Trading Markets With Canonical Momenta and Adaptive Simulated Annealing

Lester Ingber

ANALYSIS VERSUS INTUITION

Too often the management of complex systems is ill-served by not utilizing the best tools available. For example, requirements set by decision-makers often are not formulated in the same language as constructs formulated by powerful mathematical formalisms, and so the products of analyses are not properly or maximally utilized, even if and when they come close to faithfully representing the powerful intuitions they are supposed to model. In turn, even powerful mathematical constructs are ill-served, especially when dealing with multivariate nonlinear complex systems, when these formalisms are butchered into quasi-linear approximations to satisfy constraints of numerical algorithms familiar to particular analysts, but which tend to destroy the power of the intuitive constructs developed by decision-makers. These problems are present in many disciplines, including trading in financial markets. In this context, we can consider the trader as the decision maker on the nature of market data, sometimes also carrying the additional role of his or her own analyst.

For at least a large class of systems, these problems can be bypassed by using a blend of an intuitive and powerful mathematical-physics formalism to generate "canonical momenta" indicators (CMI), which are used by AI-type rule-based models of management of complex systems. Typically, both the formalism generating the CMI and the rule-based models have quite nonlinear constructs, and they must be "trained" or fit to data subsequent to testing on "out-of-sample" data, before they can be used effectively for "real-time" production runs. To handle these fits of nonlinear models of real-world data, a generic powerful optimization code, Adaptive Simulated Annealing (ASA), has been developed.

The starting point for this mathematical-physics approach starts with an economic justification for a statistical mechanics of financial markets (SMFM), modeling some financial systems as sets of coupled stochastic differential equations, e.g., as a function of short-term and long-term interest rates. I have demonstrated how some modern calculus developed by other mathematical physicists in the late 1970's can be used to study such systems. Over a decade ago, I proposed that inefficiencies in markets might be understood by such a SMFM approach. However, it wasn't until I developed some powerful tools of global optimization, ASA, that I was able to numerically cope with the nonlinear algebra presented

presented by SMFM. Recently, I have demonstrated that ASA and this calculus can be used to precisely articulate and calculate momenta of these systems, thereby permitting accurate momenta indicators to be derived. This is only half the story, as these indicators must be integrated into trading systems, and recent papers have demonstrated as well how this can be accomplished.

In the context presented here, of trading on financial markets, ASA is used recursively in both the "inner-shell" mathematical-physics model and in the "outer-shell" AI-type model. In the following sections, a brief and simplified description of CMI is given, ASA is discussed, and an application to trading is presented on cash and futures S&P 500 markets. In the conclusion, some future directions are given for finance as well as to other disciplines.

CANONICAL MOMENTA INDICATORS (CMI)

The concept of momentum is at least intuitively appreciated by all traders. Many traders use some algorithm to calculate the momenta of markets they are trading, e.g., perhaps to use as supplemental indicators to confirm other indicators to act on trades.

Markets increasingly are becoming inter-dependent, effectively defining a larger collective multivariate market. Many traders account for such circumstances by at least following indicators of other markets in addition to those they are explicitly trading. Clearly, it would be beneficial to have accurate measures of such inter-dependencies, beyond statistical correlations, to have indicators that measure the *importance* of inter-dependencies of the dynamic evolution of the markets. However, it also would be useful if such information could be presented in an understandable intuitive manner, without altering any detailed content. "Canonical momenta" can satisfy this wish-list, and a detailed application to trading is described below.

Consider a stochastic differential equation (SDE) like

$$\dot{x} = \frac{x(t + dt) - x(t)}{dt} = f(t) + g(t)\eta(t)$$

where $\eta(t)$ is "white" (Wiener) noise. Here, $f = f(t)$, the "drift," and $g = g(t)$, the square root of the "diffusion," are written in the Itô representation, a favorite of economists. When f and g are not constant, it turns out to be quite important just where in the interval dt these are defined. When f and g are defined at the midpoint of t and $t + dt$, this is the Stratonovich representation, where the standard calculus holds. In this simple one-dimensional example, g is just the standard deviation, but in more than one dimension, g^2 becomes the covariance matrix. It turns out that this is the inverse-metric of the space as well, and it enters into the calculation of the "canonical momenta." In its simplest form, with x representing the log of price and with zero drift and constant diffusion, this is the simple Brownian motion model of random markets first proposed by Bachelier in 1900. That first simple pioneering model directly led to the "efficient market hypothesis," whereby it seemed that no one could guarantee profiting from trading, any better than one can predict the results of coin-tossing. When nonlinear structures are possible, as they now seem to be, it becomes possible to devise winning strategies, at least until all of one's competitors are using the same strategies as well.

