APPENDIX

The Markov chain and its steady state

This appendix shows that the one-step Monte Carlo statics estimation algorithm produces a series of parameter vectors with the Gibbs probability distribution given by equation (3.10), when the algorithm is run at constant T for an infinite number of iterations. To obtain this result, I describe the algorithm as a Markov chain, and then show that the standard limit theorem (stated below) holds. The basic properties of Markov chains are sketched only briefly. Readers unfamiliar with the general theory should refer to Chapter 15 of Feller (1968), Kemeny and Snell (1960), or any other relevant text.

A.1 PRELIMINARIES

I refer in this appendix only to those Markov chains having a finite number of states. A chain is *irreducible* if every state can be reached (after an arbitrary number of iterations and with some positive probability) from every other state. Any state $\mathbf{x}_i$ is said to be *periodic* if the probabilities of recurrence $p_{ii}^{(n)}$ are non-zero only for some $n > 1$ and an integral multiple of n. Otherwise, $\mathbf{x}_i$ is termed *aperiodic*. The following theorem (see, e.g., Feller, 1968; or Kemeny and Snell, 1960) will be employed in the derivation of the main result:

If a Markov chain is irreducible and aperiodic, then the limits

$$U_j = \lim_{n \to \infty} p_{ij}^{(n)} \tag{A-1}$$

exist and are independent of the initial state $\mathbf{x}_i$. *The numbers* $\{U_j\}$ *uniquely satisfy*

$$U_j > 0 , \quad \sum U_j = 1 \tag{A-2}$$

and

$$U_j = \sum_i U_i \, p_{ij} \quad . \tag{A-3}$$

The *transition-probability matrix* $\mathbf{P}$ contains the transition probabilities $\{p_{ij}\}$. $\mathbf{P}$ is called a *stochastic matrix* because each $p_{ij} \geq 0$ and each row sums to unity. Each element of the *state-probability vector* $\mathbf{u}(n) = \{u_i(n)\}$ is the probability of being in state $\mathbf{x}_i$ at time n. If $\mathbf{u}$ is a row vector, the one-step transition from $\mathbf{u}(0)$ to $\mathbf{u}(1)$ can be represented by the equation

$$\mathbf{u}(1) = \mathbf{u}(0)\,\mathbf{P} \quad ,$$

where $\mathbf{u}(0)$ contains the initial state probabilities. The state probability vector after n steps is

$$\mathbf{u}(n) = \mathbf{u}(0)\,\mathbf{P}^n \quad .$$

For chains that satisfy the theorem above, we can define

$$\mathbf{U} \equiv \lim_{n \to \infty} \mathbf{u}(0)\,\mathbf{P}^n \quad .$$

$\mathbf{U}$ is the steady-state, or equilibrium vector of the Markov chain. Rewriting equation (A-3) shows explicitly that $\mathbf{U}$ is an eigenvector of $\mathbf{P}$, with eigenvalue 1:

$$\mathbf{U} = \mathbf{U}\,\mathbf{P} \quad .$$

This relationship will frequently be used below.

To show that the statics algorithm is a Markov chain with Gibbs equilibrium probabilities, I shall first specify the structure of the transition-probability matrix $\mathbf{P}$. It will then be shown that $\mathbf{P}$ is both irreducible and aperiodic, after which it will be proven that the steady-state distribution is Gibbs.

A.2 THE TRANSITION-PROBABILITY MATRIX

The one-step statics estimation algorithm sequentially "visits" each parameter X_m (a shot or receiver static) and changes the parameter's value by choosing a random number from the probability distribution in equation (3.7). Only N distinct values for each parameter are allowed (this limitation is just an upper and lower bound, sometimes called a "shift limit"). One iteration is completed after each parameter has undergone a (possible) transition.

The transition-probability matrix $\mathbf{P}(m)$ directs each change in the value of X_m. Because there are M parameters that can each assume any of N values, there are N^M possible states of the system, and $\mathbf{P}(m)$ is an N^M by N^M matrix. Each row contains only N non-zero elements, because only N new states are directly accessible from any given state. There are M distinct transition-probability matrices $\mathbf{P}(1), \cdots, \mathbf{P}(M)$ for each of the M parameters, respectively. The matrix of transition probabilities that directs the changes from one complete iteration to another is given by the product of the M matrices:

$$\mathbf{P} = \mathbf{P}(1)\,\mathbf{P}(2)\cdots\mathbf{P}(M) \ . \tag{A-4}$$

To show that the Gibbs distribution $\{\Pi_j\}$ in equation (3.10) is the equilibrium vector for $\mathbf{P}$, it will first be shown that $\mathbf{P}$ represents an irreducible and aperiodic Markov chain. It will then be shown that $\{\Pi_j\}$ is an eigenvector with eigenvalue 1 for each $\mathbf{P}(m)$ and thus also for $\mathbf{P}$. When the theorem stated above is used, it can then be concluded that the steady state of the Markov chain is the Gibbs distribution of equation (3.10).

