

Adaptive Simulated Annealing (ASA) and Path-Integral (PATHINT) Algorithms: Generic Tools for Complex Systems

Lester Ingber

ingber@ingber.com • ingber@alumni.caltech.edu
<http://www.ingber.com/>

ADAPTIVE SIMULATED ANNEALING (ASA)

MATHEMATICAL PHYSICS

NUMERICAL PATH INTEGRATION (PATHINT)

GENERIC MESOSCOPIC NEURAL NETWORKS (MNN)

SOME APPLICATIONS

COVER PAGE	1
CONTENTS-1	2
CONTENTS-2	3
ADAPTIVE SIMULATED ANNEALING (ASA)	4
Applications	5
Hills and Valleys	6
Outline of ASA Algorithm	7
Index of Pre-Compile Tuning Parameters	8
Index of Adaptive Tuning Parameters	9
Reannealing Example	10
MATHEMATICAL PHYSICS	11
Nonlinear Nonequilibrium Multivariate Stochastic Aggregation	12
Stochastic Differential Equation (SDE)	13
Partial Differential Equation (PDE)	14
Lagrangian Probability Distribution Function (PDF)	15
Path-Integral Riemannian Geometry	16
Fitting Variance With ASA	17
Information	18
Transformations Without Itô Calculus	19
Intuitive Variables	20
Euler-Lagrange Variational Equations	21
Canonical Momenta Indicators (CMI)	22
NUMERICAL PATH INTEGRATION (PATHINT)	23
Applications	24
Outline of PATHINT Algorithm	25
Boundary Condition Sensitivity	26
Mesh Limitations	27
Probability Tree (PATHTREE)	28
Standard Binomial Tree	29
Deficiency of Standard Algorithm to Order \sqrt{dt}	30
Problems Generalizing The Standard Tree	31
Probability PATHTREE	32
Construction of PATHTREE	33
Direct Calculation of Probability	34
Numerical Derivatives of Expectation of Probability	35
Alternative First Derivative Calculation of Probability	36
PATHTREE vs PATHINT	37
Black-Scholes (BS) and Ornstein-Uhlenbeck (OU) Exam- ples	38

GENERIC MESOSCOPIC NEURAL NETWORKS (MNN)	39
Applications	40
MNN Learning	41
MNN Prediction	42
MNN Parallel Processing	43
SOME APPLICATIONS	44
Statistical Mechanics of Financial Markets (SMFM)	45
2-Factor Interest-Rate Model	46
Lagrangian Representation	47
ASA Fits	48
Bond PDE/PATHINT	49
Power-Law Model	50
Multi-Factor Volatility Model	51
PATHINT Options	52
PATHINT Two-Factor	53
x Market Indicators	54
S&P Interday Futures-Cash	55
Tick Resolution CMI Trading	56
Statistical Mechanics of Combat (SMC)	57
National Training Center — Janus(T)	58
Basic Equations of Aggregated Units	59
Long-Time Correlations Test Short-Time Models	60
CMI	61
Statistical Mechanics of Neocortical Interactions (SMNI)	62
Multiple Scales	63
Microscopic Aggregation	64
Short-Term Memory (STM)	65
Electroencephalography (EEG)	66
ASA Fits of CMI	67
Chaos in SMNI	68
Duffing Analog	69
Noise Washes Out Chaos/PATHINT	70

ADAPTIVE SIMULATED ANNEALING (ASA)

Applications

This algorithm fits empirical data to a theoretical cost function over a D -dimensional parameter space, adapting for varying sensitivities of parameters during the fit.

For several test problems, ASA has been shown to be orders of magnitude more efficient than other similar techniques. ASA has been applied to several complex systems, including specific problems in neuroscience, finance and combat systems.

Heuristic arguments have been developed to demonstrate that this algorithm is faster than the fast Cauchy annealing, $T_i = T_0/k$, and much faster than Boltzmann annealing, $T_i = T_0/\ln k$.

Hills and Valleys

It helps to visualize the problems presented by such complex systems as a geographical terrain. For example, consider a mountain range, with two “parameters,” e.g., along the North–South and East–West directions. We wish to find the lowest valley in this terrain. SA approaches this problem similar to using a bouncing ball that can bounce over mountains from valley to valley. We start at a high “temperature,” where the temperature is an SA parameter that mimics the effect of a fast moving particle in a hot object like a hot molten metal, thereby permitting the ball to make very high bounces and being able to bounce over any mountain to access any valley, given enough bounces. As the temperature is made relatively colder, the ball cannot bounce so high, and it also can settle to become trapped in relatively smaller ranges of valleys.

We imagine that our mountain range is aptly described by a “cost function.” We define probability distributions of the two directional parameters, called generating distributions since they generate possible valleys or states we are to explore. We define another distribution, called the acceptance distribution, which depends on the difference of cost functions of the present generated valley we are to explore and the last saved lowest valley. The acceptance distribution decides probabilistically whether to stay in a new lower valley or to bounce out of it. All the generating and acceptance distributions depend on temperatures.

Outline of ASA Algorithm

For parameters

$$\alpha_k^i \in [A_i, B_i],$$

sampling with the random variable x^i ,

$$x^i \in [-1, 1],$$

$$\alpha_{k+1}^i = \alpha_k^i + x^i(B_i - A_i),$$

define the generating function

$$g_T(x) = \prod_{i=1}^D \frac{1}{2 \ln(1 + 1/T_i)(|x^i| + T_i)} \equiv \prod_{i=1}^D g_T^i(x^i),$$

in terms of parameter ‘temperatures’

$$T_i = T_{i0} \exp(-c_i k^{1/D}).$$

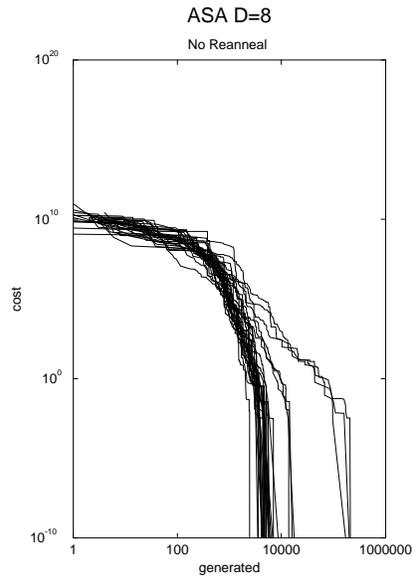
Index of Pre-Compile Tuning Parameters

OPTIONS_FILE=TRUE	ASA_RESOLUTION=FALSE
OPTIONS_FILE_DATA=TRUE	FITLOC=FALSE
RECUR_OPTIONS_FILE=FALSE	FITLOC_ROUND=TRUE
RECUR_OPTIONS_FILE_DATA=FALSE	FITLOC_PRINT=TRUE
COST_FILE=TRUE	MULTI_MIN=FALSE
ASA_LIB=FALSE	MULTI_NUMBER=2
HAVE_ANSI=TRUE	ASA_PARALLEL=FALSE
IO_PROTOTYPES=FALSE	FDLIBM_POW=FALSE
TIME_CALC=FALSE	FDLIBM_LOG=FALSE
TIME_STD=FALSE	FDLIBM_EXP=FALSE
TIME_GETRUSAGE=TRUE	USER_OUT=
INT_LONG=TRUE	ASA_PRINT=TRUE
INT_ALLOC=FALSE	ASA_OUT=
SMALL_FLOAT=1.0E-18	USER_ASA_OUT=FALSE
MIN_DOUBLE=SMALL_FLOAT	ASA_PRINT_INTERMED=TRUE
MAX_DOUBLE=1.0/SMALL_FLOAT	ASA_PRINT_MORE=FALSE
EPS_DOUBLE=SMALL_FLOAT	G_FIELD=12 & G_PRECISION=7
CHECK_EXPONENT=FALSE	ASA_SAVE=FALSE
NO_PARAM_TEMP_TEST=FALSE	ASA_SAVE_OPT=FALSE
NO_COST_TEMP_TEST=FALSE	ASA_SAVE_BACKUP=FALSE
SELF_OPTIMIZE=FALSE	ASA_PIPE=FALSE
ASA_TEST=FALSE	ASA_PIPE_FILE=FALSE
ASA_TEST_POINT=FALSE	SYSTEM_CALL=TRUE
MY_TEMPLATE=TRUE	
USER_INITIAL_COST_TEMP=FALSE	
RATIO_TEMPERATURE_SCALES=FALSE	
USER_INITIAL_PARAMETERS_TEMPS=FALSE	
DELTA_PARAMETERS=FALSE	
QUENCH_PARAMETERS=FALSE	
QUENCH_COST=FALSE	
QUENCH_PARAMETERS_SCALE=TRUE	
QUENCH_COST_SCALE=TRUE	
ASA_TEMPLATE=FALSE	
OPTIONAL_DATA_DBL=FALSE	
OPTIONAL_DATA_INT=FALSE	
OPTIONAL_DATA_PTR=FALSE	
OPTIONAL_PTR_TYPE=USER_TYPE	
USER_COST_SCHEDULE=FALSE	
USER_ACCEPT_ASYMP_EXP=FALSE	
USER_ACCEPTANCE_TEST=FALSE	
USER_GENERATING_FUNCTION=FALSE	
USER_REANNEAL_COST=FALSE	
USER_REANNEAL_PARAMETERS=FALSE	
USER_COST_FUNCTION=cost_function	
RECUR_USER_COST_FUNCTION=recur_cost_function	
MAXIMUM_REANNEAL_INDEX=50000	
REANNEAL_SCALE=10.0	
ASA_SAMPLE=FALSE	
ASA_QUEUE=FALSE	