There is another mathematically equivalent representation to the SDE representation, the Fokker-Planck partial differential equation (PDE) for p:

$$\frac{\partial p}{\partial t} = -\frac{\partial (f\ p)}{\partial x} + \frac{1}{2}\frac{\partial^2 (g^2 p)}{(\partial x)^2}$$

This is a Schrödinger-type equation, and the methods developed for statistical mechanical systems in the late 1970's are quite similar to techniques first explored for looking at quantum gravity in the 1950's. The point is that the covariance matrix g^2 enters the second partial derivative, which warps x-space. The famous Black-Scholes model for calculating options pricing can be cast as such a one-dimensional PDE with simple diffusions and drifts, where the independent random variable is essentially a measure of the short-term interest rate, e.g., as might be calculated from 90-day T-bills.

This SDE can be written in yet a third, not so well-known mathematically equivalent representation, as a conditional short-time probability distribution

$$p[x(t + dt)|x(t)] = (2\pi\ dt\ g^2)^{-1/2} \exp(-dt\ L)$$

where the "Lagrangian" L is defined by

$$L = \frac{(\dot{x} - f)^2}{2g^2}$$

The long-time evolution of p is given by the "path integral," sometimes called the Chapman-Kolmogorov equation. As finally detailed in the late 1970's, in the Stratonovich representation, L becomes the Feynman Lagrangian, and many more terms appear in L for more than one dimension when f and g are not constant; an induced Riemannian geometry becomes explicit.

The momentum is

$$\frac{\partial L}{\partial \dot{x}} = \frac{(\dot{x} - f)}{g^2}$$

If we just let f be zero, we see that L is just the "kinetic energy" in terms of "velocity" $\dot{x}$ and "mass" g^{-2}. The momentum is mass × velocity. There is a nice chapter 19 in volume III of *The Feynman Lectures in Physics*, which shows how the more well-known principle of "Force = mass × acceleration," where acceleration is the time-rate of change of momentum, is derived from the Lagrangian.

When we go to two and more variables, e.g., indexed by i, and permit the functions f_i and g_i to be dependent nonlinearly on these variables, the algebraic complexity soars. Nevertheless, still CMI can be straightforwardly derived from the Lagrangian, and the intuitive sense of CMI is retained as their being accurate measures of flows to and from evolving steady states of the system.

The calculation of the long-time evolution of these distributions most often defies any algebraic solution, and special techniques must be utilized. This is required, for example, to calculate many kinds of financial instruments, e.g., bond prices, options, derivatives, etc. People have developed numerical algorithms for each representation, i.e., for the SDE, PDE and the Lagrangian probability representations. Methods to treat the latter are usually developed around the path-integral formalism. For example, I have developed an arbitrary (subject to CPU resources!) n-dimensional code, PATHINT, based on a one-dimensional code developed by other physicists in 1983, to accurately handle the evolution of a large class of nonlinear multivariate systems, and have applied it to problems in several disciplines.

The Lagrangian representation was used to great advantage by myself and collaborators in calculating the term structure of a two-dimensional generalization of the Black-Scholes model, as a function of the short-term and long-term interest rates, the latter derived from 30-year U.S. government bonds. The ASA and PATHINT techniques promise superior predictions of interest rates and of pricing securities, but in this paper I will emphasize a different finance project that has been the subject of some recent publications, after giving a relatively non-technical description of simulated annealing.

ADAPTIVE SIMULATED ANNEALING (ASA)

In order to deal with fitting parameters or exploring sensitivities of variables, as models of systems have become more sophisticated in describing complex behavior, it has become increasingly important to retain and respect the nonlinearities inherent in these models, as they are indeed present in the complex systems they model.

Conceptual Foundations of ASA

Simulated annealing (SA) was developed in 1983 to deal with highly nonlinear problems, as an extension of a Monte-Carlo importance-sampling technique developed in 1953 for chemical physics problems. 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.

In 1984, it was established that SA possessed a proof that, by carefully controlling the rates of cooling of temperatures, it could statistically find the best minimum, e.g., the lowest valley of our example above. This was good news for people trying to solve hard problems which could not be solved by other algorithms. The bad news was that the guarantee was only good if they were willing to run SA forever. In 1987, a method of fast annealing (FA) was developed, which permitted lowering the temperature exponentially faster, thereby statistically guaranteeing that the minimum could be found in some finite time. However, that time still could be quite long. Shortly thereafter, in 1987 I developed Very Fast Simulated Reannealing (VFSR), now called Adaptive Simulated Annealing (ASA), which is exponentially faster than FA. ASA is available at no charge from my archive. It is used world-wide across many disciplines, and the feedback of many users regularly scrutinizing the source code ensures the soundness of the code as it becomes more flexible and powerful.

