Adaptive Stochastic Resonance

SANYA MITAIM AND BART KOSKO, MEMBER, IEEE

This paper shows how adaptive systems can learn to add an optimal amount of noise to some nonlinear feedback systems. Noise can improve the signal-to-noise ratio of many nonlinear dynamical systems. This "stochastic resonance" (SR) effect occurs in a wide range of physical and biological systems. The SR effect may also occur in engineering systems in signal processing, communications, and control. The noise energy can enhance the faint periodic signals or faint broadband signals that force the dynamical systems. Most SR studies assume full knowledge of a system's dynamics and its noise and signal structure. Fuzzy and other adaptive systems can learn to induce SR based only on samples from the process. These samples can tune a fuzzy system's if–then rules so that the fuzzy system approximates the dynamical system and its noise response. The paper derives the SR optimality conditions that any stochastic learning system should try to achieve. The adaptive system learns the SR effect as the system performs a stochastic gradient ascent on the signal-to-noise ratio. The stochastic learning scheme does not depend on a fuzzy system or any other adaptive system. The learning process is slow and noisy and can require heavy computation. Robust noise suppressors can improve the learning process when we can estimate the impulsiveness of the learning terms. Simulations test this SR learning scheme on the popular quartic-bistable dynamical system and on other dynamical systems. The driving noise types range from Gaussian white noise to impulsive noise to chaotic noise. Simulations suggest that fuzzy techniques and perhaps other adaptive "black box" or "intelligent" techniques can induce SR in many cases when users cannot state the exact form of the dynamical systems. The appendixes derive the basic additive fuzzy system and the neural-like learning laws that tune it.

***Keywords*—** *Adaptive signal processing, dynamical systems, fuzzy systems, neural networks, noise processing, robust statistics, stochastic resonance.*

I. Stochastic Resonance and Adaptive Function Approximation

Noise can sometimes enhance a signal as well as corrupt it. This fact 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. Fig. 1 shows how uniform pixel noise can improve our subjective perception of an image. A small level of noise sharpens the image contours and helps fill in features. Too much noise swamps the image and degrades its contours.

Stochastic resonance (SR) [11]–[13], [21], [23], [26], [68], [80], [89], [128], [161], [162], [181], [185], [186], [193], [243] 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. The classic SR signature is a signal-to-noise ratio (SNR) that is not monotone. Fig. 2 shows the SR effect for the popular quartic-bistable dynamical system [13], [26], [179]. The SNR rises to a maximum and then falls as the variance of the additive white noise grows. More complex systems may have multimodal SNR's and so show stochastic "multiresonance" [79], [240].

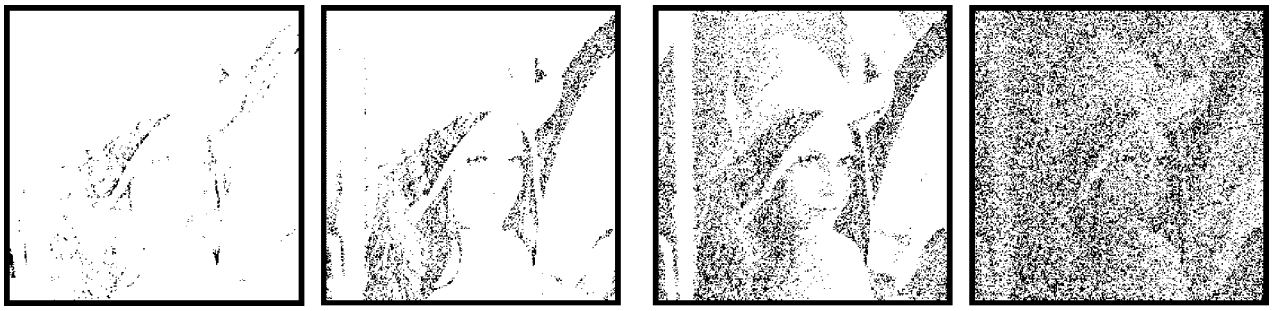
SR holds promise for the design of engineering systems in a wide range of applications. Engineers may want to shape the noise background of a fixed signal pattern to exploit the SR effect. Or they may want to adapt their signals to exploit a fixed noise background. Engineers now add noise to some systems to improve how humans perceive signals. These systems include audio compact discs [150], analog-to-digital devices [10], video images [222], schemes for visual perception [215], [216], [228], and cochlear implants [178], [182]. Some control and quantization schemes add a noise-like dither to improve system performance [10], [147], [150], [198], [222]. Additive noise can sometimes stabilize chaotic attractors [16], [77], [168]. Noise can also improve human tactile response [48], muscle contraction [42], and coordination [49]. This suggests that SR designs may improve how robots grasp objects [51] or balance themselves. SR designs might also improve how virtual or augmented reality systems [32], [106] can create or enhance the sensations of touch and balance.

SR designs might lead to better schemes to filter or multiplex the faint signals found in spread spectrum communication systems [71], [227]. 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 [142], [229], Ethernet packet flows [143], or Internet congestion [113].

Manuscript received November 1, 1997; revised April 17, 1998.

The authors are with the Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA.

Publisher Item Identifier S 0018-9219(98)07858-X.



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

Fig. 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 the popular "Lena" image of signal processing [187]: $y = g((x+n)-\Theta)$ 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 zero (black) to one (white). The threshold is $\Theta = 0.05$. We threshold the original "Lena" image to give the faint image in (a). The uniform noise n has mean $m_n = -0.02$ for images (b)–(d). The noise variance σ_n^2 grows from (b)–(d): $\sigma_n^2 = 1.67 \times 10^{-3}$ in (b), $\sigma_n^2 = 2.34 \times 10^{-2}$ in (c), and $\sigma_n^2 = 1.67 \times 10^{-1}$ in (d).

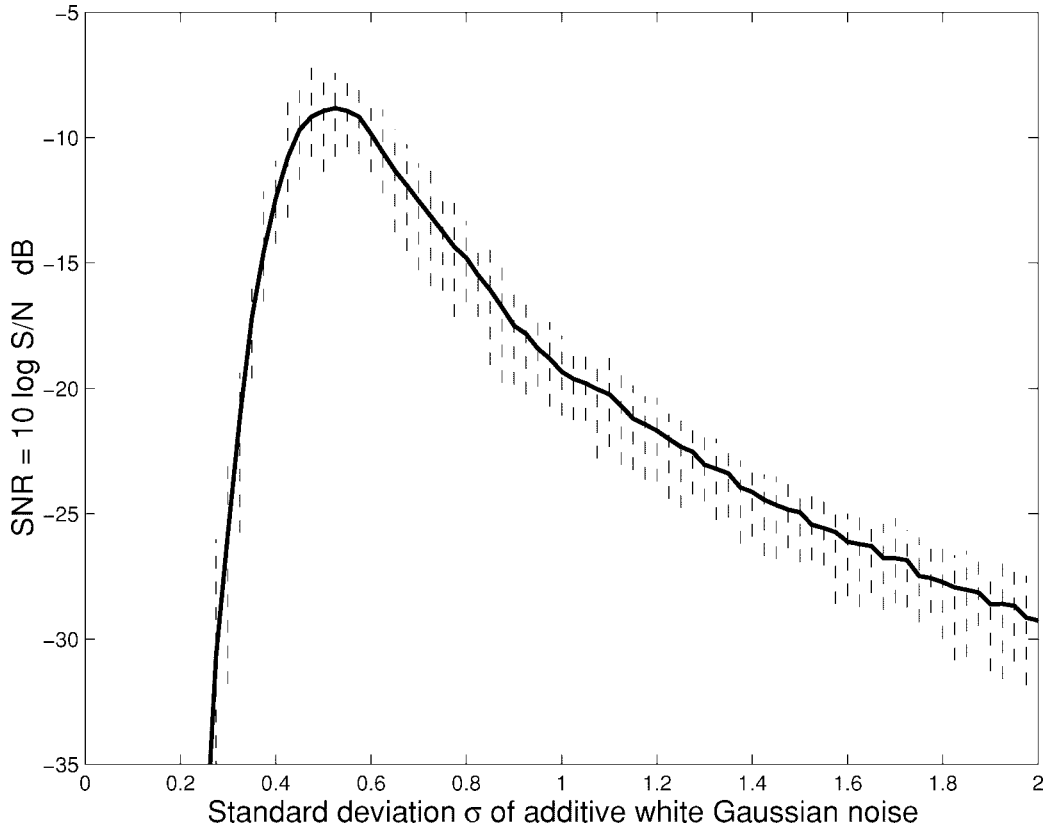


Fig. 2. The nonmonotonic signature of stochastic resonance. The graph shows the smoothed output SNR of a quartic bistable system as a function of the standard deviation of additive white Gaussian noise n. The vertical dashed lines show the absolute deviation between the smallest and largest outliers 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 $\dot{x} = f(x) + s(t) + n(t) = x - x^3 + \varepsilon \sin \omega_0 t + n(t)$. The Gaussian noise $n(t)$ adds to the external forcing narrowband signal $s(t) = \varepsilon \sin \omega_0 t$. Other systems can use multiplicative noise [9], [27], [67], [74], [78], [83] or use non-Gaussian noise [36], [38], [39], [79], [206].

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]. Scientists soon explored SR in climate models [195] to explain how noise could induce periodic ice ages [11], [12], [193], [194]. 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 [73], [194], [245]. Physicists have since found stronger evidence of SR in ring lasers [170], [236], threshold hysteretic Schmitt triggers [69], [171], Chua's electrical circuit [4], [5], bistable magnetic systems [97], electron paramagnetic resonance [81], [84], [217], magnetoelastic ribbons [230], superconducting quantum interference devices (SQUID's) [103], [117], [220], Ising systems [20], [188], [226], coupled diode resonators [151], tunnel diodes [165], [166], Josephson junctions [22], [104], optical systems [9], [61], [120], chemical systems [62], [72], [99], [105], [129], [145], [180], and quantum-mechanical systems [93]–[96], [153], [164], [205], [214], [235].

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 [58], [59], [186], [202], [208], [210], [243]. Noise helps the mechanosensors of the cricket *Acheta domestica* detect small-amplitude low-frequency air signals from predators [146], [172], [173]. 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 [47] and in the neurons in a rat's hippocampus [90]. The SR effect occurs in a wide range of models of neurons [25], [27], [44], [45], [46], [102], [207], [231] and neural networks [24], [25], [27], [29], [30], [41], [44]–[46], [114], [115], [149], [154]–[159], [183], [189], [206].

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 [35], [36], [41], [44]–[47], [102], [108], [146], [209], [231]. Below we review examples of these dynamical systems and the performance measures involved in the SR effect. There is no consensus on which signal-to-noise performance measure best measures the SR effect. The breadth of SR systems 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 model 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.

Below we explore how to learn the SR effect with adaptive systems in general and with adaptive fuzzy function approximators [132]–[136] in particular. 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 [100], [176], [136]. 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 appendixes 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 [197]. Fuzzy systems achieve their patch-covering approximation at the high cost of rule explosion [135], [136]. 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 nonfuzzy adaptive systems.

Adaptive fuzzy systems offer a balance between the structured and symbolic rule-based expert systems found in artificial intelligence [221] and the unstructured but numeric approximators found in modern neural networks [100], [101], [132]. 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 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.

This paper presents three main results. The first and central result is that a system can learn the SR effect if it performs a stochastic gradient ascent on $\mathrm{SNR} = S/N$. Then the random noise gradient $\partial\,\mathrm{SNR}/\partial\sigma$ can tune the parameters in any adaptive system through a slow type of stochastic approximation [219]. We derive these learning laws in terms of discrete Fourier transforms. The idea behind the gradient-ascent learning is that such hill climbing is nontrivial if and only if the SNR surface shows some form of SR. The second result is that the SNR first-order condition for an extremum has the ratio form $S/N = S'/N'$ for $S' = \partial S/\partial\sigma$. The term S'/N' can produce impulsive or even Cauchy noise that can destabilize the stochastic gradient ascent. Time lags in the training process can compound this impulsiveness. The third result is that a Cauchy-based noise suppressor from the theory of robust statistics can often reduce the impulsiveness of the noise gradient $\partial\,\mathrm{SNR}/\partial\sigma$ and thus improve the learning process.

The paper reviews the main math models involved in SR to date and reviews the adaptive fuzzy rule structure that can implicitly approximate these models and produce a like SR effect. The next two sections review these dynamical systems and the competing performance measures that scientists have used to detect SR in them. We used a standard $\mathrm{SNR} = S/N$ based on discrete Fourier spectra. Most SR research has focused on the quartic bistable dynamical system. We worked with that signal system in detail and also applied the stochastic learning scheme to other dynamical systems. The learning scheme converged

in most cases to the SR effect or the SNR mode in all of these systems. The SR learning scheme still converged for the quartic bistable system when we replaced the forcing additive Gaussian white noise with other additive random noise, with infinite-variance noise, and with chaotic noise from a chaotic logistic dynamical system. Sections V and VI derive the SR optimality conditions and the stochastic learning law and then test the learning scheme in SR simulations of the quartic bistable dynamical system and other dynamical systems. The appendixes derive the supervised learning laws for the fuzzy function approximator where the fuzzy sets have the shape of sinc $(\sin x/x)$ functions.

II. SR Dynamical Systems

This section reviews the main known dynamical systems that show SR. These models involve only simple nonlinearities. They also simply add a random noise term to a differential equation rather than use a formal Ito stochastic differential [43], [60], [86]. There are so far no theorems or formal taxonomies that tell which dynamical systems show SR and which do not. A dynamical system relates its input–output response through a differential equation of the form

$$\dot{x} = f(x) + u(x,t) \tag{1}$$

$$y(t) = g(x(t)). \tag{2}$$

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 depends on the state x through $y = g(x)$. The output of a simple model neuron may be a signum function: $y = \operatorname{sgn}(x)$.

A. Quartic Bistable System [13], [56], [75], [82], [109], [121], [179], [249]

The quartic bistable system is the most studied model that shows SR. It has the form

$$\dot{x} = -\frac{\partial}{\partial x}U(x,t) + s(t) + n(t) \tag{3}$$

$$= ax - bx^3 + s(t) + n(t) \tag{4}$$

for a quartic potential $U(x,t) = -(a/2)x^2 + (b/4)x^4$ with $a > 0$, $b > 0$, input signal s, and white Gaussian noise n with zero mean and variance σ^2: $E[n] = 0$ and $\operatorname{Var}(n) = \sigma^2 < \infty$. Researchers sometimes include the forcing functions s and n in the potential function: $U(x,t) = -(a/2)x^2 + (b/4)x^4 + x(t)[s(t) + n(t)]$. The unforced version of (4) 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 $U(x,t) = -(a/2)x^2 + (b/4)x^4$. Fig. 3 shows the quartic potential for $a = b = 1$. The two minima are at $x = \pm 1$. Fig. 3 shows the potential at rest and hence with no input force. Fig. 4 shows the potential $U(x,t)$ when the external sinusoidal input modulates it at each time instant t.

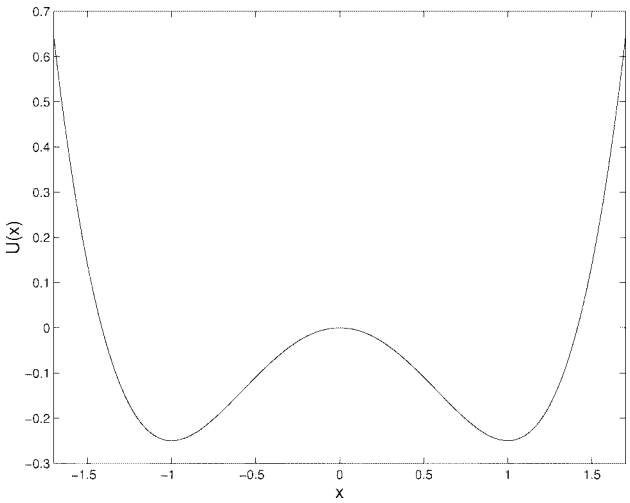


Fig. 3. Unforced quartic potential: $U(x,t) = -(1/2)x^2 + (1/4)x^4$.

B. Threshold Systems [36], [79], [88], [91], [122], [123], [201]

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

$$y(t) = \operatorname{sgn}(x(t)) = \begin{cases} -1, & \text{if } x(t) < \Theta \\ 1, & \text{if } x(t) \geq \Theta \end{cases} \tag{5}$$

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

$$y(t) = Q(x(t)) = G\Delta\left\lfloor \frac{x(t)}{\Delta} + \frac{1}{2} \right\rfloor. \tag{6}$$

A mid-riser quantizer has the form

$$y(t) = Q(x(t)) = G\Delta\left\lfloor \frac{x(t)}{\Delta} \right\rfloor + \frac{G\Delta}{2}. \tag{7}$$

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 [203].

C. Bistable Potential Neuron Model [27]

This neuron model is a bistable system of the form

$$\dot{x} = -x + (\eta_0 + n_m(t))\tanh x + n_a(t) + s(t). \tag{8}$$

The multiplicative and additive noises n_m and n_a are zero mean and uncorrelated. The term η_0 is a constant.

D. Monostable Systems [63], [64], [66], [98], [232], [238]

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

$$\ddot{x} + 2\Gamma\dot{x} + \omega_1^2 x + \gamma x^3 = s(t) + n(t) = \varepsilon\cos\omega_0 t + n(t) \tag{9}$$

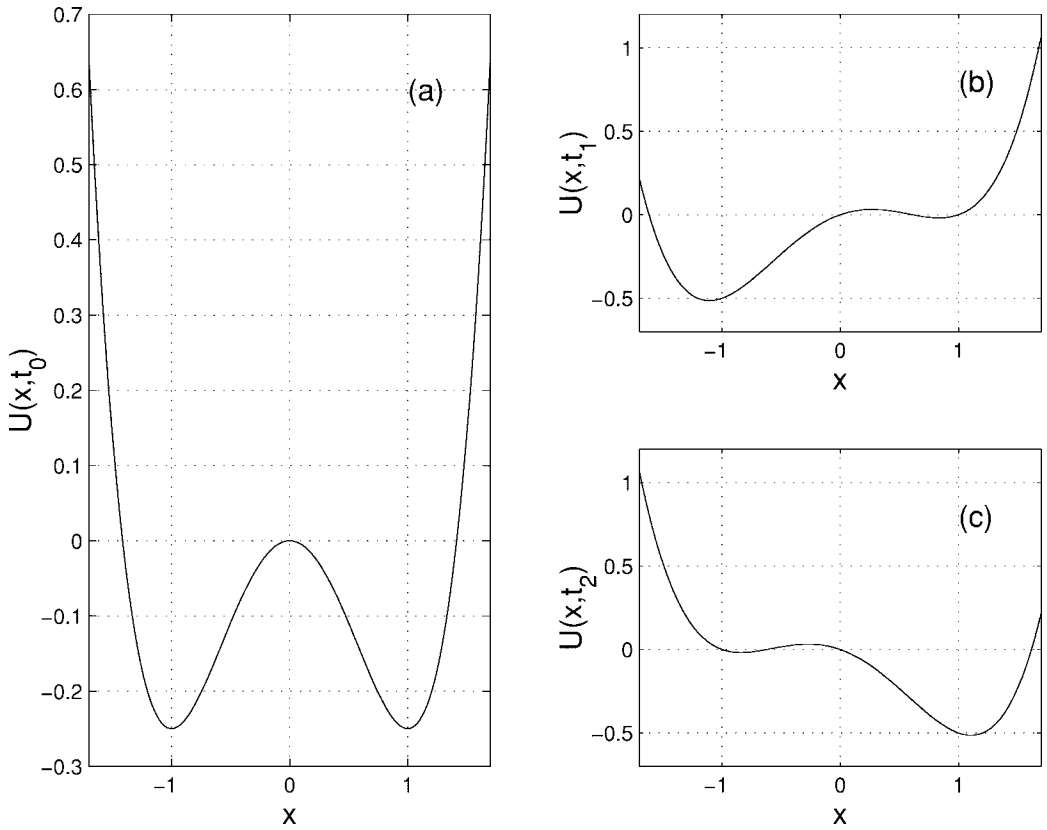


Fig. 4. Forced evolution of the noise-free quartic potential system: $U(x,t) = -(1/2)x^2 + (1/4)x^4 + (1/4)x\sin 2\pi t$. (a) Unforced potential surface at $t = 0$ when the sinusoidal forcing term is zero. (b) Surface $U(x,t)$ at time $t = (1/4)$. (c) Surface $U(x,t)$ at time $t = (3/4)$.

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.

E. Hodgkin–Huxley Neuron Model [44], [156], [209]

The Hodgin–Huxley model is among the most studied models in the neural literature.

$$C\dot{x} = -g_{\rm Na}m^3h(x - x_{\rm Na}) - g_{\rm K}p^4(x - x_{\rm K}) - g_L(x - x_L) + I + s(t) + n(t) \quad (10)$$

$$\dot{m} = \alpha_m(x)(1 - m) - \beta_m(x)m \quad (11)$$

$$\dot{h} = \alpha_h(x)(1 - h) - \beta_h(x)h \quad (12)$$

$$\dot{p} = \alpha_p(x)(1 - p) - \beta_p(x)p. \quad (13)$$

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, $x_{\rm K}$ is the potassium reversal potential, $\bar{g}_{\rm K}$ is the maximal potassium conductance, $\rho_{\rm K}$ is the potassium ion-channel density, $x_{\rm Na}$ is the sodium reversal potential, $\bar{g}_{\rm Na}$ is the maximal sodium conductance, $\rho_{\rm Na}$ is the sodium ion-channel density, I is an input current, and s is a subthreshold aperiodic input signal. These systems use a neural threshold signal function $S(x)$ that lets the neuron rest or retract after firing. SR occurs when a low level of noise n brings the input signal above the neuron's firing threshold.

F. FitzHugh–Nagumo (FHN) Neuron Model [35], [44]–[46], [102], [154], [155], [183], [207], [244]

The FHN neuron model is a two-dimensional limit cycle oscillator that has the form

$$\epsilon\dot{x} = x(x - a)(1 - x) - w + A + s(t) + n(t) \quad (14)$$

$$\dot{w} = x - w - b. \quad (15)$$

Here x is a fast (voltage) variable, w is a slow (recovery) variable, A is a constant (tonic) activation signal, s is an input signal, and n is noise. Sample constants for the SR effect are $\epsilon = 0.005$, $a = 0.5$, $A = -5/12\sqrt{3} = -0.240\,56$, and $b = 0.15$ [46].

G. Integrate-Fire Neuron Model [25], [31], [37], [39], [44], [74], [211], [231]

This neuron model has linear activation dynamics.

$$\dot{x} = \lambda(u_r - x) + \mu + s(t) + n(t) \quad (16)$$

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.

H. Array And Coupled Systems [24], [27], [29], [30], [41], [44]–[46], [92], [110], [114], [115], [124], [149], [154]–[159], [177], [183], [188]–[190], [206], [212], [215], [216]

These systems combine many units of the above systems. They include neural networks and other coupled systems. A special case is the Cohen–Grossberg ("Hopfield") [132] feedback neural network

$$C_i\dot{x}_i = -\frac{x_i}{R_i} + \sum_{j=1}^{N} m_{ij}\tanh x_j + s(t) + n(t) \quad \text{for } i = 1,\ldots,N \quad (17)$$

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 [115]. One study [115] found that the SR effect goes away for $N \geq 10$.