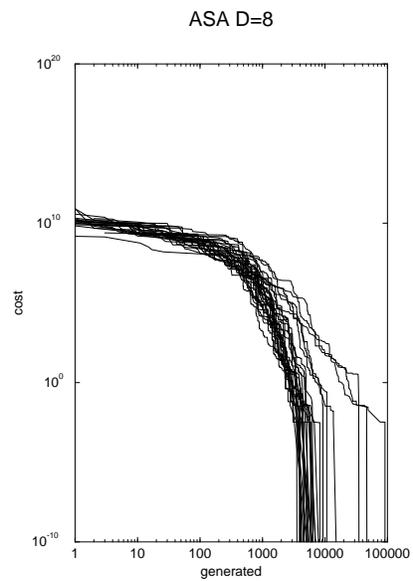
Index of Adaptive Tuning Parameters

OPTIONS->Limit_Acceptances	OPTIONS->Cost_Temp_Scale
OPTIONS->Limit_Generated	OPTIONS->Prob_Bias
OPTIONS->Limit_Invalid_Generated_States	OPTIONS->Random_Seed
OPTIONS->Accepted_To_Generated_Ratio	OPTIONS->Generating_Distrib
OPTIONS->Cost_Precision	OPTIONS->Reanneal_Cost_Function
OPTIONS->Maximum_Cost_Repeat	OPTIONS->Reanneal_Params_Function
OPTIONS->Number_Cost_Samples	OPTIONS->Bias_Acceptance
OPTIONS->Temperature_Ratio_Scale	OPTIONS->Bias_Generated
OPTIONS->Cost_Parameter_Scale_Ratio	OPTIONS->Average_Weights
OPTIONS->Temperature_Anneal_Scale	OPTIONS->Limit_Weights
OPTIONS->User_Cost_Temperature	OPTIONS->Queue_Size
OPTIONS->Include_Integer_Parameters	OPTIONS->Queue_Resolution
OPTIONS->User_Initial_Parameters	OPTIONS->Coarse_Resolution
OPTIONS->Sequential_Parameters	OPTIONS->Fit_Local
OPTIONS->Initial_Parameter_Temperature	OPTIONS->Iter_Max
OPTIONS->User_Temperature_Ratio	OPTIONS->Penalty
OPTIONS->User_Parameter_Temperature	OPTIONS->Multi_Cost
OPTIONS->Acceptance_Frequency_Modulus	OPTIONS->Multi_Params
OPTIONS->Generated_Frequency_Modulus	OPTIONS->Multi_Grid
OPTIONS->Reanneal_Cost	OPTIONS->Multi_Specify
OPTIONS->Reanneal_Parameters	OPTIONS->Gener_Mov_Avr
OPTIONS->Delta_X	OPTIONS->Gener_Block
OPTIONS->User_Delta_Parameter	OPTIONS->Gener_Block_Max
OPTIONS->User_Tangents	OPTIONS->Random_Array_Dim
OPTIONS->Curvature_0	OPTIONS->Random_Array
OPTIONS->User_Quench_Param_Scale	OPTIONS->Asa_Recursive_Level
OPTIONS->User_Quench_Cost_Scale	
OPTIONS->N_Accepted	
OPTIONS->N_Generated	
OPTIONS->Locate_Cost	
OPTIONS->Immediate_Exit	
OPTIONS->Best_Cost	
OPTIONS->Best_Parameters	
OPTIONS->Last_Cost	
OPTIONS->Last_Parameters	
OPTIONS->Asa_Data_Dim_Dbl	
OPTIONS->Asa_Data_Dbl	
OPTIONS->Asa_Data_Dim_Int	
OPTIONS->Asa_Data_Int	
OPTIONS->Asa_Data_Dim_Ptr	
OPTIONS->Asa_Data_Ptr	
OPTIONS->Asa_Out_File	
OPTIONS->Cost_Schedule	
OPTIONS->Asymp_Exp_Param	
OPTIONS->Acceptance_Test	
OPTIONS->User_Acceptance_Flag	
OPTIONS->Cost_Acceptance_Flag	
OPTIONS->Cost_Temp_Curr	
OPTIONS->Cost_Temp_Init	

Reannealing Example



Superimposed are runs for $n = 8$, the case of no reannealing, 3 trajectories each for cases of $Q = 1, 2, 3, 4, 5, 6, 7, 8, 16,$ and 24 . Although the actual final cost function values are 0, they were set to 10^{-10} for purposes of these log-log plots.



Superimposed are runs for $n = 8$, the case including reannealing, 3 trajectories each for cases of $Q = 1, 2, 3, 4, 5, 6, 7, 8, 16,$ and 24 .

MATHEMATICAL PHYSICS

Nonlinear Nonequilibrium Multivariate Stochastic Aggregation

Complex systems typically are in nonequilibrium, being driven by nonlinear and stochastic interactions described by many external and internal degrees of freedom. For these systems, classical thermodynamic descriptions typically do not apply. Many such systems are best treated by respecting some intermediate mesoscale as fundamental to drive larger macroscopic processes.

Often these mesoscopic scales are aptly described by Gaussian Markovian statistics. They naturally develop in physical and biological scales to maximally process information from microscopic scales up to macroscopic scales. Possibly this is true as well of some social systems such as financial markets.

For many physical systems this mesoscopic scale still has some audit trail back to its microscopic origins. Often, statistical deviations of drift variables lead to functional dependencies in diffusion variables.

Stochastic Differential Equation (SDE)

The Stratonovich (midpoint discretized) Langevin equations can be analyzed in terms of the Wiener process dW^i , which can be rewritten in terms of Gaussian noise $\eta^i = dW^i/dt$ if care is taken in the limit.

$$dM^G = f^G(t, M(t))dt + \hat{g}_i^G(t, M(t))dW^i ,$$

$$\dot{M}^G(t) = f^G(t, M(t)) + \hat{g}_i^G(t, M(t))\eta^i(t) ,$$

$$dW^i \rightarrow \eta^i dt ,$$

$$M = \{ M^G; G = 1, \dots, \Lambda \} ,$$

$$\eta = \{ \eta^i; i = 1, \dots, N \} .$$

$$\dot{M}^G = dM^G/dt ,$$

$$\langle \eta^j(t) \rangle_{\eta} = 0 ,$$

$$\langle \eta^j(t), \eta^j(t') \rangle_{\eta} = \delta^{jj'} \delta(t - t') ,$$

η^i represents Gaussian white noise, and moments of an arbitrary function $F(\eta)$ over this stochastic space are defined by a path-type integral over η^i , folding time increments $\theta = \Delta t$,

$$\langle F(\eta) \rangle_{\eta} = \bar{N}^{-1} \int D\eta F(\eta) \exp\left(-\frac{1}{2} \int_{t_0}^{\infty} dt \eta^i \eta^i\right) ,$$

$$\bar{N} = \int D\eta \exp\left(-\frac{1}{2} \int_{t_0}^{\infty} dt \eta^i \eta^i\right) ,$$

$$D\eta = \lim_{v \rightarrow \infty} \prod_{\alpha=0}^{v+1} \prod_{j=1}^N (2\pi\theta)^{-1/2} dW_{\alpha}^j ,$$

$$t_{\alpha} = t_0 + \alpha\theta ,$$

$$\frac{1}{2} \int dt \eta^i \eta^i = \frac{1}{2\theta} \sum_{\beta} \sum_i (W_{\beta}^i - W_{\beta-1}^i)^2 ,$$

$$\langle \eta^i \rangle_{\eta} = 0 ,$$

$$\langle \eta^i(t) \eta^j(t') \rangle_{\eta} = \delta^{ij} \delta(t - t') .$$

Partial Differential Equation (PDE)

If some boundary conditions are added as Lagrange multipliers, these enter as a “potential” V , creating a Schrödinger-type equation:

$$P_{,t} = \frac{1}{2} (g^{GG'} P)_{,GG'} - (g^G P)_{,G} + VP ,$$

$$P = \langle P_\eta \rangle_\eta ,$$

$$g^G = f^G + \frac{1}{2} \hat{g}_i^{G'} \hat{g}_{i,G}^G ,$$

$$g^{GG'} = \hat{g}_i^G \hat{g}_i^{G'} ,$$

$$(\dots)_{,G} = \partial(\dots)/\partial M^G .$$

Note that g^G replaces f^G in the SDE if the Itô (prepoint discretized) calculus is used to define that equation.

Lagrangian Probability Distribution Function (PDF)

This can be transformed to the Stratonovich representation, in terms of the Feynman Lagrangian L possessing a covariant variational principle,

$$P[M_t|M_{t_0}]dM(t) = \int \cdots \int \underline{D}M \exp\left(-\min \int_{t_0}^t dt' L\right) \delta(M(t_0) = M_0) \delta(M(t) = M_t),$$

$$\underline{D}M = \lim_{u \rightarrow \infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_G (2\pi\theta)^{-1/2} dM_\rho^G,$$

$$L(\dot{M}^G, M^G, t) = \frac{1}{2} (\dot{M}^G - h^G) g_{GG'} (\dot{M}^{G'} - h^{G'}) + \frac{1}{2} h^G_{;G} + R/6 - V,$$

$$[\cdots]_{,G} = \frac{\partial[\cdots]}{\partial M^G},$$

$$h^G = g^G - \frac{1}{2} g^{-1/2} (g^{1/2} g^{GG'})_{,G'},$$

$$g_{GG'} = (g^{GG'})^{-1},$$

$$g = \det(g_{GG'}),$$

$$h^G_{;G} = h^G_{,G} + \Gamma_{GF}^F h^G = g^{-1/2} (g^{1/2} h^G)_{,G},$$

$$\Gamma_{JK}^F \equiv g^{LF} [JK, L] = g^{LF} (g_{JL,K} + g_{KL,J} - g_{JK,L}),$$

$$R = g^{JL} R_{JL} = g^{JL} g^{JK} R_{FJKL},$$

$$R_{FJKL} = \frac{1}{2} (g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN} (\Gamma_{FK}^M \Gamma_{JL}^N - \Gamma_{FL}^M \Gamma_{JK}^N).$$

Path-Integral Riemannian Geometry

The midpoint derivation explicitly derives a Riemannian geometry induced by these statistics, with a metric defined by the inverse of the covariance matrix

$$g_{GG'} = (g^{GG'})^{-1} ,$$

and where R is the Riemannian curvature

$$R = g^{JL} R_{JL} = g^{JL} g^{JK} R_{FJKL} ,$$

An Itô prepoint discretization for the same probability distribution P gives a much simpler algebraic form,

$$M(\bar{t}_s) = M(t_s) ,$$

$$\underline{L} = \frac{1}{2} (dM^G/dt - g^G) g_{GG'} (dM^{G'}/dt - g^{G'}) - V ,$$

but the Lagrangian \underline{L} so specified does not satisfy a variational principle as useful for moderate to large noise; its associated variational principle only provides information useful in the weak-noise limit. Numerically, this often means that finer meshes are required for calculations for the prepoint representation.