A.3 IRREDUCIBILITY

A transition-probability matrix is irreducible if every state can be reached from every other state with some positive probability over some arbitrary time. A Markov chain must be irreducible if it is to have an equilibrium distribution; otherwise the system may fall into a state from which it can not enter some other states, and the system can no longer be independent of its initial configuration.

A set of states in which all members of the set are reachable (over time and with positive probability) from all other members of the set is called an *ergodic class*. Following the argument used by Fosdick (1963) for a two-dimensional (Ising) lattice, I now show that $\mathbf{P}$ is irreducible by showing that all possible states belong to the same ergodic class.

For each transition of X_m, only N possible values are allowed. This transition produces one of N new (or repeated) states. Prior to the transition of X_m, N^{M-1} different configurations of the other $M-1$ parameters are possible. When X_m changes, the other $M-1$ parameters remain constant, and the new state is one of the N possible states that contain the pre-existing configuration of the other $M-1$ parameters. These N states are all accessible to each other via the transition matrix $\mathbf{P}(m)$, but they are inaccessible from any other state using this transition matrix. Thus each $\mathbf{P}(m)$ partitions the N^M states into N^{M-1} ergodic classes, and each of these ergodic classes is a disjoint set of N states.

Now suppose that one is interested in the transition of X_m followed by the transition of X_{m+1}. Both $\mathbf{P}(m)$ and $\mathbf{P}(m+1)$ partition the states into N^{M-1} ergodic classes. Each ergodic class in $\mathbf{P}(m)$ is different from each ergodic class in $\mathbf{P}(m+1)$, but each state is contained in an ergodic class defined by both matrices. For example, $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ may be the states of an ergodic class in $\mathbf{P}(m)$. $\mathbf{P}(m+1)$ defines N ergodic classes (among others) that each contain one of these N states; the N ergodic classes are

$$\{\mathbf{x}_1, \mathbf{x}_{N+1}, \cdots, \mathbf{x}_{2N}\}$$
$$\{\mathbf{x}_2, \mathbf{x}_{2N+1}, \cdots, \mathbf{x}_{3N}\}$$
$$\vdots$$
$$\{\mathbf{x}_N, \mathbf{x}_{N^2+1}, \cdots, \mathbf{x}_{N^2+N}\}.$$

When the product $\mathbf{P}(m)\,\mathbf{P}(m+1)$ is taken, each of these N ergodic classes are linked via the class $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ in $\mathbf{P}(m)$; thus $\mathbf{P}(m)\,\mathbf{P}(m+1)$ partitions the N^M states into N^{M-2} ergodic classes. By induction, one sees that the product (A-4) links all states into a single ergodic class of N^M states. Thus, because every state can be reached from every other state, $\mathbf{P}$ is an irreducible transition-probability matrix.

A.4 APERIODICITY

A state $\mathbf{x}_i$ is periodic if its probability of recurrence, $p_{ii}^{(n)}$, is non-zero only for some $n > 1$ and an integral multiple of n. Otherwise, the state is aperiodic. All states of a system must be aperiodic if there is to be a limiting equilibrium distribution; otherwise the system will not exhibit a distribution of states that is independent of time. To show that all N^M states of the system under consideration are aperiodic, I will demonstrate that $p_{ii} = p_{ii}^{(1)}$ is non-zero for every i.

Each row of each $\mathbf{P}(m)$ contains N, and only N, non-zero elements. One of these elements is always on the diagonal, because there is always some positive probability of retaining the current value of the mth parameter. For simplicity, consider a two-parameter system with transition matrix $\mathbf{Q} = \mathbf{Q}(1)\,\mathbf{Q}(2)$. Explicitly stated, the matrix product is

$$q_{ij} = \sum_k q_{ik}(1)\, q_{kj}(2)$$

and the diagonal element is

$$q_{ii} = \sum_k q_{ik}(1)\, q_{ki}(2) \ .$$

Because $q_{ii}(1) > 0$ and $q_{ii}(2) > 0$, and all the other elements are non-negative, then

$q_{ii} > 0$. By induction, the same conclusion is true for the diagonal elements of the transition matrix for an M-parameter system. Thus, because each diagonal element of $\mathbf{P}$ is positive, all N^M states are aperiodic.