I. Chaotic Systems [3], [5], [13], [33], [52], [160], [196], [241], [242], [251]

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:

$$\ddot{x} = -\delta\dot{x} + x + x^3 + \varepsilon\sin(\omega_0 t) + n(t). \quad (18)$$

At least one researcher [87] has argued that noise-induced chaos-order transitions need not be SR.

J. Random Systems [15], [19], [28], [68], [144], [252]

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

$$r(V(t)) = r(0)\exp(V(t)). \quad (19)$$

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

Other systems show SR in the literature [7], [11], [14], [20], [55], [107], [127], [167], [169], [188], [193], [213], [225], [239], [246], [248]. Special issues of physics journals [23], [181] also present other systems that show SR. Most use the SR measures in the next section.

III. SR Performance Measures

This section reviews the most popular measures of SR. These performance measures depend on the forcing signal and noise and can vary from system to 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 [57], [125], [184], [218]

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(a(x,t)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x,t)p) \quad (20)$$

for drift term $a(x,t)$ and 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. 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. SR dynamical systems in general need not be Markov processes [78], [192].

A. Signal-to-Noise Ratio

The most common SR measure is some form of SNR [69], [75], [85], [111], [169], [249]. This seems the most intuitive measure, even though there are many ways to define SNR.

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

$$\text{SNR} = 10\log\frac{S}{N} \quad (21)$$

$$= 10\log\frac{S(\omega_0)}{N(\omega_0)}\ \text{dB}. \quad (22)$$

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 [116], [169], [249]. 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 $\{y_0, y_1, \ldots, y_{L-1}\}$ of output signal samples

$$Y[k] = \sum_{t=0}^{L-1} y_t e^{-\frac{2\pi kt}{L}}. \quad (23)$$

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], [249]

$$N = \frac{1}{2M}\sum_{j=1}^{M}\left(|Y[k_0 - j]|^2 + |Y[k_0 + j]|^2\right). \quad (24)$$

We expand this noise term in Section V to include all energy not due to the signal.

An adiabatic approximation [169] can give an explicit SNR R for the quartic bistable system in (4) with sinewave input $s(t) = \varepsilon\sin\omega_0 t$

$$R = \frac{S}{N} = \left[\frac{\sqrt{2}a\varepsilon^2 c^2}{(\sigma^2)^2}e^{-2U_0/\sigma^2}\right]\left[1 - \frac{\frac{4a^2\varepsilon^2c^2}{\pi^2(\sigma^2)^2}e^{-4U_0/\sigma^2}}{\frac{2a^2}{\pi^2}e^{-4U_0/\sigma^2} + \omega_0^2}\right] \quad (25)$$

$$\approx \frac{\sqrt{2}a\varepsilon^2c^2}{(\sigma^2)^2}\,e^{-2U_0/\sigma^2}. \quad (26)$$

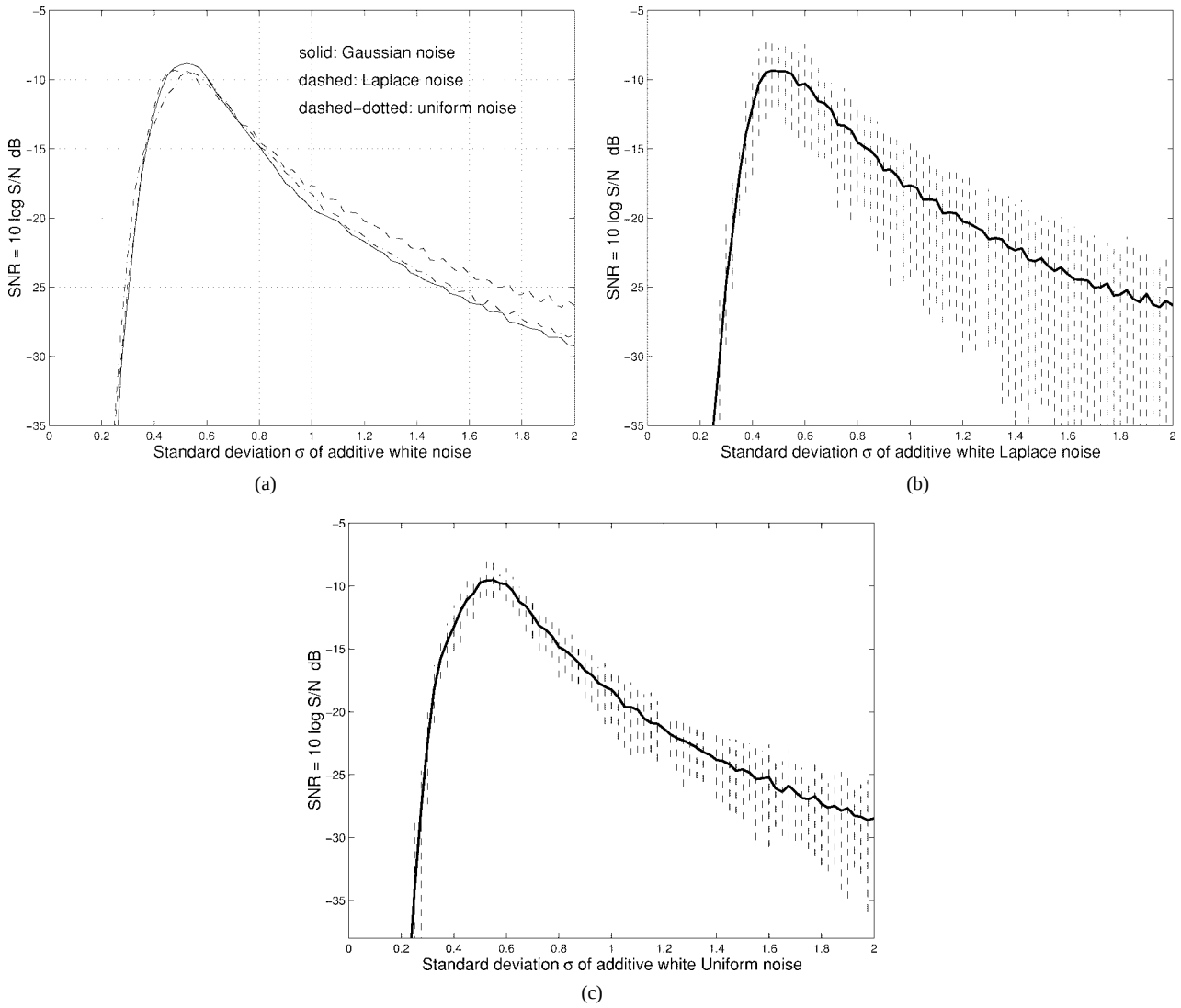
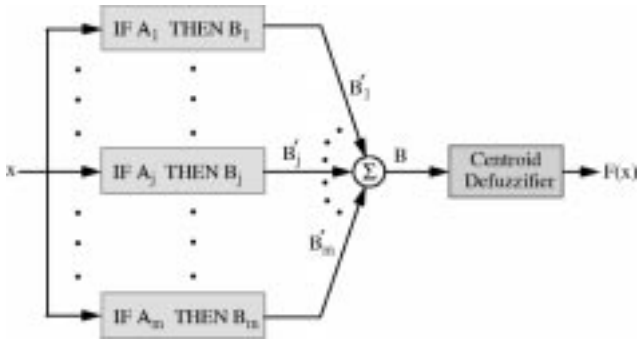


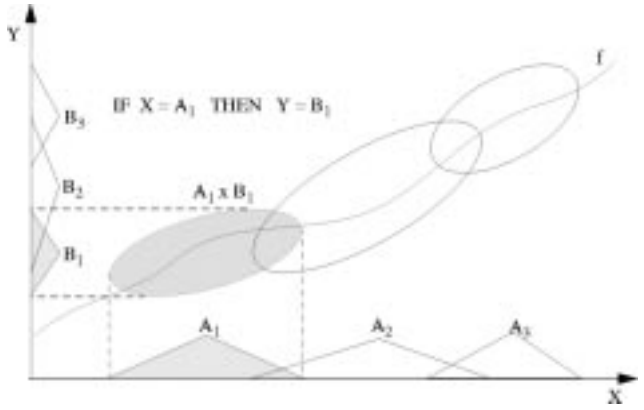
Fig. 5. SNR measure of the quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ with output $y(t) = \text{sgn}(x(t))$. The signal s is the sinewave $s(t) = \varepsilon \sin 2\pi f_0 t$ where $\varepsilon = 0.1$ and $f_0 = 0.01$ Hz. (a) SNR-noise profiles of zero-mean white noise from Gaussian, Laplace, and uniform probability densities. The simulation ran over 20 distinct noise seeds over 10 000 s with time step $\Delta T = 10\,000/1\,000\,000 = 0.01$ s 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. Fig. 2 shows a like SR profile for Gaussian noise. Fig. 18 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.

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 [139] if the signal amplitude ε is small and if its frequency is smaller than the characteristic rate or curvature at the minimum $U''(\pm c)$ [169]. The SNR approximation (26) 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 $\sigma^2 = U_0 = a^2/4b$.

There is no standard definition of system-level signal and noise in nonlinear systems. We work with an SNR that is easy to compute and that depends on standard spectral power measures in signal processing. We start with a sinewave 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. Fig. 2 shows the SR profile with this SNR measure for the quartic bistable system with forcing sinewave input signal and Gaussian noise. Fig. 5 shows the SR profiles of the quartic bistable system with forcing Gaussian, uniform, and Laplace noise. Fig. 17 shows the SR profiles of the quartic



(a)



(b)

Fig. 6. Feedforward fuzzy function approximator. (a) The parallel associative structure of the additive fuzzy system $F : R^n \rightarrow R^p$ 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 Fig. 7.

bistable system for impulsive noise with infinite variance. Fig. 18 shows the SR profile of the quartic bistable system for chaotic noise from a logistic dynamical system.

B. Cross-Correlation Measures

These "shape matchers" can measure SR when inputs are not periodic signals. Researchers coined the term "aperiodic stochastic resonance" (ASR) [41], [44], [45], [102] for such cases. They defined cross-correlation measures for the input signal s and the system response in terms of the mean transition rate r in the FHN model in (14)–(15)

$$C_0 = \max\left\{\overline{s(t)r(t+\tau)}\right\} \tag{27}$$

$$C_1 = \frac{C_0}{\left[\overline{s^2(t)}\right]^{1/2}\left\{\overline{\left[r(t) - \overline{r(t)}\right]^2}\right\}^{1/2}} \tag{28}$$

where $\overline{x}$ is the time average: $\overline{x} = \frac{1}{T}\int_0^T x(t)\, dt$.

C. 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 [55], [68], [82], [121], [250]. 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], [25], [28], [34], [76], [154]–[158], [163]. The symmetric bistable system (4) with input $s(t) = \varepsilon \sin \omega_0 t$ gives a system that tends to stay at or wander about one stable state for $T = T_0/2 = 2\pi/\omega_0$ s and then hops to a new stable state as it tracks the input.

D. 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 [31], [36], [37], [231]. The FHN neuron model (14)–(15) shows SR for aperiodic input waveforms when we measure the cross correlation between input and output or the information rate [44], [46], [102]. Noise can also sometimes maximize the mutual information [50]

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) \\ &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \end{aligned} \tag{29}$$

The mutual (Kullback) information $I(X;Y)$ and Fisher information [50] can measure SR in some neuron models [31], [191], [231]. Probability of correct detection and other statistics can also measure SR [108], [116], [231].

E. Complexity 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 [160], [247].

Other forms of SR measures also occur in the SR literature. They include the other SNR's [64], [123], [131], [148], [152], the amplification characteristic of a system like those found in electronic devices [9], [40], [94], [95], [126], susceptibility [65], [66], [177], [233], "crisis" measure in chaos [33], and prediction error of spike rates [35]. 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.

IV. Additive Fuzzy Systems and Function Approximation

This section reviews the basic structure of additive fuzzy systems. The appendixes review and extend the more

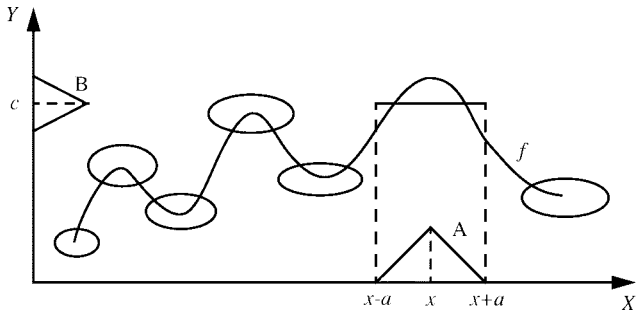


Fig. 7. Lone optimal fuzzy rule patches cover the extrema of approximand f. A lone rule defines a flat line segment that cuts the graph of the local extremum in at least two places. The mean value theorem implies that the extremum lies between these points. This can reduce much of fuzzy function approximation to the search for zeroes $\hat{x}$ of the derivative map $f' : f'(\hat{x}) = 0$.

formal math structure that underlies these adaptive function approximators.

A fuzzy system $F : R^n \to R^p$ 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 $a_j : R^n \to [0, 1]$ and $b_j : R^p \to [0, 1]$. Generalized fuzzy sets map to intervals other than $[0, 1]$. The scalar sinc set functions in Fig. 23 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 (31). The system can use the joint set function a_j or some factored form such as $a_j(x) = a_j^1(x_1) \dots a_j^n(x_n)$ or $a_j(x) = \min(a_j^1(x_1), \dots, a_j^n(x_n))$, or any other conjunctive form for input vector $x = (x_1, \dots, x_n) \in R^n$ [132].

An additive fuzzy system [132], [133] sums the "fired" then-part sets B'_j

$$B(x) = \sum_{j=1}^{m} w_j B'_j = \sum_{j=1}^{m} w_j a_j(x) B_j. \tag{30}$$

Fig. 6(a) shows the parallel fire-and-sum structure of the SAM. These nonlinear systems can uniformly approximate any continuous (or bounded measurable) function f on a compact domain [133], [136]. Engineers often apply fuzzy systems to problems of control [119] but fuzzy systems can also apply to problems of communication [200] and signal processing [130] and other fields.

Fig. 6(b) shows how three rule patches can cover part of the graph of a scalar function $f : R \to R$. 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 [135], [136]. They "patch the bumps" as in Fig. 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 SAM. Appendix A shows that taking the centroid of $B(x)$ in (30) gives the following SAM ratio [132], [133], [134], [135]:

$$F(x) = \frac{\sum_{j=1}^{m} w_j a_j(x) V_j c_j}{\sum_{j=1}^{m} w_j a_j(x) V_j} = \sum_{j=1}^{m} p_j(x) c_j. \tag{31}$$

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 $p_1(x), \dots, p_m(x)$ have the form $p_j(x) = (w_j a_j(x) V_j)/(\sum_{i=1}^{m} w_i a_i(x) V_i)$. The convex coefficients $p_j(x)$ change with each input vector x.

Fig. 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. Appendix B derives the supervised SAM learning algorithms for the sinc set functions [136], [174], [175] in Fig. 23 that we use in the SR simulations. Supervised gradient ascent changes the SAM parameters with performance data. 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). Fig. 8(e) displays the absolute error of the sinc-based fuzzy function approximation.

V. SR Learning and Equilibrium

The scalar SAM fuzzy system $F : R^n \to R$ 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. The SNR in Nonlinear Systems

Suppose a nonlinear dynamical system has a sinewave forcing function $s(t)$ of known frequency f_0 Hz. 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"

$$y_t = r_t + n_t. \tag{32}$$

The 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

$$\text{SNR} = \frac{S}{N} = \frac{S}{P - S}. \tag{33}$$

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

$$Y[k] = \sum_{t=0}^{L-1} y_t e^{-i\frac{2\pi k}{L}t}. \tag{34}$$

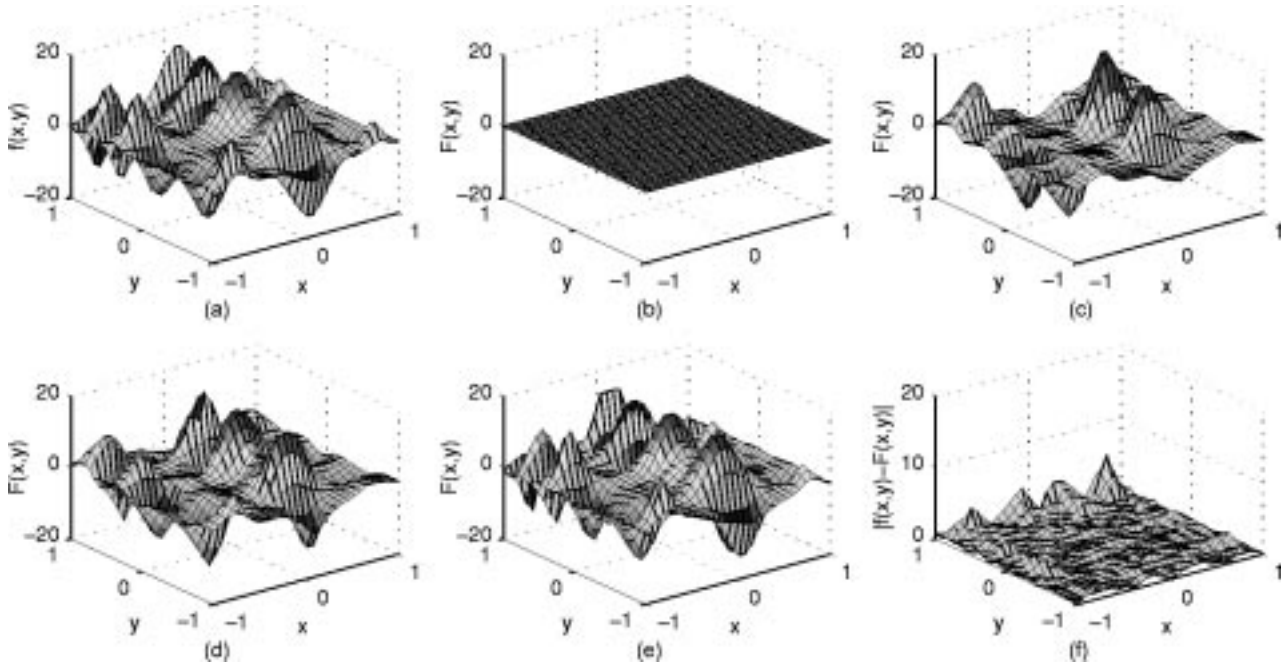


Fig. 8. Fuzzy function approximation. Two-dimensional (2-D) sinc SAM function approximation with 100 fuzzy if–then rules and supervised gradient descent learning. (a) Desired function or approximand f. (b) SAM initial phase as a flat sheet or constant approximator F. (c) SAM approximator F after it initializes its centroids to the samples: $c_j = f(m_j)$. (d) SAM approximator F after 100 epochs of learning. (e) SAM approximator Fafter 6000 epochs of learning. (f) Absolute error of the fuzzy function approximation ($|f - F|$).

We assume that the discrete frequency $k_0 = f_0 L T_s > 0$ is an integer for sampling rate $1/T_s$ and $\omega_0 = 2\pi f_0$. We also assume that there is no aliasing due to sampling. Then we can show that for large L the SNR measure in (33) tends to the standard definition of SNR as a ratio of variances.

Theorem:

$$\begin{aligned}\text{SNR} &= \frac{2|Y[k_0]|^2}{\sum_{k=0}^{L-1}|Y[k]|^2 - 2|Y[k_0]|^2} \rightarrow \frac{\sigma_r^2}{\sigma_n^2} \\ &= \frac{A^2/2}{\sigma_n^2} = \frac{\frac{1}{L}\sum_{t=0}^{L-1} r_t^2}{\frac{1}{L}\sum_{t=0}^{L-1} n_t^2}. \end{aligned} \tag{35}$$

Here $\sigma_r^2 = (1/T)\int_0^T (A\sin\omega_0 t)^2\,dt = A^2/2$ and $\sigma_n^2 = \text{Var}(n) = E[n^2]$. We need further assumptions to derive (35). First consider the "energy" in each frequency bin k of the transform $Y[k]$

$$|Y[k]|^2 = Y[k]Y^*[k] \tag{36}$$

$$= (R[k] + N[k])\,(R[k] + N[k])^* \tag{37}$$

$$= R[k]R^*[k] + R[k]N^*[k] + R^*[k]N[k] + N[k]N^*[k] \tag{38}$$

$$= |R[k]|^2 + |N[k]|^2 + 2\,\text{Re}\{R[k]N^*[k]\} \tag{39}$$

where $R[k]$ and $N[k]$ are the DFT's of r_t and n_t in (32). Suppose the sinusoidal term has the form

$$r_t = A\cos(2\pi f_0 T_s t + \phi) \tag{40}$$

for $t = 0, \ldots, L-1$. Its DFT has the form [199]

$$R[k] = \sum_{t=0}^{L-1} r_t e^{-i\frac{2\pi k}{L}t} \tag{41}$$

$$= \sum_{t=0}^{L-1} A\cos(2\pi f_0 T_s t + \phi) e^{-i\frac{2\pi k}{L}t} \tag{42}$$

$$= e^{ik\Omega(\phi - a)}\frac{A}{T_s}\left[\frac{\sin a\Omega(k-k_0)}{\Omega(k-k_0)} + \frac{\sin a\Omega(k+k_0)}{\Omega(k+k_0)}\right] \tag{43}$$

$$= e^{ik\Omega(\phi-a)}\frac{Aa}{T_s}(\delta[k-k_0] + \delta[k-(L-k_0)]) \tag{44}$$

where $k_0 = f_0 L T_s > 0$ is an integer, $\Omega = \frac{2\pi}{LT_s}$ is a frequency band, $a = \frac{LT_s}{2}$, 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

$$\sum_{k=0}^{L-1}|R[k]|^2 = |R[k_0]|^2 + |R[L-k_0]|^2 = 2|R[k_0]|^2 = 2\left(\frac{aA}{T_s}\right)^2 = \frac{L^2A^2}{2}. \tag{45}$$

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

$$\sum_{k=0}^{L-1}|N[k]|^2 = L\sum_{t=0}^{L-1}|n_t|^2 \tag{46}$$

$$\approx L(L\sigma_n^2) \tag{47}$$

$$= L^2\sigma_n^2. \tag{48}$$

The ergodicity of n_t gives (47). Now consider the total output spectrum P

$$P = \sum_{k=0}^{L-1} |Y[k]|^2 \tag{49}$$

$$= |Y[k_0]|^2 + |Y[L-k_0]|^2 + \sum_{k=0, k\neq k_0, L-k_0}^{L-1} |Y[k]|^2 \tag{50}$$

$$= 2|Y[k_0]|^2 + \sum_{k=0, k\neq k_0, L-k_0}^{L-1} |N[k]|^2 \tag{51}$$

$$= 2|R[k_0]|^2 + 2|N[k_0]|^2 + 4\,\mathrm{Re}\{R[k_0]N^*[k_0]\} + \sum_{k=0, k\neq k_0, L-k_0}^{L-1} |N[k]|^2 \tag{52}$$

$$= 2|R[k_0]|^2 + \left(\sum_{k=0}^{L-1} |N[k]|^2\right) + 4\,\mathrm{Re}\{R[k_0]N^*[k_0]\}. \tag{53}$$

