

# SIMULATED ENTROPY AND GLOBAL MINIMIZATION

A.M. Sultan\* and A.B. Templeman\*\*

\*General Dynamics Land Systems, 38500 Mound Road, Sterling Heights, Michigan 48310, USA  
[sultanm@gdls.com](mailto:sultanm@gdls.com)

\*\*Department of Civil Engineering, University of Liverpool  
P.O. Box 147, Liverpool L69 3BX, UK  
[abt@liverpool.ac.uk](mailto:abt@liverpool.ac.uk)

**ABSTRACT:** Non-linear programming deals with the problem of optimizing an objective function in the presence of equality and inequality constraints. Most of the classical deterministic methods encounter two major problems: (i) the solution obtained is heavily dependent on the starting solution: and (ii) the methods often converge to inferior local optima. This paper examines the relationship between simulated annealing and Shannon's entropy with respect to minimization, which introduces for a new subject called simulated entropy (**SE**). As a consequence, two **SE** techniques are derived which provide a simple means of seeking the global minimum of constrained minimization problems.

## **INTRODUCTION**

Given some function  $F(X)$  and some constraint functions, the process of locating a minimum value of  $F(X)$  commences with no numerical information whatsoever. An initial point is then chosen and information is calculated about the objective and constraint functions, typically their numerical values and gradients at the design point. This numerical information is then used in some deterministic mathematical programming algorithm to infer where the next trial point should be placed, so as to get closer to the constrained optimum of the problem. The new trial generates more information from which another point is inferred and eventually the solution is reached by this process of gathering better and better information and using it in an inference-based algorithm.

Almost all such optimization algorithms use some form of geometrical inference to generate a sequence of improving trial points. The functions in the problem are interpreted as geometrical hyper-surfaces with contours, slopes and gradients. Actually, we never have sufficient information to be able to plot the geometry, except for the simplest problems, but it is convenient to imagine that these hyper-geometrical shapes exist because it helps us to visualize what a numerical search algorithm is doing. Our general knowledge of geometry can be used to develop solution strategies which can use in geometrical way any numerical information which is generated about the problem. We pretend for convenience, that everything in the problem before use has a totally deterministic geometric representation. However, it is equally valid to argue that the problem should be interpreted non-deterministically: if we have made evaluations of the objective and constraint functions at some point this represents concrete information, but the introduction of some geometrical prediction strategy to estimate where to place the next trial represents, according to Jaynes', "an arbitrary assumption of information which by hypothesis we do not have". Actually, we have no grounds for assuming that the particular problem we are solving will conform precisely to the geometrical strategy we have introduced. The use of any particular geometrical strategy in this situation is analogous to the use of a particular probability distribution to represent data from some random process. According to the **MEP** both assumptions are wrong and introduce bias.

The only correct strategy is to employ the **MEP** to infer where the next trial should be placed, using only the concrete data available.

It is important to realize that the geometrical interpretations placed upon optimization processes are only interpretations. There is nothing sacrosanct about them. It is perfectly reasonable to discard this deterministic, geometric interpretation of optimization and attempt to develop new optimization algorithm based upon totally non-geometrical, non-deterministic concepts or geometrical, non-deterministic concepts. To do so, one must understand the basics of statistical mechanics.

Statistical mechanics is study of the behavior of very large systems of interacting components, such as atoms in a fluid, in thermal equilibrium at a finite temperature. Suppose that the configuration of the system is identified with the set of spatial positions of the components. If the system is in thermal equilibrium at a given temperature T, then the probability  $\pi_T(s)$  that the system is in a given configuration s depends upon the energy E(s) of the configuration and follows the Boltzmann distribution:

$$\pi_T(s) = \frac{e^{\frac{-E(s)}{kT}}}{\sum_{w \in S} e^{\frac{-E(w)}{kT}}}$$

Where k is Boltzmann's constant and S is the set of all possible configurations. In studying systems of particles, one often seeks to determine the nature of the low energy states, for example, whether freezing produces crystalline or glassy solids. Very low energy configuration is not common, when considering the set of all configurations. However, at low temperatures they predominate, because of the nature of the Boltzmann distribution. To achieve low-energy configurations, it is not sufficient to simply lower the temperature. One can use an annealing process, where the temperature of the system is elevated, and then gradually lowered, spending enough time at each temperature to reach thermal equilibrium. If in sufficient time is spent at each temperature, especially near the freezing point, then the probability of attaining a very low energy configuration is greatly reduced. This is known as the annealing process.

Simulated Annealing (**SE**), a new class of algorithms was first introduced by the authors (Sultan, 1990), is a stochastic computational technique derived from statistical mechanics for finding near globally-minimum-cost solutions to large optimization problems. In general, finding the global minimum value of an objective function with many degrees of freedom subject to conflicting constraints is an NP- complete (no-deterministic polynomial) problem, since the objective function will tend to have many local minima. A procedure for solving hard optimization problems should sample value of the objective function in such a way to have a high probability of finding a near-optimal solution and should also lend itself to efficient implementation. Consider a discrete minimization problem with N feasible solutions numbered in ascending order of their objective function values (F; I= 1,.....,N). Iterative search methods specify a neighborhood  $N_i \in (1, \dots, N)$  for each solution, and proceed to its lowest valued neighbor, provided this

switch results in a strict improvement. The methods repeat this step until no improving neighbor can be found, i.e., until they reach a local optimum. The classical deterministic methods encounter two major problems:

- (1) the solution obtained is heavily dependent on the starting solution; and
- (2) the methods often converge to inferior local optima.

