Variational Quantum Monte Carlo: Algorithm and Application to Helium

Ricky Nave

April 2020

Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse. –Alan Sokal

Quantum many-body physics is a field rich in both interesting physics and computational difficulty. Here we briefly lay out the variational quantum monte carlo (VQMC) method, starting with a short reminder of the classical Monte Carlo and the Metropolis-Hastings algorithm, and then moving on to a concrete application of VQMC to estimating the ground state energy of Helium.

1 Monte Carlo

In many areas of physics, statistics, and machine learning we are confronted with the need to evaluate the expectation of some function of a random variable:

$$E = E_\pi[f(x)] = \int f(x)\pi(x)dx \tag{1}$$

If this is a high-dimensional integral but the variance of the random variable isn't too high, we can efficiently approximate this integral by taking i.i.d samples $X^{(k)}$ from the distribution π:

$$E \approx \widehat{E_M} = \frac{1}{M}\sum_{k=1}^{M} f(X^{(k)}) \tag{2}$$

This is an unbiased estimator with asymptotic variance[1] $Var_\pi[f(x)]/M$ according to the Central Limit Theorem (CLT).

In many situations, we run into issues with the procedure just described because the distribution π is of the form $\pi(x) = q(x)/Z$ where Z (known as the partition function) is impractically difficult to compute. In such cases, we may choose to run a Markov chain monte carlo (MCMC) simulation instead. The idea is to choose a starting point and then generate a sequence of random samples via a Markov chain. Of course, to be useful, this Markov chain must incorporate enough information about the distribution π to ensure that eq. 2 is still an estimator for the integral we want to estimate. If the function q is practical to compute, a common choice is to use the Metropolis-Hastings algorithm to generate our samples.

[1] Of course, it is a bit suspect to say we can constrain the convergence of one variable (the estimator) in terms of another variable which is only guaranteed to be accurate in the asymptotic limit. Nevertheless, Monte Carlo algorithms almost never come with any hard guarantees, and collective historical experience suggests that making do with these kind of almost-circular analyses is better than nothing.

1.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is a way to generate a particular MCMC based only on relative probabilities so that the expensive partition function computation is avoided entirely. In MH algorithm, we have a transition proposal probability kernel T and a transition acceptance probability A, which are related as:

$$A(x \to x') = \min\left\{1,\ \frac{T(x \to x')}{T(x' \to x)}\frac{\pi(x')}{\pi(x)}\right\} \tag{3}$$

Unfortunately, there are two complications with this approach: (1) the choice of initial position biases $\widehat{E_M}$ (initialization bias), and (2) samples generated by this MCMC are correlated and hence the asymptotic variance of the estimator $\widehat{E_M}$ is no longer that of the CLT.

The good news is that under some standard assumptions on π and T, we are guaranteed that for $X^{(0)}$ chosen according to any starting distribution, we are guaranteed to have $X^{(k)}$ converge in distribution to π as $k \to \infty$ and that $\widehat{E_M}$ is an asymptotically unbiased estimator for E as $M \to \infty$ (see [1] or [2], ch. 13 for more precision).

The bad news is that finding provable bounds on the characteristic time needed to reduce initialization bias (known as the exponential autocorrelation time, denoted τ_{exp}) or on the discount factor needed to compensate for the reduced effectiveness of correlated samples (known as the integrated autocorrelation time for the integrand f, denoted $\tau_{int,f}$) are both nearly always impossible [1]. One of the major reasons for these difficulties is that it is extremely difficult to theoretically prove anythng about metastability for general integrands. Often, Monte Carlo simulations will be run with several different chains. This can be practically beneficial in reducing the likelihood of getting trapped near metastable solutions, but the technique provides zero theoretical guarantees in general.

It is very important to at least try estimating τ_{exp} and $\tau_{int,f}$ for any serious Monte Carlo so that you can estimate statistical error bounds even if those error bounds are not as theoretically guaranteed as we would like. In this short project, however, I will not go into this analysis since I merely want to showcase a simple application of variational quantum monte carlo and since a century of experimental work has provided a very accurate ground truth value for the ground state energy of helium (78.98 eV, or 2.902 hartrees).