Then (53) and (39) give

$$P - 2|Y[k_0]|^2 = \sum_{k=0}^{L-1} |N[k]|^2 - 2|N[k_0]|^2. \tag{54}$$

Then the SNR structure in (33) follows:

$$\mathrm{SNR} = \frac{S}{N} = \frac{S}{P-S} \tag{55}$$

$$= \frac{2|R[k_0]|^2 + 2|N[k_0]|^2 + 4\,\mathrm{Re}\{R[k_0]N^*[k_0]\}}{\left(\sum_{k=0}^{L-1} |N[k]|^2\right) - 2|N[k_0]|^2} \tag{56}$$

$$\approx \frac{2|R[k_0]|^2}{\sum_{k=0}^{L-1} |N[k]|^2} = \frac{L^2A^2/2}{L^2\sigma_n^2} = \frac{A^2/2}{\sigma_n^2} \tag{57}$$

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 (57) 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

$$E[|Y[k_0]|^2] = E[|R[k_0]|^2 + |N[k_0]|^2 + 2\,\mathrm{Re}\{R[k_0]N^*[k_0]\}] \tag{58}$$

$$= E[|R[k_0]|^2] + E[|N[k_0]|^2] + 2\,\mathrm{Re}\{E[R[k_0]]E[N^*[k_0]]\} \tag{59}$$

$$= E[|R[k_0]|^2] + E\left[\left(\sum_{t=0}^{L-1} n_t e^{-i\frac{2\pi k}{L}t}\right) \times \left(\sum_{\tau=0}^{L-1} n_\tau e^{-i\frac{2\pi k}{L}\tau}\right)^*\right] + 2\,\mathrm{Re}\left\{R[k_0]E\left[\sum_{\tau=0}^{L-1} n_\tau e^{-i\frac{2\pi k}{L}\tau}\right]^*\right\} \tag{60}$$

$$= |R[k_0]|^2 + \sum_{t=0}^{L-1}\sum_{\tau=0}^{L-1} E[n_t n_\tau] e^{-i\frac{2\pi k}{L}(t-\tau)} + 2\,\mathrm{Re}\left\{E[R[k_0]]\left(\sum_{\tau=0}^{L-1} E[n_\tau] e^{i\frac{2\pi k}{L}\tau}\right)\right\} \tag{61}$$

$$= |R[k_0]|^2 + \sum_{t=0}^{L-1}\sum_{\tau=0}^{L-1} \sigma_n^2 \delta[t-\tau] \cdot e^{-i\frac{2\pi k}{L}(t-\tau)} + 0 \tag{62}$$

$$= |R[k_0]|^2 + L\sigma_n^2 \tag{63}$$

$$= L^2\frac{A^2}{4} + L\sigma_n^2 \tag{64}$$

and

$$E[P - 2|Y[k_0]|^2] = E\left[\sum_{k=0}^{L-1} |N[k]|^2 - 2|N[k_0]|^2\right] \tag{65}$$

$$= \sum_{k=0}^{L-1} L\sigma_n^2 - 2L\sigma_n^2 \tag{66}$$

$$= L^2\sigma_n^2 - 2L\sigma_n^2. \tag{67}$$

Putting (64) and (67) into (33) gives

$$\mathrm{SNR} = \frac{2E[|Y[k_0]|^2]}{E[P - 2|Y[k_0]|^2]} \tag{68}$$

$$= \frac{2\left(L^2\frac{A^2}{4} + L\sigma_n^2\right)}{L^2\sigma_n^2 - 2L\sigma_n^2}. \tag{69}$$

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

B. Supervised Gradient Learning and SR Optimality

An adaptive system can learn an SR noise pattern that maximizes a dynamical system's SNR. The learning law updates a parameter m_j of a SAM fuzzy system (or of any other adaptive system) at time step n with the deterministic law

$$m_j(n+1) = m_j(n) + \mu_n \frac{\partial E[\mathrm{SNR}]}{\partial m_j} \tag{70}$$

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: $E[\mathrm{SNR}] \approx \mathrm{SNR}$. This gives the *stochastic* gradient learning law

$$m_j(n+1) = m_j(n) + \mu_n \frac{\partial \mathrm{SNR}}{\partial m_j} \tag{71}$$

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

$$\frac{\partial \mathrm{SNR}}{\partial m_j} = \frac{\partial \mathrm{SNR}}{\partial \sigma}\frac{\partial \sigma}{\partial m_j}. \tag{72}$$

Here σ is the noise level or standard deviation of the forcing noise term $n(t)$. We want the SAM or other adaptive system F 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

$$\frac{\partial \mathrm{SNR}}{\partial m_j} = \frac{\partial \mathrm{SNR}}{\partial \sigma}\frac{\partial F}{\partial m_j}. \tag{73}$$

The term $\partial F/\partial m_j$ shows how any adaptive system F depends on its jth parameter m_j. We again assume that

the chain rule holds to get

$$\frac{\partial \mathrm{SNR}}{\partial \sigma} = \frac{\partial \mathrm{SNR}}{\partial S}\frac{\partial S}{\partial \sigma} + \frac{\partial \mathrm{SNR}}{\partial N}\frac{\partial N}{\partial \sigma}. \tag{74}$$

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

$$\frac{\partial \mathrm{SNR}}{\partial S} = \frac{\partial}{\partial S}\frac{S}{N} = \frac{1}{N} \tag{75}$$

$$\frac{\partial \mathrm{SNR}}{\partial N} = \frac{\partial}{\partial N}\frac{S}{N} = -\frac{S}{N^2} = -\frac{\mathrm{SNR}}{N}. \tag{76}$$

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

$$\frac{\partial \mathrm{SNR}}{\partial S} = \frac{\partial}{\partial S}10\log\frac{S}{N} = (10\log e)\frac{1}{S} \tag{77}$$

$$\frac{\partial \mathrm{SNR}}{\partial N} = \frac{\partial}{\partial N}10\log\frac{S}{N} = -(10\log e)\frac{1}{N}. \tag{78}$$

We next put (75)–(78) into (74) to get the log term that drives SR learning

$$\frac{\partial \mathrm{SNR}}{\partial \sigma} = \begin{cases} \frac{1}{N}\frac{\partial S}{\partial \sigma} - \frac{\mathrm{SNR}}{N}\frac{\partial N}{\partial \sigma}, & \text{if } \mathrm{SNR} = \frac{S}{N} \\ (10\log e)\left(\frac{1}{S}\frac{\partial S}{\partial \sigma} - \frac{1}{N}\frac{\partial N}{\partial \sigma}\right), & \text{if } \mathrm{SNR} = 10\log\frac{S}{N}. \end{cases} \tag{79}$$

The right side of (79) leads to the first-order condition for an SNR extremum

$$\frac{1}{S}\frac{\partial S}{\partial \sigma} - \frac{1}{N}\frac{\partial N}{\partial \sigma} = 0 \tag{80}$$

or simply

$$\frac{S}{N} = \frac{S'}{N'}. \tag{81}$$

We can rewrite this optimality condition as

$$\left.\frac{S}{N}\right|_{\sigma_{\mathrm{opt}}} = \left.\frac{\partial S/\partial \sigma}{\partial N/\partial \sigma}\right|_{\sigma_{\mathrm{opt}}} \tag{82}$$

when the partial derivatives of S and N with respect to σ are not zero at $\sigma = \sigma_{\mathrm{opt}}$. Equations (80) and (82) give a necessary condition for the SR maximum. The result (82) 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 [140] 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 (81) and (82) hold only in a stochastic sense for sufficiently well-behaved random processes.

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

$$0 > \frac{\partial^2 \mathrm{SNR}}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\frac{\partial \mathrm{SNR}}{\partial \sigma} \tag{83}$$

$$= \frac{\partial}{\partial \sigma}(10\log e)\left[\frac{1}{S}\frac{\partial S}{\partial \sigma} - \frac{1}{N}\frac{\partial N}{\partial \sigma}\right] \tag{84}$$

$$= (10\log e)\left[\left(\frac{1}{S}\frac{\partial^2 S}{\partial \sigma^2} + \frac{\partial S}{\partial \sigma}\left(-\frac{1}{S^2}\frac{\partial S}{\partial \sigma}\right)\right) - \left(\frac{1}{N}\frac{\partial^2 N}{\partial \sigma^2} + \frac{\partial N}{\partial \sigma}\left(-\frac{1}{N^2}\frac{\partial N}{\partial \sigma}\right)\right)\right] \tag{85}$$

$$= (10\log e)\left[\frac{1}{S}\frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{S^2}\left(\frac{\partial S}{\partial \sigma}\right)^2 - \frac{1}{N}\frac{\partial^2 N}{\partial \sigma^2} + \frac{1}{N^2}\left(\frac{\partial N}{\partial \sigma}\right)^2\right] \tag{86}$$

$$= (10\log e)\left[\frac{1}{S}\frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{N}\frac{\partial^2 N}{\partial \sigma^2}\right] \tag{87}$$

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

$$\frac{1}{S}\frac{\partial^2 S}{\partial \sigma^2} - \frac{1}{N}\frac{\partial^2 N}{\partial \sigma^2} < 0. \tag{88}$$

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.

Much of the noisiness and complexity of the random learning law (71) stems from the probability structure that underlies the random optimality "error" process $\mathcal{E}$

$$\mathcal{E} = \frac{S}{N} - \frac{\partial S/\partial \sigma}{\partial N/\partial \sigma} \tag{89}$$

near the optimum noise $\sigma = \sigma_{\mathrm{opt}}$. The probability density of $\mathcal{E}$ 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 process $\mathcal{E}$ for the quartic bistable system in Fig. 9. The results suggest that in some cases the density of $\mathcal{E}$ 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 $e^{-|\omega|^\alpha}$ [18], [70], [223], [224]. The parameter α lies in $0 < \alpha \leq 2$ and gives the Gaussian random variable when $\alpha = 2$ or $\phi(\omega) = e^{-\omega^2}$. It gives the thicker-tailed Cauchy bell curve when $\alpha = 1$ or $\phi(\omega) = e^{-|\omega|}$. 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 $\mathcal{E}$-based learning grows as α falls. Note also that the ratio X/Y is Cauchy if X and Y are jointly Gaussian [70], [137], [141], [204]. Our simulations found that the impulsiveness of $\mathcal{E}$ stemmed at least in part from the step size of the successive DFT's in (92).

We now derive the SR learning laws in terms of DFT's. We can approximate $\partial S/\partial \sigma$ and $\partial N/\partial \sigma$ with a ratio of time differences at each iteration n

$$\frac{\partial S_n}{\partial \sigma_n} \approx \frac{\Delta S_n}{\Delta \sigma_n} = \frac{S_n - S_{n-1}}{\sigma_n - \sigma_{n-1}} \tag{90}$$

$$\frac{\partial N_n}{\partial \sigma_n} \approx \frac{\Delta N_n}{\Delta \sigma_n} = \frac{N_n - N_{n-1}}{\sigma_n - \sigma_{n-1}}. \tag{91}$$

The math model in (1)–(2) gives the exact learning laws. Recall that the L-point DFT [199] for a sequence of states $\{y_t\}$ has the form

$$Y_n[k] = \sum_{l=0}^{L-1} y_{l+(n+1-L)} e^{-i\frac{2\pi k}{L}l} \quad k = 0, \ldots, L-1. \quad (92)$$

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

$$\frac{\partial S_n}{\partial \sigma} = \sum_{j=n+1-L}^{n} \frac{\partial S_n}{\partial y_j}\frac{\partial y_j}{\partial \sigma}$$

and

$$\frac{\partial N_n}{\partial \sigma} = \sum_{j=n+1-L}^{n} \frac{\partial N_n}{\partial y_j}\frac{\partial y_j}{\partial \sigma}. \quad (93)$$

We first derive $\partial S_n/\partial y_j$ and $\partial N_n/\partial y_j$ in (93). Consider the partial derivative of $|Y_n[k]|^2$ with respect to y at time step j

$$\frac{\partial}{\partial y_j}|Y_n[k]|^2 = \frac{\partial}{\partial y_j} Y_n[k]Y_n^*[k] \quad (94)$$

$$= Y_n[k]\frac{\partial}{\partial y_j}Y_n^*[k] + Y_n^*[k]\frac{\partial}{\partial y_j}Y_n[k] \quad (95)$$

$$= Y_n[k]e^{i\frac{2\pi k}{L}(j-(n+1-L))} + Y_n^*[k]e^{-i\frac{2\pi k}{L}(j-(n+1-L))} \quad (96)$$

$$= 2\mathrm{Re}\{Y_n[k]e^{i\frac{2\pi k}{L}(j-(n+1-L))}\} \quad (97)$$

$$= 2\mathrm{Re}\{Y_n[k]\}\cos\left(\frac{2\pi k}{L}(j-(n+1-L))\right) - 2\mathrm{Im}\{Y_n[k]\}\sin\left(\frac{2\pi k}{L}(j-(n+1-L))\right). \quad (98)$$

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

$$\frac{\partial S_n}{\partial y_j} = 4\,\mathrm{Re}\{Y_n[k_0]\}\cos\left(\frac{2\pi k_0}{L}(j-(n+1-L))\right) - 4\,\mathrm{Im}\{Y_n[k_0]\}\sin\left(\frac{2\pi k_0}{L}(j-(n+1-L))\right). \quad (99)$$

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

$$\frac{\partial N_n}{\partial y_j} = \frac{\partial}{\partial y_j}(P_n - S_n) \quad (100)$$

$$= \frac{\partial}{\partial y_j}\sum_{k=0}^{L-1}|Y_n[k]|^2 - \frac{\partial S_n}{\partial y_j} \quad (101)$$

$$= \frac{\partial}{\partial y_j} L\sum_{t=0}^{L-1} y_t^2 - \frac{\partial S_n}{\partial y_j} \quad \text{from Parseval's relation} \quad (102)$$

$$= 2Ly_j - \frac{\partial S_n}{\partial y_j}. \quad (103)$$

We can consider the term $\partial y_j/\partial\sigma$ in (93) as a sample of $\partial y/\partial\sigma$ at the time step j.

Recall the math model of the dynamical system (1)–(2) and let $G(x,u,t) = f(x) + u(x,t)$. Assume that $u(x,t) = s(t) + n(t) = s(t) + \sigma w(t)$ for the zero-mean white noise process $w(t)$ with unit variance $E[w^2] = 1$. So the model becomes

$$\dot{x} = G(x,s,\sigma,w) = f(x) + s(t) + \sigma w(t) \quad (104)$$

$$y(t) = g(x(t)). \quad (105)$$

The chain rule gives

$$\frac{\partial y}{\partial \sigma} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial \sigma}. \quad (106)$$

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

$$\frac{d\eta}{dt} = \frac{d}{dt}\left(\frac{\partial x}{\partial \sigma}\right) = \frac{\partial \frac{dx}{dt}}{\partial \sigma} = \frac{\partial G(x,s,\sigma,w)}{\partial \sigma} \quad (107)$$

$$= \frac{\partial G}{\partial x}\frac{\partial x}{\partial \sigma} + \frac{\partial G}{\partial \sigma} = \frac{\partial G}{\partial x}\eta(t) + \frac{\partial G}{\partial \sigma}. \quad (108)$$

The last derivative $\partial G/\sigma$ results from G's explicit dependence on σ. So the additive case $G(x,s,\sigma,w) = f(x) + s(t) + \sigma w(t)$ gives

$$\frac{\partial G}{\partial x} = \frac{\partial f}{\partial x} \quad (109)$$

$$\frac{\partial G}{\partial \sigma} = \frac{\partial}{\partial \sigma}[f(x) + s(t) + \sigma w(t)] = w(t). \quad (110)$$

We need to simulate the evolution (108) for $\partial x/\partial\sigma$ and obtain $\partial y/\partial\sigma$ from (106). Then we put (99), (103), and $\frac{\partial y}{\partial \sigma}$ into (93) to get the stochastic gradient learning law

$$\sigma(n+1) = \sigma(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} \quad (111)$$

$$= \sigma(n) + \mu_n\left(\frac{\partial \mathrm{SNR}_n}{\partial S_n}\frac{\partial S_n}{\partial \sigma} + \frac{\partial \mathrm{SNR}_n}{\partial N_n}\frac{\partial N_n}{\partial \sigma}\right) \quad (112)$$

$$= \sigma(n) + \mu_n\left(\frac{1}{S_n}\sum_{l=n+1-L}^{n}\frac{\partial S_n}{\partial y_l}\frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n}\sum_{l=n+1-L}^{n}\frac{\partial N_n}{\partial y_l}\frac{\partial y_l}{\partial \sigma}\right). \quad (113)$$

Here we omit the constant factor $10\log e$ from (75)–(78) or view it as part of the learning rate μ_n in (113). 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

$$m_j(n+1) = m_j(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial m_j} \quad (114)$$

$$= m_j(n) + \mu_n\left(\frac{\partial \mathrm{SNR}_n}{\partial S_n}\frac{\partial S_n}{\partial m_j} + \frac{\partial \mathrm{SNR}_n}{\partial N_n}\frac{\partial N_n}{\partial m_j}\right) \quad (115)$$

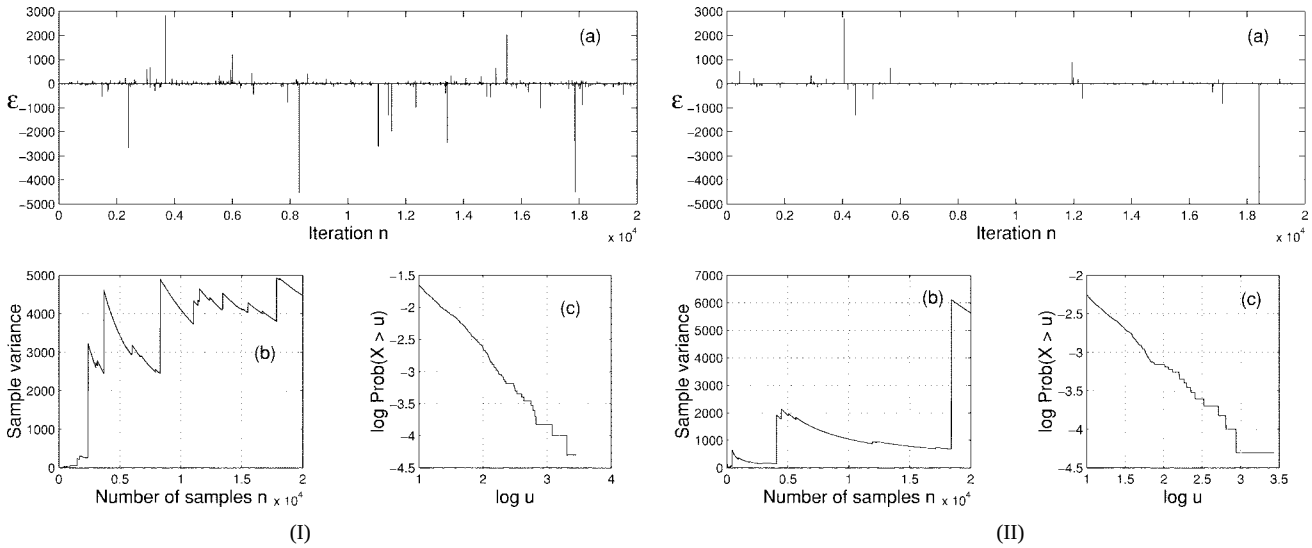


Fig. 9. Visual display of samples from the equilibrium term $\mathcal{E}_n = (S_n/N_n) - ((\partial S_n/\partial\sigma)/(\partial N_n/\partial\sigma))$. (a) Cauchy-like impulsive samples of $\mathcal{E}_n$ at each iteration n for the discretized version of the quartic bistable system $\dot{x} = x - x^3 + \varepsilon \sin 2\pi f_0 t + n(t)$where $\varepsilon = 0.1$ and $f_0 = 0.01$ Hz. The system outputs are (I) $y_t = x_t$ and (II) $y_t = \mathrm{sgn}(x_t)$. 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 $\log \mathrm{Prob}(X > u)$ 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.

