utility function. The utility function is unique up to a linear transformation [64, 121, 241]. So we can in theory replace the ordinal claim "I like object A at least as much as I like object B" with some cardinal relation $u(A) \geq u(B)$ and vice versa. The utility function $u : \mathcal{O} \rightarrow R$ converts the ordinal preference structure into a numerical utility surface in an object space $\mathcal{O}$ of low or high dimension [64, 121, 241]. The user likes the surface's peak objects and dislikes its valley objects.

We use neural fuzzy systems to learn the user's profile or utility surface as a set of adaptive fuzzy if-then rules. The rules compress the profile into modular units. The rules grow the profile from a first set of sample data or question-answer queries and change the profile's shape as the agent samples more preference data. The modular structure of the rules lets the user add or delete knowledge chunks or heuristics.

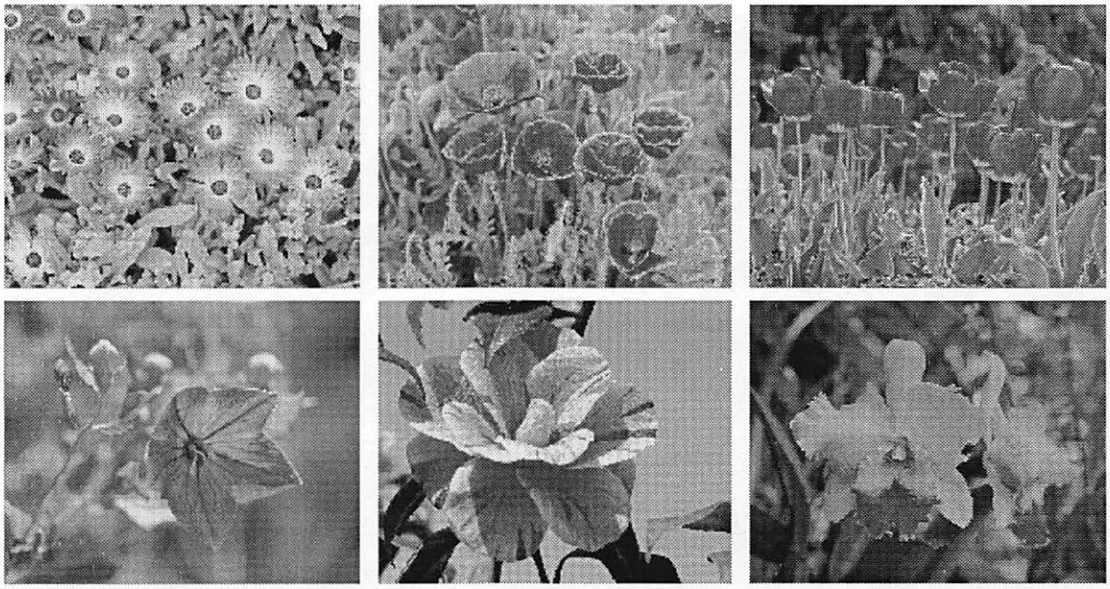


Figure B.2: Search Objects. Samples of flower images in the test database. (With permission: Hitachi Viewseum, Copyright ©1995, 1996, 1997, Hitachi, America, Ltd. All rights reserved.)

These fuzzy systems are universal approximators [159] but they suffer from exponential rule explosion in high dimension [161]. Their first set of rules give a quick but rough approximation of the user's profile. Each rule defines a fuzzy patch or subset of the object space (or product object space). Mean-square optimal rules cover the extrema or bumps of the profile surface [161]. Then other rule patches tend to quickly fill in between these bumps as learning unfolds. Figure B.2 shows some of the flower test images we used to form a 4-D feature space of objects. Figure 4.8 shows how a neural fuzzy system with 100 rules approximates a 2-D profile surface. The utility profiles grow finer as the user states more numerical ranks for test objects or pattern clusters. Rule explosion remains the chief limit to this approach.

We also combine neural learning and fuzzy set theory to search for preferred objects. We cast this search problem as one of fuzzy similarity matching and define a new measure for the task and show how supervised learning updates this measure. The user gives the system matching degrees in the unit interval for a test space of sunset images. Supervised gradient descent tunes the measure and defines a similarity surface over the sunset object space. Similar objects have nearly the same utility but objects with the same utility need not be similar. Other systems

might combine the "smart" techniques of fuzzy profile learning with fuzzy object matching to aid in the agent search process.

