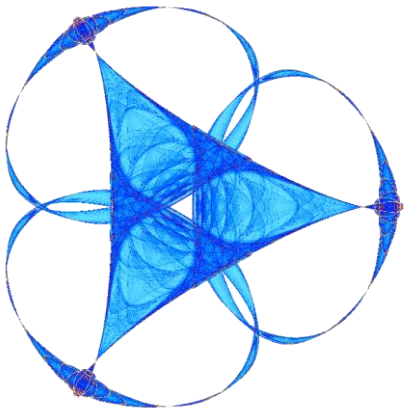
RELEVANT ASPECTS IN THE MULTIPLE MINIMAL METHOD IN BIOTECHNOLOGY

By

Ezio Marchi

IMA Preprint Series #2438

(September 2014)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS
UNIVERSITY OF MINNESOTA
400 Lind Hall
207 Church Street S.E.
Minneapolis, Minnesota 55455-0436
Phone: 612-624-6066 Fax: 612-626-7370
URL: http://www.ima.umn.edu

Relevant aspects in the multiple minimal method in biotechnology

By

Ezio Marchi *.

*Emeritus Professor of the National University of San Luis. Argentina. Founder and First Director of the IMASL, UNSL-CONICET.
Email: eziomarchi1940@gmail.com

Introduction

The Monte Carlo Method introduced by Ulam and Von Neumman as powerfull aproximated is constantly used for the scientists, physicists, chemisists and others in order to studythe multiple minimum function.

The modern version of the Monte Carlo method was invented in the late 1940s by Stanislaw Ulam, while he was working on nuclear weapons projects at the Los Alamos National Laboratory. It was named by Nicholas Metropolis, after the Monte Carlo Casino, where Ulam's uncle often gambled. Immediately after Ulam's breakthrough, John von Neumann understood its importance and programmed the ENIAC computer to carry out Monte Carlo calculations.

A Monte Carlo-minimization method has been developed to overcome the multiple-minima problem. The Metropolis Monte Carlo sampling, assisted by energy minimization, surmounts intervening barriers in moving through successive discrete local minima in the multidimensional energy surface. The method has located the lowest-energy minimum thus far reported for the brain pentapeptide [Met5]enkephalin in the absence of water. Presumably it is the global minimum-energy structure. This supports the concept that protein folding may be a Markov process. In the presence of water, the molecules appear to exist as an ensemble of different conformations.[1]

The aplication of the method have been successful as for example in a new approach to the multiple-minima problem in protein folding is presented. It is assumed that the molecule is driven toward the native structure by three types of mechanism. The first one involves an optimization of the electrostatic interactions, whereby the molecule evolves toward conformations in which the charge distribution becomes energetically more favorable. The second mechanism involves a Monte Carlo–energy minimization approach, and the third one is a backtrack mechanism that acts in the opposite direction, increasing the energy—the third type of movement provides a means to perturb the molecule when it is trapped in a stable but energetically unfavorable local energy minimum. We describe the implementation of a model based on these mechanisms, and illustrates its effectiveness by computations on different arbitrary starting conformations of a terminally blocked 19-residue chain of poly(l-alanine) for which the global minimum apparently corresponds to the right-handed a-helix. In all cases, the global minimum was attained, even when the starting conformation was a left-handed a-helix. In the latter case, the trajectory of conformations passed through partially melted forms of the left-handed a-helix (because of electrostatic defects at the ends), and then through the formation of structures leading to the more stable right-handed a-helix.

The electrostatically driven Monte Carlo (EDMC) method has been greatly improved by adding a series of new features, including a procedure for cluster analysis of the accepted conformations. This information is used to guide the search for the global energy minimum. Alternative procedures for generating perturbed conformations to sample the conformational space were also included. These procedures enhance the efficiency of the method by generating a larger number of low-energy conformations.[2]