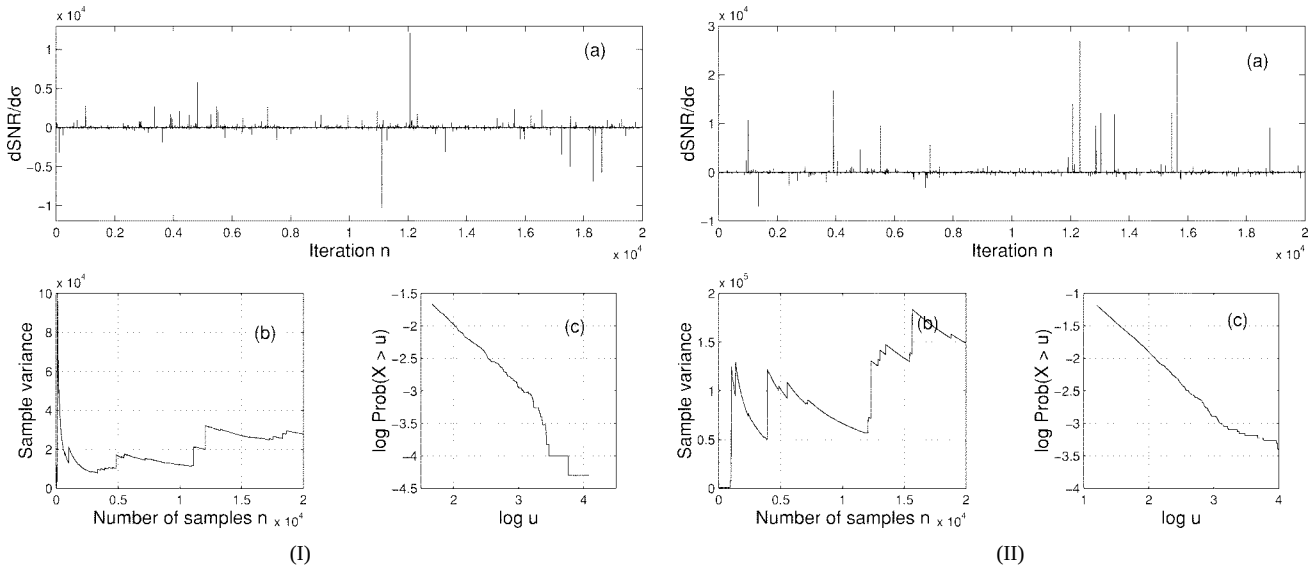
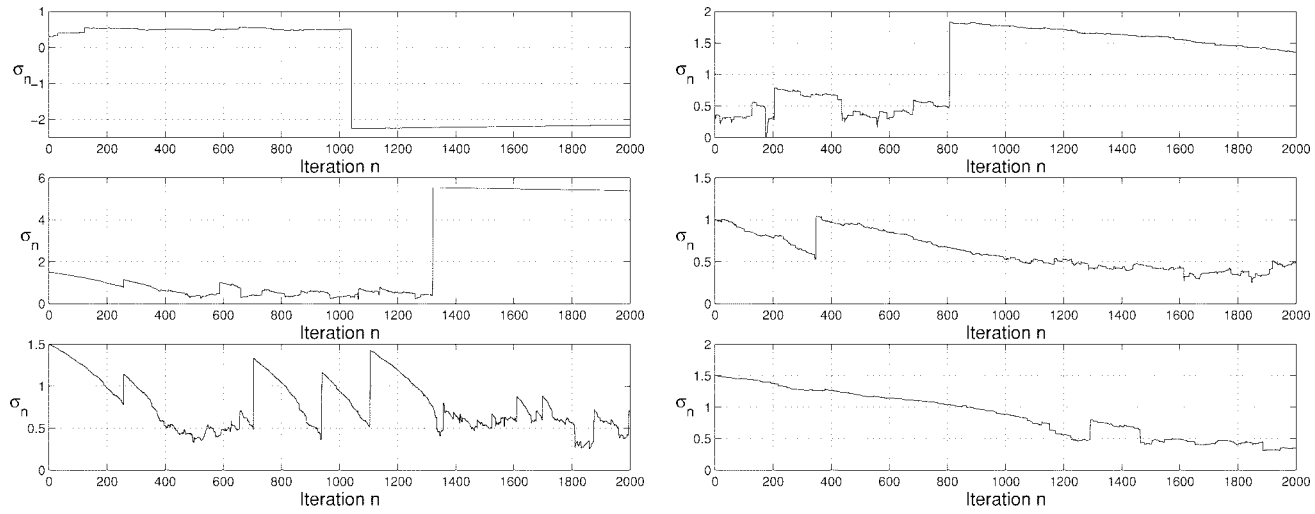
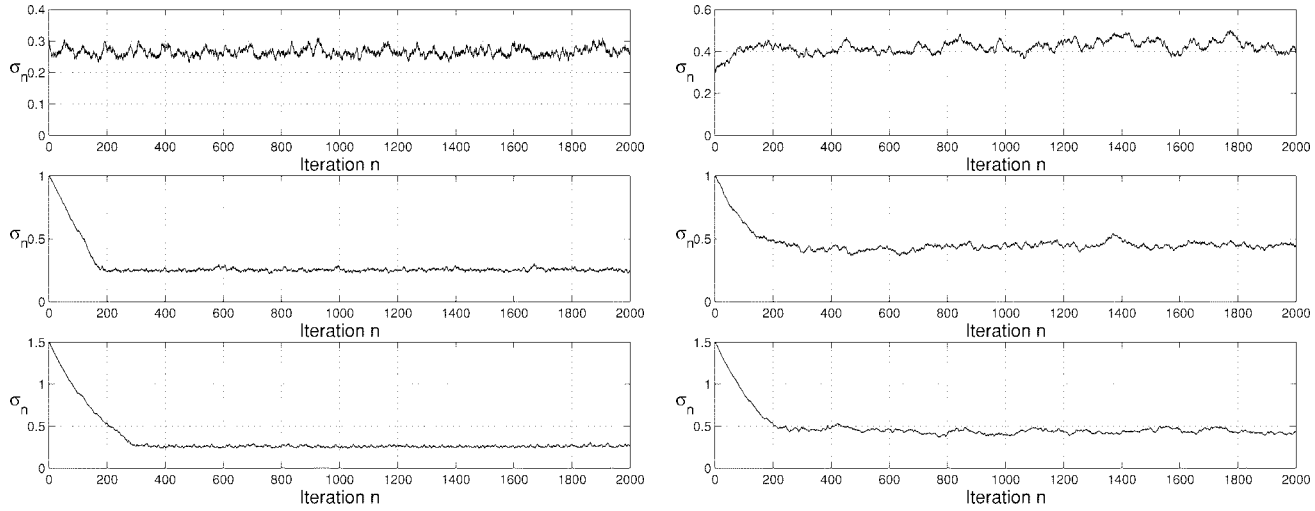


Fig. 10. Visual display of $\partial \mathrm{SNR}_n/\partial\sigma = (1/S_n)(\partial S_n/\partial\sigma) - (1/N_n)(\partial N_n/\partial\sigma)$ for the quartic bistable system $\dot{x} = x - x^3 + s(t) + n(t)$ where $s(t) = \varepsilon \sin 2\pi f$ with $\varepsilon = 0.1$ and $f = 0.01$ Hz. The system has linear output $y(t) = x(t)$ in (I) and binary output $y(t) = \mathrm{sgn}(x(t))$ in the (II). The noise variances are the constants $\sigma_n^2 = 0.25$. (a) Cauchy-like samples of $\partial \mathrm{SNR}_n/\partial\sigma$ 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, otherwise 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 $\log \mathrm{Prob}(X > u)$ 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.



(a) (b)

Fig. 11. Learning paths for the quartic bistable system with output (a) $y(t) = x(t)$ and (b) $y(t) = \mathrm{sgn}(x(t))$. The learning law takes the form (113). The optimal noise level is $\sigma \approx 0.5$ for both cases. The impulsiveness of the learning term $\partial \mathrm{SNR}/\partial\sigma$ destabilizes the learning process near the optimal noise level.



(a) (b)

Fig. 12. Learning paths for the quartic bistable system with output $y(t) = x(t)$. The learning law has the form (132). Optimal noise levels are (a) $\sigma \approx 0.35$ and (b) $\sigma \approx 0.5$. The learning paths converge close to the optimal levels.

$$= m_j(n) + \mu_n \left(\frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial F} \frac{\partial F}{\partial m_j} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial F} \cdot \frac{\partial F}{\partial m_j} \right). \quad (116)$$

We get (113) if σ replaces F and m_j. Appendix B derives the last partial derivative $\partial F/\partial m_j$ in the chain-rule expansion (73) 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 $\partial F/\partial m_j$. Formal stochastic approximation [219] further requires that the learning rate μ_n must decrease slowly but not too slowly

$$\sum_{n=1}^{\infty} \mu_n^2 < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \mu_n = \infty. \quad (117)$$

Linear decay terms $\mu_n = 1/n$ obey (117). We used small but constant learning rates in most simulations.

VI. SR Learning: Simulation Results

This section shows how the stochastic SR learning laws in Section V tend to find the optimal noise levels in many dynamical systems. 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 $\partial \mathrm{SNR}_n/\partial\sigma_n$. We used a Cauchy

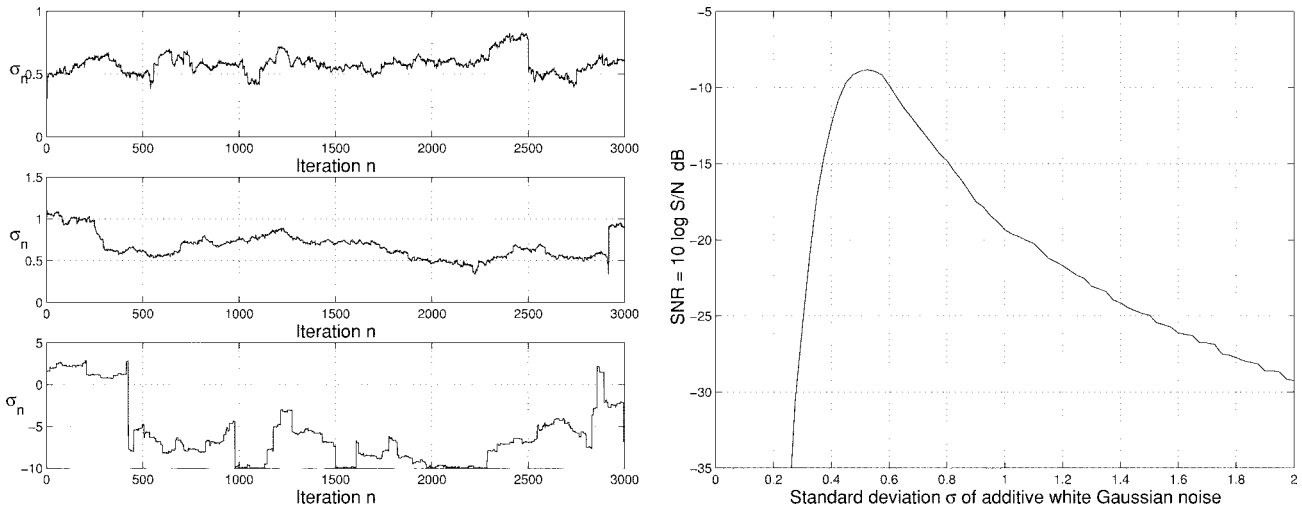


Fig. 13. Impulsive effects on learning paths of noise intensity σ_n. The quartic bistable system has the form $\dot{x} = x - x^3 + s(t) + n(t)$ with binary output $y(t) = \mathrm{sgn}(x(t))$ and initial condition $x(0) = -1$. The input sinusoid signal function is $s(t) = 0.1 \sin 2\pi(0.01)t$. (a) The 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 $\partial \mathrm{SNR}_n / \partial \sigma$ in the approximate SR learning law (137).

noise suppressor from the theory of robust statistics [112] to stabilize the learning process. Then sample paths of σ_n converged 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 applied the SNR measure to the quartic bistable and other dynamical systems with sinusoidal inputs. Future research may extend SR learning to wideband input signals. Fig. 24(a) shows how the optimum noise level varies for each input sinewave in the quartic bistable system. The learning process samples the system's input–output response as it learns the optimum noise. It does not make direct use of the equation that underlies the system.

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 [136], [200].

Consider the forced dynamical system in (1)–(2) with initial condition $x(0)$. We set up a discrete computer simulation with the stochastic version of Euler's method (the Euler–Maruyama scheme) [53], [86], [115]

$$x_{t+1} = x_t + \Delta T(f(x_t) + s_t) + \sigma\sqrt{\Delta T} w_t \quad (118)$$

$$y_t = g(x_t) \quad (119)$$

with initial condition $x_0 = x(0)$. Here the zero-mean white 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 [86], [115], [184]. 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 (118)–(119). The subsampling rate for the quartic bistable system is $1:50$. We ignored all aliasing effects.

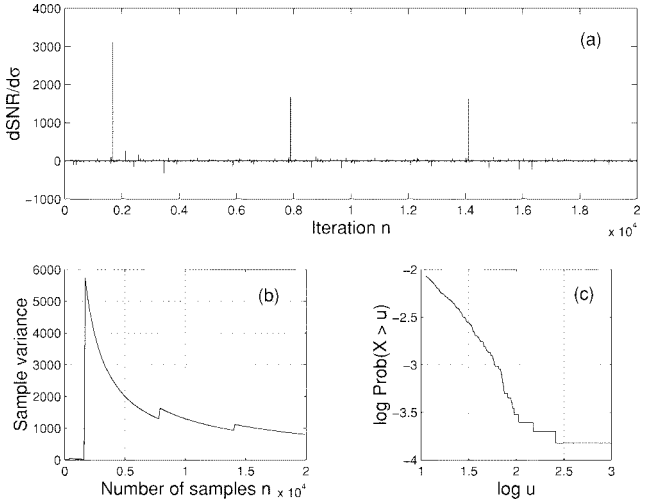


Fig. 14. Visual display of sample statistics of approximated $\partial \mathrm{SNR}_n / \partial \sigma$. (a) Cauchy-like samples of $\partial \mathrm{SNR}_n / \partial \sigma$ at each iteration n for quartic bistable system with sinusoidal input of amplitude $\varepsilon = 0.1$ and frequency $f_0 = 0.01$ Hz. We compute $\partial \mathrm{SNR}_n / \partial \sigma$ at each iteration from $\partial \mathrm{SNR}_n / \partial \sigma \approx [(S_n - S_{n-1})/S_n - (N_n - N_{n-1})/N_n] \, \mathrm{sgn}(\sigma_n - \sigma_{n-1})$ in (136). We vary the noise level σ_n between $\sigma_n = 0.50$ and $\sigma_n = 0.51$ so that $\mathrm{sgn}(\sigma_n - \sigma_{n-1})$ changes values between 1 and -1. The plot shows impulsiveness of the random variable $\partial \mathrm{SNR}_n / \partial \sigma$. (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 $\log \mathrm{Prob}(X > u)$ 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 (131) to the approximate SR gradient $\partial \mathrm{SNR}_n / \partial \sigma$ in (136) as well as to the exact SR gradient in (129).

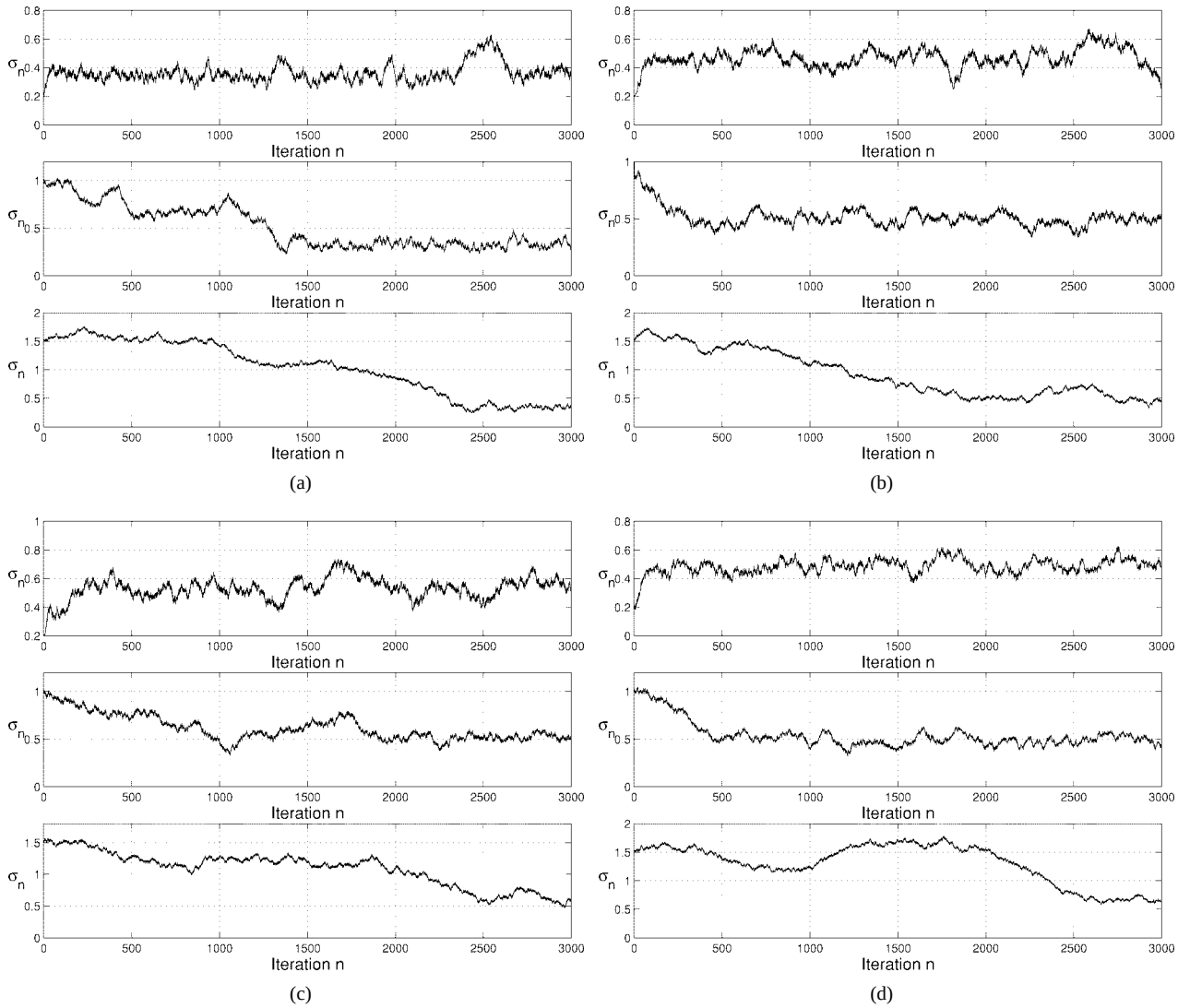


Fig. 15. Learning paths of σ_n with the Cauchy noise suppressor $\phi(z) = 2z/(1+z^2)$ for the quartic bistable system with binary threshold output $y_t = \mathrm{sgn}(x_t)$. The term $\phi(\partial \mathrm{SNR}_n/\partial\sigma)$replaces $\partial \mathrm{SNR}_n/\partial\sigma$ in the SR learning law (133). 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) $s(t) = 0.2 \sin 2\pi(0.01)t$.

A. SR Test Case: The Quartic Bistable System

We tested the quartic bistable system (4) 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 [185]

$$\dot{x} = x - x^3 + s(t) + n(t) \tag{120}$$

$$y(t) = \mathrm{sgn}(x(t)) \tag{121}$$

or $y(t) = x(t)$ in the linear-output case. The sinewave 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 (118)–(119):

$$x_{t+1} = x_t + \Delta T\left(x_t - x_t^3 + \varepsilon \sin 2\pi f_0 \Delta T t\right) + \sigma\sqrt{\Delta T} w_t \tag{122}$$

$$y_t = \mathrm{sgn}(x_t) \quad \text{or} \quad y_t = x_t \tag{123}$$

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 $T_{n+1} - T_n = 2000$ s for the simulations of the quartic bistable system. The sampling period was $T_s = 0.976$ s. 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 tested the quartic bistable system with the chaotic noise from the logistic map. Figs. 2, 5, and 18 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.

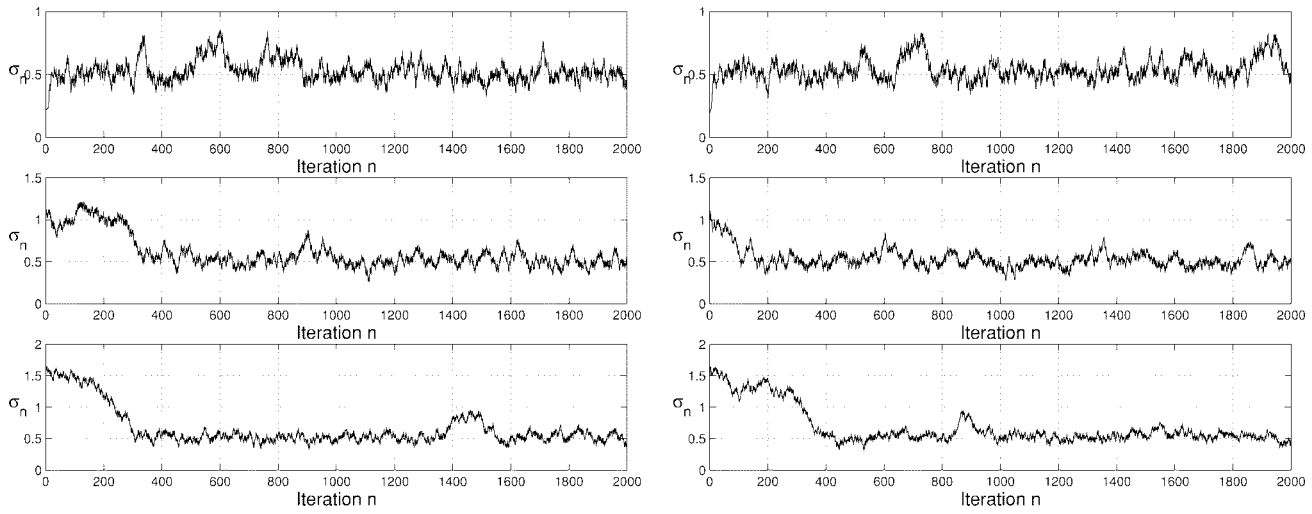


Fig. 16. Learning paths of σ_n for other noise densities in the quartic bistable system with binary output $y_t = \mathrm{sgn}(x_t)$. The input signal is $s(t) = 0.1 \sin 2\pi(0.01)t$. The optimal noise lies near $\sigma = 0.5$ for both cases of (a) Laplace noise and (b) uniform noise.

The Jacobian of the quartic bistable system has the form

$$\frac{\partial G}{\partial x} = \frac{\partial}{\partial x}\left[x - x^3 + s(t) + \sigma w(t)\right] \tag{124}$$

$$= 1 - 3x^2. \tag{125}$$

Then the partial derivative $\partial G/\partial \sigma = w(t)$ from (110) gives the evolution of $\eta(t) = \partial x/\partial \sigma$ for the quartic bistable system

$$\dot{\eta} = (1 - 3x^2)\eta(t) + w(t). \tag{126}$$

Its discrete version has the form

$$\eta(t+1) = \eta(t) + \Delta T\left(1 - 3x_t^2\right)\eta(t) + \sqrt{\Delta T} w_t. \tag{127}$$

We used the initial condition $\partial x_0/\partial \sigma = 0$ in simulations. Then we get $\partial y/\partial \sigma$ from (106) for use in the learning law (113). The linear output $y = g(x) = x$ has $\partial g/\partial x = 1$. We can approximate a binary output as $g(x) = \mathrm{sgn}(x) \approx \tanh(cx)$ for a large positive $c \gg 0$. Then $\partial g/\partial x = c(1 - \tanh^2(cx))$.

The equilibrium term $\mathcal{E}_n$ in (89) helps gauge the noisiness of the learning process. We compute $\mathcal{E}_n$ at each iteration n from

$$\mathcal{E}_n = \frac{S_n}{N_n} - \frac{\partial S_n}{\partial \sigma} \Big/ \frac{\partial N_n}{\partial \sigma}. \tag{128}$$

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. Fig. 9 shows $\mathcal{E}_n$ samples from the quartic bistable system (122)–(123) with Gaussian noise $n(t) = \sigma w(t)$. The convergence of variance test [223] confirms that $\mathcal{E}_n$ had infinite variance in our simulations. The log-tail test [223] 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 [70], [204] or if they obey certain more general statistical conditions [137], [141]. This suggests that much of the impulsive nature of $\mathcal{E}_n$ and hence of the learning process may stem from the ratio of derivatives in (128).

We also simulate the random gradient $\partial \mathrm{SNR}_n/\partial \sigma$ with the partial derivatives from (99), (103), and $\partial y/\partial \sigma$ from (108)

$$\frac{\partial \mathrm{SNR}_n}{\partial \sigma} = \frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma}. \tag{129}$$

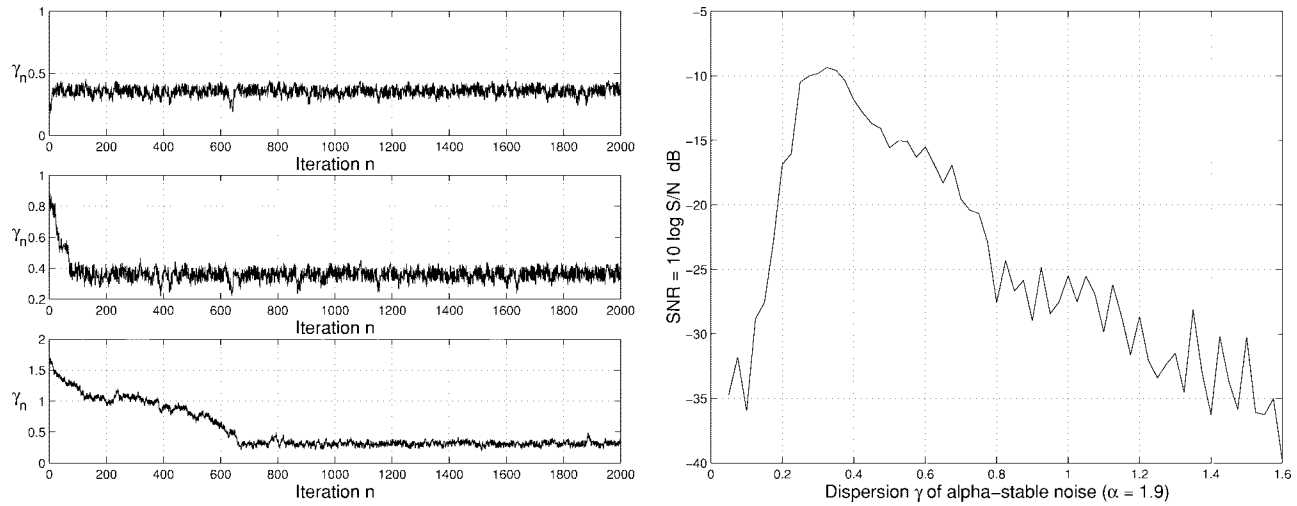
The simulations confirm that the random gradient $\partial \mathrm{SNR}_n/\partial \sigma$ is often impulsive and can destabilize the learning process (113) at or near the optimal noise level. The impulsiveness of $\partial \mathrm{SNR}_n/\partial \sigma$ in Fig. 10 suggests that $\partial \mathrm{SNR}_n/\partial \sigma$ may have an alpha-stable probability density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\partial \mathrm{SNR}_n/\partial \sigma$ again has an approximate Cauchy distribution.

