On the Statistical Mechanics of (Un)Constrained Stochastic Hopfield and 'Elastic' Neural Networks

Jan van den Berg
berg@cs.few.eur.nl

Jan C. Bioch
bioch@cs.few.eur.nl

Erasmus University Rotterdam, Dept. of Computer Science,
Room H4–13, P.O. Box 1738, 3000 DR Rotterdam

December 22, 1994

Abstract

Stochastic binary Hopfield models are viewed from the angle of statistical mechanics. After an analysis of the unconstrained model using *mean field* theory, a similar investigation is applied to a constrained model yielding comparable general explicit formulas of the free energy. Conditions are given for which some of the free energy expressions are Lyapunov functions of the corresponding differential equations. Both stochastic models appear to coincide with a specific continuous model. Physically, the models are related to spin and Potts glass models. Also, a 'complementary' free energy function of both the unconstrained and the constrained model is derived. The analysis culminates in a very *general framework* for analyzing constrained and unconstrained Hopfield neural networks: the stationary points of the corresponding free energy appears to coincide exactly with the set of equilibrium conditions of the corresponding continuous Hopfield neural network.

Moreover, the relationship with 'elastic net' algorithms is analyzed: it is proved that this class of algorithms cannot be derived from the theory of statistical mechanics (as sometimes is supposed), but should be considered as a special 'penalty method', namely as one with *dynamical penalty weights*. We mention some experimental results and discuss implications for the use of the various models in resolving constrained optimization problems.

1 Motivation and Results

The relationship between statistical mechanics and stochastic neural networks has been studied intensively (see e.g., [5, 13]). In particular, it appeared to be fruitful to transfer the mathematical techniques from the theory of spin glasses to the analysis of neural networks. Using these techniques, it is for example possible to analyze the capacity (i.e., the number of storable patterns) of stochastic Hopfield networks. Our interest in the subject was aroused after finding an expression of the energy of the thermal noise of a binary stochastic Hopfield network [16]. Later on, we discovered that the same expression was already mentioned in an article by Simic [14]. Simic's derivations of expressions of the free energy of some neural networks for solving the Travelling Salesman Problem (TSP) are very succinct. He proposes statistical mechanics as the underlying theory of 'elastic' and 'neural' optimizations. In one of his analyses, some constraints of the problem are enforced 'strongly' by summing over those configurations which obey these constraints. Physically, this model can be viewed as a so-called Potts glass model. Related work on TSP and other combinatorial optimization problems was done some years earlier by Van den Bout and Miller [2] and by Peterson and Söderberg [10], but they applied other update rules and paid more attention to practical issues.

In this paper, the various contributions are considered and extended in the framework of *general* Hopfield models. In section 2, we shall start by shortly describing the classical Hopfield networks. Then, by applying a (slightly modified version of Simic's) statistical mechanics approach to an unconstrained stochastic binary Hopfield network, we derive two theorems concerning the free energy. They accurately clarify how the stochastic network is related to the classical continuous one. Using a *mean field approximation*, the first theorem yields the sigmoid function as transfer function for the neurons together with an explicit expression of the free energy in a natural way. In the second theorem, another free energy function is derived showing that the continuous model can be seen as a (mean field) approximation of the stochastic one. Another theorem concerns the stability of the motion equations: the free energy expression of the second theorem appears to be a Lyapunov function. In still another theorem, we introduce a 'complementary' energy expression, which also appears to be a Lyapunov function. Then, the *general framework* is presented. A new free energy expression (in terms of both the input and the output of the individual neurons) is derived whose stationary points coincide precisely with the set of equilibrium conditions of the unconstrained Hopfield model.

In section 3, Simic's modified approach is used again, this time to analyze a certain type of constrained stochastic binary Hopfield network yielding theorems of similar purport. Now, another transfer function is derived together with new explicit free energy expressions. It is demonstrated that under some dynamical conditions again, the second free energy expression of this section is a Lyapunov function. Furthermore, the constrained stochastic model in mean field approximation appears to coincide with an adapted continuous Hopfield model. Again, a complementary energy is introduced and the general framework of the constrained Hopfield model is presented.

Both the unconstrained system and the constrained system can be interpreted in the same fashion physically: if the temperature in such a system is lowered during the updating of the differential equations, then so-called mean field annealing takes place. This annealing approach (which is an approximation of 'simulated annealing') favours the probability of finding the global extremum of the original energy function. The whole system can be described by the free energy (sometimes termed the 'effective energy') of the system, which is a composition of the average original energy and the thermal noise energy. At high temperatures, the original energy function surface is 'smoothed' by the presence of the thermal noise energy. On lowering the temperature, the smoothing effect of the thermal noise gradually disappears and the free energy goes over to the original energy function.

In [14], the constrained model has been applied to 'prove' that the Durbin and Willshaw's energy function [3] of the 'elastic net' algorithm can be derived from it. In a separate subsection, we shall argue why we think this proof is not correct. First, we sum up the places of wrong derivations and conclusions, which will be mathematically underpinned in the second appendix. Next, we shall explain why in our opinion, the elastic net algorithm can be considered as a special type of 'penalty method' namely, as one with *dynamical penalty weights*. This view opens the way for a search into methods of solving combinatorial optimization problems using new, self-chosen dynamical penalty terms.

In the final section, we discuss our results and mention some surveying simulations, whose practical results are in agreement with the theory. We touch lightly on the potential capabilities of the analyzed Hopfield and elastic neural networks in resolving constrained optimization problems; e.g., we discuss why the Hopfield-Lagrange model [15] might be useful in this context. Regarding the elastic net algorithm, it is interesting to compare our idea, of it being a dynamical penalty method, with the approach of 'deformable templates' [11] which we came across recently. We finish by reflecting upon the possibilities of improving our derivations at some places.

2 Unconstrained Stochastic Hopfield Networks

2.1 The Background: Classical Hopfield Networks

In 1982, Hopfield introduced the idea of an 'energy function' into neural network theory using an asynchronous updating rule and binary units [6]. He used the following expression of the energy:

$$E(\mathbf{S}) = -\tfrac{1}{2}\sum_{ij} w_{ij} S_i S_j - I_i S_i, \tag{1}$$

where $\mathbf{S} \in \{0,1\}^n$ is the state vector $(S_1, \cdots, S_n)$ of the neural network, S_i the output value and I_i the external input of neuron i and w_{ij} represents the interconnection strength from neuron j to neuron i.