Fitting Variance With ASA

Consider a one variable problem,

$$P[M_{t+\Delta t}|M_t] = (2\pi \hat{g}^2 \Delta t)^{-1/2} \exp(-\Delta t L),$$

$$L = (\dot{M} - f)^2 / (2 \hat{g}^2),$$

with parameter-coefficients α n f and g to be fi t to data.

The cost function to be fi t to $M(t)$ data is

$$\underline{L} = L\Delta t + \frac{1}{2} \ln(2\pi\Delta t\hat{g}_t^2),$$

The nonlinear entry of g into the cost function, e.g., competing influence in the denominator of L and in the logarithm term from the prefactor in P , often enables a tight fi t to the absolute value of g . In most nonlinear regression methods, this is not possible.

Similar considerations hold for more than one variable. The calculation of the evolution of Langevin systems has been implemented in several systems using ASA. It has been used as an aid to debug the ASA fi tting codes, by fi rst generating data from coupled Langevin equations, relaxing the coeffi cients, and then fi tting this data with the effective Lagrangian cost-function algorithm to recapture the original coeffi cients within the diffusions defi ned by $g^{GG'}$.

Information

With reference to a steady state $\bar{P}(\tilde{M})$, when it exists, an analytic definition of the information gain \hat{Y} in state $\tilde{P}(\tilde{M})$ is defined by

$$\hat{Y}[\tilde{P}] = \int \cdots \int_{\underline{M}} \tilde{P} \ln(\tilde{P}/\bar{P}),$$

where a path integral is defined such that all intermediate-time values of \tilde{M} appearing in the folded short-time distributions \tilde{P} are integrated over. This is quite general for any system that can be described as Gaussian-Markovian, even if only in the short-time limit.

As time evolves, the distribution likely no longer behaves in a Gaussian manner, and the apparent simplicity of the short-time distribution must be supplanted by numerical calculations.

Transformations Without Itô Calculus

Consider

$$V[S, t + \delta t | S, t] = (2\pi(\sigma S)^2 \delta t)^{-1/2} \exp(-L\delta t),$$

$$L = \frac{(\dot{S} + rS)^2}{2(\sigma S)^2} + r,$$

$$\dot{S} = \frac{\delta S}{\delta t} = \frac{S(t + \delta t) - S(t)}{\delta t}.$$

Some care must be taken with nonconstant drifts and diffusions. For example, for purposes of calculating volatilities, it is often convenient to transform to a variable Z (S relative to some \bar{S} scale)

$$Z = \ln S.$$

The above distribution can be transformed into $V[Z, t + \delta t | Z, t]$,

$$\begin{aligned} dS_t V[S, t + \delta t | S, t] &= dZ_t V[Z, t + \delta t | Z, t] \\ &= dZ_t (2\pi\sigma^2 \delta t)^{-1/2} \exp(-L'\delta t), \end{aligned}$$

$$L'\delta t = \frac{([\exp(Z_{t+\delta t} - Z_t) - 1] + r)^2}{2\sigma^2 \delta t} + r\delta t.$$

This can be expanded into

$$\begin{aligned} L'\delta t &\approx \frac{(Z_{t+\delta t} - Z_t + \frac{1}{2}(Z_{t+\delta t} - Z_t)^2 - r\delta t)^2}{2\sigma^2 \delta t} + r\delta t \\ &\approx \frac{(Z_{t+\delta t} - Z_t - (r - \frac{1}{2}\sigma^2)\delta t)^2}{2\sigma^2 \delta t} + r\delta t, \end{aligned}$$

$$(Z_{t+\delta t} - Z_t)^2 \approx \sigma^2 \delta t,$$

where only terms of order δt have been kept, yielding

$$L' = \frac{\left(\dot{Z} - \left(r - \frac{1}{2}\sigma^2\right)\right)^2}{2\sigma^2} + r.$$

Intuitive Variables

It must be emphasized that the output need not be confined to complex algebraic forms or tables of numbers. Because L possesses a variational principle, sets of contour graphs, at different long-time epochs of the path-integral of P over its variables at all intermediate times, give a visually intuitive and accurate decision-aid to view the dynamic evolution of the scenario. For example, this Lagrangian approach permits a quantitative assessment of concepts usually only loosely defined.

Euler-Lagrange Variational Equations

The Euler-Lagrange variational equations give rise to the familiar force law

$$“F = ma” : \delta L = 0 = \frac{\partial L}{\partial M^G} - \frac{\partial}{\partial t} \frac{\partial L}{\partial (\partial M^G / \partial t)},$$

$$“Force” = \frac{\partial L}{\partial M^G},$$

$$“Mass” = g_{GG'} = \frac{\partial^2 L}{\partial (\partial M^G / \partial t) \partial (\partial M^{G'} / \partial t)},$$

where M^G are the variables and L is the Lagrangian. These physical entities provide another form of intuitive, but quantitatively precise, presentation of these analyses.

Canonical Momenta Indicators (CMI)

Canonical Momenta Indicators (CMI), defined by

$$\text{“Momentum”} = \Pi^G = \frac{\partial L}{\partial(\partial M^G / \partial t)},$$

can be used as financial indicators faithful to an underlying mathematics modeling markets as stochastic distributions.

NUMERICAL PATH INTEGRATION (PATHINT)

Applications

Given a form for L , we use the path-integral to calculate the long-time distribution of variables. This is impossible in general to calculate in closed form, and we therefore must use numerical methods. PATHINT is a code developed for calculating highly nonlinear multivariate Lagrangians.

The path-integral calculation of the long-time distribution, in addition to being a predictor of upcoming information, provides an internal check that the system can be well represented as a nonlinear Gaussian-Markovian system. The use of the path integral to compare different models is akin to comparing short- and long-time correlations. Complex boundary conditions can be cleanly incorporated into this representation, using a variant of “boundary element” techniques.

Outline of PATHINT Algorithm

The histogram procedure recognizes that the distribution can be numerically approximated to a high degree of accuracy as sum of rectangles at points M_i of height P_i and width ΔM_i . For convenience, just consider a one-dimensional system. The above path-integral representation can be rewritten, for each of its intermediate integrals, as

$$\begin{aligned} P(M; t + \Delta t) &= \int dM' [g_s^{1/2} (2\pi\Delta t)^{-1/2} \exp(-L_s \Delta t)] P(M'; t) \\ &= \int dM' G(M, M'; \Delta t) P(M'; t), \end{aligned}$$

$$P(M; t) = \sum_{i=1}^N \pi(M - M_i) P_i(t),$$

$$\pi(M - M_i) = \begin{cases} 1, & (M_i - \frac{1}{2} \Delta M_{i-1}) \leq M \leq (M_i + \frac{1}{2} \Delta M_i), \\ 0, & \text{otherwise,} \end{cases}$$

which yields

$$P_i(t + \Delta t) = T_{ij}(\Delta t) P_j(t),$$

$$T_{ij}(\Delta t) = \frac{2}{\Delta M_{i-1} + \Delta M_i} \int_{M_i - \Delta M_{i-1}/2}^{M_i + \Delta M_i/2} dM \int_{M_j - \Delta M_{j-1}/2}^{M_j + \Delta M_j/2} dM' G(M, M'; \Delta t).$$

Boundary Condition Sensitivity

For derivative boundary conditions, for better numerical accuracy, it often is necessary to generalize the histogram expansion to a trapezoidal expansion to give some shape to the histograms.

Mesh Limitations

Care must be used in developing the mesh in ΔM^G , which is strongly dependent on the diagonal elements of the diffusion matrix, e.g.,

$$\Delta M^G \approx (\Delta t g^{G||G|})^{1/2} .$$

Presently, this constrains the dependence of the covariance of each variable to be a nonlinear function of that variable, albeit arbitrarily nonlinear, in order to present a straightforward rectangular underlying mesh.

A previous paper attempted to circumvent this restriction by taking advantage of Riemannian transformations to a relatively diagonal problem.

For more than one variable, the above constraints on the mesh only suffice for diagonal elements of the $g^{GG'}$ matrix. To consider the influence on off-diagonal terms, a tiling approach should be taken to the full mesh.

Probability Tree (PATHTREE)

PATHINT motivated the development of PATHTREE, an algorithm that permits extremely fast accurate computation of probability distributions of a large class of general nonlinear diffusion processes.

The natural metric of the space is used to first lay down the mesh. The evolving local short-time distributions on this mesh are then dynamically calculated.

The short-time probability density gives the correct result up to order $O(\Delta t)$ for any final point S' , the order required to recover the corresponding partial differential equation.

Standard Binomial Tree

In a two-step binomial tree, the step up Su or step down Sd from a given node at S is chosen to match the standard deviation of the differential process. The constraints on u and d are chosen as

$$ud = 1 ,$$

If we assign probability p to the up step Su , and $q = (1 - p)$ to the down step Sd , the matched mean and variance are

$$pSu + (1 - p)Sd = \langle S(t + \Delta t) \rangle ,$$

$$S^2(pu^2 + qd^2 - (pu + qd)^2) = \langle (S(t + \Delta t) - \langle S(t + \Delta t) \rangle)^2 \rangle .$$

The right-hand-side can be derived from the stochastic model used.