We tested the learning law (113)

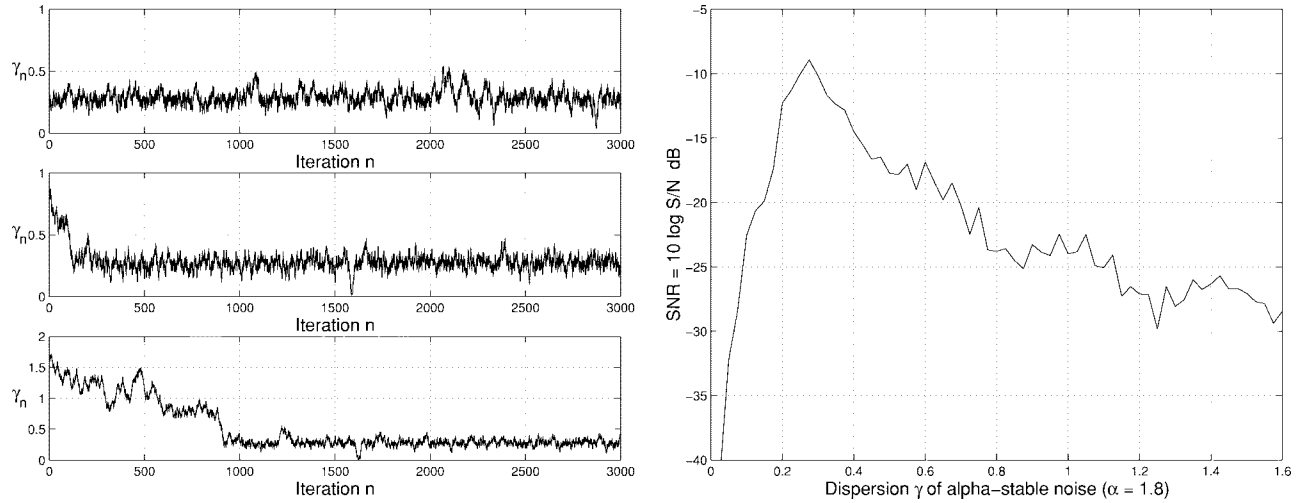
$$\sigma(n+1) = \sigma(n) + \mu_n \left(\frac{1}{S_n} \sum_{l=n+1-L}^{n} \frac{\partial S_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} - \frac{1}{N_n} \sum_{l=n+1-L}^{n} \frac{\partial N_n}{\partial y_l} \frac{\partial y_l}{\partial \sigma} \right). \tag{130}$$

Fig. 11 shows the simulation results. It displays the unstability in the learning due to the impulsiveness of the random gradient $\partial \mathrm{SNR}/\partial \sigma$.

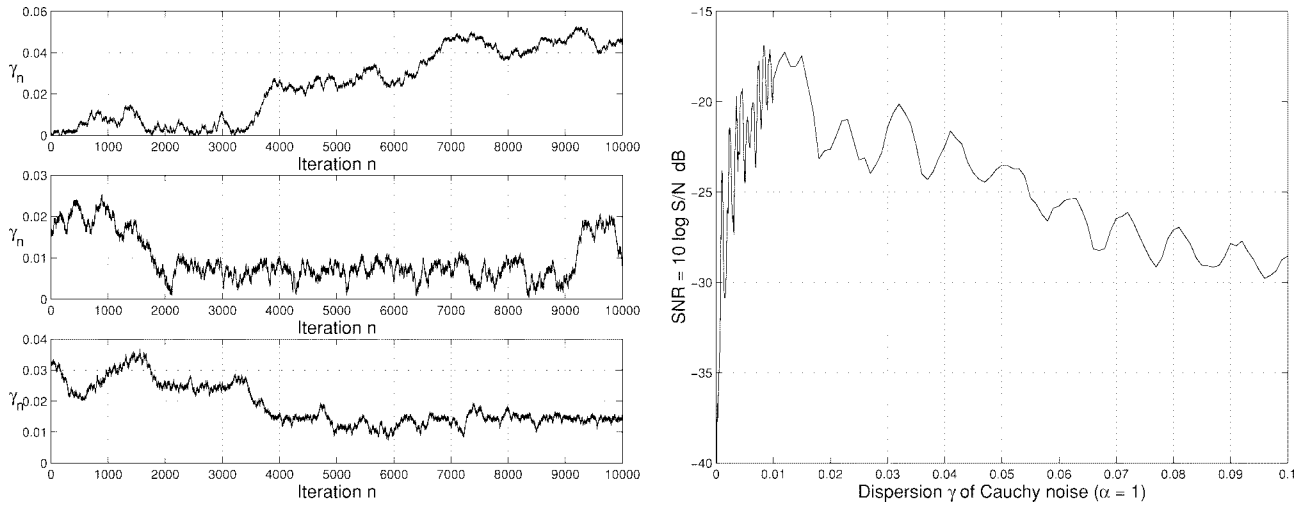
The theory of robust statistics [112] suggests one way to reduce the impulsiveness of $\partial \mathrm{SNR}_n/\partial \sigma$. We can replace the noisy random sample z_n with a Cauchy-like noise



(a)

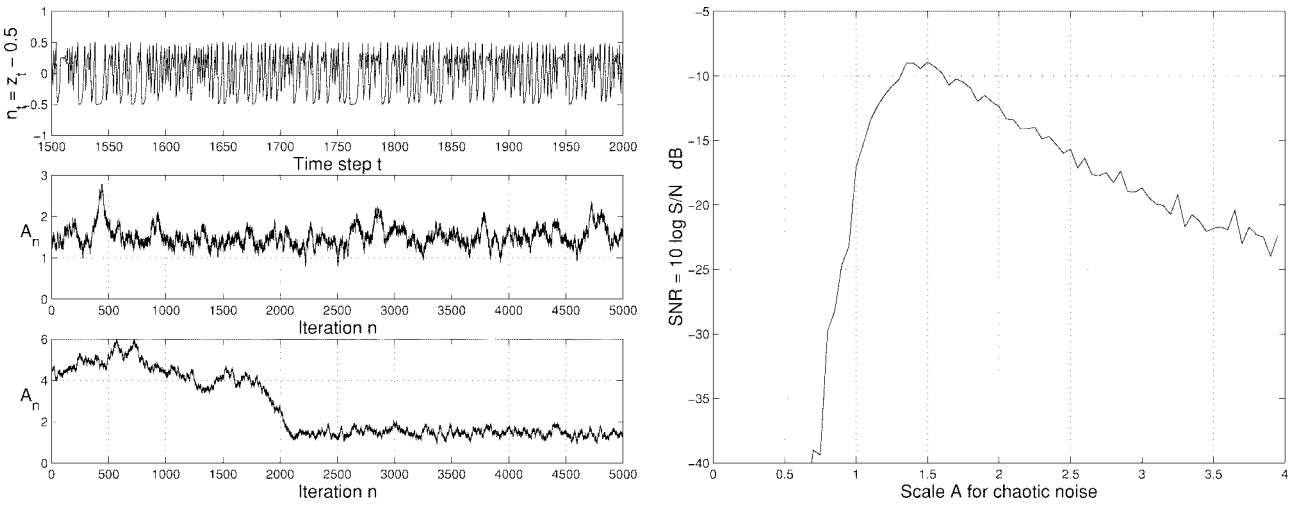


(b)



(c)

Fig. 17. Learning paths of γ_n for alpha-stable noise in the quartic bistable system with binary output $y_t = \mathrm{sgn}(x_t)$. 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 dispersion γ acts like a standard deviation and controls the width of the alpha-stable bell curve. 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 γ_n often fails to converge. Note the noisy multimodal nature of the SNR profiles.



(a) (b)

Fig. 18. Learning paths of the scaling factor A_n in chaotic noise $n_t = A_n(z_t - 1/2)$ from the logistic dynamical system $z_{t+1} = 4z_t(1 - z_t)$. The dynamical system is the quartic bistable system with binary output $y_t = \operatorname{sgn}(x_t)$. 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 graph in (a) shows a sample noise path n_t from the chaotic logistic map when $A_n = 1$.

suppressor $\phi(z_n)$ [112]

$$\phi(z_n) = \frac{2z_n}{1 + z_n^2}. \tag{131}$$

So $\phi(\partial \mathrm{SNR}_n/\partial\sigma)$ replaces the noise gradient $\partial \mathrm{SNR}_n/\partial\sigma$ in (129). This gives the robust SR learning law

$$\sigma(n+1) = \sigma(n) + \mu_n \phi\left(\frac{\partial \mathrm{SNR}_n}{\partial \sigma}\right). \tag{132}$$

Fig. 12 shows the results of the SR learning law (132) with the gradient in (129). The σ_n learning paths in (113) converge near the optimal noise level.

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 (108). This may not be practical in many cases. So we instead sample S_n and N_n and use the approximation formulas (90) and (91). This gives the learning law

$$\sigma_{n+1} = \sigma_n + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} \tag{133}$$

$$= \sigma_n + \mu_n \left(\frac{1}{S_n}\frac{\partial S_n}{\partial \sigma} - \frac{1}{N_n}\frac{\partial N_n}{\partial \sigma}\right) \tag{134}$$

$$\approx \sigma_n + \mu_n \left(\frac{1}{S_n}\frac{S_n - S_{n-1}}{\sigma_n - \sigma_{n-1}} - \frac{1}{N_n}\frac{N_n - N_{n-1}}{\sigma_n - \sigma_{n-1}}\right). \tag{135}$$

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

$$\frac{\partial \mathrm{SNR}_n}{\partial \sigma} \approx \left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n}\right) \cdot \operatorname{sgn}(\sigma_n - \sigma_{n-1}). \tag{136}$$

This approximation gives the SR learning law

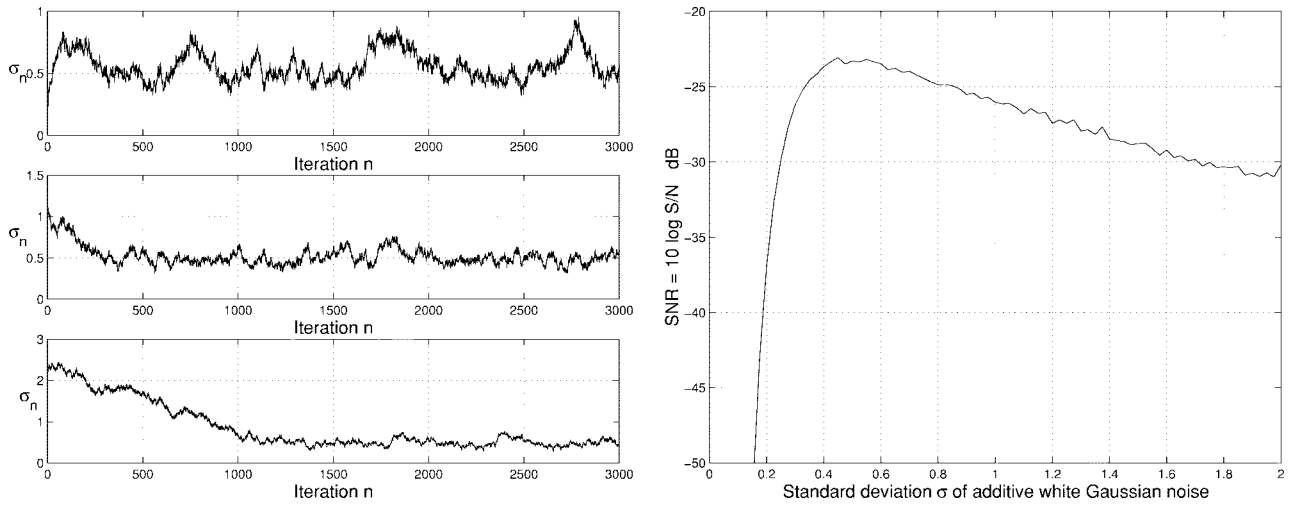
$$\sigma_{n+1} = \sigma_n + \mu_n \left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n}\right) \cdot \operatorname{sgn}(\sigma_n - \sigma_{n-1}). \tag{137}$$

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 $s(t)$.

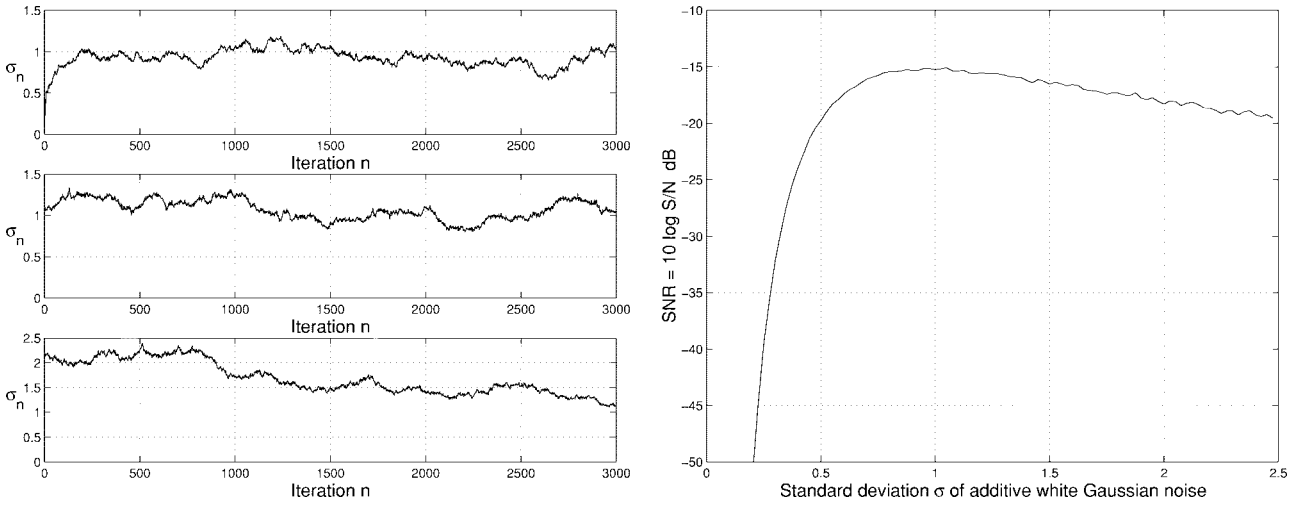
Fig. 13(a) shows sample learning paths of σ_n for the quartic bistable system and approximation (136). Fig. 13(b) shows the noise-SNR profile of the dynamical system. The σ_n learning paths converge to the optimum noise values only in some cases. The chance of path convergence is higher for larger sinewave amplitudes. The paths do not converge as often for small amplitudes. The simulations confirm that the random gradient $\partial \mathrm{SNR}_n/\partial\sigma_n$ in (136) is often impulsive and can destabilize the learning process (137) as in Fig. 13. The impulsiveness of $\partial \mathrm{SNR}_n/\partial\sigma$ in Fig. 14 suggests that $\partial \mathrm{SNR}_n/\partial\sigma$ may have an alpha-stable probability density function with parameter $\alpha < 2$. A log-tail test found that $\alpha \approx 1$. So $\partial \mathrm{SNR}_n/\partial\sigma$ in (136) also has an approximate Cauchy distribution.

We again apply the Cauchy-like noise suppressor from robust statistics [112] to reduce the impulsiveness of the approximated term $\partial \mathrm{SNR}_n/\partial\sigma$ in (136). So $\phi(\partial \mathrm{SNR}_n/\partial\sigma)$ replaces the approximation of the noise gradient $\partial \mathrm{SNR}_n/\partial\sigma$ in (136) to give the robust SR learning law

$$\sigma_{n+1} = \sigma_n + \mu_n \phi\left(\left(\frac{S_n - S_{n-1}}{S_n} - \frac{N_n - N_{n-1}}{N_n}\right) \cdot \operatorname{sgn}(\sigma_n - \sigma_{n-1})\right). \tag{138}$$



(a)



(b)

Fig. 19. SR learning paths of σ_n for the threshold system $y_t = \mathrm{sgn}(s_t + n_t - \Theta)$ where $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$. The input sinewave 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 $\Theta = 1$.

Fig. 15 shows the results of the SR learning law (138). 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. Fig. 16 shows σ_n learning paths for the quartic bistable system (122)–(123) with Laplace noise and uniform noise. We also tested the quartic bistable system with alpha-stable noise. Fig. 17 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.

We also used a chaotic time series as the forcing noise n_t in the quartic bistable dynamical system [118]. The simple and popular logistic map created the noise sequence $\{z_t\}$

$$z_{t+1} = 4z_t(1 - z_t) \tag{139}$$

from the initial value $z_0 = .123\,456\,789$ [118]. The positive sequence $\{z_t\}$ stays bounded within the unit interval: $z_t \in (0, 1)$. The chaotic noise n_t comes from

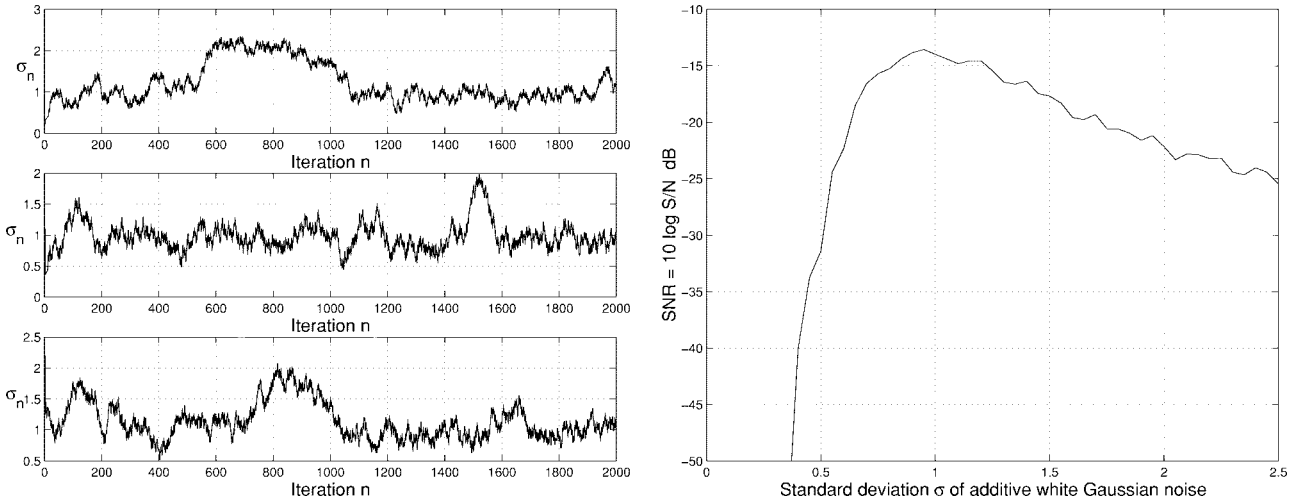
$$n_t = A\left(z_t - \frac{1}{2}\right). \tag{140}$$

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. Fig. 18 shows a sample chaotic noise sequence and shows two A learning paths on their way to stochastic convergence.

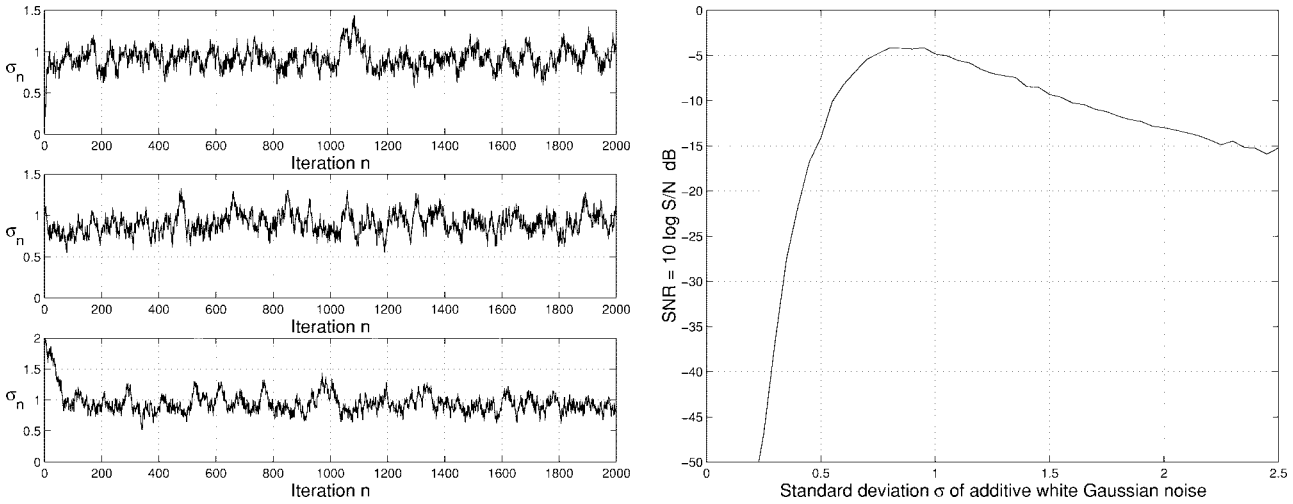
B. Other SR Test Cases

The SR learning schemes also work for other SR models. We here show only the results for zero-mean white Gaussian noise. We first tested the discrete-time threshold neuron model

$$y_t = \begin{cases} 1, & \text{if } s_t + n_t \geq \Theta \\ -1, & \text{if } s_t + n_t < \Theta \end{cases} \tag{141}$$



(a)



(b)

Fig. 20. 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 $y(t) = \mathrm{sgn}(x(t))$. The parameters of the input sinewaves are $f_0 = 0.01$ Hz and (a) $\varepsilon = 0.1$ and (b) $\varepsilon = 0.3$.

for $t = 0, 1, 2, \ldots$ The threshold Θ sets the output of the neuron. The input sinewave 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. Fig. 19 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 $T_s = \Delta T = 1$.