B.2 Agent Architecture

Figure B.3 shows our schematic view of an intelligent agent. The agent can reside in a physical world (robot) or in a virtual world (softbot) [139, 193]. The interface/sensor module transforms the information into a bit stream. The preprocessor compresses the pattern of objects or actions. The compressed patterns might be colors or textures used in image search or filtering [233, 300], keywords used in text search or e-mail classifiers or news-filtering agents [190], or object features that agents use if they bargain or negotiate [46, 252, 267, 274, 281].

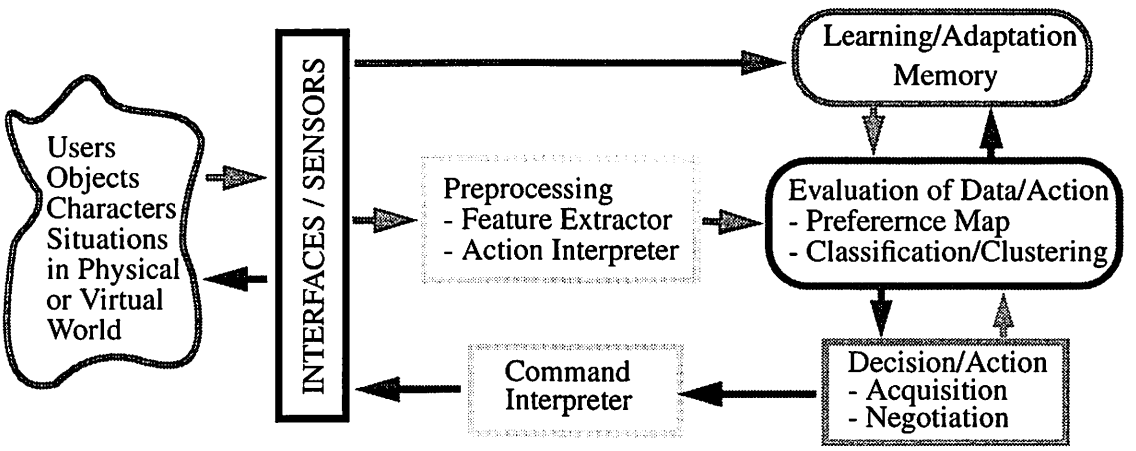


Figure B.3: Agent environment. Schematic view of an autonomous agent in a physical or virtual world. The agent interacts with objects or characters in the environment and adapts itself to better execute its goals.

A learning and memory module records the compressed patterns of the utility surface. The surface changes over time as the user gives more Q & A samples. This gives a bumpy surface that tends to better and better match the user's underlying preference map.

The decision maker module receives the data from the evaluation module and then decides what to do [192]. A classifier agent sends the control signal to that class to which the object

belongs [190]. Then an agent must decide which step to take next. The agent may need to bargain or negotiate with other agents [46, 267].

This paper deals largely with the block that computes the "value" or "worth" of an object or action. The preference map $u : \mathcal{O} \rightarrow R$ defines the value of each object. The user prefers object O_1 to object O_2 (or $O_1 \succeq O_2$ in preference notation) if and only if $u(O_1) \geq u(O_2)$. Information agents need some form of these preference maps to decide search issues on their user's behalfs [150, 156, 226, 311]. A fuzzy function approximator can give a good approximation of the preference map if the fuzzy system does not need too many rules and if the system can sample enough accurate user preference data. We also suggest a method to elicit consistent user preference data.

B.3 Profile Learning with Sunsets and Flowers

Users can define preference maps on an image space of sunsets or flowers. Each person has his own likes or dislikes that define his own fuzzy pattern of object clusters. The clusters depend on the features that define the objects. Recent work on object recognition [300] and content-based image retrieval [233] suggests that features define the "look" of the images. These features include colors, shapes, and textures. Research in machine vision seeks invariant features that can map all images into smaller clusters [34, 40, 91, 233, 256, 257, 300, 318].