In 1984, he generalized the stochastic model to a deterministic one using a continuous updating rule with continuous-valued units [7], which essentially is a *parallel gradient descent* method. Hopfield used the well known updating rule[1]

$$\dot{U}_i = -\frac{\partial E_{HM}(\mathbf{V})}{\partial V_i} = \sum_j w_{ij} V_j + I_i - U_i, \tag{2}$$

where $V_i = g(U_i)$. The corresponding energy function equals

$$\begin{aligned} E_{HM}(\mathbf{V}) &= \underbrace{-\tfrac{1}{2}\sum_{ij} w_{ij} V_i V_j - \sum_i I_i V_i}_{} + \underbrace{\sum_i \int_0^{V_i} g^{-1}(V)\,dV}_{} & (3)\\ &= \qquad E(\mathbf{V}) \qquad + \qquad E_H(\mathbf{V}) \quad , & (4) \end{aligned}$$

where $E(\mathbf{V})$ is the energy or target function to be minimized. The term $E_H(\mathbf{V})$ will be called the 'Hopfield term'. Now, $\mathbf{V} \in [0,1]^n$ is the state vector $(V_1, \cdots, V_n)$ of the neural net and V_i the output of neuron i. Furthermore, $U_i = \sum_j w_{ij} V_j + I_i$ is the total (i.e., internal plus external) input of neuron i and $g(U_i)$ the activation or transfer function. Note that $U_i = \partial E_H / \partial V_i$.

[1] There are other ways to find an equilibrium point of the neural network like $V_i^{\text{new}} = g(\sum_j w_{ij} V_j^{\text{old}} + I_i)$. However, they are not analyzed here.

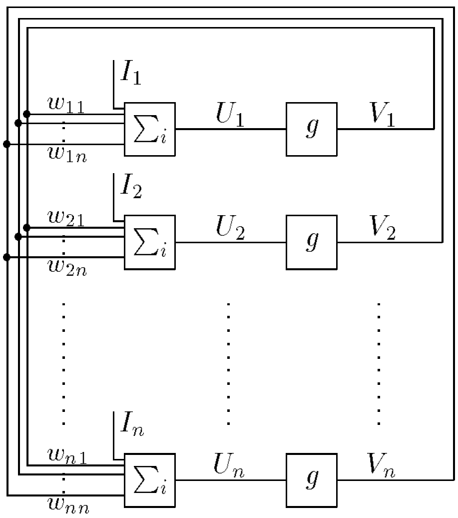

Figure 1: The classical Hopfield network with equilibrium condition: $\forall i : U_i = \sum_j w_{ij}V_j + I_i$ and $V_i = g(U_i)$.

In figure 1, a picture of the Hopfield model is given. It can be used to explain the working of the motion equations (2). After initialization, the network is generally not in an equilibrium state. Then, while keeping the relations $V_i = g(U_i)$ valid, the input values U_i are adapted in agreement with (2). The following theorem, proven by Hopfield [7], gives conditions for which an equilibrium state will eventually be reached:

Theorem 1 (Hopfield). *If $W = (w_{ij})$ is a symmetrical matrix and if $\forall i : V_i = g(U_i)$ is a monotone increasing, differentiable function, then E_{HM} is a Lyapunov function for motion equations* (2).

For the rest of this paper, we also give a 'complementary' theorem which deals with the case of a monotone decreasing transfer function:

Theorem 2. *If $W = (w_{ij})$ is a symmetrical matrix and if $\forall i : V_i = g(U_i)$ is a monotone decreasing, differentiable function, then $-E_{HM}$ is a Lyapunov function for motion equations* (2).

We confine ourselves to giving the proof of the second theorem.

Proof. Because $g_i = V(U_i)$ is monotone decreasing and differentiable, it follows that $\mathrm{dV_i}/\mathrm{dU_i} < 0$. Consequently,

$$\begin{aligned}
-\dot{E}_{HM} &= -\sum_i \frac{\partial E_{HM}}{\partial V_i}\dot{V}_i = \tfrac{1}{2}\sum_{ij} w_{ij}\dot{V}_i V_j + \tfrac{1}{2}\sum_{ij} w_{ij} V_i \dot{V}_j + \sum_i I_i \dot{V}_i - \sum_i g^{-1}(V_i)\dot{V}_i \\
&= \sum_i \dot{V}_i \big(\sum_j w_{ij}V_j + I_i - U_i\big) = \sum_i \dot{V}_i \dot{U}_i = \sum_i \frac{\mathrm{dV_i}}{\mathrm{dU_i}}(\dot{U}_i)^2 \leq 0. \qquad (5)
\end{aligned}$$

Because $-E_{HM}$ is bounded below its value decreases constantly during updating until finally a (local) minimum has been reached, where $\forall i : \dot{U}_i = 0$. So, $-E_{HM}$ is a Lyapunov function. □

Using (2), the final state condition $\forall i : \dot{U}_i = 0$ implies that $U_i = \sum_j w_{ij} V_j + I_i$, so in that state, the neural network has come to equilibrium. As transfer function the following (monotone increasing) one is often used:

$$g(U_i) = \frac{1}{1 + \exp(-\beta U_i)}. \tag{6}$$

We already mentioned that application of theorem (1) corresponds to a 'gradient descent'. It should be clear that application of theorem (2) corresponds to a 'gradient ascent' of the energy function E_{HM}.

2.2 Stochastic Hopfield Networks in Mean Field Approximation

It is possible to make the units behave stochastically [5]. E.g., taking binary units, one defines a probability of finding a neuron in one of the two states. Models of this type can be viewed from the angle of statistical mechanics [5, 13] and can be considered as spin glass models. In the statistical mechanics approach, one considers *average quantities* like the average state $\langle S_i \rangle$ of the neuron i and the average energy $\langle E(\mathbf{S}) \rangle$ of the stochastic neural network.

In [14], Simic uses a method which yields explicit expressions for both the energy function to be minimized and the entropy term. As will be shown, this makes it possible to accurately compare how stochastic networks in mean field approximation are related to their continuous counterparts. This is why we take up (a slightly modified version of) his approach and try to generalize as much as possible. Generally, the goal is to find an explicit expression for the thermodynamic 'free energy' F. This free energy is calculated by application of the formula

$$F = -T \ln(Z), \tag{7}$$

where $T = 1/\beta$ and Z_β is the so-called thermodynamic partition function:

$$Z_\beta[\mathbf{I}] = \sum_{\mathbf{S}} \exp\left(-\beta E(\mathbf{S})\right). \tag{8}$$

Considering binary neurons and applying Hopfield's energy expression (1), the partition function becomes

$$Z_\beta[\mathbf{I}] = \sum_{\mathbf{S}} \exp\left[\beta\left(\tfrac{1}{2} \sum_{ij} w_{ij} S_i S_j + \sum_i I_i S_i\right)\right]. \tag{9}$$

We can evaluate the average value $V_i = \langle S_i \rangle$ from the partition function by using the relation

$$V_i = \langle S_i \rangle = \frac{1}{\beta} \frac{\partial \ln Z_\beta[\mathbf{I}]}{\partial I_i}. \tag{10}$$

More generally, the average value of any quantity $A(\mathbf{S})$, which is a function of the system state, can be evaluated using

$$\langle A(\mathbf{S}) \rangle = \frac{1}{Z_\beta} \sum_{\mathbf{S}} A(\mathbf{S}) \exp(-\beta E(\mathbf{S})). \tag{11}$$

The main difference between Simic's and our approach concerns the way the external fields I_i are treated: Simic includes *small* 'generating fields' in the expression of the partition function [13], which are set to 0 during the derivation. We use *real* external fields I_i, which appear in the expression of the partition function as part of the energy function and which remain in the formulas.