Deficiency of Standard Algorithm to Order \sqrt{dt}

A tree is constructed that represents the time evolution of the stochastic variable S . S is assumed to take only 2 values, u , (up value), and d (down value) at moment t , given the value S at moment $t - \Delta t$. The probabilities for the up and down movements are p and q , respectively. The 4 unknowns $\{u, d, p, q\}$ are calculated by imposing the normalization of the probability and matching the first two moments conditioned by the value S at $t - \Delta t$, using the variance of the exact probability distribution $P(S, t | S_0, t_0)$. One additional condition is arbitrary and is usually used to symmetrize the tree, e.g., $ud = 1$.

If the system to be modeled is given by a differential form, e.g.,

$$dS = fdt + gdw$$

then the noise term is only given to order \sqrt{dt} .

The Ornstein-Uhlenbeck (OU) process, $f = bS$ and $g = v$, for constant b and v , is special, as some higher order dt corrections in systems described by $g \propto S^x$ are zero for $x = 0$. The Black-Scholes (BS) process, $f = bS$ and $g = \sigma S$, for constant b and σ , also is special, as it can be simply transformed to a constant-diffusion lognormal process with the same $O(dt)$ simplifications.

Problems Generalizing The Standard Tree

The main problem is that the above procedure cannot be applied to a general nonlinear diffusion process, as the algorithm involves a previous knowledge of terms of $O(\Delta t)$ in the formulas of quantities $\{u, p\}$ obtained from a finite time expansion of the exact solution P sought. Otherwise, the discrete numerical approximation obtained does not converge to the proper solution.

Probability PATHTREE

In order to obtain tree variables valid up to $O(\Delta t)$, we turn to the short-time path-integral representation of the solution of the Fokker-Planck equation, which is just the multiplicative Gaussian-Markovian distribution. In the prepoint discretization relevant to the construction of a tree,

$$P(S', t' | S, t) = \frac{1}{\sqrt{2\pi\Delta t g^2}} \exp\left(-\frac{(S' - S - f\Delta t)^2}{2g^2\Delta t}\right)$$

$$\Delta t = t' - t$$

valid for displacements S' from S ‘reasonable’ as measured by the standard deviation $g\sqrt{\Delta t}$, which is the basis for the construction of meshes in the PATHINT algorithm.

The crucial aspects of this approach are: There is no a priori need of the first moments of the exact long-time probability distribution P , as the necessary statistical information to the correct order in time is contained in the short-time propagator. The mesh in S at every time step need not recombine in the sense that the prepoint-postpoint relationship be the same among neighboring S nodes, as the short-time probability density gives the correct result up to order $O(\Delta t)$ for any final point S' . Instead, we use the natural metric of the space to first lay down our mesh, then dynamically calculate the evolving local short-time distributions on this mesh.

Construction of PATHTREE

We construct an additive PATHTREE, starting with the initial value S_0 , with successive increments

$$S_{i+1} = S_i + g\sqrt{\Delta t}, S_i > S_0$$

$$S_{i-1} = S_i - g\sqrt{\Delta t}, S_i < S_0,$$

where g is evaluated at S_i . We define the up and down probabilities p and q , resp., in an abbreviated notation, as

$$p = \frac{P(i+1|i; \Delta t)}{P(i+1|i; \Delta t) + P(i-1|i; \Delta t)}$$

$$q = 1 - p.$$

where the P 's are the short-time transition probability densities. Note that in the limit of small Δt ,

$$\lim_{\Delta t \rightarrow 0} p = \frac{1}{2}.$$

Direct Calculation of Probability

We can calculate the probability density function by first recursively computing the probabilities of reaching each node of the tree. This can be performed efficiently thanks to the Markov property. To compute the density function we need to rescale these probabilities by the distance to the neighboring nodes: the more spread the nodes are, the lower the density. First we compute the probability of reaching each final node of the tree. We do this incrementally by first computing the probabilities of reaching nodes in time slice 1, then time slice 2 and so forth. At time slice 0, we know that the middle node has probability 1 of being reached and all the others have probability 0. We compute the probability of reaching a node as a sum of two contributions from the previous time slice. We reach the node with transition pu from the node below at the previous slice, and with transition pd from the node above. Each contribution is the product of the probability at the previous node times the transition to the current node. This formula is just a discretized version of the Chapman-Kolmogorov equation

$$p(x_j, t_i) = p(x_{j-1}, t_{i-1})pu_{j-1} + p(x_{j+1}, t_{i-1})pd_{j+1} .$$

After we have computed the absolute probabilities at the final nodes, we can give a proper estimation of the density, by scaling the probabilities by the average of sizes of the two adjacent intervals, $\text{density}_i = p_i / ((S_{i+2} - S_{i-2})/2)$.

Numerical Derivatives of Expectation of Probability

The probability P at time of expiration T can be calculated as a numerical derivative with respect to strike X of a European call option, taking the risk-free rate r to be zero, given an underlying S_0 evaluated at time $t = 0$, with strike X , given other variables such as volatility σ and cost of carry b . The call is the expectation of the function $\text{Max}(S - X, 0)$.

$$P[S(T)|S(t_0)] \Big|_{S(T)=X} = P[X|S(t_0)] = \frac{\partial^2 C}{\partial X^2}$$

This calculation of the probability distribution is dependent on the same conditions necessary for any tree algorithm, i.e., that enough nodes are processed to ensure that the resultant evaluations are a good representation of the corresponding Fokker-Planck equation, and that the number of iterations in PATHTREE are sufficient for convergence.

Alternative First Derivative Calculation of Probability

An alternative method of calculating the probability P a a first-order numerical derivative, instead of as second-order derivative, with respect to X is to define a function C_H using the Heaviside step-function $H(S, X) = 1$ if $S \geq X$ and 0 otherwise, instead of the Max function at the time to expiration. This yields

$$P[S(T)|S(t_0)]_{S(T)=X} = P[X|S(t_0)] = -\frac{\partial C_H}{\partial X}$$

Sometimes this is numerically useful for sharply peaked distributions at the time of expiration, but we have found the second derivative algorithm above to work fine with a sufficient number of epochs.

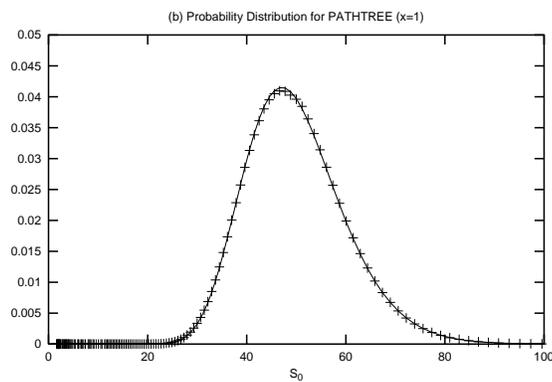
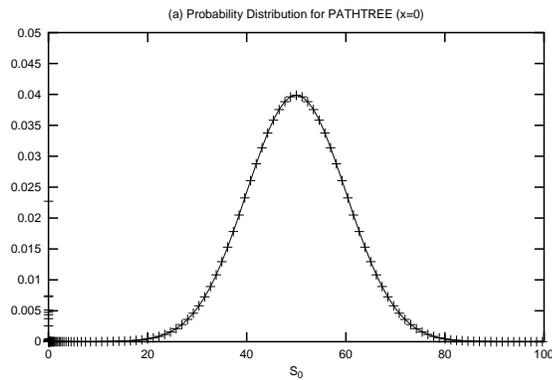
Our tests verify that the three methods give the same density. We consider the numerical-derivative calculations a very necessary baseline to determine the number of epochs required to get reasonable accuracy.

PATHTREE vs PATHINT

For PATHINT, the time and space variables are determined independently. I.e., the ranges of the space variables are best determined by first determining the reasonable spread of the distribution at the final time epoch. For PATHTREE just one parameter, the number of epochs N , determines the mesh for both time and the space variables. This typically leads to a growth of the tree, proportional to \sqrt{N} , much faster than the spread of the distribution, so that much of the calculation is not relevant.

Black-Scholes (BS) and Ornstein-Uhlenbeck (OU) Examples

The graphs below compare analytic solutions with second-derivative numerical ("+"s) PATHTREE calculations.



GENERIC MESOSCOPIC NEURAL NETWORKS (MNN)

Applications

Modern stochastic calculus permits development of alternative descriptions of path-integral Lagrangians, Fokker-Planck equations, and Langevin rate equations. The induced Riemannian geometry affords invariance of probability distribution under general nonlinear transformations.

ASA presents a powerful global optimization that has been tested in a variety of problems defined by nonlinear Lagrangians.

Parallel-processing computations can be applied to ASA as well as to a neural-network architecture.

MNN Learning

“Learning” takes place by presenting the MNN with data, and parametrizing the data in terms of the “firings,” or multivariate M^G “spins.” The “weights,” or coefficients of functions of M^G appearing in the drifts and diffusions, are fit to incoming data, considering the joint “effective” Lagrangian (including the logarithm of the prefactor in the probability distribution) as a dynamic cost function.

The cost function is a sum of effective Lagrangians from each node and over each time epoch of data.

This program of fitting coefficients in Lagrangian uses methods of adaptive simulated annealing (ASA). This maximum likelihood procedure (statistically) avoids problems of trapping in local minima, as experienced by other types of gradient and regression techniques.