Figure B.4 shows a block diagram of a neural fuzzy agent that learns a user profile in a space of images. We used a multi-dimensional histogram of an image as features for our fuzzy agent prototype. Niblack [233] and Swain [300] used color histograms to recognize images and to structure their image database retrieval systems. The histogram technique itself ignores the spatial correlation of pixels in images. This has led many researchers to suggest other local features [34, 233, 256]. We use the image dispersion σ_{ij} as an extra feature [260]:

$$\sigma_{ij} = \frac{1}{W^2}\Big[\sum_{m=-w}^{w}\sum_{n=-w}^{w}[x(i+m,j+n) - \bar{x}(i+m,j+n)]^2\Big]^{1/2} \tag{B.1}$$

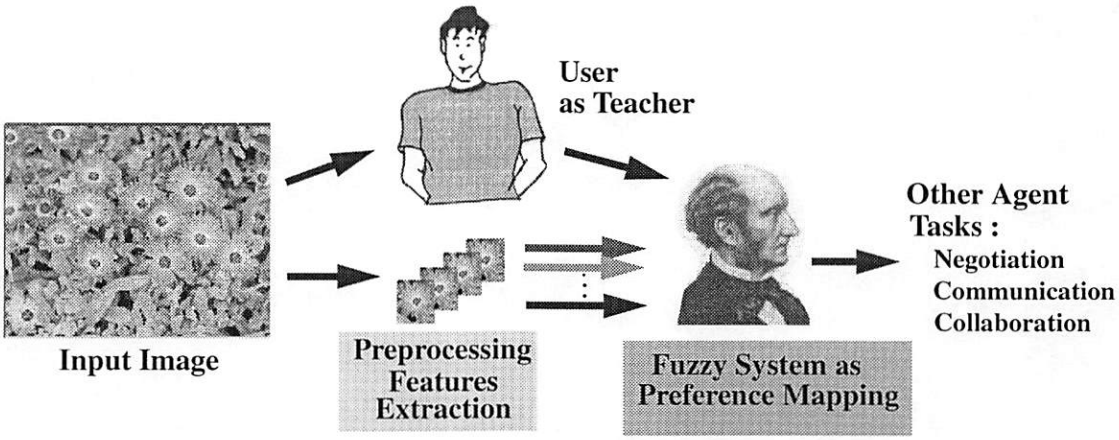


Figure B.4: Data acquisition. A fuzzy agent can learn a user's unknown preference map. The user acts as a teacher or supervisor and gives the system question-answer training samples. Then supervised gradient descent tunes the fuzzy system to better approximate the user's preference map.

where $W = 2w + 1$ and where

$$\bar{x}(i,j) = \frac{1}{W^2} \sum_{m=-w}^{w} \sum_{n=-w}^{w} x(i+m, j+n) \tag{B.2}$$

defines the sample mean in the $W \times W$ window centered at pixel location (i,j).

For each image we obtain its 4-D normalized histogram. The first three components are hue h, saturation s, and intensity v in the hue-saturation-intensity color space [260]. The other component is the standard deviation σ of the intensity component. We view this normalized 4-D histogram as an input discrete probability density function to the fuzzy system and write it in the form

$$T(h,s,v,\sigma) = \sum_{i=1}^{N_h} \sum_{j=1}^{N_s} \sum_{k=1}^{N_v} \sum_{l=1}^{N_\sigma} t_{i,j,k,l}\, \delta(h - \bar{h}_i)\, \delta(s - \bar{s}_j)\, \delta(v - \bar{v}_k)\, \delta(\sigma - \bar{\sigma}_l). \tag{B.3}$$

Here N_h, N_s, N_v, and N_σ are the number of bins on axes of hue, saturation, intensity, and standard deviation. So the total number of histogram bins is $N = N_h \times N_s \times N_v \times N_\sigma$. The term $\bar{h}_i$ is the bin center of the ith hue and likewise for $\bar{s}_j$, $\bar{v}_k$, and $\bar{\sigma}_l$. The term $t_{i,j,k,l}$ is a

normalized frequency of occurrence of the feature vector $(\bar{h}_i, \bar{s}_j, \bar{v}_k, \bar{\sigma}_l)$. We write the N-bin histogram T in the more compact form

$$T(h, s, v, \sigma) = T(x) = \sum_{n=1}^{N} t_n \delta(x - \bar{x}_n). \tag{B.4}$$

The vector $\bar{x}_n$ has the center of the histogram bin as its components: $\bar{x}_n = (\bar{h}_{i_n}, \bar{s}_{j_n}, \bar{v}_{k_n}, \bar{\sigma}_{l_n})$ as in (B.3). The normalized frequency of occurrence t_n replaces the corresponding t_{i_n, j_n, k_n, l_n} in (B.3).