Simulated annealing methods randomize the procedure to overcome these problems, allowing for occasional switches that worsen the solution. This is equivalent to annealing in condensed matter physics where a solid in a heat bath is heated up by increasing the temperature of the heat bath to a maximum value at which all particles of the solid randomly arrange themselves in the liquid phase, followed by cooling through slowly lowering the temperature of the heat bath. In this way, all particles arrange themselves in the low energy ground state of a corresponding lattice, provided the maximum temperature is sufficiently high and the cooling is carried out sufficiently slowly. At each temperature, the cooling is allowed to reach thermal equilibrium, characterized by a probability of being in such a state with energy  $E$  given by the Boltzmann distribution. To simulate the evolution to thermal equilibrium of a solid for a fixed value of the temperature, Metropolis et al. (1953) proposed a Monte Carlo Method, which generates sequences of states of the solid (Laarhoven, 1987). The Metropolis algorithm can also be used to generate sequences of configurations of a combinatorial optimization problem. In that case, the configurations assume the role of the states of a solid while the objective function and the control parameter take the roles of energy and temperature, respectively. The simulated annealing algorithm can now be viewed as a sequence of Metropolis algorithms evaluated at a sequence of decreasing values of the control parameter. It is based on randomization techniques. It also incorporates a number of aspects related to iterative improvement algorithms. Since these aspects play a major role in the understanding of the simulated annealing algorithm (Laarhoven, 1987), and – later on in this paper – the development of the simulated entropy algorithms, we first elaborate on the iterative improvement techniques.

### **ITERATIVE IMPROVEMENT ALGORITHMS**

The application of an iterative improvement algorithm presupposes the definition of configurations, a cost function and a generation mechanism, i.e. a simple prescription to generate a transition from a configuration to another by a small perturbation. The generation mechanism defines a neighborhood  $R_i$  for each configuration  $i$ . Iterative improvement is, therefore, also known as neighborhood search or local search. The algorithm can now be formulated as follows. Starting off at a given configuration, a sequence of iterations is generated, each iteration consisting of a possible transition from the current configuration to a configuration selected from the neighborhood of the current configuration. If this neighborhood has a lower cost, the current configuration is replaced by this neighbor, otherwise another neighbor is selected and compared for its cost value. The algorithm terminates when a configuration is obtained whose cost is no worse than any of its neighbors. The disadvantages of iterative improvement algorithms can be formulated as follows:

- a) By definition, iterative improvement algorithms terminate in a local minimum and there is generally no information as to the amount by which this local minimum deviates from a global minimum;
- b) The obtained local minimum depends on the initial configuration, for the choice of which generally no guidelines are available.

To avoid some of the aforementioned disadvantages, one might think of a number of alternative approaches:

- 1) Execution of the algorithm for a large number of initial configurations, say  $N$  (e.g., uniformly distributed over the set of configurations  $R$ ) at the cost of an increase in computation time; from  $N \rightarrow \infty$ , such an algorithm finds a global minimum with probability 1, if only for the fact that a global minimum is encountered as an initial configuration with probability  $1/N$  as  $N \rightarrow \infty$ ;
- 2) (Refinement of 1) use of information gained from previous runs of the algorithm to improve the choice of an initial configuration for the next run (this information relates to the structure of the set of configurations);
- 3) Introduction of a more complex generation mechanism (or, equivalently, enlargement of the neighborhoods), in order to be able to jump out of the local minima corresponding to the simple generation mechanism.
- 4) Acceptance of transitions which correspond to an increase in the cost function in a limited way (in an iterative improvement algorithm only transitions corresponding to a decrease in cost are accepted).

The first approach is a traditional way to solve combinatorial optimization problems approximately. The second and third approaches being usually strongly problem-dependent. An algorithm that follows the fourth approach was independently introduced by Kirkpatrick et al. (1982) and Cerny (1985). It is generally known as simulated annealing, Monte Carlo annealing, statistical cooling, probabilistic hill climbing, stochastic relaxation, or probabilistic exchange algorithm.

In this section, informational entropy is used to develop simulated entropy techniques for solving combinatorial constrained minimization problems. Since the first technique is based on using surrogate constraint approach, we first elaborate on it.

## SURROGATE APPROACH TO CONSTRAINED OPTIMIZATION

A surrogate problem is one in which the original constraints are replaced by only one constrain., termed the surrogate constraint. For the inequality constrained problem, primal problem:

$$(P) \quad \underset{X \in \Omega}{\text{Min}} \quad F(X)$$
$$\text{S.t:} \quad g_i(X) \leq 0 \quad i = 1, \dots, m$$

the corresponding surrogate problem has the form:

$$(S) \quad \underset{X \in \Omega}{\text{Min}} \quad F(X)$$
$$\text{S.t:} \quad \sum_{i=1}^m W_i * g_i(X) \leq 0$$

where  $W_i$  ( $i=1, \dots, m$ ) are non-negative weights, termed surrogate multipliers, which may be normalized without loss of generality by requiring:

$$\sum_{i=1}^m W_i = 1; \quad W_i \geq 0$$

Some of the relationships between the primal problem (P) and the surrogate problem (S) are (Glover, 1965):

- 1)The feasible region of problem (P) is always included by the feasible region of problem (S).
- 2)If  $X^*$  solves problem (S) and  $X^*$  solves problem (P), then  $F(X^*) \leq F(X^*)$  for all  $W \geq 0$ .
- 3)If  $X^*$  solves problem (S) and is also feasible in problem (P), then  $X^*$  also solves problem (P).
- 4)If at least one of original constraints is active at the optimum, then the surrogate constraint must be active at the optimum.

A comparison of the primal feasible region with the surrogate one shows that the primal feasible region is always a part of the surrogate feasible regions. Problem (S) can, therefore, be viewed as a relaxation problem (P).