2 Variational Quantum Monte Carlo

Suppose we are given a quantum many-body Hamiltonian $\hat{H}$ and any (not necessarily normalized) wavefunction $\psi(x)$, where $x \in X$ carries all spatial and spin degrees of freedom. If we wish to calculate the energy of the state ψ according to the Hamiltonian $\hat{H}$, we note:

$$\begin{aligned} E &= \int \psi^*(x)(H\psi)(x)dx \\ &= \int \frac{\psi^*(x)(H\psi)(x)}{|\psi(x)|^2}\frac{|\psi(x)|^2}{\int|\psi(y)|^2dy}dx \\ &= \int \mathrm{Re}\left(\frac{\psi^*(x)(H\psi)(x)}{|\psi(x)|^2}\right)\frac{|\psi(x)|^2}{\int|\psi(y)|^2dy}dx \\ &= \int E_L(x)\pi(x)dx \end{aligned} \tag{4}$$

If ψ vanishes at any point x, the proper interpretation of the intregrand is zero. Here $\pi(x)$ is the probability density

probability density:

$$\pi(x) = \frac{|\psi(x)|^2}{\int|\psi(x)|^2dx} \tag{5}$$

and $E_L(x)$ is interpreted as the local energy at configuration x:

$$E_L(x) = \mathrm{Re}\left(\frac{\psi^*(x)(H\psi)(x)}{|\psi(x)|^2}\right) \tag{6}$$

Note that we can take the real part in the integral in eq. 4 without loss of information only because $\hat{H}$ is Hermitian.

While explicit calculation of the energy E is often intractable, we can approximate it via dynamic Monte Carlo sampling using the MH algorithm according to the estimator in eq. 2.

3 Helium

In Hartree atomic units and using the Born-Oppenheimer approximation of infinite nuclear mass, the Hamiltonian of helium is

$$\hat{H} = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}} \tag{7}$$

where $r_i = \|\vec{r_i}\|_2$ is the position vector of electron i, $r_{12} = \|\vec{r_1} - \vec{r_2}\|_2$ is the interelectron distance, and $\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$ is the Laplacian for electron i.

Remark. The Hamiltonian acts on the configuration space $\mathbb{R}^6$

3.1 Variational Quantum Monte Carlo (Helium)

Consider the following trial wavefunction:

$$\psi_T(v_1, v_2; \alpha, \beta) = \exp\left(-\alpha(r_1 + r_2)\right) \exp\left(\frac{r_{12}}{2(1+\beta r_{12})}\right) \qquad v_1, v_2 \in \mathbb{R}^3 \tag{8}$$

where $v_i = (x_i, y_i, z_i) \in \mathbb{R}^3$, $r_i = \|v_i\|_2$, $r_{12} = \|v_1 - v_2\|_2$, and $\alpha, \beta > 0$ are variational parameters.[2]

We will need the partial Laplacians $\Delta_i \psi_T$, $i = 1, 2$, in order to apply our algorithm, and we do that calculation here. Let $\psi_A = \exp\left(-\alpha(r_1 + r_2)\right)$, $\psi_B = \exp\left(\frac{r_{12}}{2(1+\beta r_{12})}\right)$. A few useful partial derivatives are

$$\begin{aligned}
\frac{\partial \psi_A}{\partial x_i} &= -\frac{\alpha x_i}{r_i}\psi_A \\
\frac{\partial \psi_B}{\partial x_i} &= \psi_B \left(\frac{1}{2(1+\beta r_{12})^2}\right)\left(\frac{x_i - x_{(i+1)\bmod 2}}{r_{12}}\right) \\
\frac{\partial^2 \psi_A}{\partial x_i^2} &= \left(-\frac{\alpha}{r_i} + \frac{\alpha x_i^2}{r_i^3} + \frac{\alpha^2 x_i^2}{r_i^2}\right)\psi_A \\
&= \frac{\alpha}{r_i}\left(-1 + \frac{x_i^2}{r_i}\left(\alpha + \frac{1}{r_i}\right)\right)\psi_A \\
\frac{\partial^2 \psi_B}{\partial x_i^2} &= \psi_B \left(\frac{1}{2(1+\beta r_{12})^2}\right)\left[\left(\frac{x_i - x_{(i+1)\bmod 2)}}{r_{12}}\right)^2 \left(\frac{1}{2(1+\beta r_{12})^2} - \frac{2\beta}{1+\beta r_{12}} - \frac{1}{r_{12}}\right) + \frac{1}{r_{12}}\right]
\end{aligned} \tag{9}$$

[2]The fact that this wavefunction isn't antisymmetric with respect to the spatial coordinates of the two electrons is okay since the ground state of helium is assumed to be a singlet with antisymmetry in the spin of the two electrons. Justification for the singlet ground state assumption is typically provided by a physical argument or a perturbation theory argument.