MNN Prediction

‘Prediction’ takes advantage of a mathematically equivalent representation of the Lagrangian path-integral algorithm, i.e., a set of coupled Langevin rate-equations. The Itô (prepoint-discretized) Langevin equation is analyzed in terms of the Wiener process dW^i , which is rewritten in terms of Gaussian noise $\eta^i = dW^i/dt$ in the limit:

$$M^G(t + \Delta t) - M^G(t) = dM^G = g^G dt + \hat{g}_i^G dW^i ,$$

$$\frac{dM^G}{dt} = \dot{M}^G = g^G + \hat{g}_i^G \eta^i ,$$

$$M = \{ M^G; G = 1, \dots, \Lambda \} , \quad \eta = \{ \eta^i; i = 1, \dots, N \} ,$$

$$\langle \eta^j(t) \rangle_\eta = 0 , \quad \langle \eta^j(t), \eta^{j'}(t') \rangle_\eta = \delta^{jj'} \delta(t - t') .$$

Moments of an arbitrary function $F(\eta)$ over this stochastic space are defined by a path integral over η^i . The Lagrangian diffusions are calculated as

$$g^{GG'} = \sum_{i=1}^N \hat{g}_i^G \hat{g}_i^{G'} .$$

A coarse deterministic estimate to ‘predict’ the evolution can be applied using the most probable path

$$dM^G/dt = g^G - g^{1/2} (g^{-1/2} g^{GG'})_{,G'} .$$

PATHINT, even when parallelized, typically can be too slow for ‘predicting’ evolution of these systems. However, a new algorithm, PATHTREE holds some promise.

MNN Parallel Processing

The use of parallel processors can make this algorithm even more efficient, as ASA lends itself well to parallelization.

During “learning,” blocks of random numbers are generated in parallel, and then sequentially checked to find a generating point satisfying all boundary conditions.

Advantage is taken of the low ratio of acceptance to generated points typical in ASA, to generate blocks of cost functions, and then sequentially checked to find the next best current minimum.

Additionally, when fitting dynamic systems, e.g., the three physical systems examined to date, parallelization is attained by independently calculating each time epoch’s contribution to the cost function.

Similarly, during “prediction,” blocks of random numbers are generated to support the Langevin-equation calculations, and each node is processed in parallel. PATHINT or PATHTREE also possess features to promote fast calculations.

SOME APPLICATIONS

Statistical Mechanics of Financial Markets (SMFM)

2-Factor Interest-Rate Model

The Brennan-Schwartz model is developed in the variables of short- and long-term interest rates, assumed to follow a joint Wiener stochastic process,

$$dr = \beta_1(r, l, t)dt + \eta_1(r, l, t)dz_1 ,$$

$$dl = \beta_2(r, l, t)dt + \eta_2(r, l, t)dz_2 ,$$

where r and l are the short- and long-term rates, respectively. β_1 and β_2 are the expected instantaneous rates of change in the short-term and long-term rates respectively. η_1 and η_2 are the instantaneous standard deviations of the processes. dz_1 and dz_2 are Wiener processes, with expected values of zero and variance of dt with correlation coefficient ρ . BS simplified and reduced this system to

$$dr = (a_1 + b_1(l - r))dt + r\sigma_1 dz_1 ,$$

$$dl = l(a_2 + b_2r + c_2l)dt + l\sigma_2 dz_2 ,$$

where $\{a_1, b_1, a_2, b_2, c_2\}$ are parameters to be estimated.

Lagrangian Representation

The BS equations can be rewritten as Langevin equations (in the Itô prepoint discretization)

$$dr/dt = a_1 + b_1(l - r) + \sigma_1 r(\gamma^+ n_1 + \gamma^- n_2) ,$$

$$dl/dt = l(a_2 + b_2 r + c_2 l) + \sigma_2 l(\gamma^- n_1 + \gamma^+ n_2) ,$$

$$\gamma^\pm = \frac{1}{\sqrt{2}} [1 \pm (1 - \rho^2)^{1/2}]^{1/2} ,$$

$$n_i = (dt)^{1/2} p_i ,$$

where p_1 and p_2 are independent $[0,1]$ Gaussian distributions.

$$L = \frac{1}{2} F^\dagger \underline{g} F ,$$

$$F = \begin{pmatrix} dr/dt - (a_1 + b_1(l - r)) \\ dl/dt - l(a_2 + b_2 r + c_2 l) \end{pmatrix} ,$$

$$g = \det(\underline{g}) ,$$

$$k = 1 - \rho^2 .$$

\underline{g} , the metric in $\{r, l\}$ -space, is the inverse of the covariance matrix,

$$\underline{g}^{-1} = \begin{pmatrix} (r\sigma_1)^2 & \rho r l \sigma_1 \sigma_2 \\ \rho r l \sigma_1 \sigma_2 & (l\sigma_2)^2 \end{pmatrix} .$$

The cost function C is defined from the equivalent short-time probability distribution, P , for the above set of equations.

$$P = g^{1/2} (2\pi dt)^{-1/2} \exp(-Ldt)$$

$$= \exp(-C) ,$$

$$C = Ldt + \frac{1}{2} \ln(2\pi dt) - \ln(g) .$$

ASA Fits

Interest rates were developed from Treasury bill and bond yields during the period October 1974 through December 1979, the same period as one of the sets used by BS. Short-term rates were determined from Treasury bills with a maturity of three months (BS used 30-day maturities), and long-term rates were determined from Treasury bonds with a maturity of twenty years (BS used at least 15-year maturities).

Bond PDE/PATHINT

Some tentative PATHINT calculations were performed by another colleague on my project. It would be interesting to repeat them with this enhanced code.

Power-Law Model

There is growing evidence that the Black-Scholes lognormal distribution has been less and less descriptive of markets over the past two decades. An example of a generalization of the lognormal distribution is

$$dS/F(S, x) = \mu dt + \sigma dw_S$$

$$F(S, S_0, S_\infty, x, y) = \begin{cases} S, & S < S_0 \\ S^x S_0^{1-x}, & S_0 \leq S \leq S_\infty \\ S^y S_0^{1-x} S_\infty^{x-y}, & S > S_\infty \end{cases}$$

where S_0 and S_∞ are selected to lie outside the data region used to fit the other parameters, e.g., $S_0 = 1$ and $S_\infty = 20$ for fits to Eurodollar futures. We have used the Black-Scholes form $F = S$ inside $S < S_0$ to obtain the usual benefits, e.g., no negative prices as the distribution is naturally excluded from $S < 0$, preservation of put-call parity, etc. We have taken $y = 0$ to reflect total ignorance of markets outside the range of $S > S_\infty$.

Multi-Factor Volatility Model

Any study that geared to perform ASA fits of multivariate Lagrangians and PATHINT long-time calculations can also consider another variable σ , stochastic volatility, that can generalize the BS lognormal distribution:

$$dS/F(S, x) = \mu dt + \sigma dw_S$$

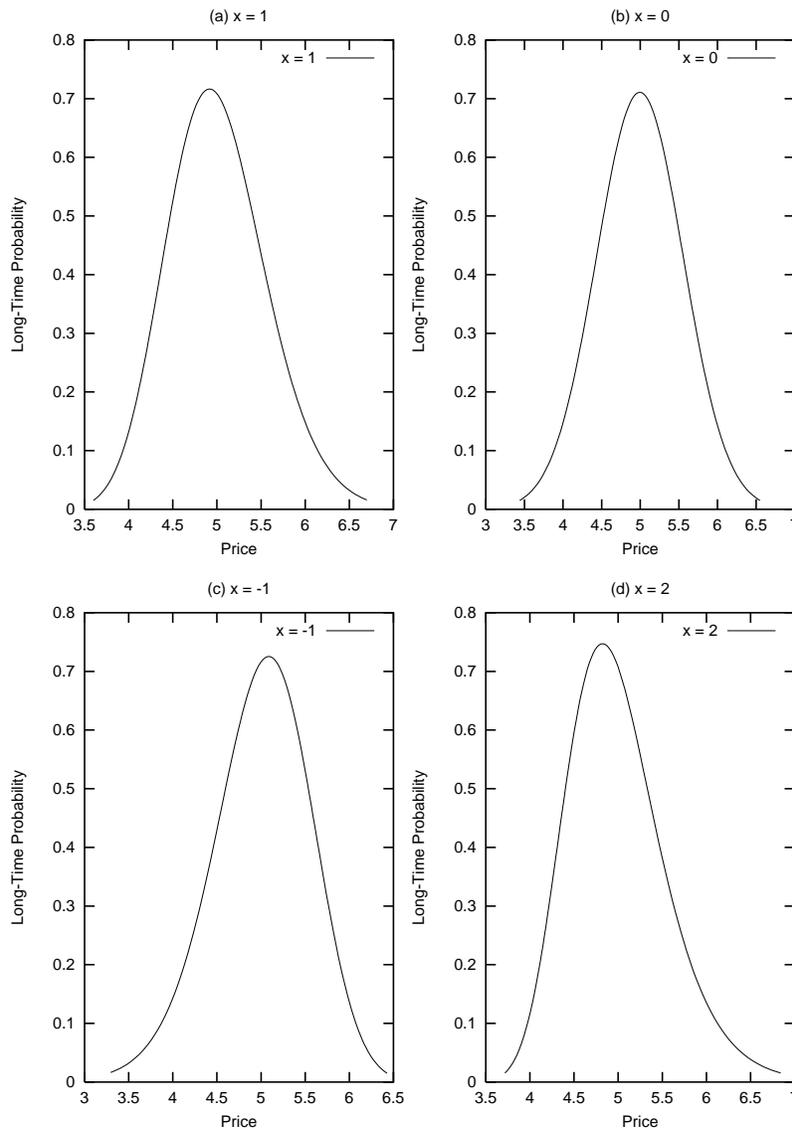
$$d\sigma = \nu + \varepsilon dw_\sigma$$

The drawback of the two-factor PATHINT code is that it is slow. However, it is accurate and robust so we can process any diffusion for general x .