This histogram T is the input to the fuzzy system. Appendix A.1 shows that this gives a generalized SAM ratio (4.80) [158, 162] as a *set* SAM system:

$$F(T) = \frac{\sum_{j=1}^{m} a_j(T) V_j c_j}{\sum_{j=1}^{m} a_j(T) V_j} = \sum_{j=1}^{m} p_j(T) c_j. \tag{B.5}$$

The convex coefficients $p_j(T) \geq 0$ and $\sum_{j=1}^{m} p_j(T) = 1$ have the form

$$p_j(T) = \frac{a_j(T) V_j}{\sum_{i=1}^{m} a_i(T) V_i}. \tag{B.6}$$

The correlation of a fuzzy set function $a_j : X \subset R^4 \rightarrow [0, 1]$ with a 4-D histogram of an image T has the form

$$a_j(T) = \int_X a_j(h, s, v, \sigma)\, T(h, s, v, \sigma) dh\, ds\, dv\, d\sigma \tag{B.7}$$

$$= \sum_{n=1}^{N} t_n\, a_j(\bar{x}_n). \tag{B.8}$$

The value $a_j(T)$ states the degree to which fuzzy set T belongs to fuzzy set A_j. The set correlation $a_j(T)$ need not lie in the unit interval. It can take on any finite nonnegative value:

$a_j(T) \in [0, \infty)$. The set SAM ratio in (B.5) still gives an output as a convex sum of the then-part set centroids c_j as the point SAM in (4.80).

We tested the fuzzy agents with 88 flowers images and 42 sunsets images. Figure B.2 shows some of the test images. We assigned subjective values to all images as numbers from 0 to 10. The value 10 stands for "It is maximally beautiful" or "I really love it." The value 0 stands for "It is minimally beautiful" or "I really hate it." The histogram bins were 8:4:4:4 for $h : s : v : \sigma$. So there were a total of 512 bins. The fuzzy system also had 512 fuzzy rules. We initialized the fuzzy agent so that it would be "indifferent" to all images (a score of 5) and trained it with supervised gradient-descent learning. The initial maximum absolute error was 5 and the mean absolute error was 2.45. The fuzzy agent converged after 40,000 epochs to our preference map and gave a score close to ours. This held for almost all test images. The maximum absolute error was 0.96 and the mean absolute error was 0.18. This error stemmed from too few features. Using more features tends to improve the system's accuracy but at the expense of greater rule complexity.

We used a histogram based on color and variance because it captured the relative amount of colors in the image that affect much of human perception [233, 300]. We can also compute histograms easily and they are translation and rotation invariant [300]. Our systems for profile learning and searching did not depend on how we chose object features. The fuzzy agent could use other inputs from this image database or from others. These input features might include shapes [233, 256], textures [34, 233, 256, 257], wavelet transforms [40, 305], or other statistical measures [256].

B.4 Adaptive Fuzzy Object Matching

This section presents fuzzy equality as a measure of similarity between objects and shows how to tune it. A search or filter agent matches objects in the databases to the query object and acts on

the match results. Supervised learning tunes the fuzzy equality measure to better approximate the user's perception of similar objects.

A fuzzy system can assist in database search in many ways. Fuzzy matching is perhaps the simplest way. The fuzzy equality measure [162] between two fuzzy sets can define the similarity between objects. The equality measure $\mathcal{E}(A, B)$ measures the degree to which fuzzy set A equals fuzzy set B. It measures how well A matches B and vice versa. Suppose fuzzy sets A and B are nonempty. Then $\mathcal{E}(A, B) = \mathcal{E}(B, A) \in [0, 1]$, $\mathcal{E}(A, A) = 1$, and $\mathcal{E}(A, \emptyset) = 0$ for the empty set $\emptyset$. The equality measure depends on the *counting* or *cardinality* [158] function c of a fuzzy set as