The trial wavefunction then has second derivatives

$$\begin{aligned}\frac{\partial^2 \psi_T}{\partial x_i^2} &= \frac{\partial}{\partial x_i}\left(\frac{\partial \psi_A}{\partial x_i}\psi_B + \psi_A \frac{\partial \psi_B}{\partial x_i}\right) \\ &= \frac{\partial^2 \psi_A}{\partial x_i^2}\psi_B + 2\left(\frac{\partial \psi_A}{\partial x_i}\frac{\partial \psi_B}{\partial x_i}\right) + \psi_A \frac{\partial^2 \psi_B}{\partial x_i^2}\end{aligned} \tag{10}$$

Exactly analogous equations hold for y_i and z_i due to the rotational invariance of ψ_A and ψ_B. We can represent the Laplacian action on the trial wavefunction by

$$\Delta_i \psi_T = (\Delta_i \psi_A)\psi_B + (\Delta_i \psi_B)\psi_A + 2(\nabla_i \psi_A) \cdot (\nabla_i \psi_B) \tag{11}$$

Our proposal probability kernel T will be such that at each time step, there is a uniform probability for each electron to have a proposed move in the cube $[-\delta, \delta]^3$.

3.2 Computational Results

With $\delta = 0.5$, the acceptance rate ranged from around 1% or less when $\alpha = 10$ to 100% when $\alpha = 0$ and about 86% when $\alpha = 0.204$. These numbers were estimated using several independent Markov chains with $M = 1000$ samples each and β valus ranging from 0 to 10. Interestingly, the effect of β seemed to have very little effect on the acceptance rate (a few percent at most in the range sampled) wheres the value of α had a strong influence on the acceptance rate.

With $\delta = 1.0$, the acceptance rate dropped to just below 100% for $\alpha = 1$ and dropped to around 0.1% for $\alpha = 10$ (using $M = 10,000$).

3.2.1 $M = 10,000$

First, I ran a naive (no compensation for initialization bias) search over a large grid with relatively few Monte Carlo steps to try to get a rough intuition of the energy landscape.

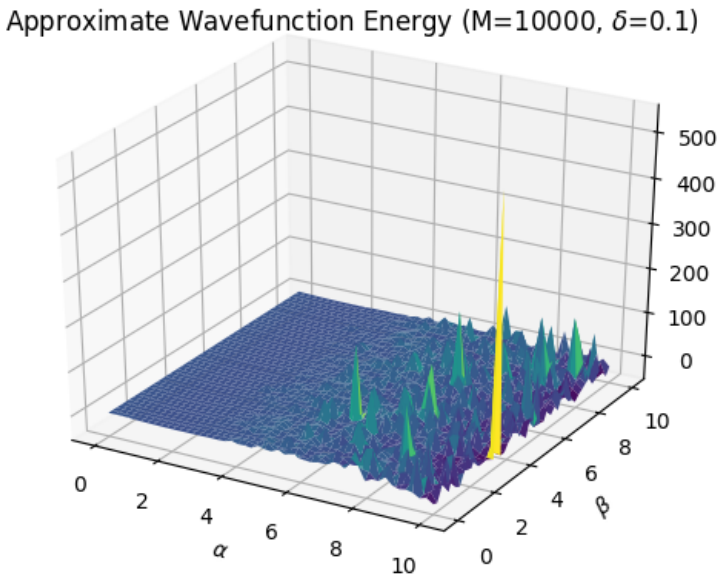


Figure 1: $M = 10,000$ Grid search over $\alpha \in [0, 10]$, $\beta \in [0, 10]$. Massive statistical fluctuations exist for larger values of α.