Theorem 3. *In mean field approximation, the free energy of unconstrained stochastic binary Hopfield networks can be stated as*

$$F_{U1}(\mathbf{V}) = \tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \tfrac{1}{\beta}\sum_i \ln\left[1+\exp\left(\beta\left(\sum_j w_{ij}V_j + I_i\right)\right)\right]. \tag{12}$$

The stationary points of F_{U1} are found at points of the state space where

$$\forall i : V_i = \frac{1}{1+\exp(-\beta(\sum_j w_{ij}V_j + I_i))}. \tag{13}$$

Proof. We give an extended sketch of the proof using some lemmas from the appendix. To be able to perform the summation in the partition function (9) the exponentials in the quadratic terms S_iS_j are turned into exponentials which are linear in the S_i's by using lemma 1. This yields

$$Z_\beta[\mathbf{I}] = \sum_{\mathbf{S}} \frac{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j + \beta\sum_i S_i(\phi_i + I_i)\right]\prod_i \mathrm{d}\phi_i}{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j\right]\prod_i \mathrm{d}\phi_i}, \tag{14}$$

where the w_{ij}^{-1}'s are the elements of the inverted matrix W^{-1}. Note that the condition of symmetry of the matrix W of lemma 1 coincides with one of the conditions for theorem 1. By expanding, for every state, the quotient of the two integrals of (14) around its saddle-point $\hat{\boldsymbol{\phi}}$ — using an n-dimensional version of lemma 2 — it is possible to evaluate exactly this expression of the partition function, i.e., one recovers formula (9). The saddle-point equation leads to the (exact) formula

$$\hat{\phi}_i = \sum_j w_{ij}S_j \qquad \text{implying that} \qquad \langle\hat{\phi}_i\rangle = \sum_j w_{ij}\langle S_j\rangle = \sum_j w_{ij}V_j, \tag{15}$$

where $\langle\hat{\phi}_i\rangle$ is the i-th component of the average of the saddle-point values of (14). Apparently, $\langle\hat{\phi}_i\rangle$ represents the average internal input of neuron i. We also may perform the summation over all states $\mathbf{S}$ in (14) yielding

$$Z_\beta[\mathbf{I}] = \frac{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j + \sum_i \ln\left(1+\exp(\beta(\phi_i + I_i))\right)\right]\prod_i \mathrm{d}\phi_i}{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j\right]\prod_i \mathrm{d}\phi_i}. \tag{16}$$

Writing

$$E(\boldsymbol{\phi},\mathbf{I}) = \tfrac{1}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j - \tfrac{1}{\beta}\sum_i \ln\left[1+\exp\left(\beta(\phi_i + I_i)\right)\right], \tag{17}$$

the saddle-point $\tilde{\boldsymbol{\phi}}$ of equation (16) is found by partial differentiation of $E(\boldsymbol{\phi},\mathbf{I})$ to the ϕ_i's:

$$\tilde{\phi}_i = \sum_j \frac{w_{ij}}{1+\exp\left(-\beta(\tilde{\phi}_j + I_j)\right)}. \tag{18}$$

On the other hand, by using lemma 3 (which uses a mean field approximation), we obtain

$$V_i \approx -\frac{\partial E(\tilde{\boldsymbol{\phi}}, \mathbf{I})}{\partial I_i} = \frac{1}{1 + \exp(-\beta(\tilde{\phi}_i + I_i))}. \tag{19}$$

By substituting (19) in the exact formula (15) we obtain the saddle-point equation (18), so in mean field approximation, the average $\langle\hat{\boldsymbol{\phi}}\rangle$ coincides with the saddle-point $\tilde{\boldsymbol{\phi}}$ of the integral over $\boldsymbol{\phi}$ in (16). Because in a first order approximation the partition function in (16) equals

$$Z_\beta = \exp(-\beta E(\tilde{\boldsymbol{\phi}}, \mathbf{I})), \tag{20}$$

one finds for the expression of the free energy (7), by also substituting (15), precisely (12). The stationary points (13) are found by resolving the equations $\partial F_{U1}/\partial V_i = 0$. □

Theorem 4. *Using the mean field approximation* (13), *the free energy of unconstrained stochastic binary Hopfield networks can also be stated as*

$$F_{U2}(\mathbf{V}) = -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \sum_i I_iV_i + \tfrac{1}{\beta}\sum_i (V_i \ln V_i + (1 - V_i)\ln(1 - V_i)). \tag{21}$$

The stationary points of F_{U2} coincide with those of F_{U1}.

Proof. The fact that (13) holds in mean field approximation can be derived by a substitution of (15) in (19) using the result that $\tilde{\phi}_i = \langle\hat{\phi}_i\rangle$. Moreover, taking $a = \beta(\sum_j w_{ij}V_j + I_i)$ and $m = V_i$, lemma 4 states:

$$\begin{aligned} \ln\left[1 + \exp\left(\beta(\sum_j w_{ij}V_j + I_i)\right)\right] = \\ -V_i \ln V_i - (1 - V_i)\ \ln(1 - V_i) + \beta(\sum_j w_{ij}V_iV_j + I_iV_i). \end{aligned} \tag{22}$$

By combining this result and equation (12) the expression (21) for $F_{U2}(\mathbf{V})$ is found. Moreover, the stationary points are found by resolving the equations $\partial F_{U2}/\partial V_i = 0$ yielding precisely (13). □

As has been shown in [16], we found the same expression (21) for the energy $E_{HM}(\mathbf{V})$ of the continuous Hopfield model, *presuming* the validity of the sigmoid (6) as transfer function: this was done by simply elaborating the integral of the Hopfield term E_H in (3). We also notice that

$$V_i = \langle S_i\rangle = 1 \times \mathrm{P}\,(S_i = 1) + 0 \times \mathrm{P}\,(S_i = 0) = \mathrm{P}\,(S_i = 1). \tag{23}$$

The discovered expression (21) for the free energy has the well known form

$$F_U(\mathbf{V}) = \langle E(\mathbf{S})\rangle - TS = E(\mathbf{V}) - TS, \tag{24}$$

where $T = 1/\beta$ and S equals the expression of the entropy of a binary neuron

$$S = -\sum_i (V_i \ln V_i + (1 - V_i)\ln(1 - V_i)). \tag{25}$$

Summarizing, we may consider the classical continuous Hopfield model with the sigmoid function as a transfer function as an approximation of the stochastic binary model based on partition function (9): the energy E_{HM} of the continuous model coincides, in mean field approximation, with the free energy expression $F_{U2}(\mathbf{V})$ of the stochastic model, where the Hopfield term E_H of the continuous model equals the thermal energy $-TS$ of the stochastic one. It is clear too, that all neurons have a mutually *independent* thermal energy contribution equal to $T(V_i \ln V_i + (1-V_i)\ln(1-V_i))$. At high temperatures, the total thermal energy dominates, yielding as an equilibrium solution of the system $\forall i : V_i \approx 0.5$, because then $-TS$ has its minimum value. Lowering the temperature corresponds to a decrease of thermal noise in the system. If this lowering is done during updating conform equation (2) one speaks of 'mean field annealing' [5]. More details about the effect of the Hopfield term can be found in [16].