PATHINT Options

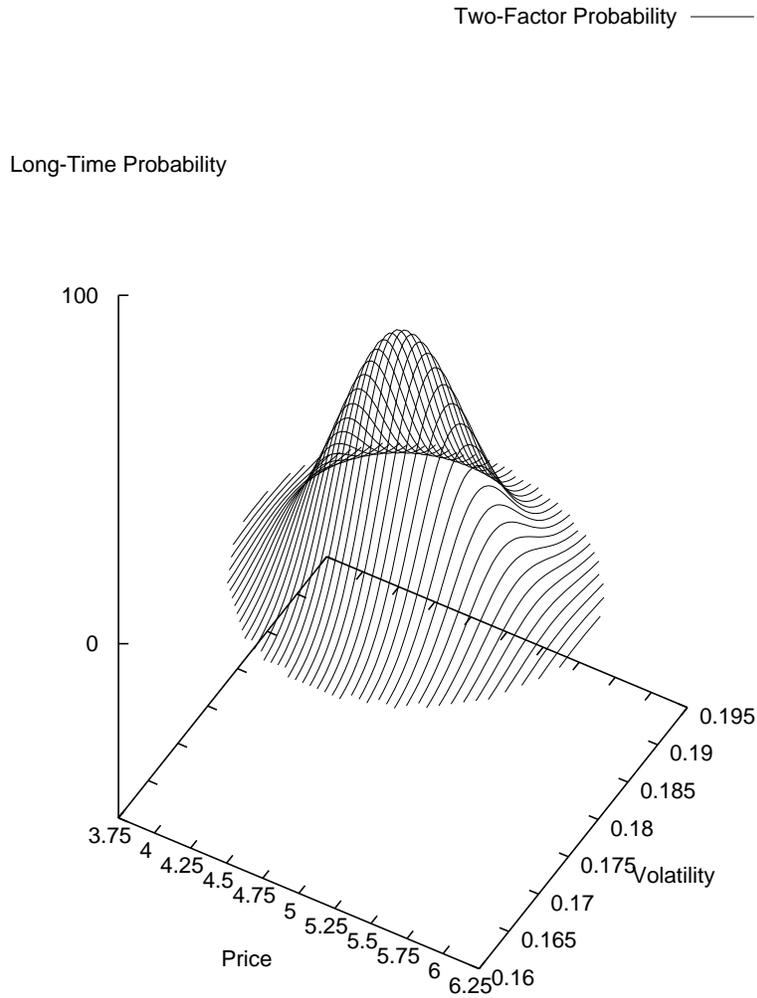
PATHINT is being used to perform European and American, one-factor and two-factor, PATHINT calculations. Some examples are $F(S, S_0, S_\infty, x, y) dz_S$ for x in $\{-1, 0, 1, 2\}$.

The short-time probability distribution at time $T = 0.5$ years for $x = 1$, the (truncated) Black-Scholes distribution. The short-time probability distribution at time $T = 0.5$ years for $x = 0$, the normal distribution. The short-time probability distribution at time $T = 0.5$ years for $x = -1$. The short-time probability distribution at time $T = 0.5$ years for $x = 2$.



PATHINT Two-Factor

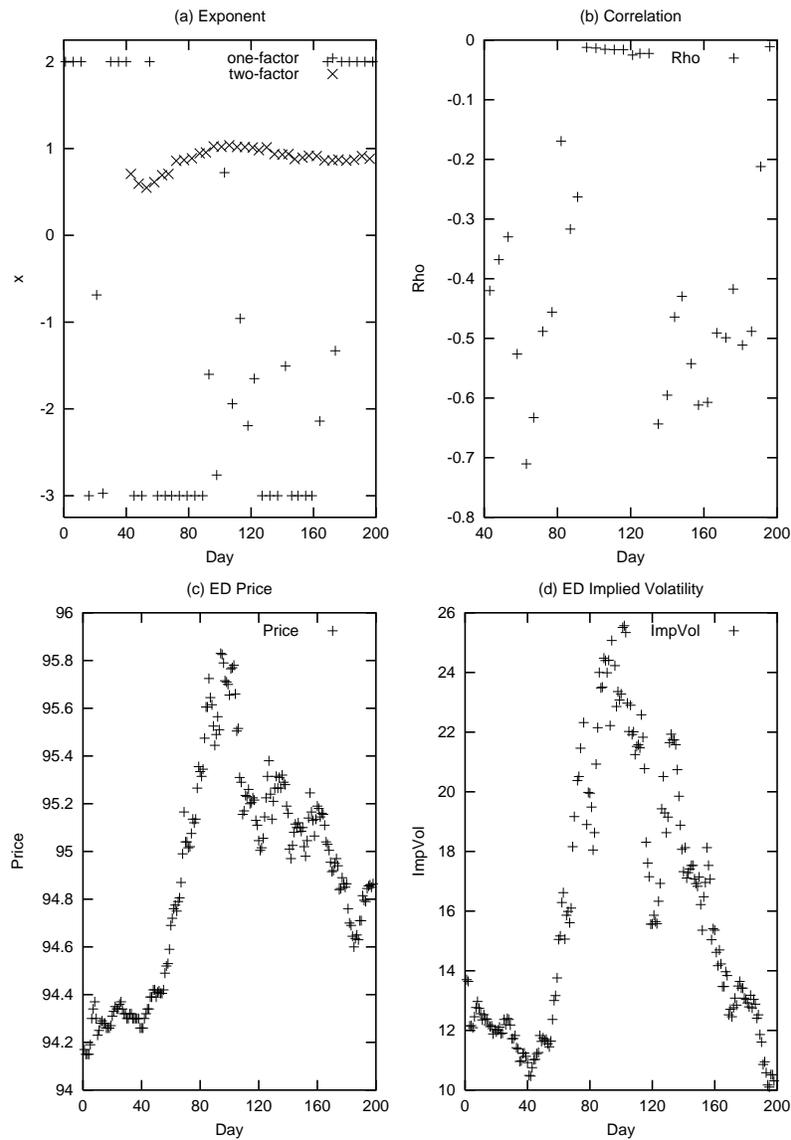
The two-factor distribution at time $T = 0.5$ years for $x = 0.7$.



x Market Indicators

We have developed x 's as indicators of different market contexts. E.g., x may be -2 for some quarter and +2 for a different quarter.

(a) Weekly two-month moving-averaged one-factor and two-factor exponents for ED contract expiring in September 1999 during the period June 1998 through March 1999. (b) Weekly two-month moving-averaged two-factor correlation ρ for this same data. (c) Raw price data used in fits for the above parameters. (d) Implied-volatility data used in fits for the above parameters.



S&P Interday Futures-Cash

CMI and ASA were blended together to form a simple trading code, TRD. An example was published on inter-day trading the S&P 500, using stops for losses on short and long trades and using CMI of the coupled cash and futures data. Data for years 1989 and 1990 was used, wherein one of the years was used to train TRD, and the other year to test TRD; then the years were reversed to establish two examples of trading on two years of quite different data.

In the 1991 study, it was noted that the sensitivity of testing trades to CMI overshadowed any sensitivity to the stops. Therefore, a second study was performed on this same data, but using only CMI. Better results were obtained, but more important, this established that the CMI themselves could lead to profitable trading, taking advantage of inefficiencies in these markets. Therefore, CMI at least can be useful supplemental indicators for other trading systems.

Tick Resolution CMI Trading

We are developing a fully automated electronic trading system using CMI on S&P with minute resolution data.

Statistical Mechanics of Combat (SMC)

National Training Center — Janus(T)

The U.S. Army National Training Center (NTC) is located at Fort Irwin, just outside Barstow, California. As of 1989, there have been about 1/4 million soldiers in 80 brigade rotations at NTC, at the level of two battalion task forces (typically about 3500 soldiers and a battalion of 15 attack helicopters), which train against two opposing force (OPFOR) battalions resident at NTC. NTC comprises about 2500 km², but the current battlefi eld scenarios range over about 5 km linear spread, with a maximum lethality range of about 3 km. NTC is gearing up for full brigade level exercises. The primary purpose of data collection during an NTC mission is to patch together an after action review (AAR) within a few hours after completion of a mission, giving feedback to a commander who typically must lead another mission soon afterward. Data from the fi eld, i.e., multiple integrated laser engagement system (MILES) devices, audio communications, OCs, and stationary and mobile video cameras, is sent via relay stations back to a central command center where this all can be recorded, correlated and abstracted for the AAR. Within a couple of weeks afterwards, a written review is sent to commanders, as part of their NTC take home package.

Janus(T) is an interactive, two-sided, closed, stochastic, ground combat computer simulation. We have expanded Janus(T) to include air and naval combat, in several projects with the author's previous thesis students at the Naval Postgraduate School (NPS).

Stochastic multivariate models were developed for both NTC and Janus, to form a common language to compare the two systems to baseline the simulation to exercise data.