We next tested the bistable potential neuron model with Gaussian white noise [27]

$$\dot{x} = -x + 2\tanh x + s(t) + n(t) \quad (142)$$

$$y(t) = \mathrm{sgn}(x(t)). \quad (143)$$

We ignored the multiplicative noise in (8). Fig. 20 shows the SR learning paths of σ_n. The sinewave input is $s(t) = \varepsilon \sin 2\pi f_0 t$ where $f_0 = 0.01$ Hz and $\varepsilon = 0.1$ and $\varepsilon = 0.3$. 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.

We next tested the forced FHN neuron model [183]. We rewrote (14)–(15) with $a = 0.5$ and with the changes of variables $x \to x + 0.5$, $w \to w - b + 0.5$, and $A \to A - b + 0.5$ [46]

$$\epsilon \dot{x} = -x\left(x^2 - \frac{1}{4}\right) - w + A + s(t) + n(t) \quad (144)$$

$$\dot{w} = x - w \quad (145)$$

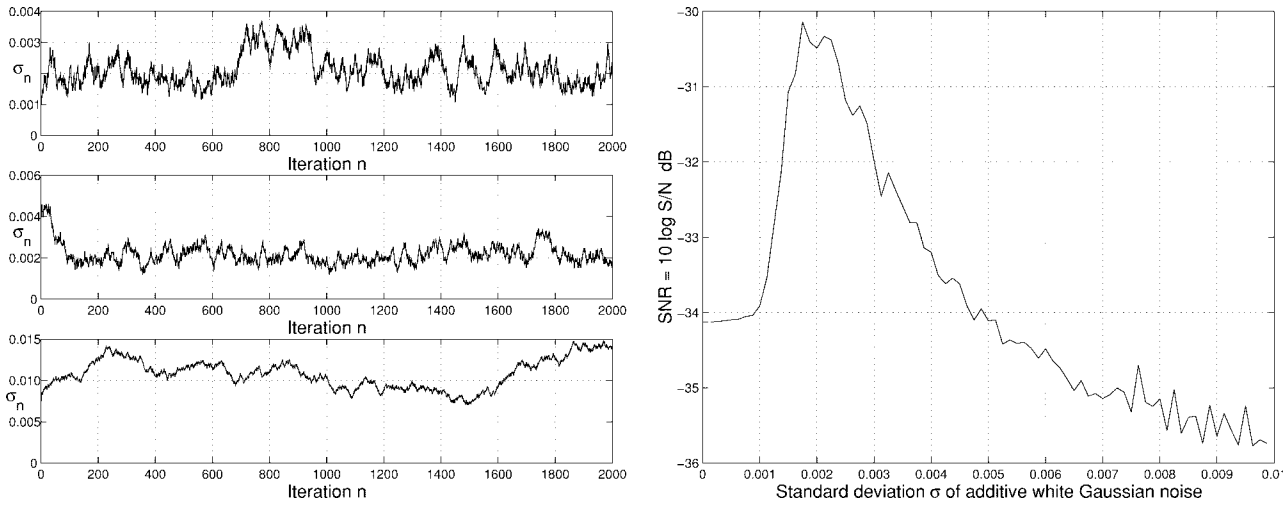
$$y(t) = x(t). \quad (146)$$

The constants are $\epsilon = 0.005$, $a = 0.5$, and $A = -(5/12\sqrt{3} + 0.07) = -.310\,56$ as in [102]. The sinewave input is $s(t) = \varepsilon \sin 2\pi f_0 t$ with $\varepsilon = 0.01$, $f_0 = 0.1$ and 0.5 Hz. The sampling period is $T_s = 0.01$ with $\Delta T = 0.001$. Fig. 21 shows the learning paths of the standard deviation σ_n of the Gaussian white noise n.

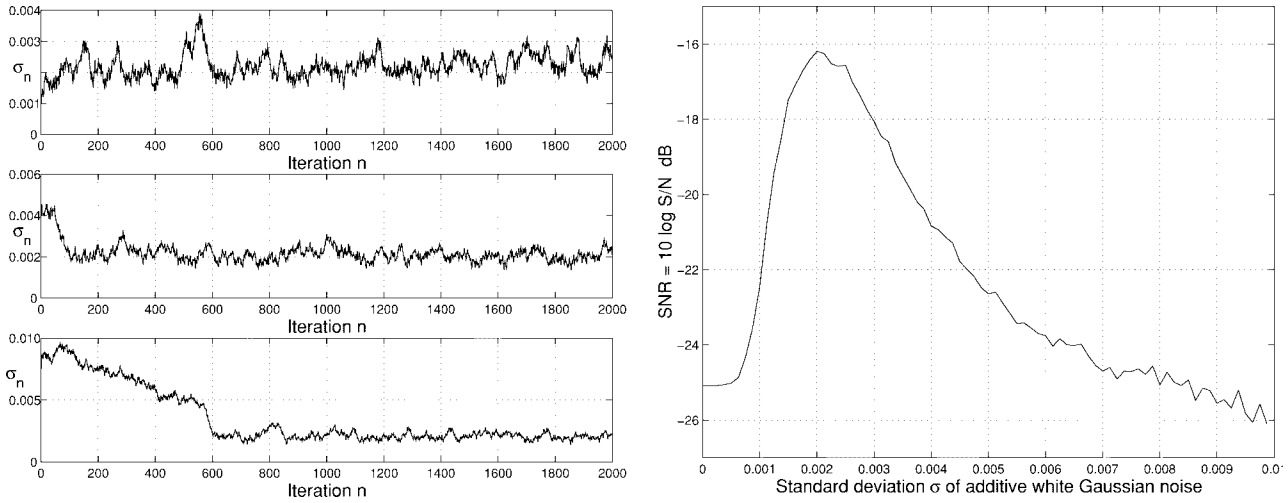
We also showed SR learning in the forced Duffing oscillator with Gaussian white noise n [196]

$$\ddot{x} = -0.15\dot{x} + x + x^3 + \varepsilon \sin(\omega_0 t) + n(t) \quad (147)$$

$$y(t) = x(t). \quad (148)$$



(a)



(b)

Fig. 21. SR learning paths of σ_n for the FHN neuron model $\epsilon\dot{x} = -x(x^2 - \frac{1}{4}) - w + A + s(t) + n(t)$ and $\dot{w} = x - w$ with output $y(t) = x(t)$. The parameters are $\epsilon = 0.005$ and $A = -(5/12\sqrt{3} + 0.07) = -.310\,56$. The sinewave 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. (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 it leads to convergence in the third learning sample in (b).

Fig. 22 shows the learning paths of σ_n for input sinewave 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 $\Delta T = 0.005$.

C. Fuzzy SR Learning: The Quartic Bistable System

We used 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 input sinewaves. The fuzzy system had as its input the 2-D vector of sinewave amplitude ε and frequency f_0. We tested the system with the fixed input initial value $x(0) = -1$. The fuzzy system itself defined a vector function $F : R^2 \rightarrow R$ and used 200 rules. The chain rule extended the learning laws in the previous sections to tune the fuzzy system's parameters m_j as in (71)

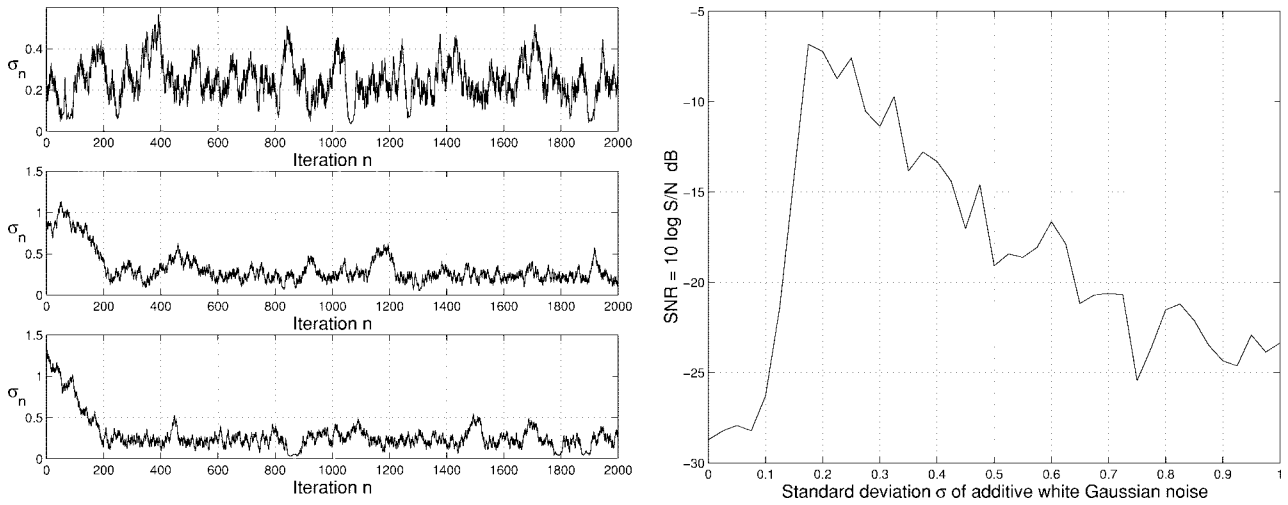
$$m_j(n+1) = m_j(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial m_j} \tag{149}$$

$$= m_j(n) + \mu_n \frac{\partial \mathrm{SNR}_n}{\partial \sigma} \frac{\partial F}{\partial m_j}. \tag{150}$$

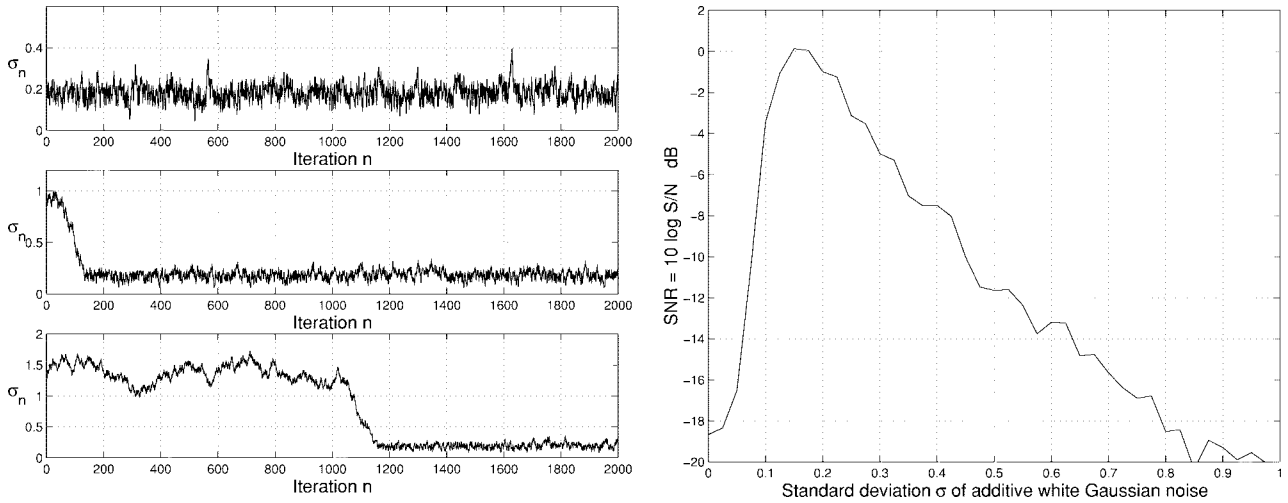
Appendix B derives the partial derivative $\frac{\partial F}{\partial m_j}$ for the sinc SAM fuzzy system that we used. The Cauchy noise suppressor gives the learning law as

$$m_j(n+1) = m_j(n) + \mu_n \phi\left(\frac{\partial \mathrm{SNR}_n}{\partial \sigma}\right) \frac{\partial F}{\partial m_j}. \tag{151}$$

Fig. 23 shows how we formed 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



(a)



(b)

Fig. 22. 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 input sinewaves are $f_0 = 0.01$ Hz and (a) $\varepsilon = 0.1$ and (b) $\varepsilon = 0.3$.

if-part sinc set. Fig. 24 shows the results of SAM learning of the optimal noise pattern for the quartic bistable system. The sinc SAM used 200 rules. Fewer rules gave a coarser approximation.

VII. 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. 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.

Fourier techniques may not extend well to the general case or broadband or nonperiodic forcing signals found in many nonlinear and nonstationary environments. Wavelet transforms [54], [138], [234], [237] may offer better ways to measure SR effects in these cases when nonperiodic signals drive nonlinear dynamical systems. Wavelet transforms can adaptively localize nonperiodic signals in both time and frequency. Fourier techniques tend to localize periodic signals either in frequency or in time. Arbitrary or random broadband signals may require new techniques to detect these signals and extract their key statistical features from their noisy dynamical backgrounds.

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.

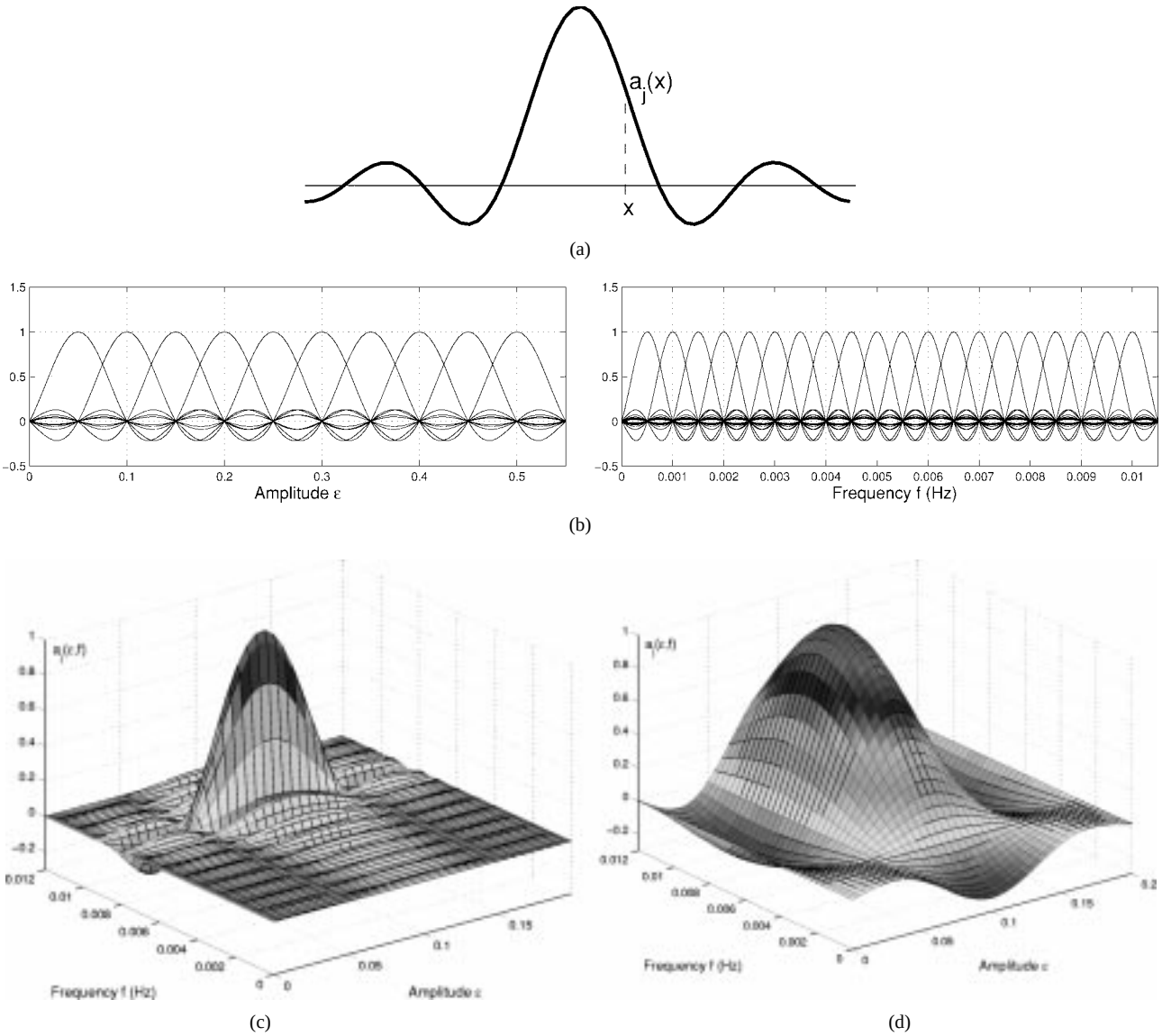


Fig. 23. If-part sinc fuzzy sets. (a) Scalar sinc set function $a_j(x) = \sin x/x$. Sinc sets are generalized fuzzy sets with "membership values" in $[-.217,1]$. Element x belongs to set A_j to degree $a_j(x)$: Degree$(x \in A_j) = a_j(x)$. (b) Initial subsets for sinewave amplitudes and frequencies. There are ten fuzzy sets for amplitudeε and 20 fuzzy sets for frequency f_0. The product of two one-dimensional (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.

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?

APPENDIX A
THE STANDARD ADDITIVE MODEL (SAM) THEOREM

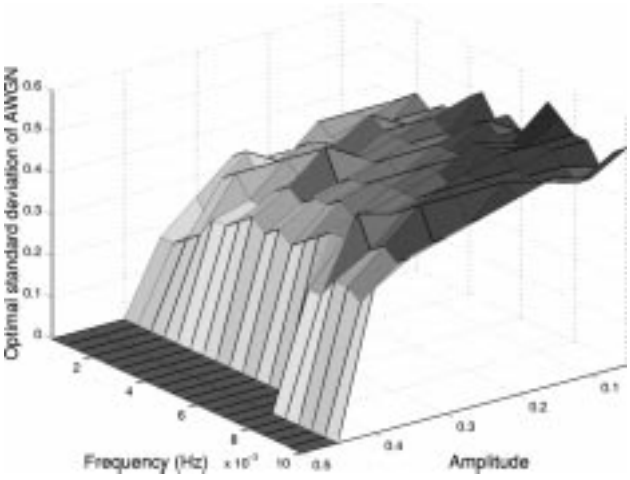
This appendix derives the basic ratio structure (31) of a SAM fuzzy system and review the local structure of optimal fuzzy rules.

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

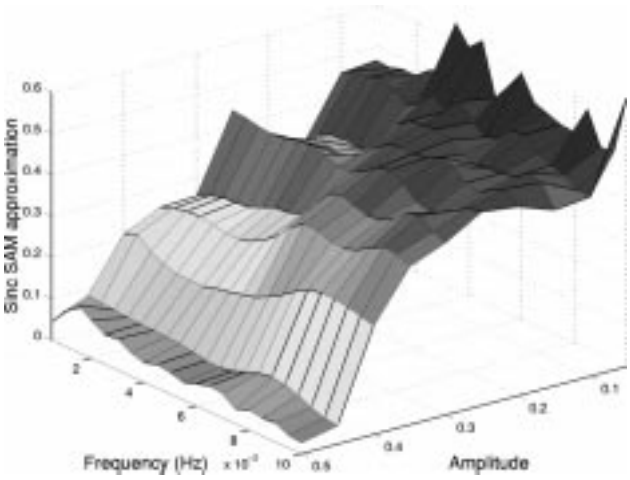
$$F(x) = \frac{\sum_{j=1}^m w_j a_j(x) V_j c_j}{\sum_{j=1}^m w_j a_j(x) V_j} = \sum_{j=1}^m p_j(x) c_j. \qquad (152)$$

The convex coefficients or discrete probability weights $p_1(x), \ldots, p_m(x)$ depend on the input x through

$$p_j(x) = \frac{w_j a_j(x) V_j}{\sum_{i=1}^m w_i a_i(x) V_i}. \qquad (153)$$



(a)



(b)

Fig. 24. Optimal noise levels in terms of the SNR for the quartic bistable system with binary output. (a) The optimum noise pattern when inputs are sinewaves with distinct amplitudes and frequencies. (b) SAM fuzzy approximation of the optimum noise after 30 epochs. The sinc SAM used 200 rules. One epoch used 20 iterations that trained 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.

V_j is the finite positive volume (or area if $p = 1$) and c_j is the centroid of then-part set B_j

$$V_j = \int_{R^p} b_j(y_1, \ldots, y_p) dy_1 \cdots dy_p > 0 \tag{154}$$

$$c_j = \frac{\int_{R^p} y b_j(y_1, \ldots, y_p) dy_1 \cdots dy_p}{\int_{R^p} b_j(y_1, \ldots, y_p) dy_1 \cdots dy_p}. \tag{155}$$

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 (154) and (155) as

$$V_j = \int_{-\infty}^{\infty} b_j(y)\, dy \tag{156}$$

$$c_j = \frac{\int_{-\infty}^{\infty} y b_j(y)\, dy}{\int_{-\infty}^{\infty} b_j(y)\, dy}. \tag{157}$$

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

$$F(x) = \text{Centroid}(B(x)) \tag{158}$$

$$= \frac{\int_{-\infty}^{\infty} y b(y)\, dy}{\int_{-\infty}^{\infty} b(y)\, dy} \tag{159}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^m w_j b'_j(y)\, dy}{\int_{-\infty}^{\infty} \sum_{j=1}^m w_j b'_j(y)\, dy} \tag{160}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^m w_j a_j(x) b_j(y)\, dy}{\int_{-\infty}^{\infty} \sum_{j=1}^m w_j a_j(x) b_j(y)\, dy} \tag{161}$$

$$= \frac{\sum_{j=1}^m w_j a_j(x) \int_{-\infty}^{\infty} y b_j(y)\, dy}{\sum_{j=1}^m w_j a_j(x) \int_{-\infty}^{\infty} b_j(y)\, dy} \tag{162}$$

$$= \frac{\sum_{j=1}^m w_j a_j(x) V_j \frac{\int_{-\infty}^{\infty} y b_j(y)\, dy}{V_j}}{\sum_{j=1}^m w_j a_j(x) V_j} \tag{163}$$

$$= \frac{\sum_{j=1}^m w_j a_j(x) V_j c_j}{\sum_{j=1}^m w_j a_j(x) V_j}. \tag{164}$$

Now we give a simple *local* description of optimal lone fuzzy rules [135], [136]. 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

$$e(x) = (f(x) - F(x))^2 = (f(x) - c)^2. \tag{165}$$

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

$$0 = \frac{\partial e(x)}{\partial x} = 2(f(x) - c)\frac{\partial f(x)}{\partial x}. \tag{166}$$

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

$$\frac{\partial e(x)}{\partial x} = 0 \Leftrightarrow \frac{\partial f(x)}{\partial x} = 0 \tag{167}$$

at $x = \hat{x}$. So the extrema of e and f coincide in this case. Fig. 7 shows how fuzzy rule patches can "patch the bumps" and so help minimize the error of approximation. □

Appendix B
SAM Gradient Learning

Supervised gradient ascent can tune all the parameters in the SAM model (31) [134], [136]. A gradient ascent learning law for a SAM parameter ξ has the form

$$\xi(t+1) = \xi(t) + \mu_t \frac{\partial P}{\partial \xi} \tag{168}$$

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 SNR defines the performance P.