The improved EDMC method has been used to explore the conformational space of a 20-residue polypeptide chain whose sequence corresponds to the membrane-bound portion of melittin. The ECEPP/3 (Empirical Conformational Energy Program for Peptides) algorithm was used to describe the conformational energy of the chain. After an exhaustive search involving 14 independent runs, the lowest energy conformation (LEC) (-91.0 kcal/mol) of the entire study was encountered in four of the runs, while conformations higher in energy by no more than 1.8 kcal/mol were found in the remaining runs with the exception of one of them (run 8). The LEC is identical to the conformation found by J. Lee et all [3] as the lowest energy conformation obtained in their study using the conformational space annealing method. These results suggest that this conformation corresponds to the global energy minimum of the ECEPP/3 potential function for this specific sequence; it also appears to be the conformation of lowest free energy. [4]

An automatic procedure is proposed for reconstruction of a protein backbone from its Ca-trace; it is based on optimization of a simplified energy function of a peptide backbone, given its a-carbon trace. The energy is expressed as a sum of the energies of interaction between backbone peptide groups that are not neighbors in the sequence, the energies of local interactions within all amino acid residues, and a harmonic penalty function accounting for the conservation of standard bond angles. The energy of peptide group interactions is calculated using the assumption that each peptide group acts as a point dipole. For local interaction energy, use is made of a two-dimensional Fourier series expansion of the energies of model terminally blocked amino acid residues, calculated with the Empirical Conformational Energy Program for Peptides (ECEPP/3) force field in the angles ?(1) and ?(2) defining the rotation of peptide groups adjacent to a Ca carbon atom about the corresponding Ca … Ca virtual-bond axes. To explore all possible rotations of peptide groups within a fixed Ca-trace, a Monte Carlo search is carried out. The initial ? angles are calculated by aligning the dipoles of the peptide groups that are close in

space, subject to the condition of favorable local interactions. After the Monte Carlo search is accomplished with the simplified energy function, the energy of the structure is minimized with the ECEPP/3 force field, with imposition of distance constraints corresponding to the initial Ca-trace geometry. The procedure was tested on model a-helices and ß-sheets, as well as on the crystal structure of the immunoglobulin binding protein (PDB code: 1IGD, an a/ß protein). In all cases, complete backbone geometry was reconstructed with a root-mean-square (rms) deviation of 0.5 Å from the all-atom target structure. [5]

Then, we summarize the computational methods for sampling the conformational space of biomacromolecules. We discuss the methods applicable to find only lowest energy conformations (global minimization of the potential-energy function) and to generate canonical ensembles (canonical Monte Carlo method and canonical molecular dynamics method and their extensions). Special attention is devoted to the use of coarse-grained models that enable simulations to be enhanced by several orders of magnitude. [6]

In statistics and in statistical physics, the Metropolis–Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the distribution (i.e., to generate a histogram), or to compute an integral (such as an expected value). Metropolis–Hastings and other MCMC algorithms are generally used for sampling from multi-dimensional distributions, especially when the number of dimensions is high. For single-dimensional distributions, other methods are usually available (e.g. adaptive rejection sampling) that can directly return independent samples from the distribution, and are free from the problem of auto-correlated samples that is inherent in MCMC methods.

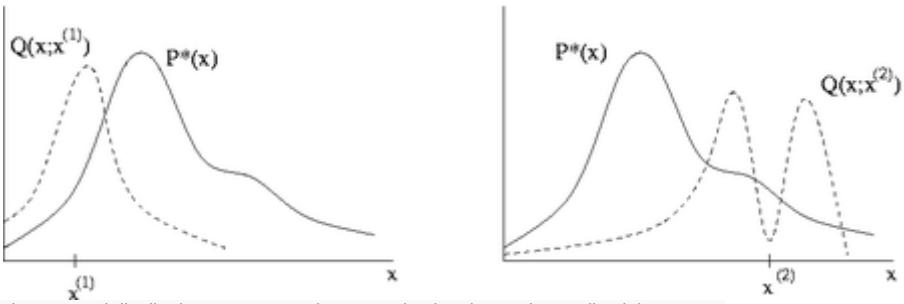
History

The algorithm was named after Nicholas Metropolis, who was an author along with Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller of the 1953 paper Equation of State Calculations by Fast Computing Machines which first proposed the algorithm for the specific case of the canonical ensemble;[7][8] and W. K. Hastings who extended it to the more general case in 1970.[9] There is controversy over the credit for discovery of the algorithm. Edward Teller states in his memoirs that the five authors of the 1953 paper worked together for "days (and nights)"[10]. M. Rosenbluth, in an oral history recorded shortly before his death credits E. Teller with posing the original problem, himself with solving it, and A.W. Rosenbluth (his wife) with programming the computer. According to M. Rosenbluth, neither Metropolis nor A.H. Teller participated in any way. Rosenbluth's account of events is supported by other contemporary recollections.[11]