Basic Equations of Aggregated Units

Consider a scenario taken from our NTC study: two red systems, red T-72 tanks (RT) and red armored personnel carriers ($RBMP$), and three blue systems, blue M1A1 and M60 tanks (BT), blue armored personnel carriers ($BAPC$), and blue tube-launched optically-tracked wire-guided missiles ($BTOW$), where RT specifies the number of red tanks at a given time t , etc. Consider the kills suffered by BT , ΔBT , e.g., within a time epoch $\Delta t \approx 5$ min. Here, the x terms represent attrition owing to point fire; the y terms represent attrition owing to area fire. Note that the algebraic forms chosen are consistent with current perceptions of aggregated large scale combat. Now consider sources of noise, e.g., that at least arise from PD, PA, PH, PK, etc. Furthermore, such noise likely has its own functional dependencies, e.g., possibly being proportional to the numbers of units involved in the combat. We write

$$\dot{BT} = \frac{\Delta BT}{\Delta} = x_{RT}^{BT} RT + y_{RT}^{BT} RT BT + x_{RBMP}^{BT} RBMP + y_{RBMP}^{BT} RBMP BT \\ + z_{BT}^{BT} BT \eta_{BT}^{BT} + z_{RT}^{BT} \eta_{RT}^{BT} + z_{RBMP}^{BT} \eta_{RBMP}^{BT}$$

$$\dot{RT} = \dots$$

$$\dot{RBMP} = \dots$$

$$\dot{BAPC} = \dots$$

$$\dot{BTOW} = \dots$$

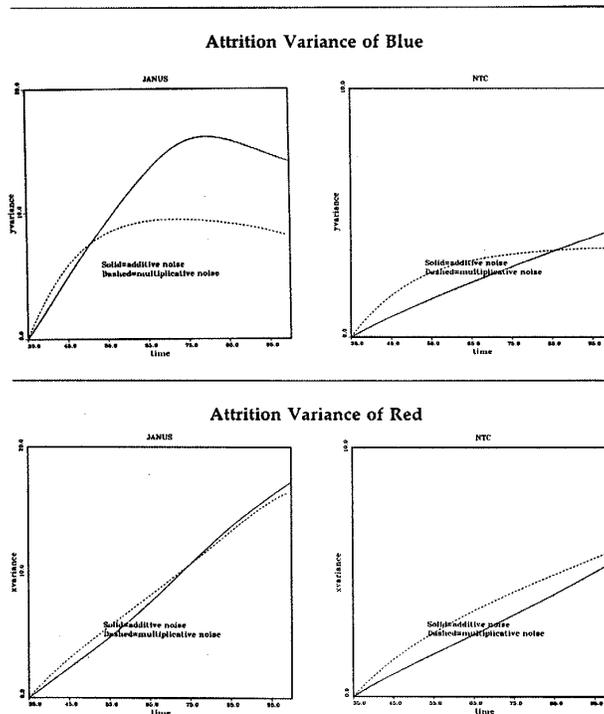
where the η represent sources of (white) noise (in the Itô prepoint discretization discussed below). The noise terms are taken to be log normal (multiplicative) noise for the diagonal terms and additive noise for the off-diagonal terms. The diagonal z term (z_{BT}^{BT}) represents uncertainty associated with the *target* BT , and the off-diagonal z terms represent uncertainty associated with the *shooters* RT and $RBMP$. The x and y are constrained such that each term is bounded by the mean of the KVS, averaged over all time and trajectories of similar scenarios; similarly, each z term is constrained to be bounded by the variance of the KVS. Equations similar to the \dot{BT} equation are also written for \dot{RT} , \dot{RBMP} , \dot{BAPC} , and \dot{BTOW} . Only x and y that reflect possible nonzero entries in the KVS are free to be used for the fitting procedure. For example, since Janus(T) does not permit direct-fire fratricide, such terms are set to zero. In most NTC scenarios, fratricide typically is negligible.

Long-Time Correlations Test Short-Time Models

Especially because we are trying to mathematically model sparse and poor data, different drift and diffusion algebraic functions can give approximately the same algebraic cost-function when fitting short-time probability distributions to data. The calculation of long-time distributions permits a better choice of the best algebraic functions, i.e., those which best follow the data through a predetermined epoch of battle. Thus, dynamic physical mechanisms, beyond simple Lanchester ‘line’ and ‘area’ firing terms, can be identified. Afterwards, if there are closely competitive algebraic functions, they can be more precisely assessed by calculating higher algebraic correlation functions from the probability distribution.

Data from 35 to 70 minutes was used for the short-time fit. The path integral used to calculate this fitted distribution from 35 minutes to beyond 70 minutes. This serves to compare long-time correlations in the mathematical model versus the data, and to help judge extrapolation past the data used for the short-time fits. The means are fit very well by this model, even in out-of-sample time periods, something that other Lanchester modelers have not achieved, especially with such empirical data. The variances strongly suggest that the additive-noise model is superior to the multiplicative-noise model.

JANUS(T) & NTC Attrition Variances



CMI

The results of Janus(T) attrition of Red and Blue units are given in the upper figure. The canonical momenta indicators (CMI) for each system are given in the lower figure.

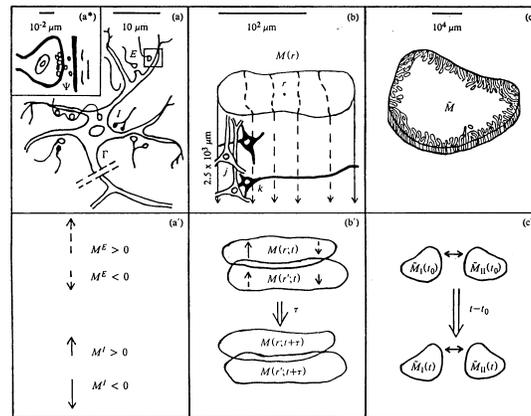
Using the particular model considered here, the CMI are seen to be complementary to the attrition rates, being somewhat more sensitive to changes in the battle than the raw data. The coefficients fit to the combat data are modifiable to fit the current ‘reality’ of system capabilities.

The CMI are more sensitive measures than the energy density, effectively the square of the CMI, or the information which also effectively is in terms of the square of the CMI (essentially integrals over quantities proportional to the energy times a factor of an exponential including the energy as an argument). This is even more important when replenishment of forces is permitted, often leading to oscillatory variables. Neither the energy or the information give details of the components as do the CMI.

Statistical Mechanics of Neocortical Interactions (SMNI)

Multiple Scales

Multiple scales are aggregated, from synaptic dynamics, to neuronal dynamics, to minicolumnar dynamics (100's of neurons). At the level of minicolumns, neocortex seems to be well described by Gaussian-Markovian dynamics.



A derivation is given of the physics of chemical inter-neuronal and electrical intra-neuronal interactions. The derivation yields a short-time probability distribution of a given neuron firing due to its just-previous interactions with other neurons. Within $\tau_j \sim 5-10$ msec, the conditional probability that neuron j fires ($\sigma_j = +1$) or does not fire ($\sigma_j = -1$), given its previous interactions with k neurons, is

$$p_{\sigma_j} \approx \Gamma \Psi \approx \frac{\exp(-\sigma_j F_j)}{\exp(F_j) + \exp(-F_j)},$$

$$F_j = \frac{V_j - \sum_k a_{jk}^* v_{jk}}{(\pi \sum_{k'} a_{jk'}^* (v_{jk'}^2 + \phi_{jk'}^2))^{1/2}},$$

$$a_{jk} = \frac{1}{2} A_{jk} (\sigma_k + 1) + B_{jk}.$$

Γ represents the ‘intra-neuronal’ probability distribution.

Microscopic Aggregation

A derived mesoscopic Lagrangian \underline{L}_M defines the short-time probability distribution of firings in a minicolumn, composed of $\sim 10^2$ neurons, given its just previous interactions with all other neurons in its macrocolumnar surround. G is used to represent excitatory (E) and inhibitory (I) contributions. \bar{G} designates contributions from both E and I .

$$\begin{aligned} P_M &= \prod_G P_M^G [M^G(r; t + \tau) | M^{\bar{G}}(r'; t)] \\ &= \sum_{\sigma_j} \delta \left(\sum_{jE} \sigma_j - M^E(r; t + \tau) \right) \delta \left(\sum_{jI} \sigma_j - M^I(r; t + \tau) \right) \prod_j^N p_{\sigma_j} \\ &\approx \prod_G (2\pi\tau g^{GG})^{-1/2} \exp(-N\tau \underline{L}_M^G), \end{aligned}$$

$$P_M \approx (2\pi\tau)^{-1/2} g^{1/2} \exp(-N\tau \underline{L}_M),$$

$$\underline{L}_M = \underline{L}_M^E + \underline{L}_M^I = (2N)^{-1} (\dot{M}^G - g^G) g_{GG'} (\dot{M}^{G'} - g^{G'}) + M^G J_G / (2N\tau) \underline{V}',$$

$$\underline{V}' = \sum_G \underline{V}''^G (\rho \nabla M^{G'})^2,$$

$$g^G = -\tau^{-1} (M^G + N^G \tanh F^G),$$

$$g^{GG'} = (g_{GG'})^{-1} = \delta_G^{G'} \tau^{-1} N^G \operatorname{sech}^2 F^G,$$

$$g = \det(g_{GG'}),$$

$$F^G = \frac{(V^G - a_G^{[G]} v_G^{[G]} N^{G'} - \frac{1}{2} A_G^{[G]} v_G^{[G]} M^{G'})}{(\pi[(v_G^{[G]})^2 + (\phi_G^{[G]})^2](a_G^{[G]} N^{G'} + \frac{1}{2} A_G^{[G]} M^{G'}))^{1/2}},$$