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

$$\frac{\partial \mathrm{SNR}}{\partial \xi_j^k} = \frac{\partial \mathrm{SNR}}{\partial F}\frac{\partial F}{\partial a_j}\frac{\partial a_j}{\partial \xi_j^k}, \quad \frac{\partial \mathrm{SNR}}{\partial c_j} = \frac{\partial \mathrm{SNR}}{\partial F}\frac{\partial F}{\partial c_j}$$

and

$$\frac{\partial \mathrm{SNR}}{\partial V_j} = \frac{\partial \mathrm{SNR}}{\partial F}\frac{\partial F}{\partial V_j}. \tag{169}$$

We have derived the partial derivative $\partial \mathrm{SNR}/\partial F = \partial \mathrm{SNR}/\partial \sigma$ in Section V-B. We next derive the partial derivatives for the SAM parameters

$$\frac{\partial F}{\partial a_j} = \frac{\left(\sum_{i=1}^m w_i a_i(x) V_i\right)(w_j V_j c_j) - w_j V_j \left(\sum_{i=1}^m w_i a_i(x) V_i c_i\right)}{\left(\sum_{i=1}^m w_i a_i(x) V_i\right)^2} \tag{170}$$

$$= \frac{[c_j - F(x)] w_j V_j}{\sum_{i=1}^m w_i a_i(x) V_i} = [c_j - F(x)]\frac{p_j(x)}{a_j(x)}. \tag{171}$$

The SAM ratio (31) gives [134]

$$\frac{\partial F}{\partial c_j} = \frac{w_j a_j(x) V_j}{\sum_{i=1}^m w_i a_i(x) V_i} = p_j(x) \tag{172}$$

and

$$\frac{\partial F}{\partial V_j} = \frac{w_j a_j(x)[c_j - F(x)]}{\sum_{i=1}^m w_i a_i(x) V_i} = \frac{p_j(x)}{V_j}[c_j - F(x)]. \tag{173}$$

Then the learning laws for the centroid and volume have the final form

$$c_j(t+1) = c_j(t) + \mu_t \frac{\partial \mathrm{SNR}}{\partial \sigma} p_j(x) \tag{174}$$

and

$$V_j(t+1) = V_j(t) + \mu_t \frac{\partial \mathrm{SNR}}{\partial \sigma}\frac{p_j(x)}{V_j}[c_j - F(x)]. \tag{175}$$

Learning laws for set parameters depend on how we define the set functions. The partial derivatives for the scalar sinc set function $a_j(x) = \sin(\frac{x-m_j}{d_j})/(\frac{x-m_j}{d_j})$ have the form

$$\frac{\partial a_j}{\partial m_j} = \begin{cases} \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{x-m_j}, & \text{for } x \neq m_j \\ 0, & \text{for } x = m_j \end{cases} \tag{176}$$

$$\frac{\partial a_j}{\partial d_j} = \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{d_j}. \tag{177}$$

So this scalar set function leads to the learning laws

$$m_j(t+1) = m_j(t) + \mu_t \frac{\partial \mathrm{SNR}}{\partial \sigma}[c_j - F(x)]\frac{p_j(x)}{a_j(x)} \times \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{x-m_j} \tag{178}$$

$$d_j(t+1) = d_j(t) + \mu_t \frac{\partial \mathrm{SNR}}{\partial \sigma}[c_j - F(x)]\frac{p_j(x)}{a_j(x)} \times \left(a_j(x) - \cos\left(\frac{x-m_j}{d_j}\right)\right)\frac{1}{d_j}. \tag{179}$$

Like results hold for the learning laws of product n-D set functions. A factored set function $a_j(x) = a_j^1(x_1)\ldots a_j^n(x_n)$ leads to a new form for the performance gradient. The gradient with respect to the parameter m_j^k of the jth set function a_j has the form

$$\frac{\partial P}{\partial m_j^k} = \frac{\partial P}{\partial F}\frac{\partial F}{\partial a_j}\frac{\partial a_j}{\partial a_j^k}\frac{\partial a_j^k}{\partial m_j^k}$$

where

$$\frac{\partial a_j}{\partial a_j^k} = \prod_{i \neq k}^n a_j^i(x_i) = \frac{a_j(x)}{a_j^k(x_k)}. \tag{180}$$

Products of the scalar sinc set functions defined the if-part fuzzy sets $A_j \subset R^n$ in the SAM approximator. Simulations have shown [174], [175] that sinc set functions tend to perform at least as well as other popular set functions in supervised fuzzy function approximation.

REFERENCES

[1] Y. S. Abu-Mostafa, "Learning from hints in neural networks," *J. Complexity*, vol. 6, pp. 192–198, 1990.

[2] ——, "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," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 183–196, 1993.

[4] V. S. Anishchenko, M. A. Safonova, and L. O. Chua, "Stochastic resonance in Chua's circuit," *Int. J. Bifurcation and Chaos*, vol. 2, no. 2, pp. 397–401, 1992.

[5] ——, "Stochastic resonance in Chua's circuit driven by amplitude or frequency modulated signals," *Int. J. 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 *Proc. 1995 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-95)*, vol. 2, pp. 1332–1335.

[7] P. Babinec, "Stochastic resonance in the Weidlich model of public opinion formation," *Phys.s Lett. A*, vol. 225, pp. 179–181, Jan. 1997.

[8] P. Baldi, "Gradient descent learning algorithm overview: A general dynamical systems perspective," *IEEE Trans. Neural Networks*, vol. 6, pp. 182–195, Jan. 1995.

[9] R. Bartussek, P. Hänggi, and P. Jung, "Stochastic resonance in optical bistable systems," *Phys. Rev. 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] ——, "A theory of stochastic resonance in climatic change," *SIAM J. Appl. Math.*, vol. 43, no. 3, pp. 565–578, June 1983.

[13] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Physics A: Math. and General*, vol. 14, pp. L453–L457, 1981.

[14] R. Benzi, A. Sutera, and A. Vulpiani, "Stochastic resonance in the Landau-Ginzberg equation," *J. Physics A: Mathematical and General*, vol. 18, pp. 2239–2245, 1985.
[15] S. M. Bezrukov and I. Vodyanoy, "Stochastic resonance in nondynamical systems without response thresholds," *Nature*, vol. 385, pp. 319–321, Jan. 1997.
[16] Y. Braiman, J. F. Lindner, and W. L. Ditto, "Taming spatiotemporal chaos with disorder," *Nature*, vol. 378, pp. 465–467, Nov. 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, Jan. 1994.
[18] L. Breiman, *Probability*. Reading, MA: Addision-Wesley, 1968.
[19] J. J. Brey, J. Casado-Pascul, and B. Sánchez, "Resonant behavior of a Poisson process driven by a periodic signal," *Phys. Rev. E*, vol. 52, no. 6, pp. 6071–6081, Dec. 1995.
[20] J. J. Brey and A. Prados, "Stochastic resonance in a one-dimension Ising model," *Phys.s Lett. A*, vol. 216, pp. 240–246, June 1996.
[21] K. S. Brown, "Noises on," *New Scientist*, vol. 150, pp. 28–31, June 1996.
[22] 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," *J. Appl. Physics*, vol. 62, no. 7, pp. 2898–2913, Oct. 1987.
[23] A. Bulsara, S. Chillemi, L. Kiss, P. V. E. McClintock, R. Mannella, F. Maresoni, K. Nicolis, and K. Wiesenfeld, Eds., *Il Nuovo Cimento, Special Issue on Fluctuation in Physics and Biology: Stochastic Resonance, Signal Processing, and Related Phenomena*, Luglio-Agosto, vol. 17 D, nos. 7–8, 1995.
[24] A. R. Bulsara, R. D. Boss, and E. W. Jacobs, "Noise effects in an electronic model of a single neuron," *Biological Cybern.*, vol. 61, pp. 211–222, 1989.
[25] 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," *Phys. Rev. E*, vol. 53, no. 4, pp. 3958–3969, Apr. 1996.
[26] A. R. Bulsara and L. Gammaitoni, "Tuning in to noise," *Phys.s Today*, pp. 39–45, Mar. 1996.
[27] A. R. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss, "Stochastic resonance in a single neuron model: Theory and analog simulation," *J. Theoretical Biology*, vol. 152, pp. 531–555, 1991.
[28] 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," *Phys. Rev. E*, vol. 49, no. 6, pp. 4989–5000, June 1994.
[29] A. R. Bulsara, A. J. Maren, and G. Schmera, "Single effective neuron: Dendritic coupling effects and stochastic resonance," *Biological Cybern.*, vol. 70, pp. 145–156, 1993.
[30] A. R. Bulsara and G. Schmera, "Stochastic resonance in globally coupled nonlinear oscillators," *Phys. Rev. E*, vol. 47, no. 5, pp. 3734–3737, May 1993.
[31] A. R. Bulsara and A. Zador, "Threshold detection of wideband signals: A noise-induced maximum in the mutual information," *Phys. Rev. E*, vol. 54, no. 3, pp. R2185–R2188, Sept. 1996.
[32] G. C. Burdea, *Force and Touch Feedback for Virtual Reality*. New York: Wiley, 1996.
[33] T. L. Carroll and L. M. Pecora, "Stochastic resonance and crises," *Phys. Rev. Lett.*, vol. 70, no. 5, pp. 576–579, Feb. 1993.
[34] J. M. Casado and M. Morillo, "Distribution of escape times in a driven stochastic model," *Phys. Rev. E*, vol. 49, no. 2, pp. 1136–1139, Feb. 1994.
[35] R. Castro and T. Sauer, "Chaotic stochastic resonance: Noise-enhanced reconstruction of attractors," *Phys. Rev. Lett.*, vol. 79, no. 6, pp. 1030–1033, Aug. 1997.
[36] F. Chapeau-Blondeau, "Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel," *Phys. Rev. E*, vol. 55, no. 2, pp. 2016–2019, Feb. 1997.
[37] F. Chapeau-Blondeau and X. Godivier, "Stochastic resonance in nonlinear transmission of Spike signals: An exact model and an application to the neuron," *Int. J. Bifurcation and Chaos*, vol. 6, no. 11, pp. 2069–2076, 1996.
[38] ——, "Theory of stochastic resonance in signal transmission by static nonlinear system," *Phys. Rev. E*, vol. 55, no. 2, pp. 1478–1495, Feb. 1997.
[39] F. Chapeau-Blondeau, X. Godivier, and N. Chambet, "Stochastic resonance in a neuron model that transmits Spike trains," *Phys. Rev. E*, vol. 53, no. 1, pp. 1273–1275, Jan. 1996.
[40] A. K. Chattah, C. B. Briozzo, O. Osenda, and M. O. Cáceres, "Signal-to-noise ratio in stochastic resonance," *Modern Phys. Lett. B*, vol. 10, no. 22, pp. 1085–1094, 1996.
[41] D. R. Chialvo, A. Longtin, and J. Müller-Gerking, "Stochastic resonance in models of neuronal ensembles," *Phys. Rev. E*, vol. 55, no. 2, pp. 1798–1808, Feb. 1997.
[42] 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," *Amer. J. Physical Med. Rehab.*, vol. 76, no. 1, pp. 14–18, Jan./Feb. 1997.
[43] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd edition. Boston: Birkhäuser, 1990.
[44] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, "Aperiodic stochastic resonance," *Phys. Rev. E*, vol. 54, no. 5, pp. 5575–5584, Nov. 1996.
[45] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Aperiodic stochastic resonance in excitable systems," *Phys. Rev. E*, vol. 52, no. 4, pp. R3321–R3324, Oct. 1995.
[46] ——, "Stochastic resonance without tuning," *Nature*, vol. 376, pp. 236–238, July 1995.
[47] J. J. Collins, T. T. Imhoff, and P. Grigg, "Noise-enhanced information transmission in rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance," *J. Neurophysiology*, vol. 76, no. 1, pp. 642–645, July 1996.
[48] ——, "Noise-enhanced tactile sensation," *Nature*, vol. 383, pp. 770, Oct. 1996.
[49] 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, Oct. 1996.
[50] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[51] J. J. Craig, *Introduction to Robotics: Mechanics and Control*, 2nd edition. Reading, MA: Addison-Wesley, 1989.
[52] A. Crisanti, M. Falcioni, G. Paladin, and A. Vulpiani, "Stochastic resonance in deterministic chaotic systems," *J. Physics A: Mathematical and General*, vol. 27, pp. L597–L603, 1994.
[53] G. Dahlquist and Å. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[54] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[55] G. Debnath, F. Moss, T. Leiber, H. Risken, and F. Maresoni, "Holes in the two-dimensional probability density of bistable systems driven by strongly colored noise," *Phys. Rev. A*, vol. 42, no. 2, pp. 703–710, July 1990.
[56] G. Debnath, T. Zhou, and F. Moss, "Remarks on stochastic resonance," *Phys. Rev. A*, vol. 39, no. 8, pp. 4323–4326, Apr. 1989.
[57] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
[58] 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, Sept. 1993.
[59] ——, "Stochastic resonance in Crayfish hydrodynamic receptors stimulated with external noise," in *AIP Conf. Proc. 285: Noise in Physical Systems and 1/f Fluctuations*, P. H. Hanel and A. L. Chung, Eds., 1993, pp. 712–715.
[60] R. Durrett, *Stochastic Calculus: A Practical Introduction*. Boca Raton, FL: CRC, 1996.
[61] M. I. Dykman, G. P. Golubev, I. K. Kaufman, D. G. Luchinsky, P. V. E. McClintock, and E. A. Zhukov, "Noise-enhanced optical heterodyning in an all-optical bistable system," *Appl. Phys. Lett.*, vol. 67, no. 3, pp. 308–310, July 1995.
[62] M. I. Dykman, T. Horita, and J. Ross, "Statistical distribution and stochastic resonance in a periodically driven chemical system," *J. Chemical Physics*, vol. 103, no. 3, pp. 966–972, July 1995.
[63] 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," *Phys. Rev. E*, vol. 54, no. 3, pp. 2366–2377, Sept. 1996.
[64] M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Nonconventional

stochastic resonance," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 479–499, Jan. 1993.
[65] ——, "Stochastic resonance: Linear response and giant nonlinearity," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 463–478, Jan. 1993.
[66] ——, "Supernarrow spectral peaks and high-frequency stochastic resonance in systems with coexisting periodic attractors," *Phys. Rev. E*, vol. 49, no. 2, pp. 1198–1215, Feb. 1994.
[67] M. I. Dykman, D. G. Luchinsky, P. V. E. McClintock, N. D. Stein, and N. G. Stocks, "Stochastic resonance for periodically modulated noise intensity," *Phys. Rev. A*, vol. 46, no. 4, pp. R1713–R1716, Aug. 1992.
[68] J.-P. Eckmann and L. E. Thomas, "Remarks on stochastic resonances," *J. Physics A: Mathematical and General*, vol. 15, pp. L261–L266, 1982.
[69] S. Fauve and F. Heslot, "Stochastic resonance in a bistable system," *Phys.s Lett. A*, vol. 97, nos. 1/2, pp. 5–7, Aug. 1983.
[70] W. Feller, *An Introduction to Probability Thoery and Its Applications*. New York: Wiley, vol. II, 1966.
[71] P. G. Flikkema, "Spread-spectrum techniques for wireless communication," *Signal Processing Mag.*, vol. 14, no. 3, pp. 26–36, 1997.
[72] A. Förster, M. Merget, and F. W. Schneider, "Stochastic resonance in chemistry 2: The peroxidase-oxidase reaction," *J. Physical Chemistry*, vol. 100, pp. 4442–4447, 1996.
[73] 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, Dec. 1997.
[74] J. Foss, F. Moss, and J. Milton, "Noise, multistability, and delayed recurrent loops," *Phys. Rev. E*, vol. 55, no. 4, pp. 4536–4543, Apr. 1997.
[75] R. F. Fox, "Stochastic resonance in a double well," *Phys. Rev. A*, vol. 39, no. 8, pp. 4148–4153, Apr. 1989.
[76] R. F. Fox and Y. N. Lu, "Analytic and numerical study of stochastic resonance," *Phys. Rev. E*, vol. 48, no. 5, pp. 3390–3398, Nov. 1993.
[77] W. J. Freeman, H.-J. Chang, B. C. Burke, P. A. Rose, and J. Badler, "Taming chaos: Stabilization of aperiodic attractors by noise," *IEEE Trans. Circuits Systems I*, vol. 44, pp. 989–996, Oct. 1997.
[78] A. Fuliński, "Relaxation, noise-induced transitions, and stochastic resonance driven by Non-Markovian dichotomic noise," *Phys. Rev. E*, vol. 52, no. 4, pp. 4523–4526, Oct. 1995.
[79] L. Gammaitoni, "Stochastic resonance in multi-threshold systems," *Phys.s Lett. A*, vol. 208, pp. 315–322, Dec. 1995.
[80] ——, "Stochastic resonance," *Rev. Modern Physics*, vol. 70, no. 1, pp. 223–287, Jan. 1998.
[81] L. Gammaitoni, F. Maresoni, M. Martinelli, L. Pardi, and S. Santucci, "Phase shifts in bistable EPR systems at stochastic resonance," *Phys.s Lett. A*, vol. 158, no. 9, pp. 449–452, Sept. 1991.
[82] L. Gammaitoni, F. Maresoni, E. Menichella-Saetta, and S. Santucci, "Stochastic resonance in bistable systems," *Phys. Rev. Lett.*, vol. 62, no. 4, pp. 349–352, Jan. 1989.
[83] L. Gammaitoni, F. Maresoni, E. Menichella-Saetta, and S. Santucci, "Multiplicative stochastic resonance," *Phys. Rev. E*, vol. 49, no. 6, pp. 4878–4881, June 1994.
[84] L. Gammaitoni, M. Martinelli, L. Pardi, and S. Santucci, "Observation of stochastic resonance in bistable electron-paramagnetic-resonance systems," *Phys. Rev. Lett.*, vol. 67, no. 13, pp. 1799–1802, Sept. 1991.
[85] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, F. Maresoni, and C. Pressilla, "Periodically time-modulated bistable systems: Stochastic resonance," *Phys. Rev. A*, vol. 40, no. 4, pp. 2114–2119, Aug. 1989.
[86] T. C. Gard, *Introduction to Stochastic Differential Equations*. New York: Marcel Dekker, 1988.
[87] F. Gassmann, "Noise-induced chaos-order transitions," *Phys. Rev. E*, vol. 55, no. 3, pp. 2215–2221, Mar. 1997.
[88] Z. Gingl, L. B. Kiss, and F. Moss, "Non-dynamical stochastic resonance: Theory and experiments with white and arbitrarily colored noise," *Europhysics Lett.*, vol. 29, no. 3, pp. 191–196, Jan. 1995.
[89] J. Glanz, "Mastering the nonlinear brain," *Science*, vol. 277, pp. 1758–1760, Sept. 1997.
[90] 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," *Phys. Rev. Lett.*, vol. 77, no. 19, pp. 4098–4101, Nov. 1996.
[91] 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.
[92] 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," *Phys. Rev. A*, vol. 46, no. 6, pp. 3243–3249, Sept. 1992.
[93] M. Grifoni, "Dynamics of the dissipative two-state system under AC modulation of bias and coupling energy," *Phys. Rev. E*, vol. 54, no. 4, pp. R3086–R3089, Oct. 1996.
[94] M. Grifoni and P. Hänggi, "Nonlinear quantum stochastic resonance," *Phys. Rev. E*, vol. 54, no. 2, pp. 1390–1401, Aug. 1996.
[95] M. Grifoni, L. Hartmann, S. Berchtold, and P. Hänggi, "Quantum tunneling and stochastic resonance," *Phys. Rev. E*, vol. 53, no. 6, pp. 5890–5898, June 1996.
[96] M. Grifoni, M. Sassetti, P. Hänggi, and U. Weiss, "Cooperative effects in the nonlinearly driven Spin-Boson system," *Phys. Rev. E*, vol. 52, no. 4, pp. 3596–3607, Oct. 1995.
[97] A. N. Grigorenko and P. I. Nikitin, "Stochastic resonance in a bistable magnetic system," *IEEE Trans. Magn.*, vol. 31, pp. 2491–2493, Sept. 1995.
[98] A. N. Grigorenko, S. I. Nikitin, and G. V. Roschepkin, "Stochastic resonance at higher harmonics in monostable systems," *Phys. Rev. E*, vol. 56, no. 5, pp. R4907–4910, Nov. 1997.
[99] A. Guderian, G. Dechert, K.-P. Zeyer, and F. W. Schneider, "Stochastic resonance in chemistry. 1. The Belousov-Zhabitinsku reaction," *J. Physical Chemistry*, vol. 100, pp. 4437–4441, 1996.
[100] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Mew York: McMillan, 1994.
[101] R. Hecht-Nielsen, *Neurocomputing*. Reading, MA: Addison-Wesley, 1990.
[102] C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich, "Information measures quantifying aperiodic stochastic resonance," *Phys. Rev. E*, vol. 54, no. 3, pp. R2228–R2231, 1996.
[103] 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*, Luglio-Agosto, vol. 17 D, nos. 7–8, pp. 811–817, 1995.
[104] 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," *J. Appl. Physics*, vol. 77, no. 6, pp. 2582–2590, Mar. 1995.
[105] W. Hohmann, J. Müller, and F. W. Schneider, "Stochastic resonance in chemistry. 3. The minimal-bromate reaction," *J. Physical Chemistry*, vol. 100, pp. 5388–4492, 1996.
[106] R. Hollands, *The Virtual Reality Home Brewer's Handbook*. New York: Wiley, 1996.
[107] G. Hu, T. Ditzinger, C. Z. Ning, and H. Haken, "Stochastic resonance without external periodic force," *Phys. Rev. Lett.*, vol. 71, no. 6, pp. 807–810, Aug. 1993.
[108] 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," *Phys. Rev. A*, vol. 46, no. 6, pp. 3250–3254, Sept. 1992.
[109] G. Hu, H. Haken, and C. Z. Ning, "A study of stochastic resonance without adiabatic approximation," *Phys. Lett. A*, vol. 172, nos. 1/2, pp. 21–28, Dec. 1992.
[110] G. Hu, H. Haken, and F. Xie, "Stochastic resonance with sensitive frequency dependence in globally coupled continuous systems," *Phys. Rev. Lett.*, vol. 77, no. 10, pp. 1925–1928, Sept. 1996.
[111] G. Hu, G. Nicolis, and N. Nicolis, "Periodically forced Fokker-Planck equation and stochastic resonance," *Phys. Rev. A*, vol. 42, no. 4, pp. 2030–2041, Aug. 1990.
[112] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[113] B. A. Huberman and R. M. Lukose, "Social dilemmas and internet congestion," *Science*, vol. 277, pp. 535–537, July 1997.
[114] M. E. Inchiosa and A. R. Bulsara, "Coupling enhances stochastic resonance in nonlinear dynamic elements driven by a sinusoidal plus noise," *Phys.s Lett. A*, vol. 200, pp. 283–288, Apr. 1995.
[115] ——, "Nonlinear dynamic elements with noisy sinusoidal forcing: Enhancing response via nonlinear coupling," *Phys. Rev. E*, vol. 52, no. 1, pp. 327–339, July 1995.