**SIMULATED ENTROPY**

While annealing define a process of heating a solid and cooling it slowly, entropy in an isolated system tends to a maximum so that this variable is a criterion for the direction in which processes can take place. In simulated entropy (SE), a process which tends to a maximum entropy must be defined first. The two main results to be proved in this section are as follows:

- a) For positive and decreasing values of P, minimizing the objective function F(x) over variables  $x \in X$  subject to the entropy-based surrogate constraint function:

$$\sum_{i=1}^m \lambda_i \cdot g_i(x) - P \sum_{i=1}^m \lambda_i \cdot \ln(\lambda_i) = 0$$

converges, hopefully, to the global minimum of the primal problem (P).

- b) For positive and decreasing values of P, minimizing the objective function F(x) over variables  $x \in X$  subject to the entropy-based constraints functions:

$$\lambda_i \cdot g_i(x) - P \lambda_i \cdot \ln(\lambda_i) \leq 0; \quad i = 1, \dots, m$$

converges, hopefully, to the global minimum of the problem (P). Theorem 1 formally states the first of these results.

**Theorem 1**

The vector  $x^*$  which seeks the global minimum of the single-criteria optimization problem:

$$(P) \quad \underset{X \in \Omega}{Min} \quad F(X)$$

$$S.t: \quad g_i(X) \leq 0 \quad i = 1, \dots, m$$

is generated by solving the entropy-based surrogate constraint problem, simulated entropy problem:

$$(SE) \quad \left. \begin{array}{l} \underset{x \in X}{Min} F(X) \\ S.t: \sum_{i=1}^m \lambda_i \cdot g_i(X) - P \sum_{i=1}^m \lambda_i \cdot \ln(\lambda_i) = 0 \end{array} \right\}$$

Where  $g$  is a vector of dimensionless constraint functions, with positive values of the parameter  $P$  decreasing towards zero and:

$$\lambda_i = \frac{e^{\frac{1}{P}g_i(X)}}{\sum_{i=1}^m e^{\frac{1}{P}g_i(X)}} \quad (2)$$

**Proof:**

This requires the use of Cauchy's inequality (the arithmetic-geometric mean inequality), Hardy (1934):

$$\sum_{i=1}^m U_i \geq \prod_{i=1}^m \left( \frac{U_i}{\lambda_i} \right)^{\lambda_i}$$

or

$$\left( \sum_{i=1}^m U_i^{\frac{1}{P}} \right)^P \geq \left( \prod_{i=1}^m \left( \frac{U_i^{\frac{1}{P}}}{\lambda_i} \right)^{\lambda_i} \right)^P; P > 0 \quad (3)$$

Inequality (3) shows that the  $(1/P - \text{th})$  norm of the set  $U$  increases monotonically as its order,  $(1/P)$ , increases. It tends to its limit as  $P$  tends towards zero:

$$\lim_{P \rightarrow 0} \left( \sum_{i=1}^m U_i^{\frac{1}{P}} \right)^P = \text{Max}_{i \in m} \langle U_i \rangle \quad (4)$$

Let  $U_i = e^{g_i(X)} \quad (5)$

Since the constraints  $g$  are dimensionless, the  $U$  defined by equation (5) are all positive numbers and their substitution into result (4), gives:

$$\lim_{P \rightarrow 0} \left( \sum_{i=1}^m e^{\frac{1}{P}g_i(X)} \right)^P = \text{Max}_{i \in m} \langle e^{g_i(X)} \rangle \quad (6)$$

Taking natural logarithms of both sides and noting that:

$$\ln \lim (\bullet) = \lim \ln (\bullet)$$

$$\ln \text{Max} (\bullet) = \text{Max} \ln (\bullet)$$

equation (6) becomes:

$$\lim_{P \rightarrow 0} P \bullet \ln \left( \sum_{i=1}^m e^{\frac{1}{P} \bullet g_i(X)} \right) = \underset{i \in m}{\text{Max}} < e^{g_i(X)} > \quad (7)$$

Result (7) shows that the constrained minimization problem, (P), can be solved by minimizing  $F(X)$  subject to the aggregated constraints function; AC:

(SE)

$$\left. \begin{array}{l} \underset{x \in X}{\text{Min}} F(X) \\ \text{S.t.} : P \bullet \ln \left( \sum_{i=1}^m e^{\frac{1}{P} \bullet g_i(X)} \right) = 0 \end{array} \right\} \quad (8)$$

as  $P$  in the range  $0 < P \leq \infty$  decreasing towards zero. Taking natural logarithms of both sides of inequality (3) with  $U$  defined by:

$$U_i = e^{\frac{1}{P} \bullet g_i(X)}$$

gives:

$$P \bullet \ln \left( \sum_{i=1}^m e^{\frac{1}{P} \bullet g_i(X)} \right) \geq \sum_{i=1}^m \lambda_i \bullet g_i(X) - P \sum_{i=1}^m \lambda_i \bullet \ln(\lambda_i) \quad (10)$$

which becomes an equality when the variables  $\lambda$  takes the values given by equation (2) on substitution of equation (9), and theorem (1) is proved.

The RHS consists of the vector constraint function in its surrogate form (S) but with an additional entropy term which is a function only of the multipliers  $\lambda$  and the parameter  $P$ . In its equality form with multipliers given by equation (2), relationship (10) shows that the entropy of the multipliers measures the difference between the aggregated constraint function (8) and the surrogate constraint function (S).

It is clear that if we minimize  $F(X)$  over variables  $x \in X$  subject to either side of inequality (10) with multipliers given by equation (2) and  $P$  becoming decreasingly small and positive, the entropy term in (10) will tend towards zero.