The Metropolis–Hastings algorithm can draw samples from any probability distribution P(x), provided you can compute the value of a function f(x) which is proportional to the density of P. The lax requirement that f(x) should be merely proportional to the density, rather than exactly equal to it, makes the Metropolis–Hastings algorithm particularly useful, because calculating the necessary normalization factor is often extremely difficult in practice.

The Metropolis–Hastings algorithm works by generating a sequence of sample values in such a way that, as more and more sample values are produced, the distribution of values more closely approximates the desired distribution, P(x). These sample values are produced iteratively, with the distribution of the next sample being dependent only on the current sample value (thus making the sequence of samples into a Markov chain). Specifically, at each iteration, the algorithm picks a candidate for the next sample value based on the current sample value. Then, with some probability, the candidate is either accepted (in which case the candidate value is used in the next iteration) or rejected (in which case the candidate value is discarded, and current value is reused in the next iteration)-the probability of acceptance is determined by comparing the likelihoods of the current and candidate sample values with respect to the desired distribution P(x).

For the purpose of illustration, the Metropolis algorithm, a special case of the Metropolis–Hastings algorithm where the proposal function is symmetric, is described below.



The proposal distribution *Q* proposes the next point that the random walk might move to

Metropolis algorithm (symmetric proposal distribution)

Let f(x) be a function that is proportional to the desired probability distribution P(x).

Initialization: Choose an arbitrary point x0 to be the first sample, and choose an arbitrary probability density Q(x|y) which suggests a candidate for the next sample value x, given the previous sample value y. For the Metropolis algorithm, Q must be symmetric; in other words, it must satisfy Q(x|y) = Q(y|x). A usual choice is to let Q(x|y) be a Gaussian distribution centered at y, so that points closer to y are more likely to be visited next—making the sequence of samples into a random walk. The function Q is referred to as the proposal density or jumping distribution.
For each iteration t:
Generate a candidate x' for the next sample by picking from the distribution Q(x'|x_t).
Calculate the acceptance ratio a = f(x')/f(xt), which will be used to decide whether to accept or reject the candidate. Because f is proportional to the density of P, we have that a = f(x')/f(xt) = P(x')/P(xt).
If a = 1, then the candidate is more likely than xt; automatically accept the candidate by setting xt+1 = x'. Otherwise, accept the candidate with probability a; if the candidate is rejected, set xt+1 = xt, instead.
This algorithm proceeds by randomly attempting to move about the sample space, sometimes accepting the moves and sometimes remaining in place. Note that the acceptance ratio \alpha indicates how probable the new proposed sample is with respect to the current sample, according to the distribution \displaystyle P(x). If we attempt to move to a point that is more probable than the existing point (i.e. a point in a higher-density region of \displaystyle P(x)), we will always accept the move. However, if we attempt to move to a less probable point, we will sometimes reject the move, and the more the relative drop in probability, the more likely we are to reject the new point. Thus, we will tend to stay in (and return large numbers of samples from) high-density regions of \displaystyle P(x), while only occasionally visiting low-density regions. Intuitively, this is why this algorithm works, and returns samples that follow the desired distribution \displaystyle P(x).