$$\mathcal{E}(A, B) = \text{Degree}(A = B) = \frac{c(A \cap B)}{c(A \cup B)} = \frac{\int \min(a(x), b(x))\, dx}{\int \max(a(x), b(x))\, dx} \tag{B.9}$$

where

$$c(A) = \sum_{i=1}^{N} a_i \quad \text{or} \quad c(A) = \int_{R^n} a(x)\, dx \tag{B.10}$$

for an integrable fuzzy set function $a : X \to [0, 1]$. The fuzzy equality measure rests on the theory of fuzzy sets as points in unit hypercubes or fuzzy cubes. Appendix A.3 reviews this unit-cube geometry of discrete fuzzy sets [158, 162].

Consider an example. Let $a = (.8 \quad .4 \quad 0)$ and $b = (.1 \quad .5 \quad .2)$ be discrete set functions for fuzzy sets A and B in $X = \{x_1, x_2, x_3\}$. So the set function or fit vector $a = (a_1 \quad a_2 \quad a_3)$ defines the fuzzy set A as $a_1 = a(x_1) = .8$, $a_2 = a(x_2) = .4$, and $a_3 = a(x_3) = 0$. The fit vector b defines the fuzzy set B as $b_1 = b(x_1) = .1$, $b_2 = b(x_2) = .5$, and $b_3 = b(x_3) = .2$. Then fuzzy set A equals fuzzy set B to degree one-third:

$$\mathcal{E}(A, B) = \text{Degree}(A = B) = \frac{c(A \cap B)}{c(A \cup B)} \tag{B.11}$$

$$= \frac{\sum_{i=1}^{3} \min(a_i, b_i)}{\sum_{i=1}^{3} \max(a_i, b_i)} \tag{B.12}$$

$$= \frac{.1 + .4 + 0}{.8 + .5 + .2} = \frac{1}{3}. \tag{B.13}$$

A fuzzy system maps two objects (or their two vectors of "features") to the output fuzzy sets A and B. Then the equality measure gives a value near 1 if the two objects match well or "look alike." It gives a value near 0 if they match poorly.

We use the same histogram features as in the prior section to match images. Let T_A and T_B be the histograms of two images. Again we view these two normalized N-bin histograms as discrete probability density functions whose domain $X = \{\bar{x}_1, \ldots, \bar{x}_N\}$ is a set of vectors $\bar{x}_i$ that define the bin centers. This gives the same form as in (B.4). Then we compute the correlation of a set function a_j with two histograms T_A and T_B as in (B.8) with

$$A_j = a_j(T_A) = \sum_{n=1}^{N} T_A(\bar{x}_n)\, a_j(\bar{x}_n) \tag{B.14}$$

$$B_j = a_j(T_B) = \sum_{n=1}^{N} T_B(\bar{x}_n)\, a_j(\bar{x}_n). \tag{B.15}$$

This gives two m-D vectors of set values $(A_1, \ldots, A_m)$ and $(B_1, \ldots, B_m)$ from m fuzzy rules. The standard additive structure of fuzzy systems suggests that the output fuzzy set should equal the sum of the scaled then-part sets [162]. So we define the then-part sets to be the same as the if-part sets. So the output fuzzy sets A and B from the histograms T_A and T_B have the form

$$A(x) = \sum_{j=1}^{m} A_j\, a_j(x) \tag{B.16}$$

$$B(x) = \sum_{j=1}^{m} B_j\, a_j(x) \tag{B.17}$$

where $x = (h, s, v, \sigma) \in X$. The input to the system is an N-bin histogram on the discrete domain $X = \{\bar{x}_1, \ldots, \bar{x}_N\}$. Then we can view the output sets A and B as discrete sets and rewrite (B.16) - (B.17) as