To explain the theory, we give a simple example. Suppose the function to be minimized equals $E(S) = 2S^2$. Then, the corresponding free energy expressions (from theorems 3 and 4) equal

$$F1(V) = -2V^2 - \frac{1}{\beta}\ln\left(1 + \exp(-4\beta V)\right), \quad (26)$$

$$F2(V) = 2V^2 + \frac{1}{\beta}(V \ln V + (1-V)\ln(1-V)). \quad (27)$$

A diagram of these functions is shown in figure 2.2 for three values of β. The coincidence

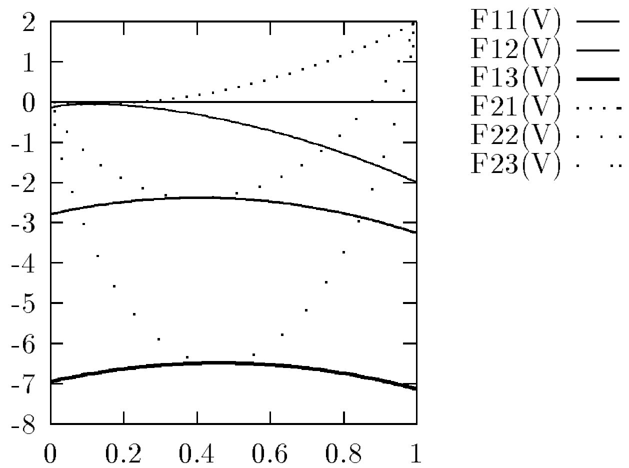

Figure 2: The two free energy expressions $F1$ and $F2$ for various values of β

of the stationary points should be clear together with the expected effect of parameter β.

Because E_{HM} and F_{U2} coincide we shall speak of *the* free energy of binary Hopfield networks denoting both the free energy of the stochastic network (in mean field approximation) and the energy of the continuous one. We proceed by giving a simple theorem about stability of the motion equations:

Theorem 5. *Using* (6) *as a transfer function, the energy F_{U2} is a Lyapunov function for the motion equations* (2).

Proof. Application of the technique of gradient descent yields

$$\dot{U}_i = -\frac{\partial F_{U2}(\mathbf{V})}{\partial V_i} = \sum_j w_{ij} V_j + I_i - U_i, \qquad (28)$$

so, the motion equations (2) correspond to the free energy function (21). Because the transfer function g defined by (6) is monotone increasing and differentiable, theorem 1 can be applied. □

The above given theorems can be modified in some ways. First, in practice the function to be minimized has a sign opposite to the sign of equation (1). This can be considered as a replacement of w_{ij} by $-w_{ij}$ and of I_i by $-I_i$. A similar effect is produced if the parameter β is replaced by $-\beta$. Let us investigate some of the consequences of the last replacement. In theorem 3, we simply perform the substitution. By this, the function (6) is transferred into

$$\forall i : V_i = \frac{1}{1 + \exp(\beta U_i)} \qquad (29)$$

making it monotone decreasing. This affects theorem 5 and actuated us to introduce a so-called *complementary* energy. The modified version of theorem 5 can be stated as:

Theorem 6. *Using* (29) *as a transfer function, the complementary energy*

$$F_{UC}(\mathbf{V}) = \tfrac{1}{2} \sum_{ij} w_{ij} V_i V_j + \sum_i I_i V_i + \tfrac{1}{\beta} \sum_i (V_i \ln V_i + (1 - V_i) \ln(1 - V_i)) \qquad (30)$$

is a Lyapunov function for the motion equations (2).

Proof. Because (29) is monotone decreasing and differentiable, we see that $\mathrm{dV_i/dU_i} < 0$. Consequently,

$$\dot{F}_{UC} = \sum_i \frac{\partial F_{UC}}{\partial V_i} \dot{V}_i = \sum_i (\sum_j w_{ij} V_j + I_i - U_i) \dot{V}_i = \sum_i \frac{\mathrm{d}V_i}{\mathrm{d}U_i} (\dot{U}_i)^2 \leq 0. \qquad (31)$$

Because F_{UC} is bounded below its value decreases constantly until a (local) minimum has been reached, where $\dot{U}_i = 0$. □

2.3 The General Framework

In this subsection, we introduce a general view on binary Hopfield networks which puts the previous analysis in a wider context, and which appears to be crucial in the constrained case of the next section.

Theorem 7. *The energy of unconstrained binary Hopfield networks can also be stated as*

$$F_{U3}(\mathbf{U}, \mathbf{V}) = -\tfrac{1}{2} \sum_{ij} w_{ij} V_i V_j - \sum_i I_i V_i + \sum_i U_i V_i - \tfrac{1}{\beta} \sum_i \ln(1 + \exp(\beta U_i)). \qquad (32)$$

The stationary points of F_{U3} *are found at points where*

$$\forall i : V_i = \frac{1}{1 + \exp(-\beta U_i)} \ \wedge \ U_i = \sum_j w_{ij} V_j + I_i. \qquad (33)$$

Proof. Application of lemma 4 in its original form to the energy function F_{U2} of theorem 4 immediately yields the free energy expression (32). Resolving the system of equations $\forall i : \partial F_{U3}/\partial V_i = 0, \partial F_{U3}/\partial U_i = 0$ yields the equations (33) as solutions. □

The interesting thing of theorem 7 is the fact that the stationary points of F_{U3} exactly coincide with the conditions of equilibrium of the classical continuous Hopfield model. Knowing this, various methods can be chosen to find the equilibrium points [5]. One of them, of course, consists of Hopfield's updating rules (2). As function F_{U2} is one, so F_{U3} appears to be a Lyapunov function of these motion equations:

Theorem 8. *Using* (6) *as a transfer function, the energy* F_{U3} *is a Lyapunov function for the motion equations* (2).