Compared with an algorithm like adaptive rejection sampling that directly generates independent samples from a distribution, Metropolis–Hastings and other MCMC algorithms have a number of disadvantages:

The samples are correlated. Even though over the long term they do correctly follow \displaystyle P(x), a set of nearby samples will be correlated with each other and not correctly reflect the distribution. This means that if we want a set of independent samples, we have to throw away the majority of samples and only take every nth sample, for some value of n (typically determined by examining the auto-correlation between adjacent samples). Auto-correlation can be reduced by increasing the jumping width (the average size of a jump, which is related to the variance of the jumping distribution), but this will also increase the likelihood of rejection of the proposed jump. Too large or too small a jumping size will lead to a slow-mixing Markov chain, i.e. a highly correlated set of samples, so that a very large number of samples will be needed to get a reasonable estimate of any desired property of the distribution.
Although the Markov chain eventually converges to the desired distribution, the initial samples may follow a very different distribution, especially if the starting point is in a region of low density. As a result, a burn-in period is typically necessary, where an initial number of samples (e.g. the first 1,000 or so) are thrown away.
On the other hand, most simple rejection sampling methods suffer from the "curse of dimensionality", where the probability of rejection increases exponentially as a function of the number of dimensions. Metropolis–Hastings, along with other MCMC methods, do not have this problem to such a degree, and thus are often the only solutions available when the number of dimensions of the distribution to be sampled is high. As a result, MCMC methods are often the methods of choice for producing samples from hierarchical Bayesian models and other high-dimensional statistical models used nowadays in many disciplines.

In multivariate distributions, the classic Metropolis–Hastings algorithm as described above involves choosing a new multi-dimensional sample point. When the number of dimensions is high, finding the right jumping distribution to use can be difficult, as the different individual dimensions behave in very different ways, and the jumping width (see above) must be "just right" for all dimensions at once to avoid excessively slow mixing. An alternative approach that often works better in such situations, known as Gibbs sampling, involves choosing a new sample for each dimension separately from the others, rather than choosing a sample for all dimensions at once. This is especially applicable when the multivariate distribution is composed out of a set of individual random variables in which each variable is conditioned on only a small number of other variables, as is the case in most typical hierarchical models. The individual variables are then sampled one at a time, with each variable conditioned on the most recent values of all the others. Various algorithms can be used to choose these individual samples, depending on the exact form of the multivariate distribution: some possibilities are adaptive rejection sampling, a one-dimensional Metropolis–Hastings step, or slice sampling.[]

Adaptive rejection sampling

For many distributions, finding a proposal distribution that includes the given distribution without a lot of wasted space is difficult. An extension of rejection sampling that can be used to overcome this difficulty and efficiently sample from a large number of distributions (provided that they are log-concave, which is in fact the case for most of the common distributions) is known as adaptive rejection sampling. The basic idea is to model the proposal distribution using a set of piecewise exponential distributions (i.e. segments of one or more exponential distributions, attached end to end). This can be easier visualized in log space (i.e. by looking at the logarithm of the distribution). The logarithm of an exponential distribution is a straight line, and hence this method essentially involves enclosing the logarithm of the density in a series of line segments. This is the source of the log-concave

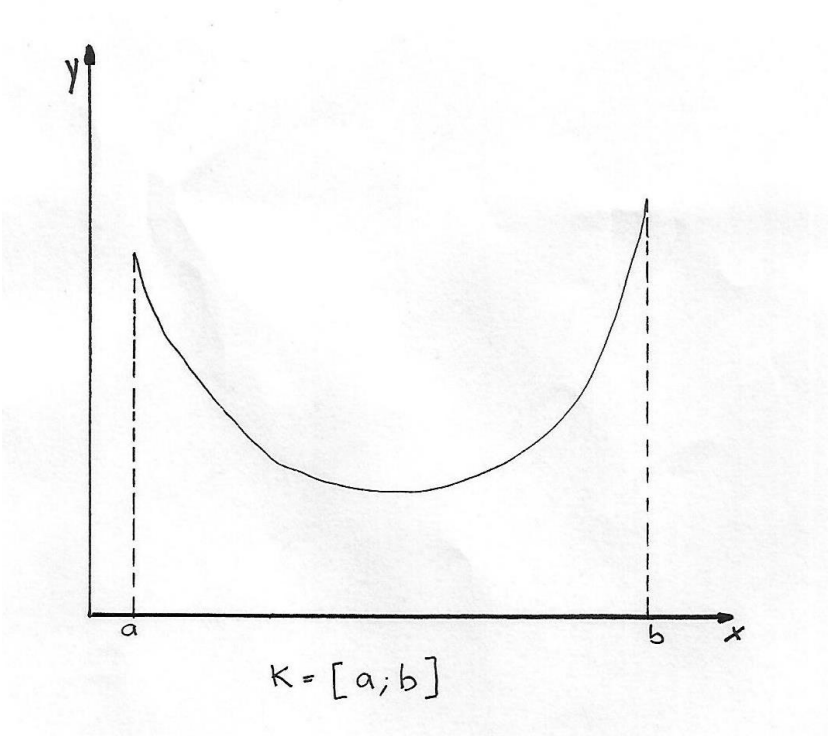
restriction: if a distribution is log-concave, then its logarithm is concave (shaped like an upside-down U), meaning that a line segment tangent to the curve will always pass over the curve. The method essentially involves successively determining an envelope of straight-line segments that approximates the logarithm better and better while still remaining above the curve, starting with a fixed number of segments (possibly just a single tangent line). Any time we choose a point that is rejected, we tighten the envelope with another line segment that is tangent to the curve at the point with the same x-coordinate as the chosen point.[17]