The second main result of this section still remains to be proved and Theorem (2) formally states this.

**Theorem 2:**

The vector  $x^*$  which seeks the global minimum of the single-optimization problem, (P), is generated by solving the entropy-based constrained problem:

$$\begin{array}{l}
 \text{(SE)} \\
 \left. \begin{array}{l}
 \text{Min } F(X) \\
 \text{S.t. : } \sum_{i=1}^m \lambda_i \cdot g_i(X) - P \sum_{i=1}^m \lambda_i \cdot \ln(\lambda_i) = 0
 \end{array} \right\}
 \end{array}$$

with positive values of the parameter P decreasing towards zero and:

$$\lambda_i = \frac{e^{\frac{1}{P}g_i(X)}}{\sum_{i=1}^m e^{\frac{1}{P}g_i(X)}} \quad (2)$$

**Proof:**

This requires recalling the entropy-based surrogate constrained problem:

$$\begin{array}{l}
 \text{(SE)} \\
 \left. \begin{array}{l}
 \text{Min } F(X) \\
 \text{S.t. : } \sum_{i=1}^m \lambda_i \cdot g_i(X) - P \sum_{i=1}^m \lambda_i \cdot \ln(\lambda_i) = 0
 \end{array} \right\} \quad (1)
 \end{array}$$

with  $\lambda_i(X)$  is given by (2). Without loss of generality, (1) may be rewritten as follows:

$$\begin{array}{l}
 \left. \begin{array}{l}
 \text{Min } F(X) \\
 \text{S.t. : } \lambda_i \cdot g_i(X) - P \lambda_i \cdot \ln(\lambda_i) = 0; \forall i \in m
 \end{array} \right\} \quad (12)
 \end{array}$$

substituting (2) into (12) and assuming that:

$$\sum_{i=1}^m e^{\left(\frac{1}{P}g_i(X)\right)} = y(X, P)$$

yields to:

$$\left. \begin{array}{l}
 \text{Min } F(X) \\
 \text{S.t. : } P \cdot e^{\frac{1}{P}g_i(X)} \cdot \frac{\ln y(X, P)}{y(X, P)} \leq 0; \forall i \in m
 \end{array} \right\}$$

decreasing P towards zero yields to:

$$\left. \begin{array}{l} \text{Min}_{x \in X} F(x) \\ \text{S.t. : } \text{Max} < g_i(x) > \leq 0; \forall i \in m \end{array} \right\} \quad (13)$$

Result (13) shows that primal constrained minimization problem, (P), can be solved by minimizing  $F(x)$  subject to the sub-aggregated constraints; SAC:

$$\left. \begin{array}{l} \text{(SE)} \\ \text{Min}_{x \in X} F(x) \\ \text{S.t. : } P \bullet \lambda_i(x) \bullet \ln \sum_{i=1}^m e^{\frac{1}{P} g_i(x)} \leq 0; \forall i \in m \end{array} \right\} \quad (14)$$

as  $P$  in the range  $0 \leq P \leq \infty$  decreasing towards zero. This is also true when (12) is used with  $\lambda_i(x)$  given by (2). This completes the proof.

There are several problems in statistics can be formulated so that the desired solution is the global minimum of some explicitly defined objective function. In many cases the number of candidate solutions increases exponentially with the size of the problem making exhaustive search impossible, but descent procedures, devised to reduce the number of solutions examined, can terminate with local minima. In the previous section we have described two simulated entropy algorithms which can escape local minima. The simulated entropy algorithms (SE) are stochastic search procedures which seek the minimum of some deterministic objective function. It applies small perturbations to the current solution. A deterministic descent method will always go to the minimum in whose basin it finds itself, whereas the simulated entropy techniques can climb out of one basin into the next. Because of its ability to make small uphill steps, the simulated entropy techniques avoid being trapped in local minima. However, it is far from apparent at first, that one could not always do at least as well by running a deterministic algorithm many times, with randomly chosen starting points.

Although simulated entropy algorithm details vary when applied to different problems, we always use the same basic simulated entropy idea:

*At each temperature, the process must tends to a maximum entropy*

We define a descent algorithm to generate a solution like the method of feasible directions; a perturbation scheme to perturb the current solution using the informational entropy defined before; and, analogous to the decreasing temperature regime of the physical cooling process, a sequence of control parameters  $P$  which starts at some initially high value of  $P_0$ , decreases by  $dP$  to a final near zero value  $P_f$ . Indeed, it is not apparent so far, why we should vary  $P$  at all: why not just choose  $P=P_f$  in the first place? Two temperature (control parameter, or entropy parameter) scheduling mechanism can be used, either one, in the prototype entropier. Temperature is exponentially reduced using the formula (Kirkpatrick, 1983):

$$P_i = (P_1 / P_0)^i P_0; i = 1, \dots, I$$

Temperature scheduling is completely defined by specifying the temperature at which simulated entropy to begin,  $P_0$ , and the ratio of subsequent to initial temperature,  $P_1 / P_0$ . The second scheduling mechanism, where the temperature is reduced at intervals it uses the formula:

$$P_i = P_0 - \Delta P; i = 1, \dots, I$$

where  $\Delta P$  is the step size to be defined in addition to the temperature at which simulated entropy to begin,  $P_0$ . The simulated entropy techniques developed in this paper repeat the following steps until at least one of the original constraints is active, assuming we are given value  $P_0$ ,  $\Delta P$  or  $P_1 / P_0$  of the central parameter and a starting point:

- 1) Using an iterative technique, generate a neighbor  $F_i(X)$  of the current solution  $F_{i-1}(X)$ . This is a potential candidate for  $F_{i+1}(X)$ .
- 2) Set  $P_i$  to  $(P_0 - \Delta P)$  or  $[P_1 / P_0]^i P_0$  and replaces  $i$  with  $i+1$ .