Proof. Knowing that the transfer function (6) holds and that it is a monotone increasing and differentiable function, it follows that

$$\dot{F}_{U3} = \sum_i \frac{\partial F_{U3}}{\partial V_i}\dot{V}_i + \sum_i \frac{\partial F_{U3}}{\partial U_i}\dot{U}_i \tag{34}$$

$$= \sum_i \left(-\sum_j w_{ij}V_j - I_i + U_i\right)\dot{V}_i + \sum_i \left(V_i - \frac{1}{1+\exp(-\beta U_i)}\right)\dot{U}_i \tag{35}$$

$$= \sum_i \frac{\mathrm{d}V_i}{\mathrm{d}U_i}(\dot{U}_i)^2 \leq 0. \tag{36}$$

In [16, 7], it is proven that for finite values of β the extrema of the energy are never found in the corners of the hypercube $[0,1]^n$ implying that the extrema correspond to *finite* values of U_i which makes F_{U3} bounded below. Therefore, execution of the motion equations (2) constantly decreases the value of F_{U3} until $\forall i : \dot{U}_i = 0$ and a (local) minimum has been reached. □

3 Constrained Stochastic Hopfield Networks

3.1 Methods of Constraint Enforcement

Among other things, Hopfield models are applied to *constrained* optimization problems. The most widely used approach concerns the so-called penalty method, where 'penalty terms' are added to the original energy function [5, 8, 18]. These terms penalize violation of constraints. In practice, it is hard to determine optimal weight values of the penalty terms. Another way to treat the constraints is to use Lagrange multipliers [16]. Then, the constrained optimization problem is converted into an unconstrained extremization one. The correct values of the multipliers are determined by the system itself by performing a gradient ascent. Still another way to deal with the constraints consists of changing the properties of the neural net [2, 10]. Mostly, this is done by restricting the space of allowed states. Instead of allowing the neurons to be 'on' and 'off' independently, only such states are admitted where exactly one of the neurons is 'on'. Physicists call this type of models Potts glasses. We shall analyze this type of networks in the following subsection.

3.2 Constrained Stochastic Networks in Mean Field Approximation

We perform a similar analysis as in the case of unconstrained networks. We consider a binary Hopfield network with stochastic neurons subject to the constraint:

$$\sum_j S_j = 1. \tag{37}$$

This constraint implies that only one of all neurons may be 'on', all the others being 'off'. Therefore, the original state space $\{0,1\}^n$ has been strongly reduced to a constrained one. To put it clearly, the reduced space consists of the admissible n states $(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,0,\ldots,1)$.

Theorem 9. *In mean field approximation, the free energy of stochastic binary Hopfield networks submitted to the constraint* (37) *can be stated as*

$$F_{C1}(\mathbf{V}) = \tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \tfrac{1}{\beta}\ln\big[\sum_i \exp\big(\beta(\sum_j w_{ij}V_j + I_i)\big)\big]. \tag{38}$$

The stationary points of F_{C1} are found at points of the state space where

$$\forall i : V_i = \frac{\exp\left(\beta(\sum_j w_{ij}V_j + I_i)\right)}{\sum_l \exp\left(\beta(\sum_j w_{lj}V_j + I_l)\right)}. \tag{39}$$

Proof. The proof follows the same scheme as the proof of theorem 3. For the same reasons, the exact equation (15) holds. On the other hand, summation over the states of the constrained state space now yields, by using

$$\sum_{\mathbf{S}} \exp\Big(\beta\sum_i S_i(\phi_i + I_i)\Big) = \exp\Big(\ln\sum_i \exp\left(\beta(\phi_i + I_i)\right)\Big), \tag{40}$$

the following expression for the partition function:

$$Z_\beta[\mathbf{I}] = \frac{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j + \ln\sum_i \exp(\beta(\phi_i + I_i))\right]\prod_i \mathrm{d}\phi_i}{\int \exp\left[-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j\right]\prod_i \mathrm{d}\phi_i}. \tag{41}$$

By writing

$$E(\boldsymbol{\phi}, \mathbf{I}) = \tfrac{1}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j - \tfrac{1}{\beta}\ln\sum_i \exp(\beta(\phi_i + I_i)), \tag{42}$$

partial differentiation of $E(\boldsymbol{\phi}, \mathbf{I})$ this time leads to the saddle-point

$$\tilde{\phi}_i = \sum_j w_{ij}\frac{\exp\left(\beta(\tilde{\phi}_i + I_i)\right)}{\sum_l \exp\left(\beta(\tilde{\phi}_l + I_l)\right)}. \tag{43}$$

Applying lemma 3, we find

$$V_i \approx -\frac{\partial E(\tilde{\boldsymbol{\phi}}, \mathbf{I})}{\partial I_i} = \frac{\exp\left(\beta(\tilde{\phi}_i + I_i)\right)}{\sum_l \exp\left(\beta(\tilde{\phi}_i + I_l)\right)}. \tag{44}$$

If we now replace (44) in the exact formula (15) we again obtain the result that, in mean field approximation, $\langle\hat{\boldsymbol{\phi}}\rangle$ coincides with the saddle-point $\tilde{\boldsymbol{\phi}}$. Application of the approximation (20) leads to (38) and partial differentiation of F_{C1} leads to the stationary points (39). □

We already mentioned that $\mathbf{V} \in [0,1]^n$. The constrained subspace $\mathcal{C}$ is defined as the subspace of $[0,1]^n$ for which $\sum_i V_i = 1$.

Theorem 10. *Using the mean field approximation* (39), *the free energy of stochastic binary Hopfield networks submitted to the constraint* (37) *can also be stated as*

$$F_{C2}(\mathbf{V}) = -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \sum_i I_iV_i + \tfrac{1}{\beta}\sum_i V_i \ln V_i. \tag{45}$$

The stationary points of F_{C2}, considered as function over the constrained space $\mathcal{C}$, coincide with the (global) stationary points F_{C1}.

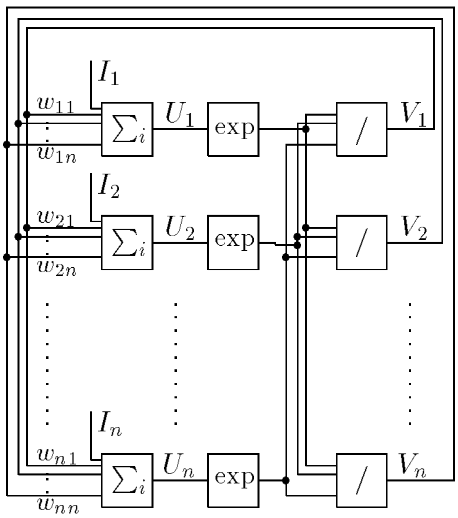

Figure 3: The adapted Hopfield network with equilibrium condition: $\forall i : U_i = \sum_j w_{ij}V_j + I_i$ and $V_i = \exp(U_i)/\sum_l \exp(U_l)$.