In principle relevant mathematical solution

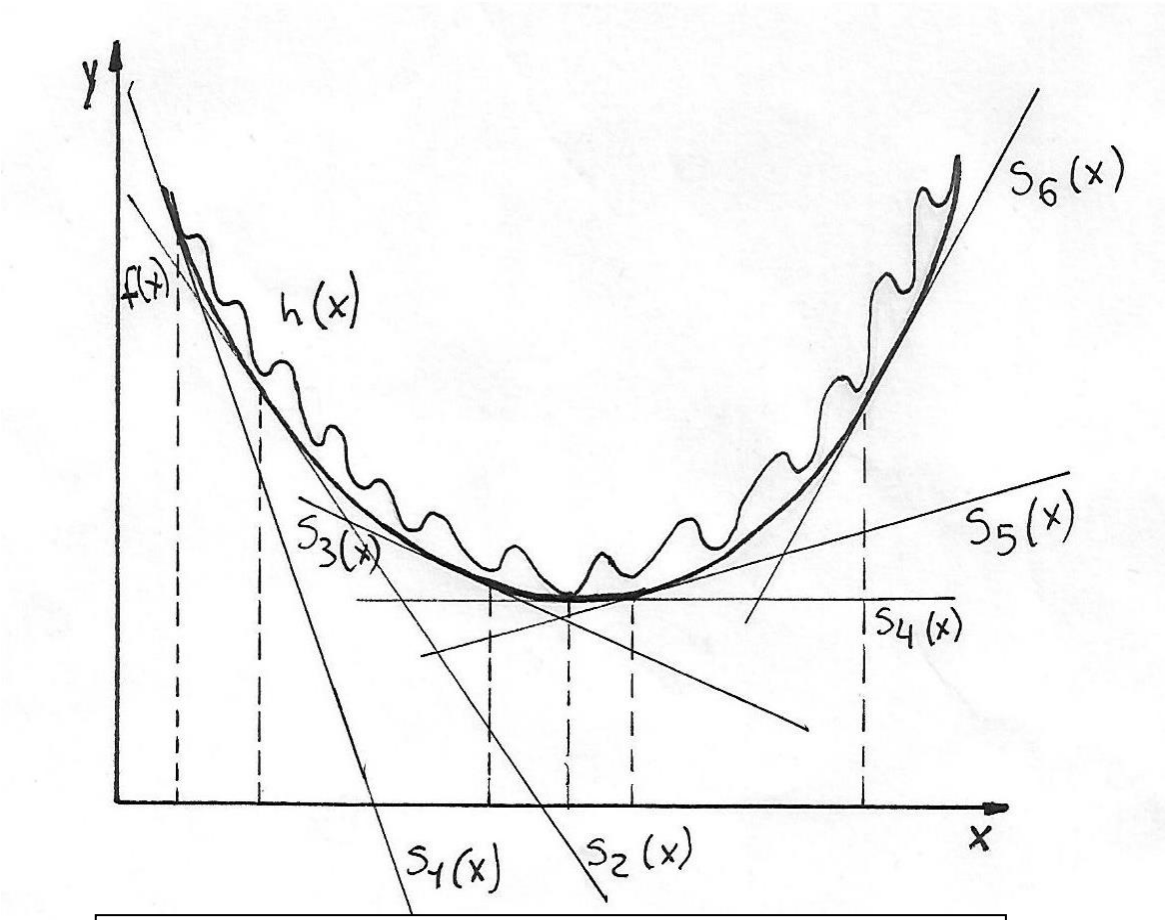
Wherever is the function considerate for an a physical, chemical, biochemical, etc. point of you by taking the envolvent see Rockafellar[13] of the care under consideration we can have another function with the same absolute minimum at the one before. Now the problem is to see if is easier to study or to compute the subsequence minimum of the original function or to construct the envolvent and to get immediately. This is an open question and is going to be study in the future for each or any real function particularly. We are going to explain this in by a graphic point of you. Remember that for mathematical way a real function has a domain and a range. It is real if the range and the real number. If the domain is a convex set: this is to say for any $0 \leq \lambda \leq 1$, and any x, y belonging to the domain then $\lambda x+(1-\lambda)y$, is an element of the domain. This can be generalized to a Euclidean space: R^N. A function defined on a non empty convex and closed [12] set $K \subseteq R^N$ it is call to be convex if for each x, y and $0 \leq \lambda \leq 1$ then