Theorem 1 which implies two forms of the first simulated entropy technique seeks the global minimum if it is in a convex region, while Theorem 2 which implies two forms of the second simulated entropy technique seeks the global minimum if it is in a concave or convex region. The new developed algorithms and their non-entropy forms, are summarized in Table (1).

### SIMULATED ENTROPY METHODS

No.	Name	Form
1.	The Entropy-Based Surrogate Constraint, ESC	$\sum_{i=1}^m \lambda_i \bullet g_i(X) - P \sum_{i=1}^m \lambda_i \bullet \ln(\lambda_i) = 0$
2.	The Aggregated Constraint, AC	$P * \sum_{i=1}^m \exp[(1/P) \bullet g_i(X)] = 0$
3.	The Entropy-Based Constrained, EC	$\lambda_i \bullet g_i(X) - P \bullet \lambda_i \bullet \ln(\lambda_i) \leq 0; i = 1, \dots, m$
4.	The Sub-Aggregated Constrained, SAC	$P \bullet \lambda_i \sum_{i=1}^m \exp[(1/P) \bullet g_i(X)] \leq 0; i = 1, \dots, m$
with $\sum_{i=1}^m \lambda_i = 1;$		$\lambda_i = e^{(1/P) \bullet g_i(X)} / \left( \sum_{i=1}^m e^{(1/P) \bullet g_i(X)} \right); \infty \geq P \geq 0$

**Table (1)**

## NUMERICAL EXAMPLES

Two numerical examples are solved using the new entropy-based simulated entropy techniques developed before. General discussion showing the characteristics of these techniques is included also.

### Example (1) (Li, 1987)

Minimize:

$$F(X) = (X_1 - 2)^2 + (X_2 - 1)^2$$

$$g_1(X) \equiv X_1^2 - X_2 \leq 0$$

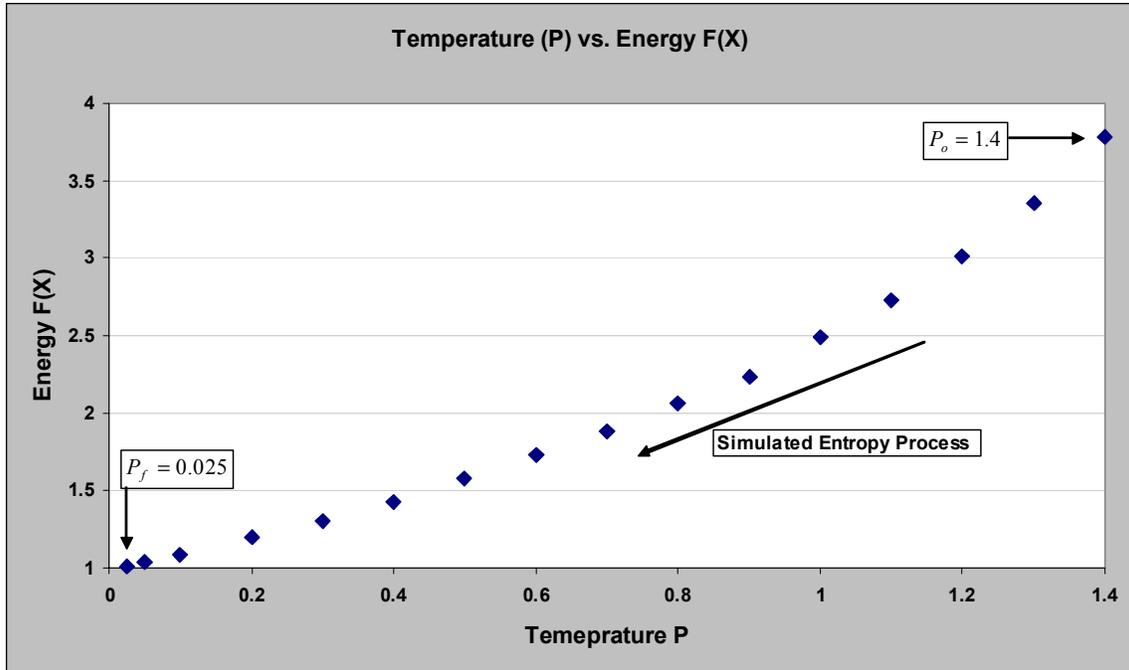
S.t.:  $g_2(X) \equiv X_1 + X_2 - 2 \leq 0$