Proof. The fact that in mean field approximation equations (39) hold[2] can be proven in the same way as in the unconstrained case. Moreover, with the appropriate substitutions, lemma 5 states:

$$\ln \sum_i \exp\left(\beta(\sum_j w_{ij}V_j + I_i)\right) = -\sum_i V_i \ln V_i + \beta(\sum_{ij} w_{ij}V_iV_j + \sum_i I_iV_i). \tag{46}$$

By combining this result and equation (38) the expression (45) for $F_{C2}(\mathbf{V})$ is found. In order to find the constrained stationary points, a Lagrange multiplier term is added to (45)

[2] In [10], they have been applied in the iterative way that was mentioned in footnote 1.

giving

$$F_{C3}(\mathbf{V}) = -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \sum_i I_iV_i + \tfrac{1}{\beta}\sum_i V_i \ln V_i + \lambda(\sum_i V_i - 1). \qquad (47)$$

By resolving the system of $(n+1)$ equations $\partial F_{C3}/\partial V_i = 0$ and $\partial F_{C3}/\partial\lambda = 0$, we see that the stationary points of F_{C3} are found at state points where (39) holds. □

We note that

$$V_i = \langle S_i\rangle = \mathrm{P}\,(S_i = 1 \wedge \forall j \neq i : S_j = 0). \qquad (48)$$

Furthermore, we see that this time again, the free energy equation (45) has the form (24), where $S = \sum_i V_i \ln V_i$ equals the expression of the entropy of an n-fold source. But, contrary to what we concluded in the unconstrained case, we now see that the neurons have a mutually *dependent* contribution (of $V_i \ln V_i$) to the thermal noise. This is due to the fact that we force them to be mutually dependent by imposing $\sum_i S_i = 1$. The free energy expression $F_{C2}(\mathbf{V})$ of the constrained stochastic binary model coincides, in mean field approximation, with the energy expression E_{HM} of an *adapted* continuous Hopfield model if we take as a transfer function (which follows from equation (39))

$$\forall i : V_i = g(U_i) = \frac{\exp(\beta U_i)}{\sum_l \exp(\beta U_l)} \qquad (49)$$

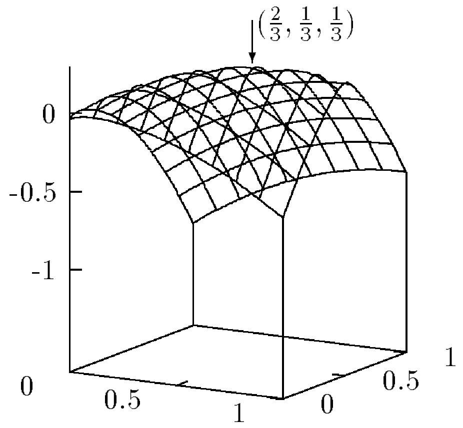

Fig. 4. The free energy expression $FC1$ with global extremum

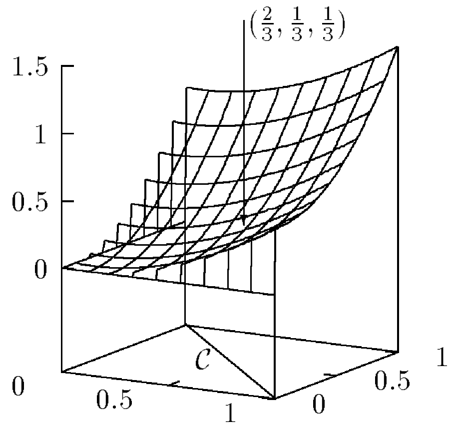

Fig. 5. The free energy expression $FC2$ with constrained extremum

and if we associate the Hopfield term $E_H(\mathbf{V})$ with $-TS = \frac{1}{\beta}\sum_i V_i \ln V_i$. We notice that in this case

$$\tfrac{1}{\beta}\sum_i V_i \ln V_i \neq \sum_i \int_0^{V_i} g^{-1}(V)\mathrm{d}V, \qquad (50)$$

where g equals the transfer function (49). The reason that the inequality (50) holds is that in this case, V_i is a function of $U_1, U_2, \ldots, U_n$ and not of U_i alone. A similar physical interpretation of the model can be given. At high temperatures, the thermal energy dominates the total energy. This yields as equilibrium solution of the system $\forall i : V_i = 1/n$, because then $-TS$ has its constrained minimum value. Lowering the temperature corresponds to a decrease of thermal noise in the system and mean field annealing can be applied.

As in the unconstrained case, we give an example of the theory. Suppose the function to be minimized is

$$E(\mathbf{S}) = \tfrac{1}{2}(S_1^2 + 2S_2^2) \qquad \text{subject to} \;\; S_1 + S_2 = 1, \tag{51}$$

then the corresponding free energy expressions (from theorems 9 and 10) equal

$$FC1(V_1, V_2) = -\tfrac{1}{2}(V_1^2 + 2V_2^2) - \tfrac{1}{\beta}\ln[\exp(-\beta V_1) + \exp(-2\beta V_2)], \tag{52}$$

$$FC2(V_1, V_2) = \tfrac{1}{2}(V_1^2 + 2V_2^2) + \tfrac{1}{\beta}(V_1 \ln V_1 + V_2 \ln V_2). \tag{53}$$

A diagram of these functions is shown in figures 2 and 3, with $\beta = 20$, which corresponds to a low noise level. The arrow denotes the point $(\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$, which is the global, respectively constrained stationary point if noise is neglected. The constrained subspace $\mathcal{C}$ consists of the subspace of $[0,1]^2$ for which $V_1 + V_2 = 1$. In figure 6, $FC1$ and $FC2$ are shown over the constrained subspace $\mathcal{C}$.

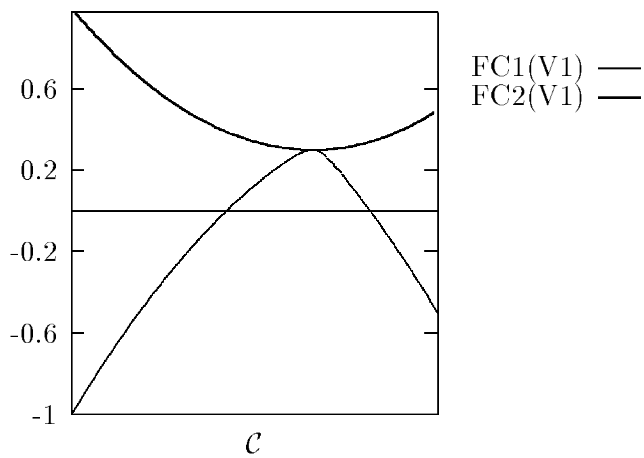

Figure 6: The energy expressions $FC1$ and $FC2$ in the constrained space $\mathcal{C}$

.

Like in the unconstrained case, we shall speak of *the* energy (of binary constrained Hopfield networks) in the rest of this section. The question arises whether we again can prove stability of the motion equations (2).