$$\lambda f(x)+(1-\lambda)f(y) \geq f(\lambda x+(1-\lambda)y)$$

It is said that is concave if –f is convex.



This is an example of a convex function which is very easy to see that has only a minimum value. It may happen that the places where the minimum is an archived are multiple but this is in all of them the function has the same value. In another way the set of minimum is a convex set. This is an example for a convex function defined on a compact set in the real but it is the same if the domain is a compact set in a Euclidean space.



Minimize f(x)(x∈K) =Minimize h(x)(x∈K)

A.M. Geoffrion, Perspectives on optimization, October 1971.[17]
Outer linearization of a convex fuction

In many areas of theeretical and applied mathematics when it is important to fine the absolute minimun of the function when the function is known or not. Generally analytic expression of the one can reach one minimun by some algorithm and then it is extremally difficult and there are may be attempt to pass to a better minimun. We solve the problem considering the envelope of all the linear or concave functions that the fuction under consideration.
Consider a compact convex and (closed) non empty set $K \subset R^N$, an a continuos fuction
$h: K \longrightarrow R$, where are all the real members then take the set of all the concave functions g

$$L(h)= \{g: g(x) \leq h(x)\ \forall\ x \in K\} = \{S: \text{linear and } S(x) \leq h(x)\ \forall\ x\}$$

Then take the $\sup_{g \in L(h)} g(x) = f(x) = \max_{g \in L(h)} g(x)$

It is clear that "h f" is concave, and it reaches the absolute minimum value of the fuction f.

It is clear that the minimun of "f" and "h" are the same see [13]Rockafellar. Next if we have an aproximation by linear function of the fuction "h" then if we search a minimun we have for the aproximation for the envelope "f" a convex polyhedron we can aplly the material present in the mathematical way by Rockafellar[15]. Moreover a new method for optimitazion can be or may be powerfull for the such posible aplication. See Marchi and Matons[16]. An alternative method would be the classical simplex method.

Final remarks

We would like to say that now we have solved mathematically the envolvent now we need to have a lot of study to see what is much better that is to say to try to apply many computation in order to get better local or absolute

minimal or to fine the envolvent. We strongly remark that the probem in the high dimension or many dimensions it is the same. This is a work for the future for the mathematicians and physicists, chemisists, etc.
In this days we have solve many many interesting problems related with the previous remarks. This is related with some aspects of the fundamental theory of convex and network flows and monotropic oprimization. In this area one of the worldwild leader is Professor R. T. Rockafellar.
Another important poin is to introduce the theory of Morse theory in this rich and important subject of quantum chemetry, byochemetry, ect.