$$A(\bar{x}_n) = \sum_{j=1}^{m} A_j\, a_j(\bar{x}_n) \tag{B.18}$$

$$B(\bar{x}_n) = \sum_{j=1}^{m} B_j\, a_j(\bar{x}_n) \tag{B.19}$$

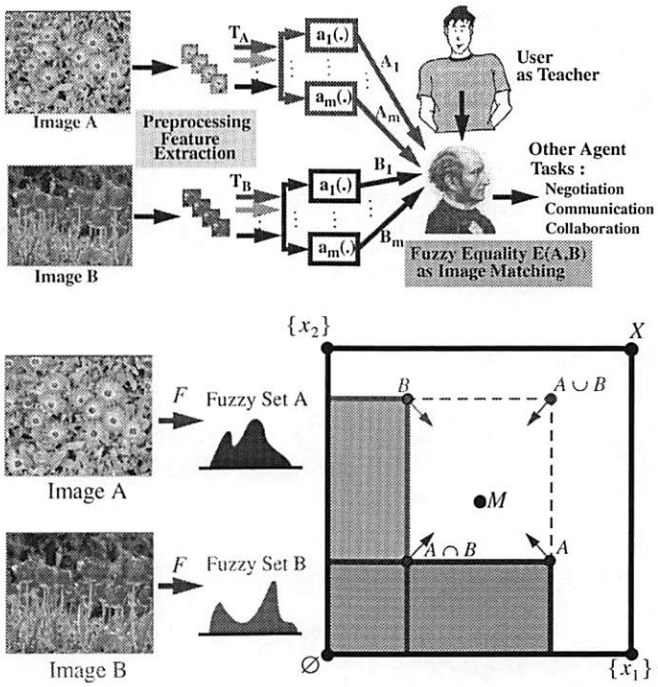


Figure B.5: Adaptive fuzzy search. Fuzzy equality measures the likeness of two objects A and B. Supervised learning tunes the fuzzy equality measure $\mathcal{E}(A, B)$ inside the fuzzy-cube state space to better approximate the user's perception of similar images. The equality measure grows to unity as the A and B set points approach each other. The cube midpoint M is the maximally fuzzy set where $\mathcal{E}(M, M^c) = 1$. Binary sets V lie at the 2^n cube vertices and they alone give $\mathcal{E}(V, V^c) = 0$.

for $n = 1, \ldots, N$. Then the fuzzy equality (in the discrete case) in (B.9) measures the degree to which fuzzy set A equals or matches fuzzy set B:

$$\mathcal{E}(A, B) = \frac{\sum_{i=1}^{N} \min(A(\bar{x}_i), B(\bar{x}_i))}{\sum_{i=1}^{N} \max(A(\bar{x}_i), B(\bar{x}_i))}. \tag{B.20}$$

This in turns measures the "similarity" between two images.

The similarity measure depends on how we define the m fuzzy rules. Tuning or learning schemes can move or reshape the fuzzy sets to approximate desired matching values. Appendix A.3 derives the learning laws that tune the set-function parameters in $\mathcal{E}$.

Figure B.5 shows a block diagram of how a fuzzy agent matches images. The simulation used a 4-D version of the 1-D Laplace set function $a_j(x) = \exp\{-\left|\frac{x - m_j}{d_j}\right|\}$ in (A.85) - (A.86). We trained the fuzzy matching system on a space of sunset images with the histogram intersection in [300]:

$$S(H, I) = \frac{\sum_{i=1}^{N} \min(H_i, I_i)}{\sum_{i=1}^{N} H_i}. \tag{B.21}$$

The fuzzy system gave a rough approximation of the histogram intersection. We may not be able to find a closed-form formula for matching in the general case. Then the fuzzy matching process might learn from Q & A sessions or from other user feedback.

B.5 The Agent-User Interface: The Q & A Bottleneck

How does an agent get numerical values for sample objects? What questions should the agent ask the user in a Q & A session? How many objects must a user rank? These questions reveal the practical weakness of any search system that depends on numbers. Cardinal data eases numerical processing but comes at the expense of a question-answer bottleneck.

This section reviews some of the techniques used in decision theory to rank objects. We show how to apply the technique to obtain numerical values for all sample objects. Other criteria can help agents ask users new questions in Q & A sessions.

Suppose a user states a subjective numerical value for each sample object in a list. There are problems with this absolute valuation beyond its artificial nature and the sheer inconvenience it forces on the user. Miller [205] observed that the largest number of objects that our minds can process at one time is the "magic number" 7 ± 2. We tend to forget how we have ranked objects at the top of the list when we rank objects at the bottom of the list. Relative rankings can increase the "capacity" of our information processing [205]. Then techniques in decision theory

allow us to rank objects with only *pairwise* comparisons [56, 150, 151, 278]. We can compute these relative object weights and convert them to the user's absolute weights.