Theorem 11. *Using* (49) *as a transfer function, the energy* F_{C2} *is a Lyapunov function for the motion equations* (2) *if during updating the Jacobian matrix of* $\mathbf{V}(U_1, U_2, \ldots, U_n)$ *becomes and then remains positive definite.*

Proof. The use of transfer function (49) guarantees, that the solution is sought in the constrained space $\mathcal{C}$. Using lemma 5, we conclude[3] that

$$\frac{\partial}{\partial V_i}\Big(\tfrac{1}{\beta}\sum_i V_i \ln V_i\Big) = \frac{\partial}{\partial V_i}\Big(\sum_i U_i V_i - \tfrac{1}{\beta}\ln\left(1+\exp(\beta U_i)\right)\Big) = U_i, \tag{54}$$

implying that

$$\dot{U}_i = -\frac{\partial F_{C2}(\mathbf{V})}{\partial V_i} = \sum_j w_{ij} V_j + I_i - U_i. \tag{55}$$

Therefore, the motion equations (2) correspond to the free energy function (45). Now, assuming that the Jacobian matrix J of $\mathbf{V}(U_1, U_2, \ldots, U_n)$ becomes positive definite and then remains so during the updating, we can proceed in a similar way as was done in the proof of theorem 6:

$$\begin{aligned} \dot{F}_{C2} &= \sum_i \frac{\partial F_{C2}}{\partial V_i}\dot{V}_i = -\sum_i \big(\sum_j w_{ij}V_j + I_i - U_i\big)\dot{V}_i \\ &= -\sum_i \dot{U}_i \dot{V}_i = -\sum_i \dot{U}_i \sum_j \frac{\partial V_i}{\partial U_j}\dot{U}_j = -\dot{\mathbf{U}}^T J \dot{\mathbf{U}} \leq 0. \end{aligned} \tag{56}$$

The fact that F_{C2} is bounded below completes the proof. □

Whether in general the condition holds that the matrix J will become and remain positive definite, is not easy to say. It turns out (see lemma 6), that all diagonal elements of this matrix are positive, while all non-diagonal elements are negative. Therefore, we decided to do some experiments which are described in section 4. The theorems of this section can be modified in a similar way as the theorems of subsection 2.2. The replacement of β by $-\beta$ changes the transfer function (49) into

$$\forall i : V_i = \frac{\exp(-\beta U_i)}{\sum_l \exp(-\beta U_l)}, \tag{57}$$

changing the sign of all elements of the Jacobian. Under this condition, the previous theorem should be modified into

Theorem 12. *Using* (57) *as a transfer function*[4], *the complementary energy*

$$F_{CC}(\mathbf{V}) = \tfrac{1}{2}\sum_{ij} w_{ij} V_i V_j + \sum_i I_i V_i + \tfrac{1}{\beta}\sum_i V_i \ln V_i \tag{58}$$

is a Lyapunov function for the motion equations (2) *if during updating the Jacobian matrix of* $\mathbf{V}(U_1, U_2, \ldots, U_n)$ *becomes and then remains negative definite.*

Proof. The proof can be done in the same way as the proof of the previous theorem. □

[3] The correctness of this approach becomes more clear in the next subsection about 'The General Framework'.

[4] In [2], this transfer function (with $U_i = \sum_j w_{ij} V_j + I_i$) has been applied in the iterative way that was mentioned in footnote 1.

3.3 The General Framework

In this subsection, we introduce the general view on the binary constrained Hopfield model, putting the previous analysis in a broader context.

Theorem 13. *Using* (49) *as a transfer function, the energy of binary Hopfield networks submitted to the constraint* (37) *can also be stated as*

$$F_{C3}(\mathbf{U},\mathbf{V}) = -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \sum_i I_iV_i + \sum_i V_iU_i - \tfrac{1}{\beta}\ln\big(\sum_i \exp(\beta U_i)\big). \tag{59}$$

The stationary points of F_{C3} *are found at points where*

$$\forall i : V_i = \frac{\exp(\beta U_i)}{\sum_l \exp(\beta U_l)} \;\wedge\; U_i = \sum_j w_{ij}V_j + I_i. \tag{60}$$

Proof. Using lemma 5, the proof can be done in the same way as that of theorem 7. □

Again, we see the interesting phenomenon that the stationary points of an energy function (here, F_{C3}) coincide with the conditions of equilibrium of a Hopfield neural network (here, the constrained model as defined in the beginning of this section). Moreover, function F_{C3} too appears to be a Lyapunov function of the motion equations (2):

Theorem 14. *Using* (49) *as a transfer function, the energy* F_{C3} *is a Lyapunov function for the motion equations* (2) *if during updating the Jacobian matrix of* $\mathbf{V}(U_1, U_2, \ldots, U_n)$ *becomes and then remains positive definite.*

Proof. Assuming that the conditions of the theorem hold we may say:

$$\begin{aligned}
\dot{F}_{C3} &= \sum_i \frac{\partial F_{C3}}{\partial V_i}\dot{V}_i + \sum_i \frac{\partial F_{C3}}{\partial U_i}\dot{U}_i & (61)\\
&= \sum_i \big(-\sum_j w_{ij}V_j - I_i + U_i\big)\dot{V}_i + \sum_i \big(V_i - \frac{\exp(\beta U_i)}{\sum_l \exp(\beta U_l)}\big)\dot{U}_i & (62)\\
&= \sum_i \dot{U}_i \sum_j \frac{\partial V_i}{\partial U_j}(\dot{U}_j) = -\dot{\mathbf{U}}^T J \dot{\mathbf{U}} \le 0. & (63)
\end{aligned}$$

Because F_{C3} is supposed to be bounded below, its value decreases constantly until a (local) minimum has been reached. □

3.4 About the Relation with Elastic Nets

In [14], Simic reveals an interesting result concerning the relation between 'elastic' and 'neural' optimizations. Using the statistical mechanics approach, he 'derives' the Durbin-Willshaw Lyapunov function [3] of the elastic net for solving the TSP, which equals

$$F_{DW}(\mathbf{x}) = \sum_i \tfrac{1}{2} \mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid^2 - \tfrac{1}{\beta}\sum_p \ln \sum_j \exp\big(\tfrac{-\beta^2}{2} \mid \mathbf{x}_p - \mathbf{x}^j \mid^2\big). \tag{64}$$

The basis for this is obvious: both the statistical energy expression (38) and the elastic energy expression (64) are composed of an energy (or cost) term plus a $\ln[\sum \exp()]$-term.

Nevertheless, we think his derivation is false for several reasons. After stating our objections (the mathematical underpinning of which can be found in Appendix B), we argue why we think the elastic net algorithm is a *dynamical penalty* method, where among other things, the penalty weights are dynamically changed by lowering the temperature.
The objections are:

- The Taylor series expansion mentioned on page 97 of [14] is incorrect. The penalty term with weight $\alpha/4$ (see (105) in our Appendix B) must have a minus-sign instead of a plus-sign. Moreover, for high values of β (corresponding to low values of the temperature) the approximation of the expansion does not hold.