$$X_1, X_2 \geq 0$$

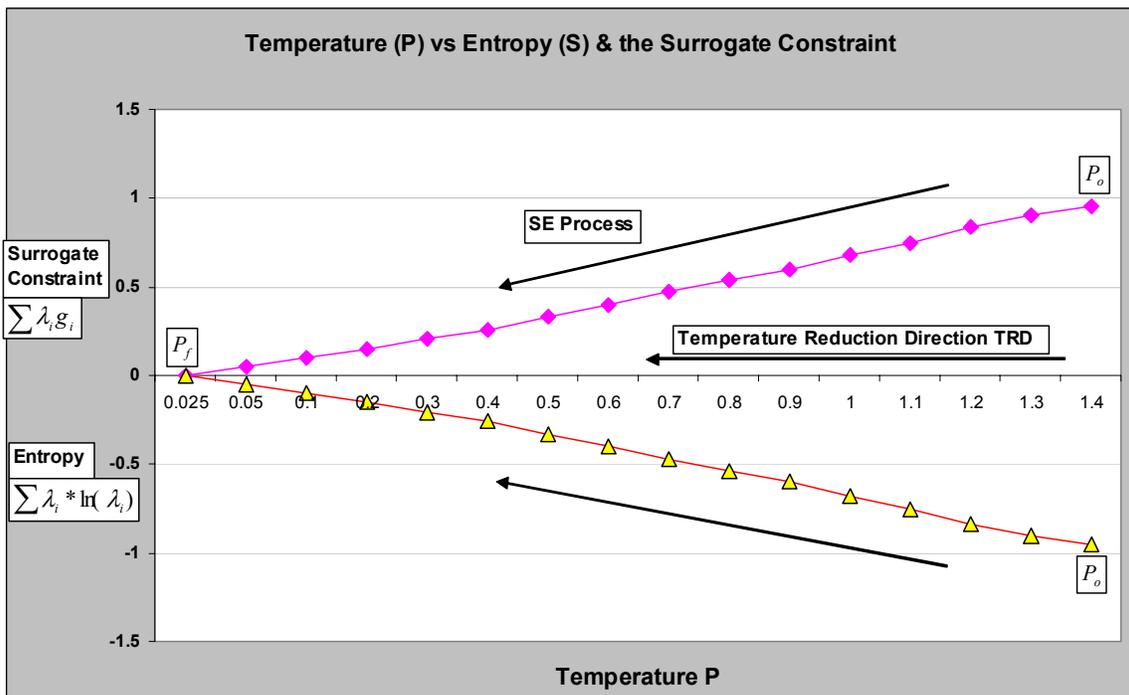
The results are summarized in Table (2) and Fig. (1).

P	F	$X_1$	$X_2$	$g_1$	$g_2$
1.4	3.7791	.0563	.9656	-.96245	-.97808
1.3	3.3557	.1685	.9654	-.937	-.8662
1.2	3.0168	.2634	.9698	-.9004	-.7669
1.1	2.7322	.3472	.9762	-.8556	-.6766
1.0	2.4862	.4233	.9838	-.8046	-.5928
0.9	2.2387	.5084	.8824	-.6239	-.6092
0.8	2.0592	.5697	.8836	-.5590	-.5467
0.7	1.8872	.6311	.8848	-.4866	-.4841
0.6	1.7286	.6902	.8857	-.4096	-.4241
0.5	1.5835	.7468	.8859	-.3282	-.3673
0.4	1.4256	.8062	.9798	-.3299	-.2140
0.3	1.3076	.8580	.9418	-.2057	-.2003
0.2	1.2018	.9052	.9432	-.1237	-.1516
0.1	1.0826	.9596	.9848	-.064	-.0555
0.05	1.0345	.9829	1.0009	-.0349	-.0162
0.025	1.0104	.9948	.9962	-.0066	-.0089

**Table (2)**  
**The Global Minimum by the ESC Method of Example (1)**



(a)



(b)

Fig. (1) Example (1)  
 (a) Temperature P vs Energy F(X)  
 (b) Temperature P vs Entropy & the Surrogate Constraint

As noted, we started the simulated entropy process at a maximum temperature of 1.4 which is correspondent to a maximum energy of 3.7791 and minimum values of constraints  $g_1(x)$  and  $g_2(x)$ . Lowering the temperature continuously towards zero along the process, the energy of the system  $F(x)$  decreases continuously along to its minimum and the corresponding constraints values increase continuously towards zero. The simulated entropy process stops when at least one original constraint approaches zero.

**Example (2) (Hock et al. 1981)**

Minimize:

$$F(X) = X_1^2 + 0.5X_2^2 + X_3^2 + 0.5 * X_4^2 - X_1X_3 + X_3X_4 - X_1 - 3X_2 + X_3 - X_4$$

S.t.:

$$g_1(X) \equiv 5 - X_1 - 2X_2 - X_3 - X_4 \geq 0$$

$$g_2(X) \equiv 4 - 3X_1 - X_2 - 2X_3 + X_4 \geq 0$$

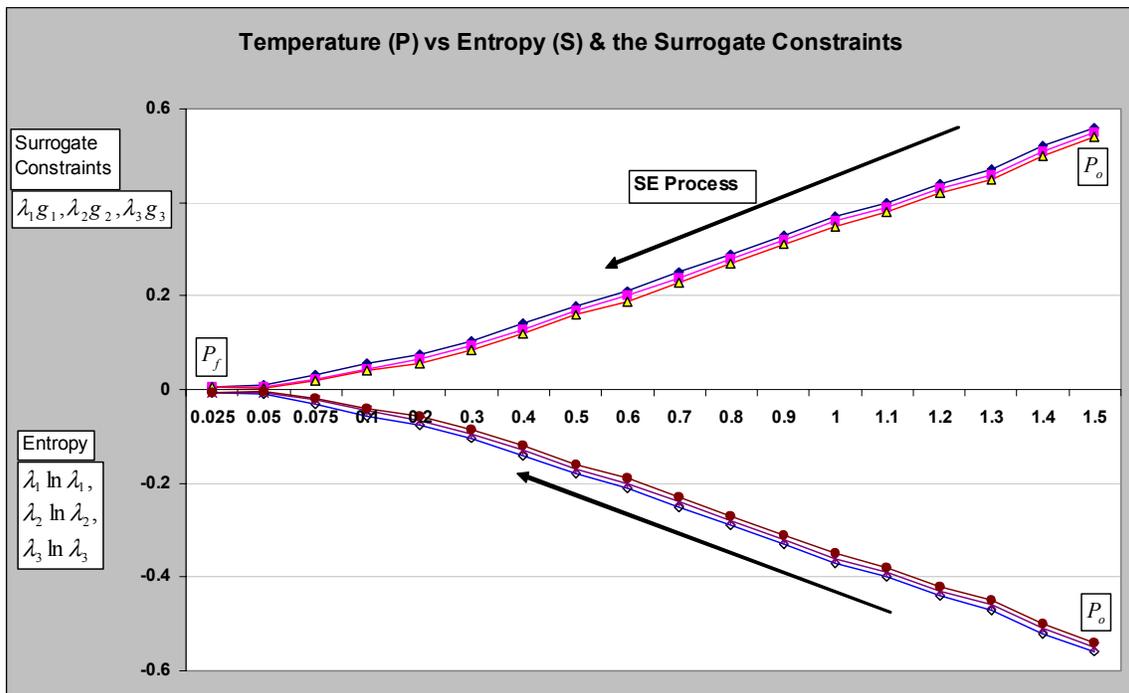
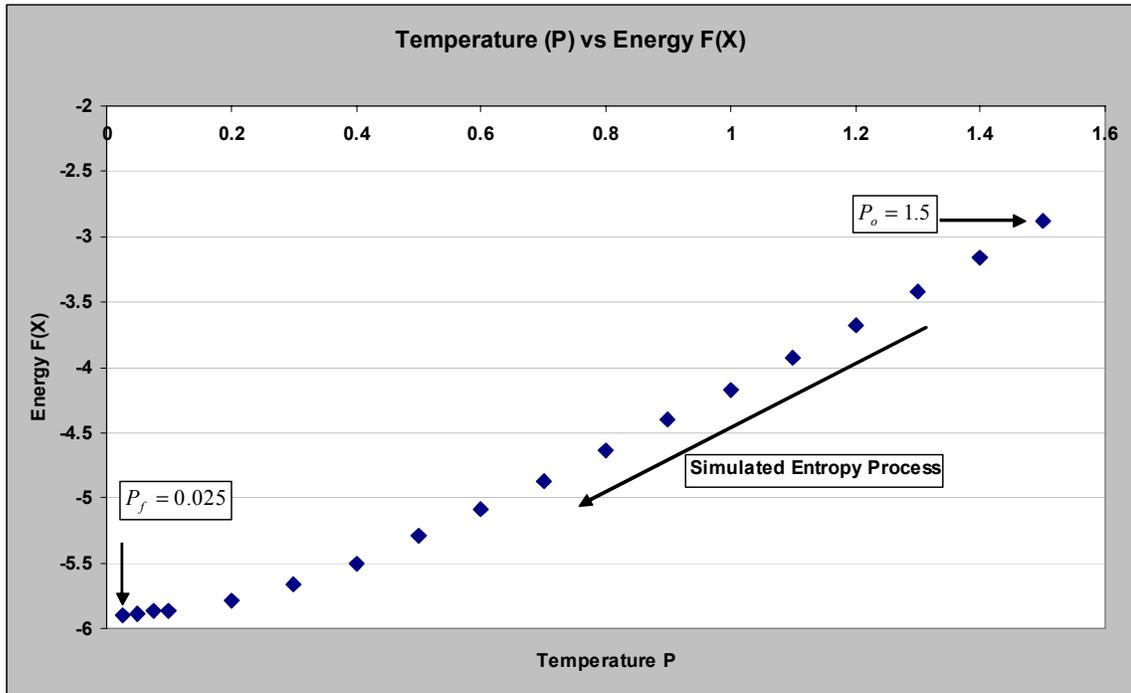
$$g_3(X) \equiv X_2 + 4X_3 - 1.5 \geq 0$$

$$X_i \geq 0; i = 1, \dots, 4$$

The results are summarized in Table (3) and Fig. (2). Here, the minimum energy obtained by the SE methods,  $F(X) = -5.8971$ , is less than the value given by Hock et al. (1981),  $F(X) = -4.682$ . This assures the necessity of using the simulated entropy techniques to secure our search for the global minimum.

P	F	$X_1$	$X_2$	$X_3$	$X_4$	$g_1$	$g_2$	$g_3$
1.5	-2.8795	0.00011	1.4091	.3661	.29904	-1.5165	-2.1574	-1.3735
1.4	-3.162	0.0001	1.5283	.31264	.28463	-1.3461	-2.1308	-1.2788
1.3	-3.4176	0.041	1.5785	.273	.29587	-1.2331	-2.0483	-1.1705
1.2	-3.6759	0.10919	1.6194	.2575	.30198	-1.0924	-1.84	-1.1495
1.1	-3.9249	0.14549	1.6635	.20588	.31367	-1.0079	-1.8019	-0.9871
1.0	-4.1706	0.2857	1.7041	.18542	.32017	-0.8776	-1.6195	-0.9458
0.9	-4.4028	0.26179	1.74	.15398	.3313	-0.7729	-1.4979	-0.8559
0.8	-4.6366	0.31235	1.7787	.11222	.32702	-0.691	-1.3868	-0.7276
0.7	-4.8679	0.37211	1.8238	.091488	.33322	-0.5556	-1.2102	-0.6897
0.6	-5.0853	0.4255	1.8601	.05629	.33269	-0.4654	-1.036	-0.5852
0.5	-5.3948	0.48263	1.8925	.024625	.33426	-0.3734	-0.9446	-0.491
0.4	-5.5005	0.5483	1.9251	.0001	.3293	-0.272	-0.7591	-0.4255
0.3	-5.6671	0.61991	1.9468	.0001	.33583	-0.1506	-0.5291	-0.4472
0.2	-5.7836	0.6751	1.9639	.000863	.33104	-0.0651	-0.3401	-0.4674
0.1	-5.8623	0.7306	1.9620	.0001	.3229	-0.0223	-0.1688	-0.4624
0.075	-5.8662	0.72921	1.9635	.0001	.3279	-0.0157	-0.1765	-0.4639
0.05	-5.8867	0.73933	1.9814	.0001	.29016	-0.007702	-0.0906	-0.4818
0.025	-5.8971	0.74389	1.9911	.0001	.27119	-0.000258	-0.0482	-0.4915