Saaty's analytic hierarchy process (AHP) [278, 279, 280] can find the numerical values. AHP computes the relative weights $w = (w_1, \ldots, w_n)$ of n objects from their pairwise comparisons. Let a_{ij} be a ratio scale [278] of comparison of object O_i and O_j. Then a reciprocal matrix $A = [a_{ij}]$ has its elements of the form $a_{ij} = 1/a_{ji}$ for $i, j = 1, \ldots, n$. So the diagonal entries are always unity: $a_{ii} = 1$ for all $i = 1, \ldots, n$. The claim "I like object O_1 twice as much as I like object O_2" gives $a_{12} = 2$. Its reciprocal $a_{21} = 1/2$ gives the claim "I like object O_2 half as much as I like object O_1." The principle eigenvector w of a matrix A obeys the equation $Aw = \lambda_{\max} w$. The Perron-Frobenius theorem of matrix algebra ensures that $\lambda_{\max}$ is the unique maximum (positive) eigenvalue of A [88]. The components of w are always positive and allow us to recover the relative object weights [278].

Suppose reciprocal matrices A and B for objects O_1, O_2, O_3, and O_4 have the values

A	O_1	O_2	O_3	O_4	w
O_1	1	$\frac{1}{2}$	5	1	0.238
O_2	2	1	10	2	0.476
O_3	$\frac{1}{5}$	$\frac{1}{10}$	1	$\frac{1}{5}$	0.048
O_4	1	$\frac{1}{2}$	5	1	0.238

B	O_1	O_2	O_3	O_4	v
O_1	1	$\frac{1}{2}$	5	1	0.242
O_2	2	1	7	3	0.494
O_3	$\frac{1}{5}$	$\frac{1}{7}$	1	$\frac{1}{4}$	0.056
O_4	1	$\frac{1}{3}$	4	1	0.208

The matrix A contains many claims such as "I like object O_2 twice as much as I like object O_1 and ten times as much as I like object O_3. I like the object O_1 as much as I like O_4." Users may prefer to say that they like an object O_2 ten times as much as they like O_3 than to say that they like O_3 one tenth as much as they like O_2. Agents can offer users both options.

The matrix A is "consistent" but B is not. A matrix is consistent when its elements a_{ik} obey $a_{ik} = a_{ij}a_{jk}$ for all $i, j, k = 1, \ldots, n$ [278]. Each row of a consistent matrix is a multiple of the first row. Consistency also implies transitivity: The claim $a_{23} = a_{21}a_{13}$ implies that if we prefer O_2 to O_1 ($O_1 \succ O_2$) and if we prefer O_1 to O_3 ($O_1 \succ O_3$) then we prefer O_2 to

O_3 ($O_2 \succ O_3$). But pairwise comparisons are often inconsistent. So the weight matrix may look more like B than like A. The maximum eigenvalue $\lambda_{\max}$ obeys $\lambda_{\max} \geq n$. The equality $\lambda_{\max} = n$ holds if and only if the reciprocal matrix A is consistent. The consistency measure $\mu = (\lambda_{\max} - n)/(n-1)$ can help the agent decide whether it needs to ask a user to verify the rankings [278, 280].

The principle eigenvectors $w = (w_1, w_2, w_3, w_4)$ and $v = (v_1, v_2, v_3, v_4)$ above reflect the relative weights of objects from comparison matrices A and B. The preference order from A is $O_2 \succ O_1 \sim O_4 \succ O_3$ where $\sim$ denotes the indifference preference between two objects. The preference order from B is $O_2 \succ O_1 \succ O_4 \succ O_3$. Then a linear (affine) transformation L relates the weights to the user's subjective values: $u(O_i) = L(w_i) = cw_i + d$ for some $c > 0$ and some $d \in R$. This holds because we assume that w measures the relative utility of objects and because a linear transformation preserves the structure of a utility function [64, 121, 241].