The “elementary” discussion of SA given here really is not so elementary. I have answered thousands of queries on ASA, and most often I find it most helpful to examine the direct output of ASA — not the algebra, logic, or data that embody a particular cost function — to tune the code based on how it is “bouncing” around.

I often hear people worry about over-fitting their parameters, of getting “too good a fit.” However, the problem usually is that the underlying model is not a very good description of their system, and so different degrees of fitting can give quite different results. When general nonlinear structures are permitted to enter models, and when they can be properly processed by global optimization techniques as ASA, then the issue of fitting becomes more an issue of developing the best possible model.

ASA Applications to Finance

An article in the *Wall Street Journal* in 1993 brought ASA to the attention of the finance community, and it now is used regularly in many financial institutions. A few examples can be mentioned here. The asa_papers file in my archive references several financial and economics projects. For example, some economists use ASA to fit models of manufacturing capacity, labor tolerance, and interest rates. I have published papers using ASA to fit two-variable interest-rate models (coupled long-term and short-term interest rates) to several years of bond data. I have consulted for a large bank that used ASA to fit a class of features of a set of complex derivatives to similar features of a simpler portfolio, so they could conveniently trade on the simpler portfolio. I have consulted for several traders, using ASA to find optimal parameters of trading models, e.g., parameters of moving-average indicators. Below, I give another example of using ASA for a trading model.

NONLINEAR MULTIVARIABLE TRADING

CMI and ASA have been blended together to form a simple trading code, TRD. An example was published in 1996 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. (See Figure 1.)

In this first study, calculated in 1991, 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.

The model contains an inner-shell and an outer-shell, both of which need to be optimally fit to data. The inner shell develops CMI, as discussed above, as trading indicators. Many traders use such indicators as price, volume, etc., to trade, but here I wanted to explore the use of CMI to see if a “true” quantitative measure of momenta could be used. This required that the CMI be fit to data, e.g., to define quantities that themselves are functions of price. The cost function for the CMI is directly related to the Lagrangian.

The outer shell is the set of trading rules, defined as moving averages of the momenta indicators over various sized windows. This is pretty much like many simple trading rules, but here ASA is used to find the optimal sizes of the windows and of the thresholds triggering trading actions. Here, the thresholds are in terms of the CMI.

The CMI are fit to a year’s worth of data, but they are continually fine-tuned within the widest moving window used in the outer shell. The cost function for the trading rules is the net profit over a year of data. Over the course of a year, every day a trading decision is made on the CMI, but only after the CMI are tuned using optimization over the widest window. This defines the need for recursive optimization.

Inner-Shell CMI

In the late 1980's, I developed a practical maximum likelihood numerical tool for fitting parameters in nonlinear Lagrangians using VFSR/ASA, eventually applying these techniques to financial problems. For the S&P studies, I used a two-variable model of end-of-day cash and futures, $c(r)$ and $f(r)$, for day r, taking the variables to be ratios between days, e.g., $C(r) = c(r)/c(r-1)$, $F(r) = f(r)/f(r-1)$. These ratio-variables were used to develop coupled SDEs,

$$\dot{C} = f_C^C C + f_C^F F + g_1^C \eta^1 + g_2^C \eta^2$$

$$\dot{F} = f_F^C C + f_F^F F + g_1^F \eta^1 + g_2^F \eta^2$$

where all eight f and g parameters were taken to be constants, and the two η's were independent sources of Gaussian-Markovian noise. This set of SDEs were recast into a Lagrangian representation to define a single cost function, whose parameters were fit by ASA to data. This seems to be the simplest example that can illustrate the techniques, that requires ASA (other quasi-local algorithms get stuck in local minima), and that has the power to deal with the actual data. Any higher degree of nonlinearity for a larger set of variables can easily be coded and does not impede the running of the code, albeit such changes could produce new systems requiring different tunings of ASA. The parameters in the model were fit to one year of data. From the fit Lagrangian, a zeroth order approximate of the CMI are derived.

Outer-Shell TRD

A simple outer-shell AI-type model for trading was defined for the TRD code. A wide and a narrow window were defined, whose widths were parameters of TRD. These windows defined epochs over which moving averages of CMI were calculated for both the C and F variables. For each window, a short and a long "threshold" parameter were defined. If the CMI of both C and F were above the thresholds in both the wide and narrow windows, then a long trade was executed or maintained for the futures market. Similarly, if the CMI of C and F fell below the negative of these threshold parameters in the two windows, a short trade was executed or maintained. Otherwise, no trade was executed.

Thus, the six parameters of the outer-shell were the widths of the two windows and the two threshold parameters for each of the two variables. Each day, the CMI were determined by an inner-shell optimization: Over the length of the wide window, using the zeroth-order prior fit as a first guess, two of the CMI parameters were refit to the data in the window. At first, ASA was used recursively to establish the best fit, but it was determined for this system that only small perturbations of the CMI were regularly required, and so thereafter a faster quasi-local code was used.