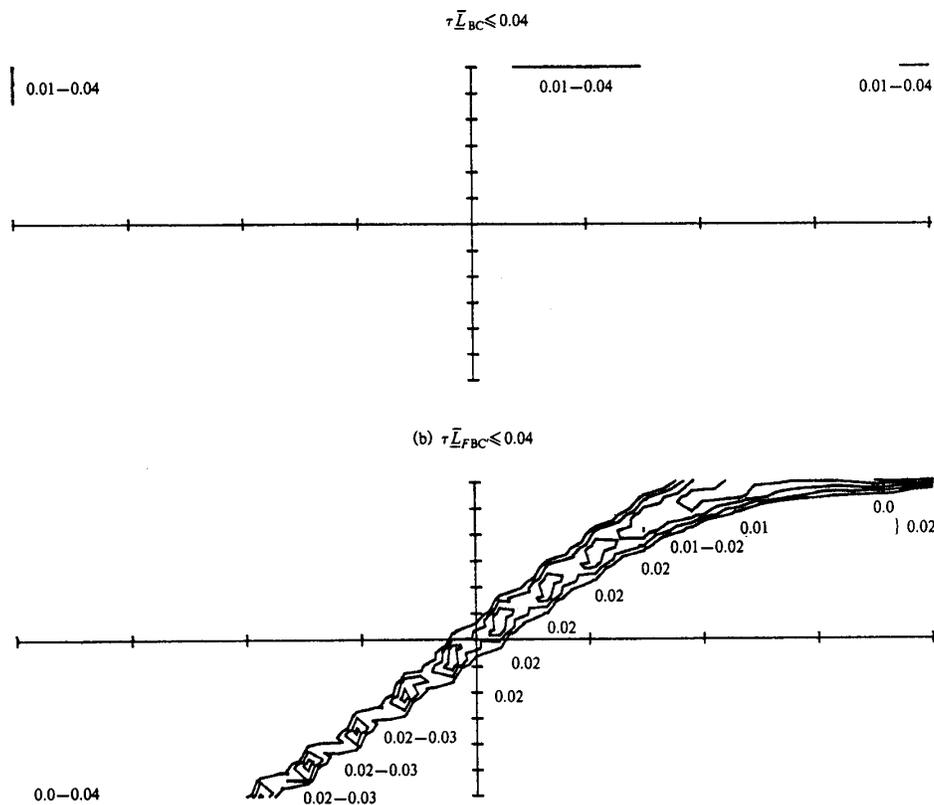
$$a_{G'}^G = \frac{1}{2} A_{G'}^G + B_{G'}^G.$$

Short-Term Memory (STM)

We choose empirical ranges of synaptic parameters corresponding to a predominately excitatory case (EC), predominately inhibitory case (IC), and a balanced case (BC) in between. For each case, also consider a “centering mechanism” (EC’, IC’, BC’), whereby some synaptic parameter is internally manipulated, e.g., some chemical neuromodulation or imposition of patterns of firing, such that there is a maximal efficiency of matching of afferent and efferent firing:

$$M^G \approx M^{*G} \approx 0 .$$

This sets conditions on other possible minima of the *static* Lagrangian \bar{L} .



Electroencephalography (EEG)

A coarse calculation begins by considering the Lagrangian \underline{L}_F , the Feynman midpoint-discretized Lagrangian. The Euler-Lagrange variational equation associated with \underline{L}_F leads to a set of 12 coupled first-order differential equations, with coefficients nonlinear in M^G , in the 12 variables $\{ M^G, \dot{M}^G, \ddot{M}^G, \nabla M^G, \nabla^2 M^G \}$ in $(r; t)$ space. In the neighborhood of extrema $\ll \bar{M}^G \gg$, \underline{L}_F can be expanded as a Ginzburg-Landau polynomial, i.e., in powers of M^E and M^I . To investigate first-order linear oscillatory states, only powers up to 2 in each variable are kept, and from this the variational principle leads to a relatively simple set of coupled linear differential equations with constant coefficients:

$$\begin{aligned}
 0 &= \delta \underline{L}_F = \underline{L}_{F, G;t} - \delta_G \underline{L}_F \\
 &\approx - \underline{f}_{|G|} \ddot{M}^{|G|} + \underline{f}_{|G|}^1 \dot{M}^{G^*} - \underline{g}_{|G|} \nabla^2 M^{|G|} + \underline{b}_{|G|} M^{|G|} + \underline{b} M^{G^*}, \quad G^* \neq G, \\
 (\dots)_{,G;t} &= (\dots)_{,GG'} \dot{M}^{G'} + (\dots)_{,GG'} \ddot{M}^{G'}, \\
 \underline{M}^G &= M^G - \ll \bar{M}^G \gg, \quad \underline{f}_{|E|}^1 = -\underline{f}_{|I|}^1 \equiv \underline{f}.
 \end{aligned}$$

These equations are then Fourier transformed and the resulting dispersion relation is examined to determine for which values of the synaptic parameters and of the normalized wave-number ξ , the conjugate variable to r , can oscillatory states, $\omega(\xi)$, persist.

For instance, a typical example is specified by extrinsic sources $J_E = -2.63$ and $J_I = 4.94$, $N^E = 125$, $N^I = 25$, $V^G = 10$ mV, $A_E^G = 1.75$, $A_I^G = 1.25$, $B_{G'}^G = 0.25$, and $v_{G'}^G = \phi_{G'}^G = 0.1$ mV. The synaptic parameters are within observed ranges, and the J_G 's are just those values required to solve the Euler-Lagrange equations at the selected values of M^G . The global minimum is at $\bar{M}^E = 25$ and $\bar{M}^I = 5$. This set of conditions yields (dispersive) dispersion relations

$$\omega\tau = \pm \{ -1.86 + 2.38(\xi\rho)^2; -1.25i + 1.51i(\xi\rho)^2 \},$$

where $\xi = |\xi|$. The propagation velocity defined by $d\omega/d\xi$ is ~ 1 cm/sec, taking typical wavenumbers ξ to correspond to macrocolumnar distances $\sim 30\rho$. Calculated frequencies ω are on the order of EEG frequencies $\sim 10^2$ sec⁻¹, equivalent to $\nu = \omega/(2\pi) = 16$ cps (Hz). These mesoscopic propagation velocities permit processing over several minicolumns $\sim 10^{-1}$ cm, simultaneous with the processing of mesoscopic interactions over tens of centimeters via association fibers with propagation velocities $\sim 600\text{--}900$ cm/sec. I.e., both intraregional and interregional information processing can occur within $\sim 10^{-1}$ sec.

ASA Fits of CMI

These momenta indicators should be considered as supplemental to other clinical indicators. This is how they are being used in financial trading systems. The CMI are more sensitive measures of neocortical activity than other invariants such as the energy density, effectively the square of the CMI, or the information which also effectively is in terms of the square of the CMI. Neither the energy or the information give details of the components as do the CMI. EEG is measuring a quite oscillatory system and the relative signs of such activity are quite important. Each set of results is presented with 6 figures, labeled as [{alcoholic|control}, {stimulus 1|match|no-match}, subject, {potential|momenta}], where match or no-match was performed for stimulus 2 after 3.2 sec of a presentation of stimulus 1. For each subjects run, after fitting 28 parameters with ASA, epoch by epoch averages are developed of the raw data and of the multivariate SMNI canonical momenta. Below is a comparison between an alcoholic and control subject under the match paradigm.

Chaos in SMNI

Duffing Analog

Some aspects of EEG can be approximately cast as a model of chaos, the Duffing oscillator.

$$\begin{aligned} \dot{x} &= f(x, t) , \\ f &= -\alpha \dot{x} - \omega_0^2 x + B \cos t . \end{aligned}$$

This can be recast as

$$\begin{aligned} \dot{x} &= y , \\ \dot{y} &= f(x, t) , \\ f &= -\alpha y - \omega_0^2 x + B \cos t . \end{aligned}$$

Note that this is equivalent to a 3-dimensional autonomous set of equations, e.g., replacing $\cos t$ by $\cos z$, defining $\dot{z} = \beta$, and taking β to be some constant.

We studied a model embedding this deterministic Duffing system in moderate noise, e.g., as exists in such models as SMNI. Independent Gaussian-Markovian (“white”) noise is added to both \dot{x} and \dot{y} , η_i^j , where the variables are represented by $i = \{x, y\}$ and the noise terms are represented by $j = \{1, 2\}$,

$$\begin{aligned} \dot{x} &= y + \hat{g}_x^1 \eta_1 , \\ \dot{y} &= f(x, t) + \hat{g}_y^2 \eta_2 , \\ \langle \eta^j(t) \rangle_\eta &= 0 , \\ \langle \eta^j(t), \eta^{j'}(t') \rangle_\eta &= \delta^{jj'} \delta(t - t') . \end{aligned}$$

In this study, we take moderate noise and simply set $\hat{g}_i^j = 1.0 \delta_i^j$.

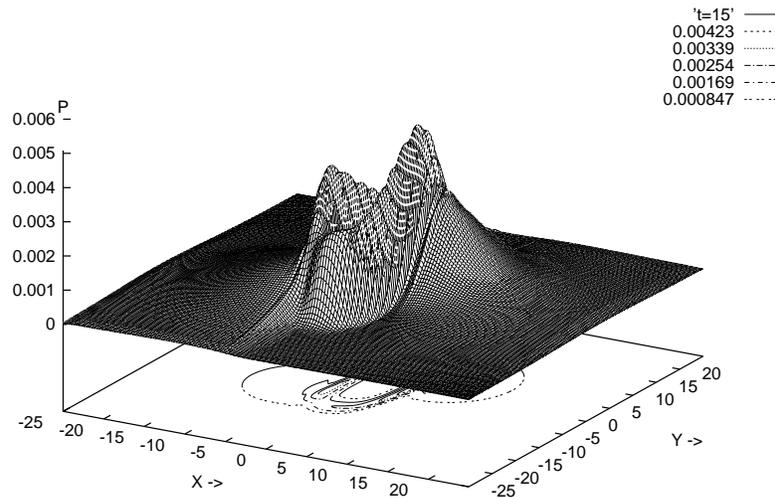
The equivalent short-time conditional probability distribution P , in terms of its Lagrangian L , corresponding to these Langevin rate-equations is

$$\begin{aligned} P[x, y; t + \Delta t | x, y, t] &= \frac{1}{(2\pi\Delta t)(\hat{g}^{11} \hat{g}^{22})} \exp(-L\Delta t) , \\ L &= \frac{(\dot{x} - y)^2}{2(\hat{g}^{11})} + \frac{(\dot{y} - f)^2}{2(\hat{g}^{22})} . \end{aligned}$$

Noise Washes Out Chaos/PATHINT

No differences were seen in the stochastic system, comparing regions of Duffing parameters that give rise to chaotic and non-chaotic solutions. More calculations must be performed for longer durations to draw more definitive conclusions.

Path Integral Evolution of Non-Chaotic Stochastic Duffing Oscillator



Path Integral Evolution of Chaotic Stochastic Duffing Oscillator