3.2.2 $M = 100,000$ with variable delta

Next, I refined my search to $\alpha \in [0, 5]$, $\beta \in [0, 5]$. I used a variable δ chosen to optimize the approximated acceptance rate being close to 0.5 from 10 randomly chosen options. I discovered via later exploration that I was not using anywhere near enough samples to estimate the acceptance rate accurately so the effect it had is unclear.

In this search, I identified a min energy $E = -2.894$ at $\alpha = 1.8367$ and $\beta = 0.204$. This is actually extremely close to the experimentally measured value of the helium ground state.

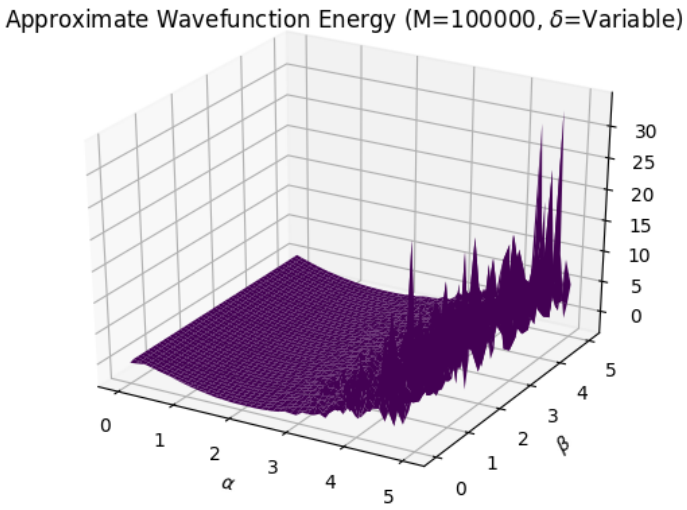


Figure 2: Grid search for $\alpha \in [0, 5]$, $\beta \in [0, 5]$ with $M = 100,000$ Monte Carlo samples used to estimate the energy at each grid point.

3.3 Refined Grid Search

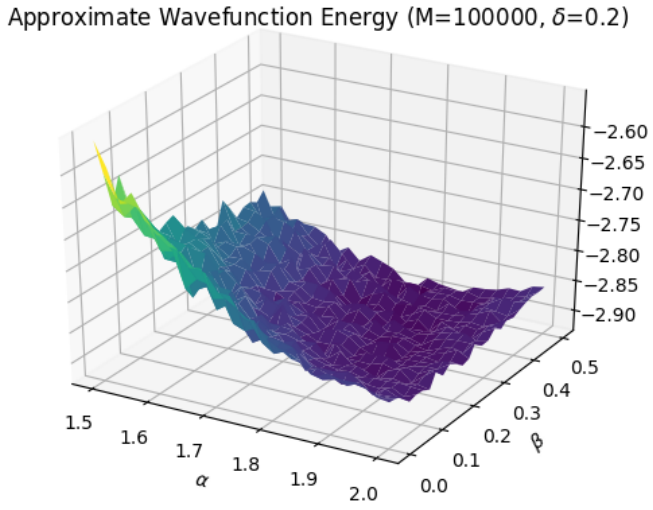


Figure 3: $M = 100,000$, $\delta = 0.2$, minimum energy $E = -2.926$ with optimal parameters $\alpha = 1.792$, $\beta = 0.458$. The search was over a 25 x 25 grid, a from 1.5 to 2, b from 0 to 0.5.

Here you can see the still visible statistical fluctuation, indicating the need for a more thorough convergence analysis. Note that despite this simulation being over a smaller region of parameter space, chance decreed that this simulation actually yield a minimum energy which is a worse estimate of the ground state energy than our previous simulation.