Stepping through the trading decisions each trading day of a year's data determined the yearly net profit/loss as the single value of the outer-shell cost function. ASA then importance-sampled the CMI parameter space to determine the largest net profit, determining the final CMI parameters in the training set.

The CMI parameter values in TRD were then used to trade for an out-of-sample year of data. The inner-shell optimization was performed each day as before. Details of this successful procedure are given in the technical papers. Reprints are available via WWW or FTP from my archive.

CMI Utility

In addition to CMI being a natural coordinate system to study dynamically evolving multivariate systems, the value of the CMI has several noteworthy aspects as used in TRD. Although only one variable, the futures S&P 500, was actually traded (the code can accommodate trading on multiple markets), the multivariable coupling to the cash market entered in three important ways: (1) The inner-shell fits were to the coupled system, requiring a global optimization of all parameters in both markets to define the time evolution of the futures market. (2) The canonical momenta for the futures market is in terms of the partial derivative of the full Lagrangian; the dependency on the cash market enters both as a function of the relative value of the off-diagonal to diagonal terms in the metric (the inverse of the covariance matrix), as well as a contribution to the drifts and diffusions from this market. (3) The canonical momenta of both markets were used as technical indicators for trading the futures market. While it is common for traders to look at information in markets other than those they are actually trading, this particular kind of indicator also has the feature of including such information in a more detailed manner.

FUTURE DIRECTIONS

It certainly seems to be true that new techniques of quantitative analysis are now proving and will prove to be more powerful than popular techniques of AI-type technical analysis such as we now see embedded in off-the-shelf trading systems. However, I also believe that the experiences and intuitions built into these techniques of technical analysis are invaluable to complement such quantitative analysis. Above, I have given a simple TRD example of the utility of marrying quantitative analysis and popular techniques of technical analysis, especially when the two systems are considered as a synergetic system and recursive optimization is performed to ensure that they work together maximally.

There are at least two categories of future projects. The first category is to test TRD on more markets, and at other temporal scales, e.g., using data collected per tick, minute, day, week, month, etc.

The second category of future projects is to apply these techniques to other disciplines, with the expectation of getting more feedback to improve these approaches for all the systems so studied. For example, in a twist of the usual technology transfer, of applying techniques from academia to finance, I have started a project applying this methodology to EEG diagnoses, performing recursive ASA optimization of “canonical momenta” indicators of subject’s/patient’s EEG (using a neuroscience model I have developed over the past 15 years), nested in parameterized customized clinician’s rules.

SUGGESTED READING

R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison Wesley, Reading, MA, 1963).

L. Ingber, ‘‘Statistical mechanics of nonlinear nonequilibrium financial markets,’’ *Math. Modelling* **5**, 343-361 (1984). [URL http://www.ingber.com/markets84_statmech.ps.Z]

L. Ingber, ‘‘Statistical mechanical aids to calculating term structure models,’’ *Phys. Rev. A* **42**, 7057-7064 (1990). [URL http://www.ingber.com/markets90_interest.ps.Z]

L. Ingber, ‘‘Adaptive Simulated Annealing (ASA),’’ [http://www.ingber.com/ASA-shar, ASA-shar.Z, ASA.tar.Z, ASA.tar.gz, ASA.zip], Lester Ingber Research, McLean, VA, 1993.

L. Ingber, ‘‘Canonical momenta indicators of financial markets and neocortical EEG,’’ in *International Conference on Neural Information Processing (ICONIP’96)* (Springer, New York, 1996), p. 777-784. [Invited paper to the 1996 International Conference on Neural Information Processing (ICONIP’96), Hong Kong, 24-27 September 1996. URL http://www.ingber.com/markets96_momenta.ps.Z]

L. Ingber, ‘‘Statistical mechanics of nonlinear nonequilibrium financial markets: Applications to optimized trading,’’ *Mathl. Computer Modelling* **23**, 101-121 (1996). [URL http://www.ingber.com/markets96_trading.ps.Z]

L. Ingber, ‘‘Statistical mechanics of neocortical interactions: Applications of canonical momenta indicators to electroencephalography,’’ (submitted) (1997). [URL http://www.ingber.com/smni97_cmi.ps.Z]

S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, ‘‘Optimization by simulated annealing,’’ *Science* **220**, 671-680 (1983).

F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Dordrecht, The Netherlands, 1982).

Prof. Lester Ingber can be contacted by writing to Lester Ingber Research, P.O. Box 857, McLean, VA 22101, or by electronic mail to ingber@ingber.com. The latest ASA code and other reprints can be retrieved via WWW at URL http://www.ingber.com/, as well as by anonymous FTP from ftp.ingber.com.

Figure 1. S&P 500 cash and futures prices.