[116] ____, "Signal detection statistics of stochastic resonators," *Phys. Rev. E*, vol. 53, no. 3, pp. R2021–R2024, Mar. 1996.
[117] M. E. Inchiosa, A. R. Bulsara, and L. Gammaitoni, "Higher-order resonant behavior in asymmetric nonlinear stochastic systems," *Phys. Rev. E*, vol. 55, no. 4, pp. 4049–4056, Apr. 1997.
[118] E. Ippen, J. Lindner, and W. L. Ditto, "Chaotic resonance: A simulation," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 437–450, Jan. 1993.
[119] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neurofuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Englewoods Cliffs, NJ: Prentice-Hall, 1996.
[120] B. M. Jost and B. E. A. Saleh, "Signal-to-noise ratio improvement by stochastic resonance in a unidirectional photorefractive ring resonator," *Optics Lett.*, vol. 21, no. 4, pp. 287–289, Feb. 1996.
[121] P. Jung, "Thermal activation in bistable systems under external periodic forces," *Zeitschrift für Physik B*, vol. 76, pp. 521–535, 1989.
[122] P. Jung, "Threshold devices: Fractal noise and neural talk," *Phys. Review E*, vol. 50, no. 4, pp. 2513–2522, Oct. 1994.
[123] ____, "Stochastic resonance and optimal design of threshold detectors," *Phys.s Lett. A*, vol. 207, pp. 93–104, Oct. 1995.
[124] P. Jung, U. Behn, E. Pantazelou, and F. Moss, "Collective response in globally coupled bistable systems," *Phys. Rev. A*, vol. 46, no. 4, pp. R1709–R1712, Aug. 1992.
[125] P. Jung and P. Hänggi, "Resonantly driven Brownian motion: Basic concepts and exact results," *Phys. Rev. A*, vol. 41, no. 6, pp. 2977–2988, Mar. 1990.
[126] P. Jung and P. Hänggi, "Amplification of small signals via stochastic resonance," *Phys. Rev. A*, vol. 44, no. 12, pp. 8032–8042, Dec. 1991.
[127] P. Jung and P. Talkner, "Suppression of higher harmonics at noise induced resonances," *Phys. Rev. E*, vol. 51, no. 3, pp. 2640–2643, Mar. 1995.
[128] P. Jung and K. Wiesenfeld, "Too quiet to hear a whisper," *Nature*, vol. 385, p. 291, Jan. 1997.
[129] S. Kádár, J. Wang, and K. Showalter, "Noise-supported travelling waves in sub-excitable media," *Nature*, vol. 391, pp. 770–772, Feb. 1998.
[130] H. M. Kim and B. Kosko, "Neural fuzzy motion estimation and compensation," *IEEE Trans. Signal Processing*, vol. 45, pp. 2515–2532, Oct. 1997.
[131] L. B. Kiss, "Possible breakthrough: Significant improvement of signal to noise ratio by stochastic resonance," in *AIP Conf. Proc. 375: Chaotic, Fractal, and Nonlinear Signal Processing, 1995*, R. A. Katz, Ed., pp. 382–396.
[132] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewoods Cliffs, NJ: Prentice-Hall, 1991.
[133] ____, "Fuzzy systems as universal approximators," *IEEE Trans. Comput.*, vol. 43, pp. 1329–1333, Nov. 1994.
[134] ____, "Combining fuzzy systems," in *Proc. IEEE Int. C. Fuzzy Systems (IEEE FUZZ-95)*, pp. 1855–1863.
[135] ____, "Optimal fuzzy rules cover extrema," *Int. J. Intell. Syst.*, vol. 10, no. 2, pp. 249–255, Feb. 1995.
[136] ____, *Fuzzy Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[137] I. Kotlasi, "On random variables whose quotient follows the Cauchy law," *Colloquium Mathematicum*, vol. VII, pp. 277–284, 1960.
[138] J. Kovačević and I. Daubechies, Eds., *Proc. IEEE*, vol. 84, no. 4, 1996.
[139] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Phys.*, vol. VII, no. 4, pp. 284–304, Apr. 1940.
[140] D. M. Kreps, *A Course in Microeconomic Theory*. Princeton, NJ: Princeton Univ. Press, 1990.
[141] R. G. Laha, "On a class of distribution functions where the quotients follows the Cauchy law," *Trans. Amer. Math. Soc.*, vol. 93, pp. 205–215, Nov. 1959.
[142] W. C. Y. Lee, *Mobile Cellular Telecommunications: Analog and Digital Systems*. New York: McGraw-Hill, 1995.
[143] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic," in *Comput. Commun. Rev.: Proc. SIGCOMM-93*, vol. 23, pp. 183–193.
[144] D. S. Leonard, "Stochastic resonance in a random walk," *Phys. Rev. A*, vol. 46, no. 10, pp. 6742–6744, Nov. 1992.
[145] D. S. Leonard and L. E. Reichl, "Stochastic resonance in a chemical reaction," *Phys. Rev. E*, vol. 49, no. 2, pp. 1734–1737, Feb. 1994.
[146] 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, Mar. 1996.
[147] P. H. Lewis and C. Yang, *Basic Control Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[148] 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," *Phys. Rev. E*, vol. 51, no. 5, pp. 3964–3967, May 1995.
[149] 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," *Phys. Rev. E*, vol. 53, no. 3, pp. 2081–2086, Mar. 1996.
[150] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: A theoretical survey," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 355–374, May 1992.
[151] M. Löcher, G. A. Johnson, and E. R. Hunt, "Spatiotemporal stochastic resonance in a system of coupled diode resonators," *Phys. Rev. Lett.*, vol. 77, no. 23, pp. 4698–4701, Dec. 1996.
[152] K. Loerincz, Z. Gingl, and L. B. Kiss, "A stochastic resonator is able to greatly improve signal-to-noise ratio," *Phys. Lett. A*, vol. 224, pp. 63–67, Dec. 1996.
[153] R. Löfstedt and S. N. Coppersmith, "Quantum stochastic resonance," *Phys. Rev. Lett.*, vol. 72, no. 13, pp. 1947–1950, Mar. 1994.
[154] A. Longtin, "Stochastic resonance in neuron models," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 309–327, Jan. 1993.
[155] ____, "Synchronization of the stochastic FitzHugh–Nagumo equations to periodic forcing," *Il Nuovo Cimento*, Luglio-Agosto, vol. 17 D, nos. 7–8, pp. 835–846, 1995.
[156] ____, "Autonomous stochastic resonance in bursting neurons," *Phys. Rev. E*, vol. 55, no. 1, pp. 868–876, Jan. 1997.
[157] A. Longtin, A. R. Bulsara, and F. Moss, "Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons," *Phys. Rev. Lett.*, vol. 67, no. 5, pp. 656–659, July 1991.
[158] A. Longtin, A. R. Bulsara, D. Pierson, and F. Moss, "Bistability and the dynamics of periodically forced sensory neurons," *Biological Cybern.*, vol. 70, pp. 569–578, 1994.
[159] A. Longtin and K. Hinzer, "Encoding with bursting, subthreshold oscillations, and noise in mammalian cold receptors," *Neural Computation*, vol. 8, pp. 215–255, 1996.
[160] V. Loreto, G. Paladin, and A. Vulpiani, "Concept of complexity in random dynamical systems," *Phys. Rev. E*, vol. 53, no. 3, pp. 2087–2098, Mar. 1996.
[161] J. Maddox, "Toward the brain-computer's code?," *Nature*, vol. 352, p. 469, Aug. 1991.
[162] ____, "Bringing more order out of noisiness," *Nature*, vol. 369, p. 271, May 1994.
[163] M. C. Mahato and S. R. Shenoy, "Hysteresis loss and stochastic resonance: A numerical study of a double-well potential," *Phys. Rev. E*, vol. 50, no. 4, pp. 2503–2512, Oct. 1994.
[164] D. E. Makarov and N. Makri, "Stochastic resonance and nonlinear response in double-quantum-well structures," *Phys. Rev. B*, vol. 52, no. 4, pp. R2257–R2260, July 1995.
[165] R. N. Mantegna and B. Spagnolo, "Stochastic resonance in a tunnel diode," *Phys. Rev. E*, vol. 49, no. 3, pp. R1792–R1795, Mar. 1994.
[166] R. N. Mantegna and B. Spagnolo, "Stochastic resonance in a tunnel diode in the presence of white or colored noise," *Il Nuovo Cimento*, Luglio-Agosto, vol. 17 D, nos. 7–8, pp. 873–881, 1995.
[167] F. Maresoni, L. Gammaitoni, and A. R. Bulsara, "Spatiotemporal stochastic resonance in a ϕ^4 model of Kink-Antikink nucleation," *Phys. Rev. Lett.*, vol. 76, no. 15, pp. 2609–2612, Apr. 1996.
[168] K. Matsumoto and I. Tsuda, "Noise-induced order," *J. Statist. Physics*, vol. 31, no. 1, pp. 87–106, 1983.
[169] B. McNamara and K. Wiesenfeld, "Theory of stochastic resonance," *Phys. Rev. A*, vol. 39, no. 9, pp. 4854–4869, May 1989.
[170] B. McNamara, K. Wiesenfeld, and R. Roy, "Observation of stochastic resonance in a ring laser," *Phys. Rev. Lett.*, vol. 60, no. 25, pp. 2626–2629, June 1988.
[171] V. I. Melnikov, "Schmitt trigger: A solvable model of stochastic

resonance," *Phys. Rev. E*, vol. 48, no. 4, pp. 2481–2489, Oct. 1993.
[172] 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," *J. Neurophysiology*, vol. 66, no. 5, pp. 1680–1689, Nov. 1991.
[173] ——, "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," *J. Neurophysiology*, vol. 66, no. 5, pp. 1690–1703, Nov. 1991.
[174] S. Mitaim and B. Kosko, "What is the best shape for a fuzzy set in function approximation?," in *Proc. 5th IEEE Int. Conf. Fuzzy Systems (FUZZ-96)*, vol. 2, pp. 1237–1243.
[175] ——, "Adaptive joint fuzzy sets for function approximation," in *Proc. 1997 IEEE Int. Conf. Neural Networks (ICNN-97)*, vol. 1, pp. 537–542.
[176] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing unit," *Neural Computation*, no. 2, pp. 281–294, 1989.
[177] M. Morillo, J. Gómez-Ordo nez, and J. M. Casado, "Stochastic resonance in a mean-field model of cooperative behavior," *Phys. Rev. E*, vol. 52, no. 1, pp. 316–320, July 1995.
[178] 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, Aug. 1996.
[179] F. Moss, "Stochastic resonance: From the ice ages to the monkey's ear," in *Contemporary Problems in Statistical Physics*, G. H. Weiss, Ed. Philadelphia, PA: SIAM, 1994, ch. 5, pp. 205–253
[180] ——, "Noisy waves," *Nature*, vol. 391, pp. 743–744, Feb. 1998.
[181] F. Moss, A. Bulsara, and M. Shlesinger, Eds.,*J. Statist. Physics, Special Issue on Stochastic Resonance in Physics and Biology (Proc. NATO Advanced Research Workshop)*, Jan. 1993, vol. 70, nos. 1/2.
[182] F. Moss, F. Chiou-Tan, and R. Klinke, "Will there be noise in their ears?," *Nature Medicine*, vol. 2, no. 8, pp. 860–862, Aug. 1996.
[183] F. Moss, J. K. Douglass, L. Wilkens, D. Pierson, and E. Pantazelou, "Stochastic resonance in an electronic FitzHugh–Nagumo model," in *Ann. New York Academy of Sciences Volume 706: Stochastic Processes in Astrophysics*, J. R. Buchler and H. E. Kandrup, Eds., pp. 26–41, 1993.
[184] F. Moss and P. V. E. McClintock, Eds., *Noise in Nonlinear Dynamical Systems*, vol. I–III. Cambridge, UK: Cambridge University Press, 1989.
[185] F. Moss, D. Pierson, and D. O'Gorman, "Stochastic resonance: Tutorial and update," *Int. J. Bifurcation and Chaos*, vol. 4, no. 6, pp. 1383–1397, 1994.
[186] F. Moss and K. Wiesenfeld, "The benefits of background noise," *Scientific Amer.*, vol. 273, no. 2, pp. 66–69, Aug. 1995.
[187] D. C. Munson, "A note on lena," *IEEE Trans. Image Processing*, vol. 5, p. 3, Jan. 1996.
[188] Z. Néda, "Stochastic resonance in Ising systems," *Phys. Rev. E*, vol. 51, no. 6, pp. 5315–5317, June 1995.
[189] A. Neiman and L. Schimansky-Geier, "Stochastic resonance in two coupled bistable systems," *Phys. Lett. A*, vol. 197, pp. 379–386, Feb. 1995.
[190] A. Neiman, L. Schimansky-Geier, and F. Moss, "Linear response theory applied to stochastic resonance in models of ensembles of oscillators," *Phys. Rev. E*, vol. 56, no. 1, pp. R9–R12, July 1997.
[191] A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, "Dynamic entropies applied to stochastic resonance," *Phys. Rev. Lett.*, vol. 76, no. 23, pp. 4299–4302, June 1996.
[192] A. Neiman and W. Sung, "Memory effects on stochastic resonance," *Phys.s Lett. A*, vol. 223, pp. 341–347, Dec. 1996.
[193] C. Nicolis, "Stochastic aspects of climatic transitions—Response to a periodic forcing," *Tellus*, vol. 34, pp. 1–9, 1982.
[194] ——, "Long-term climatic transitions and stochastic resonance," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 4–14, Jan. 1993.
[195] C. Nicolis and G. Nicolis, "Stochastic aspects of climate transitions—Additive fluctuations," *Tellus*, vol. 33, pp. 225–234, 1981.
[196] G. Nicolis, C. Nicolis, and D. McKernan, "Stochastic resonance in chaotic dynamics," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 125–139, Jan. 1993.
[197] 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.
[198] K. Ogata, *Modern Control Engineering*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[199] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.
[200] P. J. Pacini and B. Kosko, "Adaptive fuzzy frequency hopper," *IEEE Trans. Commun.*, vol. 43, no. 6, pp. 2111–2117, June 1995.
[201] A. Restrepo (Palacios), L. F. Zuluaga, and L. E. Pino, "Optimal noise levels for stochastic resonance," in *Proc. 1997 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-97)*, vol. III, pp. 2365–2368.
[202] 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," *Int. J. Bifurcation and Chaos*, vol. 5, no. 1, pp. 101–108, 1995.
[203] H. C. Papadopoulos and G. W. Wornell, "A class of stochastic resonance systems for signal processing applications," in *Proc. 1996 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-96)*, pp. 1617–1620.
[204] A. Papoulis, *Probability and Statistics*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[205] T. P. Pareek, M. C. Mahato, and A. M. Jayannavar, "Stochastic resonance and nonlinear response in a dissipative quantum two-state system," *Phys. Rev. B*, vol. 55, no. 15, pp. 9318–9321, Apr. 1997.
[206] B. R. Parnas, "Noise and neuronal populations conspire to encode simple waveforms reliably," *IEEE Trans. BioMed. Eng.*, vol. 43, pp. 313–318, Mar. 1996.
[207] X. Pei, K. Bachmann, and F. Moss, "The detection threshold, noise and stochastic resonance in the Fitzhugh-Nagumo neuron model," *Phys. Lett. A*, vol. 206, pp. 61–65, Oct. 1995.
[208] X. Pei, L. Wilkens, and F. Moss, "Light enhances hydrodynamic signaling in the multimodal caudal photoreceptor interneurons of the Crayfish," *J. Neurophysiology*, vol. 76, no. 5, pp. 3002–3011, Nov. 1996.
[209] X. Pei, L. Wilkens, and F. Moss, "Noise-mediated Spike timing precision from aperiodic stimuli in an array of Hodgkin-Huxley-type neurons," *Phys. Rev. Lett.*, vol. 77, no. 22, pp. 4679–4682, Nov. 1996.
[210] 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 Conf. Proc. 285: Noise in Physical Systems and 1/f Fluctuations*, P. H. Hanel and A. L. Chung, Eds., 1993, pp. 731–734.
[211] H. E. Plesser and S. Tanaka, "Stochastic resonance in a model neuron with reset," *Phys.s Lett. A*, vol. 225, pp. 228–234, Feb. 1997.
[212] W.-J. Rappel and A. Karma, "Noise-induced coherence in neural networks," *Phys. Rev. Lett.*, vol. 77, no. 15, pp. 3256–3259, Oct. 1996.
[213] W.-J. Rappel and S. H. Strogatz, "Stochastic resonance in an autonomous system with a nonuniform limit cycle," *Phys. Rev. E*, vol. 50, no. 4, pp. 3249–3250, Oct. 1994.
[214] D. O. Reale, A. K. Pattanayak, and W. C. Schieve, "Semiquantal corrections to stochastic resonance," *Phys. Rev. E*, vol. 51, no. 4, pp. 2925–2932, Apr. 1995.
[215] M. Riani and E. Simonotto, "Stochastic resonance in the perceptual interpretation of ambiguous figures: A neural network model," *Phys. Rev. Lett.*, vol. 72, no. 19, pp. 3120–3123, May 1994.
[216] M. Riani and E. Simonotto, "Periodic perturbation of ambiguous figure: A neural network model and a nonsimulated experiment," *Il Nuovo Cimento*, Luglio-Agosto, vol. 17 D, nos. 7–8, pp. 903–913, 1995.
[217] T. F. Ricci and C. Scherer, "Linear response and stochastic resonance of superparamagnets," *J. Statist. Physics*, vol. 86, no. 3/4, pp. 803–819, Feb. 1997.
[218] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Application*. New York: Springer-Verlag, 1984.
[219] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[220] R. Rouse, S. Han, and J. E. Lukens, "Flux amplification using stochastic superconducting quantum interference devices," *Appl. Physics Letters*, vol. 66, no. 1, pp. 108–110, Jan. 1995.
[221] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[222] D. W. E. Schobben, R. A. Beuker, and W. Oomen, "Dither and data compression," *IEEE Trans. Signal Processing*, vol. 45, pp. 2097–2101, Aug. 1997.
[223] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, pp. 984–1010, July 1993.
[224] ____, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.
[225] V. A. Shneidman, P. Jung, and P. Hänggi, "Power spectrum of a driven bistable system," *Europhysics Lett.*, vol. 26, no. 8, pp. 571–576, June 1994.
[226] 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," *J. Appl. Physics*, vol. 81, no. 8, pp. 5597–5599, Apr. 1997.
[227] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*. New York: McGraw-Hill, 1994.
[228] E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss, "Visual perception of stochastic resonance," *Phys. Rev. Lett.*, vol. 78, no. 6, pp. 1186–1189, Feb. 1997.
[229] C. Smith and C. Gervelis, *Cellular System Design and Optimization*. New York: McGraw-Hill, 1996.
[230] M. L. Spano, M. Wun-Fogle, and W. L. Ditto, "Experimental observation of stochastic resonance in a magnetoelastic ribbon," *Phys. Rev. A*, vol. 46, no. 8, pp. 5253–5256, Oct. 1992.
[231] 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.
[232] N. G. Stocks, N. D. Stein, and P. V. E. McClintock, "Stochastic resonance in monostable systems," *J. Physics A: Math. General*, vol. 26, pp. L385–L390, 1993.
[233] N. G. Stocks, N. D. Stein, S. M. Soskin, and P. V. E. McClintock, "Zero-dispersion stochastic resonance," *J. Physics A: Math. General*, vol. 25, pp. L1119–L1125, 1992.
[234] G. Strang and T. Q. Nguyen, *Wavelets and Filter Banks*. Cambridge, UK: Wellesley-Cambridge Press, 1996.
[235] M. Thorwart and P. Jung, "Dynamical hysteresis in bistable quantum systems," *Phys. Rev. Lett.*, vol. 78, no. 13, pp. 2503–2506, Mar. 1997.
[236] G. Vemuri and R. Roy, "Stochastic resonance in a bistable ring laser," *Phys. Rev. A*, vol. 39, no. 9, pp. 4668–4674, May 1989.
[237] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.
[238] J. M. G. Vilar and J. M. Rubí, "Divergent signal-to-noise ratio and stochastic resonance in monostable systems," *Phys. Rev. Lett.*, vol. 77, no. 14, pp. 2863–2866, Sept. 1996.
[239] ____, "Spatiotemporal stochastic resonance in the Swift-Hohenberg equation," *Phys. Rev. Lett.*, vol. 78, no. 15, pp. 2886–2889, Apr. 1997.
[240] ____, "Stochastic multiresonance," *Phys. Rev. Lett.*, vol. 78, no. 15, pp. 2882–2885, Apr. 1997.
[241] S. T. Vohra and F. Bucholtz, "Observation of stochastic resonance near a subcritical bifurcation," *J. Statist. Physics*, vol. 70, nos. 1/2, pp. 413–421, Jan. 1993.
[242] S. T. Vohra and L. Fabiny, "Induced stochastic resonance near a subcritical bifurcation," *Phys. Rev. E*, vol. 50, no. 4, pp. R2391–2394, Oct. 1994.
[243] K. Wiesenfeld and F. Moss, "Stochastic resonance and the benefits of noise: From ice ages to Crayfish and SQUID's," *Nature*, vol. 373, pp. 33–36, Jan. 1995.
[244] K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss, "Stochastic resonance on a circle," *Phys. Rev. Lett.*, vol. 72, pp. 2125–2129, Apr. 1994.
[245] 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, Oct. 1992.
[246] H. S. Wio, "Stochastic resonance in a spatially extended system," *Phys. Rev. E*, vol. 54, no. 4, pp. R3075–3078, Oct. 1996.
[247] A. Witt, A. Neiman, and J. Kurths, "Characterizing the dynamics of stochastic bistable systems by measures of complexity," *Phys. Rev. E*, vol. 55, no. 5, pp. 5050–5059, May 1997.
[248] W. Yang, M. Ding, and G. Hu, "Trajectory (phase) selection in multistable systems: Stochastic resonance, signal bias, and the effect of signal phase," *Phys. Rev. Lett.*, vol. 74, no. 20, pp. 3955–3958, May 1995.
[249] T. Zhou and F. Moss, "Analog simulations of stochastic resonance," *Phys. Rev. A*, vol. 41, no. 8, pp. 4255–4264, Apr. 1990.
[250] T. Zhou, F. Moss, and P. Jung, "Escape-time distributions of a periodically modulated bistable system with noise," *Phys. Rev. A*, vol. 42, no. 6, pp. 3161–3169, Sept. 1990.
[251] S. Zozor and P.-O. Amblard, "Stochastic resonance in a discrete time nonlinear SETAR(1,2,0,0) model," in *Proc. 1997 IEEE Workshop Higher-Order-Statistics*, July 1997.
[252] U. Zürcher and C. R. Doering, "Thermally activated escape over fluctuating barriers," *Phys. Rev. E*, vol. 47, no. 6, pp. 3862–3869, June 1993.

Sanya Mitaim was born in Bangkok, Thailand, in 1970. He received the B.Eng. degree in control engineering from King Mongkut's Institute of Technology, Ladkrabang, Thailand in 1990 and the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1992. He is currently a Ph.D. candidate in electrical engineering at the University of Southern California.

He was a Lecturer at Thammasat University, Thailand, from 1990 to 1991. His research interests include neural and fuzzy techniques in nonlinear signal processing, software agents, and noise processing.

Bart Kosko (Member, IEEE), for a photograph and biography, see this issue, p. 2120.