An agent must interact with the user to get the transform coefficients c and d. The agent picks any two objects that have different weights and asks the user to give non-negative weights for both objects. Then the agent can solve for the coefficients c and d and find the rest of the object weights. Arbitrary positive values for two mid-rank objects can give negative values for low-rank objects. So the highest-rank and lowest-rank objects are often the best choice. Suppose for matrix B that $u(O_2) = 10$ and $u(O_3) = 1$. Then $c = 20.55$ and $d = -0.15$. This gives $u(O_1) = cv_1 + d = 20.55 \times 0.242 - 0.15 = 4.82$ and $u(O_4) = 20.55 \times 0.208 - 0.15 = 4.12$.

Consistency gives a linear ranking complexity. A consistent user needs to give only $n-1$ pairwise rankings to construct a matrix. Then we can deduce the other entries from $a_{ij} = a_{ik}a_{kj}$ for all $i, j, k = 1, \ldots, n$. But humans are seldom consistent. So an agent may need the user to give $\frac{1}{2}(n^2 - n)$ rankings to form a matrix. This may not be practical for large n. The agent can estimate the matrix entries with the geometric mean of all paths in the matrix from the first n rankings [115]. He can use other criteria [115, 208] to ask the user for additional pairwise rankings. This can reduce the number of rankings from $\frac{1}{2}(n^2 - n)$ rankings to on the order of n rankings.

A final remark concerns scaling fuzzy profiles. A linear transformation L both preserves the structure of a preference map [64, 121, 241] and preserves how the agent's neural fuzzy system models the preference map. This always holds for the SAM fuzzy system F. Suppose the linear transformation L consists of the matrix C and the vector d. Then (4.80) gives

$$L(F(x)) = CF(x) + d = C\Big(\sum_{j=1}^{m} p_j(x)\, c_j\Big) + d \tag{B.22}$$

$$= \sum_{j=1}^{m} p_j(x)\,(C\, c_j) + d \tag{B.23}$$

$$= \sum_{j=1}^{m} p_j(x)\,(C\, c_j + d) \qquad \text{since} \sum_{j=1}^{m} p_j(x) = 1 \tag{B.24}$$

$$= \sum_{j=1}^{m} p_j(x)\, c_j' \tag{B.25}$$

where $c_j' = Cc_j + d = L(c_j)$. So an agent can change the profile to match the user's new scale or new set of data. A neural fuzzy agent may also want to rescale its profiles for large hierarchical systems where each system level has its own ratio scale.

B.6 Conclusion

Neural fuzzy systems can assist agents in many ways. We have shown how these adaptive function approximators can both help learn a user's preference map and help choose preferred search objects cast as features of low dimension. The color histogram we used did not give a complete set of features. Other neural fuzzy systems can more fully combine these two fuzzy tasks to aid in agent database search. Future research may depend on advances in pattern recognition and machine vision. Neural fuzzy systems might also assist agents when agents bargain [46, 267, 274] or cooperate [66, 225] with other agents. Then an agent may try to learn a second or third user's profile as well as learn its master's profile.

Agents could also help neural fuzzy systems approximate functions from training samples. Today most neural fuzzy systems work with just one fuzzy system and one supervised or unsupervised learning law. Rule explosion in high dimensions may force the user to replace the lone fuzzy system with several smaller systems. Agents can help combine these fuzzy systems [160, 162] if they pick and change the weights or rankings of each system based on sample data or domain knowledge. Agents can also pick which learning law to use or which set of parameters to use as the system tunes its rules on-line. Still more complex hybrids can use nested agents within multi-system function approximators and use the approximators to help higher-level agents learn profiles and search databases and perhaps perform other agent tasks.

The neural fuzzy agent needs to improve how it acquires knowledge [153, 283]. The agent should not ask the user too many questions. The agent needs to learn the user's profile fast enough before it tires the user. Efficient agents would make the user state rankings that are at most linear in the number of search objects or search-object clusters. Our system asks the user a large number of numerical questions even though the user may not want to give and perhaps cannot give precise numerical answers to these questions. Researchers have long searched for techniques that can lessen the number of numerical questions the system must ask the user [153, 283]. The bootstrap and other statistical methods [78] may offer more efficient ways for an adaptive agent to sample its user and its environment. Ordinal or chunking techniques [168, 205, 232] may also ease the burden of preference acquisition. But all such techniques tend to increase the complexity of the neural and fuzzy systems.