A.5 THE STEADY STATE

I now explicitly construct $\mathbf{P}(m)$ for the one-step heat-bath method. Unlike the two-step Metropolis approach, the random changes of a parameter's value that this method performs do *not* depend on the parameter's current value. The method can be characterized as a Markov chain, however, because the current values of the *neighboring* parameters (those shot and receiver statics within a cablelength) determine the conditional probability distribution for any particular parameter.

The transition probabilities that govern the m th parameter are given by

$$p_{ij}(m) = \begin{cases} \dfrac{\exp\{-E(\mathbf{x}_j)/T\}}{\sum\limits_{j \in A_i(m)} \exp\{-E(\mathbf{x}_j)/T\}} & j \in A_i(m) \\ 0 & \text{otherwise} \end{cases} \tag{A-5}$$

where $A_i(m)$ is the set of N indices j such that $\mathbf{x}_j = \mathbf{x}_i$ everywhere except (possibly) at x_m. It will be shown that

$$\Pi_j = \sum_{i \in A} \Pi_i \; p_{ij}(m) \;, \tag{A-6}$$

where Π_j is given by the Gibbs probability distribution of equation (3.10) and $A = \{1, 2, \cdots, N^M\}$. This relationship will establish that $\{\Pi_j\}$ is an eigenvector with eigenvalue 1 for each $\mathbf{P}(m)$.

Equation (A-5) says that, for a given i, $p_{ij}(m)$ is non-zero only if $j \in A_i(m)$. Likewise, for a given j, $p_{ij}(m)$ is non-zero only if $i \in A_j(m)$. Thus one may write

$$\sum_{i \in A} \Pi_i \; p_{ij}(m) = \sum_{i \in A_j(m)} \Pi_i \; p_{ij}(m) \;.$$

Substituting equation (3.10) for Π_i and equation (A-5) for $p_{ij}(m)$ in the right-hand side above, one obtains

$$\sum_{i \in \mathbf{A}} \Pi_i \; p_{ij}(m) = \sum_{i \in A_j(m)} \frac{\exp\{-E(\mathbf{x}_i)/T\}}{\sum\limits_{i \in \mathbf{A}} \exp\{-E(\mathbf{x}_i)/T\}} \; \frac{\exp\{-E(\mathbf{x}_j)/T\}}{\sum\limits_{j \in A_i(m)} \exp\{-E(\mathbf{x}_j)/T\}} \;.$$

Reversing the order of the numerators and bringing the outside summation inside then yields

$$\sum_{i \in \mathbf{A}} \Pi_i \; p_{ij}(m) \;=\; \frac{\exp\{-E(\mathbf{x}_j)/T\}}{\sum_{i \in \mathbf{A}} \exp\{-E(\mathbf{x}_i)/T\}} \; \frac{\sum_{i \in A_j(m)} \exp\{-E(\mathbf{x}_i)/T\}}{\sum_{j \in A_i(m)} \exp\{-E(\mathbf{x}_j)/T\}} \; .$$

Now note that when $j \in A_i(m)$, $A_j(m) = A_i(m)$. Thus the second term on the right-hand side cancels to yield

$$\sum_{i \in \mathbf{A}} \Pi_i \; p_{ij}(m) \;=\; \frac{\exp\{-E(\mathbf{x}_j)/T\}}{\sum_{i \in \mathbf{A}} \exp\{-E(\mathbf{x}_i)/T\}} \; ,$$

which is equal to Π_j. Thus $\{\Pi_j\}$ is an eigenvector with eigenvalue 1 for $\mathbf{P}(m)$.

Inspection of equation (A-4) shows readily that $\{\Pi_j\}$ is also an eigenvector with eigenvalue 1 for $\mathbf{P}$, and that $U_j = \Pi_j$ satisfies equations (A-2) and (A-3) of the limit theorem. Thus, because $\mathbf{P}$ is irreducible and aperiodic, $\{\Pi_j\}$ is the steady-state distribution of the Markov chain.

References

Aki, K. and Richards, P., 1980, Quantitative seismology: W. H. Freeman and Co.

Binder, K., Ed., 1984, Applications of the Monte Carlo method in statistical physics: Springer-Verlag.

Binder, K., Ed., 1979, Monte Carlo methods in statistical physics: Springer-Verlag.

Box, G.E.P., and Tiao, G.C., 1973, Bayesian inference in statistical analysis: Addison-Wesley Publishing Co.