3.3.1 Further refined grid search

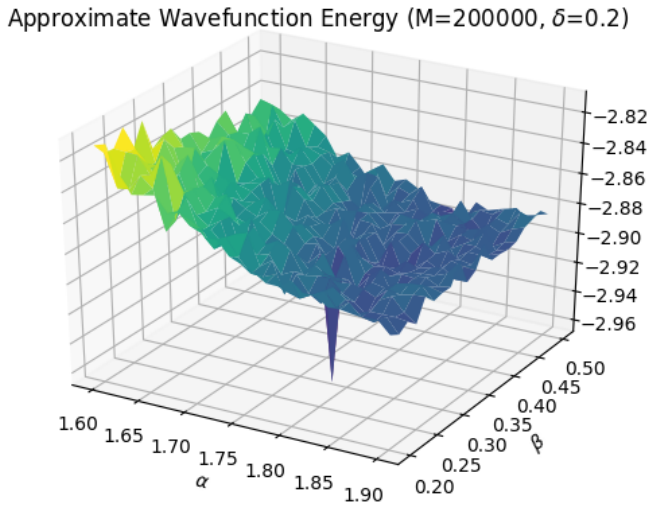


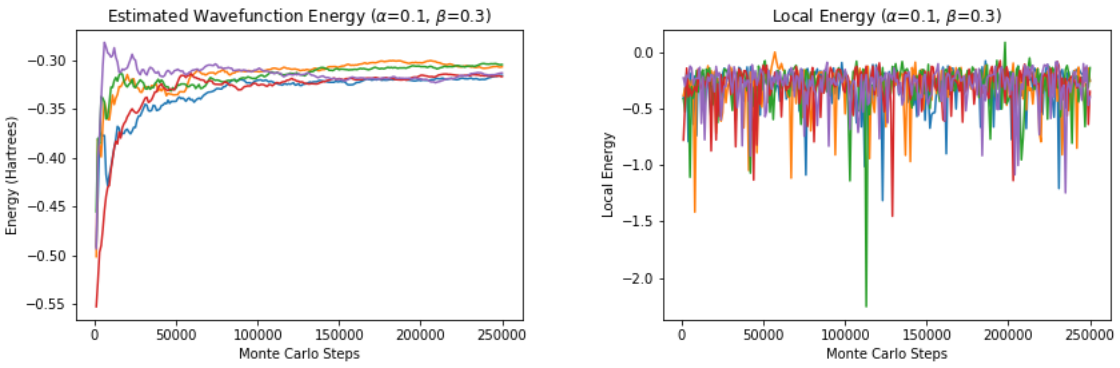
Figure 4: $M = 200,000$, α 1.6 to 1.9 with 25 bins, β 0.2 to 0.5 with 20 bins, $\delta = 0.2$. The results were minimum energy $E = -2.965$ at $\alpha = 1.7875$, $\beta = 0.2947$.

This simulation really highlights the dangers of Monte Carlo methods. The number of samples was doubled relative to the last simulation and yet the minimum energy estimate is three times farther away from the true ground state than the last simulation. This exemplifies that the proximity of our last two simulations to the true value were mere chance and that the true statistical uncertainty in our monte carlo simulations is actually still quite high, even with hundreds of thousands of samples per chain.

3.4 MCMC Convergence Analysis

In this section, we show some plots of wavefunction energy estimates and local energy values at various sampled points through different monte carlo chains to help give visual indications of chain convergence (or lack thereof). In some examples, we throw out a number of initial samples in an attempt to reduce initialization bias. However, since we haven't even tried to approximate τ_{exp}, the number of samples thrown out is totally unjustified and comes with no theoretical guarantee

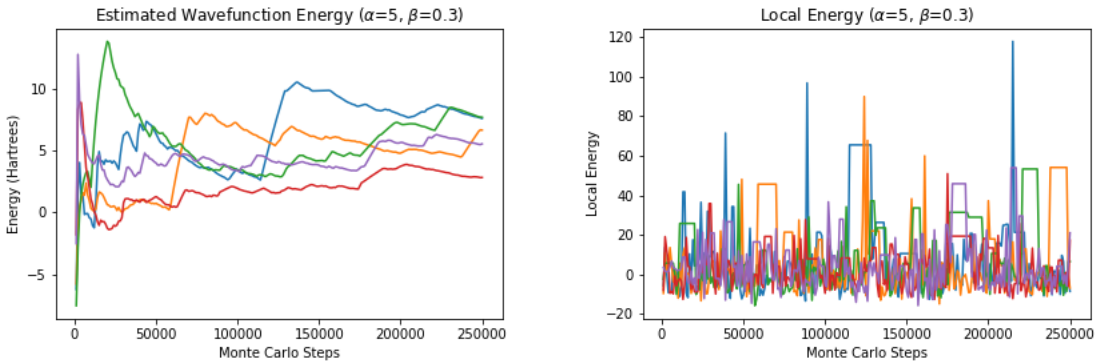
3.4.1 Without Compensating for Initialization Bias