**Table (3)**  
**The Global Minimum by the EC Method of Example (2)**



**Fig. (2) Example (2)**  
**(b) Temperature P vs Energy F(X)**  
**(b) Temperature P vs Entropy & the Surrogate Constraints**

## DISCUSSION

Two of the most important factors which influence the convergence are the descent method and the way the sequence of control parameter  $P_i$  is chosen. The justification for using the simulated entropy algorithm with  $P > 0$  is that we may not be able to tell a priori whether a given descent method will produce potential local optima or not. In other words, using the simulated entropy algorithm with  $P > 0$  provides extra security against bad or uncertain descent method.

We have given examples which show that it is not possible to ensure convergence in less than exponential time for a general problem. However, we have not developed any rigorous theory concerning the class of problems for which the typical behavior is agreed. Nevertheless, many applications of the simulated entropy algorithms, example (1) for example, have displayed good performance of the method in terms of a comparison with results obtained by using a descent method alone, when applied to the same problem.

We have applied the method to different mathematical examples and demonstrated how the algorithm may be applied. While these applications do indicate that the simulated entropy idea is likely to be a useful one, a good deal of further work, possibly in the form of a large scale simulation, is required to study the effect of alternative perturbation schemes and a satisfactory method of using feedback to determine the control.

For each example, the temperature versus energy relationship is plotted. It shows that simulated entropy takes place at initial maximum temperature  $P_0$  at which energy is maximum force.

Temperature reduction continues slowly towards  $P_f$  at which energy is minimum. Also, the temperature versus the weighted constraints, the upper vertical axis, and entropy, the lower vertical axis is plotted. It shows that the two are symmetric with respect to the temperature axis alongside the simulated entropy process.

At the minimum and final temperature  $P_f$ , very close to zero, the weighted constraints and entropy coincide at a value very close to zero also, i.e. the equilibrium configuration is reached and the simulated entropy process is terminated.

In summary, we see that, for most applications, given a particular starting solution and a sufficient number of transitions, the simulated entropy algorithm, with  $P > 0$ , is more likely to obtain a good solution as opposed to a local optimum, than the descent method when it is used by itself ( $P = 0$ ). At first glance these results suggest that simulated entropy algorithm with  $P > 0$  is to be preferred to the descent method by itself ( $P = 0$ ). However, one question which must be answered is whether the number of iterations,  $I$ , required by the descent method by itself, to first observe the corresponding global or new local optima is significantly less than those, say  $I_{P>0}$ , required by the simulated entropy algorithm, with  $P > 0$ . For, if this is true, it may well be possible to carry out several applications of the descent methods ( $P = 0$ ), each from a different starting point and for  $I_{P=0}$  iterations, so that the total number of iterations required is less than the  $I_{P>0}$  iterations required in one application of the algorithm with  $P > 0$ .

We have described a stochastic minimization procedure previously and tested its applicability in this section by solving different minimization problems. The methods, which the author chose to call simulated entropy incorporates a scheme for simulating the equilibrium behavior of atoms at constant temperature, in a procedure analogous to chemical entropy. It has the attraction that it is general yet extremely easy to apply. The two simulated entropy methods use a descent method but it can accept small objective function increases. The procedure can be made to behave as simulated annealing by defining randomly a topology via a perturbation scheme, so that neighbors may be generated from any given solution.

We have found that for some applications  $I_{P=0}$  is comparable to  $I_{P>0}$ ,  $I_{P=0}$  is considerably less than  $I_{P>0}$ . This means that it may be preferable to make several applications of the descent method ( $P = 0$ ), each for a small number of iterations and from a different random start. However, note that, in this second case, there is still a chance that the global optimum may not be among the set of local optima.

Finally, the two simulated entropy techniques developed and employed here provide a secure method by means of which designers and mathematicians can detect their minima. However, as with all global optimization-seeking methods, computational costs and time requirements are high.

## **REFERENCES**

Cerny, V., Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm, J. Opt. Theory Appl., 45(41-51), 1985.

Glover, F., Multiphase Dual Algorithm for the Zero-One Integer Programming Problem, Oper. Res., Vol. 13, pp. 879-919.

Hock, W. and Schittokowski, K., Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1981.

Hardy, G.H., Littlewood, J.E., and Polya, G., Inequalities, Cambridge: Cambridge University Press, 1934.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., Optimization by Simulated Annealing, SCIENCE, Vol. 220, No. 4598, 13 May, 1983.

Laarhoven, P.J.M. Van and Aarts, E.H.L., Simulated Annealing: Theory and Applications, D. Reidel Publishing Company, Dordrecht, Holland, 1987.

Li, X.S., Entropy and Optimization, Ph.D. Thesis, University of Liverpool, 1987.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E., Equation and State Calculations by Fast Computing Machines, J. of Chem. Physics, 21(1087-1092), 1953.

Sultan, A.M., Entropic Vector Optimization and Simulated Entropy: Theory and Applications, Ph.D. Thesis, Liverpool University, 1990.

Sultan, A.M. and Templeman, A.B., Generation of Pareto Solutions by Entropy-Based Methods, Multi-Objective Programming and Goal Programming: Theory and Applications, Lecture Notes in Economics and Mathematical Systems, 432.