Bracewell, R., 1978, The Fourier transform and its applications: McGraw-Hill Book Co.

Claerbout, J.F., 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co.

Claerbout, J.F., 1985, Imaging the Earth's interior: Blackwell Scientific Publications.

Creutz, M., 1984, Quarks, gluons, and lattices: Cambridge University Press.

Dobrin, M., 1976, Introduction to geophysical prospecting: McGraw-Hill Book Co.

Donoho, D.L., 1981, On minimum entropy deconvolution, *in* Findley, D., Ed., Applied time series analysis II: Academic Press, 565-608.

Feller, W., 1968, An introduction to probability theory and its applications: 3rd ed., v. 1, John Wiley and Sons.

Fosdick, L.D., 1963, Monte Carlo computations on the Ising lattice: Methods in Computational Physics, **1**, 245-280.

Fowler, P., 1985, Migration velocity analysis by optimization: linear theory: Stanford Exploration Project Rep. 44, 1-21.

Garey, M.R. and Johnson, D.S., 1979, Computers and intractability: W.H. Freeman and Co.

Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-6**, 721-741.

Hammersley, J.M. and Handscomb, D.C., 1964, Monte Carlo methods: Chapman and Hall.

Hinton, G. and Sejnowski, T., 1983, Optimal perceptual inference: Proceedings IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, 448-453.

Hockney, R.W., and Jesshope, C.R., 1981, Parallel computers: Adam Hilger Ltd.

Johnson, J.H., Yancey, M.S., and D'Onfro, P.S., 1983, Application of a statics solution, Wyoming Overthrust, *in* Bally, A.W., Ed., Seismic expression of structural styles: v. 3, The American Association of Petroleum Geologists, 3.4.1-35 - 3.4.1-40.

Kemeny, J. and Snell, J., 1960, Finite Markov chains: D. Van Nostrand and Co., Inc.

Kindermann, R. and Snell, J.L., 1980, Markov random fields and their applications: American Mathematical Society.

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

Lin, S., 1975, Heuristic programming as an aid to network design: Networks, **5**, 33-43.

Lines, L.R. and Treitel, S., 1984, Tutorial: a review of least-squares inversion and its application to geophysical problems: Geophysical Prospecting, **32**, 159-186.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of state calculations by fast computing machines: Journal of Chemical Physics, **21**, 1087-1092.

Moussouris, J., 1974, Gibbs and Markov random systems with constraints: Journal of Statistical Physics, **10**, 11-33.

Parker, R.L., 1977, Understanding inverse theory: Ann. Rev. Earth Planet. Sci., **5**, 35-64.

Rebbi, C., 1984, Monte Carlo calculations in lattice gauge theories: *in* Binder, K., Ed., Applications of the Monte Carlo method: Springer-Verlag, 277-298.

Reif, F., 1965, Fundamentals of statistical and thermal physics: McGraw-Hill Book Co.

Ronen, S. and Claerbout, J., 1985, Residual statics estimation by stack-power maximization, Geophysics, **50**, no. 12 (in press).

Rothman, D., 1985, Velocity estimation by simulated annealing: problems and prospects: Stanford Exploration Project Rep. 44, 47-54.

Sheriff, R.E., 1973, Encyclopedic dictionary of exploration geophysics: Society of Exploration Geophysicists.

Stanley, H.E., 1971, Introduction to phase transitions and critical phenomena: Oxford University Press.

Sword, C., 1983, The generalized frequency-dependent surface-consistent problem: Stanford Exploration Project Rep. 35, 19-42.

Taner, M.T., Koehler, F., and Alhilali, K.A., 1974, Estimation and correction of near-surface time anomalies: Geophysics, **39**, 441-463.

Tarantola, A. and Valette, B., 1982, Inverse problems = Quest for information: Journal of Geophysics, **50**, 159-170.

Toldi, J., 1985, Velocity analysis without picking: Ph.D. thesis, Stanford University, Stanford Exploration Project Rep. 43.

Tribolet, J., 1979, Seismic applications of homomorphic signal processing: Prentice Hall, Inc.

Vichniac, G.Y., 1984, Simulating physics with cellular automata: *in* Farmer, D., Toffoli, T., and Wolfram, S., Eds., Cellular automata: North Holland Physics Publishing, 96-116.

Waters, K.H., 1981, Reflection seismology: John Wiley and Sons.

Wiggins, R., Larner, K., and Wisecup, D., 1976, Residual statics analysis as a general linear inverse problem: Geophysics, **41**, 922-938.

Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21-35.

Wiggins, R., 1985, Entropy guided deconvolution: preprint.