Acknowledgements: We acknowledge the comment of Professor Adam Liwo and the finantial help of the Faculty of Physics, Mathematics and Natural Sciences of the UNSL, and the Institute for Mathematics and its Applications with funds provided by the National Science Foundation."

References

[1][Scheraga - Monte Carlo-minimization approach to the multiple-minima problem in protein folding – Proc Natl Acad Sci U S A. Oct 1987 ; 84(19): 6611–6615.]
[Ripoll, D. R. and Scheraga, H. A. (1988), On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte]
[Carlo method—tests on poly(L-alanine). Biopolymers, 27: 1283–1303. doi: 10.1002/bip.360270808]
[2][Ripoll, D. R., Liwo, A. and Scheraga, H. A. (1998), New developments of the electrostatically driven monte carlo method: Test on the membrane-bound portion of melittin. Biopolymers, 46: 117–126. doi: 10.1002/(SICI)1097-0282(199808)46:2<117::AID-BIP6>3.0.CO;2-P]
[3] [J. Lee, H.A. Scheraga, and S. Rackovsky [(1998) “Conformational Analysis of the 20-Residue Membrane-Bound Portion of Melittin by Conformational Space Annealing,” Biopolymers, Vol. 46, pp. 103–115]
[4] [© 1998 John Wiley & Sons, Inc. Biopoly 46: 117–126, 1998]
[5] [Kazmierkiewicz, R., Liwo, A. and Scheraga, H. A. (2002), Energy-based reconstruction of a protein backbone from its a-carbon trace by a Monte-Carlo method. J. Comput. Chem., 23: 715–723. doi: 10.1002/jcc.10068] [© 2002 Wiley Periodicals, Inc. J Comput Chem 7: 715–723, 2002]
[6][Computational techniques for efficient conformational sampling of proteins Adam Liwo1, Cezary Czaplewski1, Stanislaw Oldziej1, 2, Harold A Scheraga1, 2008.]
[7] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.(1953). "Equations of State Calculations by Fast Computing Machines".Journal of Chemical Physics 21 (6): 1087–1092.Bibcode:1953JChPh..21.1087M. doi:10.1063/1.1699114.
[8]Jump up^ Rosenthal, Jeffrey (March 2004). "W.K. Hastings, Statistician and Developer of the Metropolis-Hastings Algorithm". Retrieved 2009-06-02.
[9] Hastings, W.K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". Biometrika 57 (1): 97–109.doi:10.1093/biomet/57.1.97. JSTOR 2334940. Zbl 0219.65008.
[10] Teller, Edward. Memoirs: A Twentieth-Century Journey in Science and Politics. Perseus Publishing, 2001, p. 328
[11][J.E. Gubernatis (2005). "Marshall Rosenbluth and the Metropolis Algorithm". Physics of Plasmas 12 (5): 057303.Bibcode:2005PhPl...12e7303G. doi:10.1063/1.1887186.]
[12][A.M. Geoffrion, Perspectives on optimization, October 1971.]
[13][Rockafellar R.T. Convex analisis, Princeton University 1970.]
[14][A.M. Geoffrion, Perspectives on optimization, October 1971.]
[15][Rockafellar R.T. Network Flows and Monotropic optimization 1984 John Wiley & Sons.]
[16][Marchi E. and M.Matons The face projection in linear programmig. Journal Advance Mathematics an Aplication Vol. 2, 2012]
[17][Neal, Radford M. (2003). "Slice Sampling". Annals of Statistics 31 (3): 705–767. doi:10.1214/aos/1056562461. MR 1994729.Zbl 1051.65007.
[18] [Bishop, Christopher (2006). "11.4: Slice sampling". Pattern Recognition and Machine Learning. Springer. ISBN 0-387-31073-8.]
[19][.Milnor,L:MNorse Theory, Princeton Univers<ity Press, L970]
[20]Milnor, John (1963). *Morse Theory*. Princeton University Press. ISBN 0-691-08008-9. A classic advanced reference in mathematics and mathematical physics.
[21]Milnor, John (1965). Lectures on the h-Cobordism theorem.
[22] Frank H. Clarke “Optimization and nonsmooth analysis”
[23]Steven R. Lay “Convex sets and their applications”