(a) Caption

(b) Caption

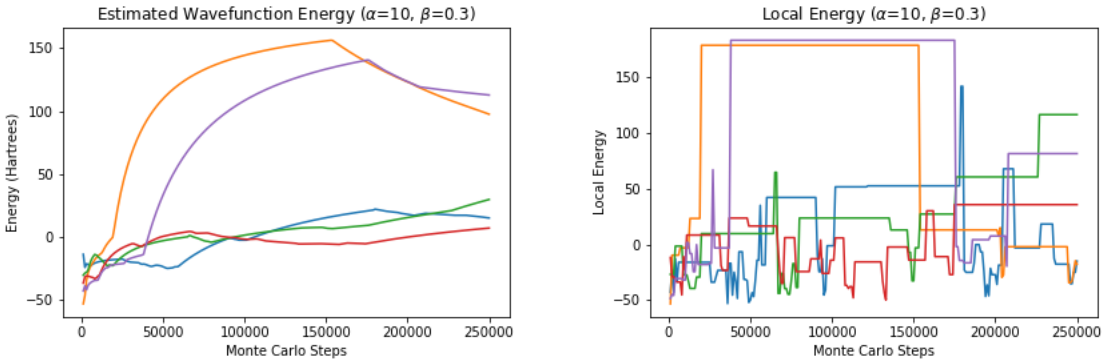
Figure 5: Five independent Metropolis-Hastings MCMC simulations (each shown in a different color), each with $M = 250,000$ steps, $\delta = 0.2$, $\alpha = 5$, and $\beta = 0.3$.



(a) Total wavefunction energy

(b) Sampled local energy

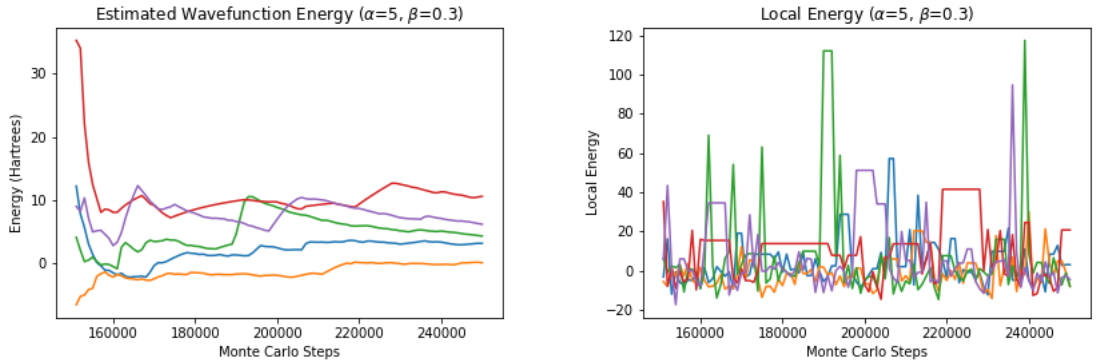
Figure 6: Five independent Metropolis-Hastings MCMC simulations (each shown in a different color), each with $M = 250,000$ steps, $\delta = 0.2$, $\alpha = 5$, and $\beta = 0.3$.



(a) Caption (b) Caption

Figure 7: Five independent Metropolis-Hastings MCMC simulations (each shown in a different color), each with $M = 250,000$ steps, $\delta = 0.2$, $\alpha = 10$, and $\beta = 0.3$.

3.4.2 With Compensation for Initialization Bias



(a) Caption (b) Caption

Figure 8: Five independent Metropolis-Hastings MCMC simulations (each shown in a different color), each with $M = 250,000$ steps, $\delta = 0.2$, $\alpha = 5$, and $\beta = 0.3$. Here, the first $150,000$ Monte Carlo steps were discarded based on visual inspection of figure 6(a)

References

[1] Alan Sokal. Monte carlo methods in statistical physics: foundations and algorithms, 1996.

[2] Richard Tweedie and Sean Meyn. *Mrkov Chains and Stochastic Stability.* Springer-Verlag, London, 1993.