- The decomposition (106) of the particle trajectory leading to the Durbin and Willshaw's elastic energy expression, is not applied correctly.

- The last, but possible most important objection: in the statistical mechanics analysis of subsection 3.2, it has been proved that the equilibrium equations of the constrained neural network correspond to stationary points of a corresponding free energy expression. The effect of the $\ln[\sum \exp()]$-term is such that irrespective of the value of the thermal noise, the extrema *automatically* lie within the constrained space. However, in case of the elastic net this condition is not fulfilled. Instead, a *competition* takes place between on the one side, the energy term to be minimized and on the other side, the $\ln[\sum \exp()]$-term which promotes fulfillment of the constraints.

Our conclusion is the following. The last observation about the competition between minimizing the target function and the fulfillment of the constraints, reminds one of the traditional penalty method. The penalty method is usually applied with *quadratic* penalty weights in such a way, that any minimum of the sum of penalty terms corresponds to a 'feasible' solution of the problem [16]. Observing the elastic net algorithm, we conclude that the $\ln[\sum \exp()]$-terms are approximately quadratic and, moreover, that their minima correspond to feasible solutions. This is exactly why the method sometimes works (and why it sometimes, like the penalty method, does not!). However, in contrast to the classical penalty method (where fixed weights are used), here, the penalty weights are dynamically changed during the lowering of the temperature. This actuated us to term the elastic net algorithm a *dynamical penalty* method. The correspondence between the two different methods may be summarized as follows: in the statistical mechanics approach, the 'smoothing effect' of the thermal energy gradually disappears on lowering the temperature, while in the elastic net algorithm, the 'feasibility promoting' effect gradually diminishes on lowering the temperature (although it is to be hoped that the final solution is still feasible). This new view on the elastic net algorithm opens the way to a generalization of the elastic net algorithm to a *dynamical penalty method*: in resolving constraint optimization problems, it should be possible to apply 'problem dependent' dynamical penalty terms, whose influence gradually disappears on lowering the temperature.

4 Review, Experimental Results and Outlook

Reviewing the analyses above, we conclude that either the unconstrained stochastic binary Hopfield network or the treated constrained one, behaves, in mean field approximation, as a specific continuous Hopfield network. In both cases, a corresponding free energy expression can be derived, as well as a complementary version, all with an explicit expression for the energy of the thermal noise. Moreover, both models can be even better understood in a more general framework: in that approach, the energy function has stationary points which coincide with the whole set of equilibrium conditions of the corresponding Hopfield neural network.

To verify the theory about the constrained network, we performed some simple experiments. We tried for example to

$$\text{minimize} \quad V_1^2 + 2V_2^2 + 3V_3^2 + 4V_4^2 \qquad \text{subject to :} \; V_1 + V_2 + V_3 + V_4 = 1. \tag{65}$$

Applying the motion equations (2) in combination with the transfer function (39) as well as with (49) and taking random initializations we found the correct solution in all cases. With $\beta = 20$, which corresponds to a low thermal noise level, the solution $V_1 = 0.471$, $V_2 = 0.244$, $V_3 = 0.163$, $V_4 = 0.122$ is found, which corresponds to the location of the constrained minimum. By taking $\beta = 0.0001$, the equilibrium solution $V_1 = 0.250$, $V_2 = 0.250$, $V_3 = 0.250$, $V_4 = 0.250$ appears, which shows the expected effect of a high thermal noise level.

It is remarkable, that the motion equations (2) of the unconstrained model may still be applied using the constrained model. This raises the question, whether it is generally allowed to change the properties of stochastic Hopfield networks by redefining the transfer function g of the neurons, while adhering the update rule (2). In fact, we think this is simply a generalization of the theorems 11 and 12, yielding a generalized formula (3). This is an interesting subject for future research.

The new view on elastic networks also deserves attention. The observation of this being a dynamical penalty method suggests a research effort of analyzing the effect of existing and new dynamical penalty weights. A separate paper is in preparation, which specifically deals with this subject. Some basic results concerning the analysis of elastic networks can be found in [4]. Recently, we received a paper where the elastic net algorithm is derived from statistical mechanics in a different way using another cost function as a starting point [11]. It looks interesting to compare that approach of so-called deformable templates with our view.

In subsection 3.2, we considered the stochastic neural network submitted to the constraints (37). In resolving constrained optimization problems, one often meets problems with several groups of neurons, each group being submitted to these constraints. If those groups interfere there is no simple solution, because this interference introduces new constraints and the derivations of the previous section no longer hold. In fact, one needs another partition function and a new derivation. This is usually a tough task. But, if this is successful, other constrained optimization problems are within the reach of artificial neural networks.

If, on the other hand, the groups of neurons do not interfere, there is no problem and the given theorems can easily be generalized. This is e.g. the case in the following formulation of the Travelling Salesman Problem:

$$\text{minimize } E(\mathbf{S}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} S_{ij} d_{ik} S_{kj+1}, \tag{66}$$

subject to:

$$\forall i : \sum_{k=1}^{n} S_{ik} - 1 = 0 \quad \text{and} \quad \forall j : \sum_{k=1}^{n} S_{kj} - 1 = 0, \tag{67}$$

where $S_{ij} \in \{0,1\}$. Because the n constraints of the first group are independent, the theory about the constrained network can be used provided that we enforce the constraints of the second group in a different way. In [2, 10] this is done by using a penalty method. We propose to use the Hopfield-Lagrange model [16], because there, the multipliers are determined automatically by the model itself. We have planned to do these experiments in the near future.

Reviewing our derivations, it should be clear that some of them can be sharpened. E.g., we have used the general Hopfield model as the framework of analysis, where the neural network consists of a square of connected neurons, each neuron S_i having one index i. The derived theory is applied on more complex neural networks with two indices like (i and p) for neurons S_i^p (see Appendix B). Of course, this step requires a justification. Similarly, the one-dimensional version of lemma 2 should be generalized to an n-dimensional one. Last but not least, the 'mean field approximation' of lemma 3 can be supported with a more thorough mathematical analysis.

5 Acknowledgements

We thank Marc Leipoldt for reading this article and suggesting changes to the English. We thank Reino de Boer for the same and for his advice on the LaTeX document preparation system.

A Appendix

Lemma 1. *If A is a symmetrical and non-singular matrix then*

$$\exp(\tfrac{\beta}{2}\mathbf{x}^T A\mathbf{x}) = \frac{\int \exp(-\frac{\beta}{2}\boldsymbol{\phi}^T A^{-1}\boldsymbol{\phi} \pm \beta\boldsymbol{\phi}^T\mathbf{x}) \prod_i \mathrm{d}\phi_i}{\int \exp(-\frac{\beta}{2}\boldsymbol{\phi}^T A^{-1}\boldsymbol{\phi}) \prod_i \mathrm{d}\phi_i}. \tag{68}$$

Proof. The lemma is a generalization of the following trick

$$\exp(\tfrac{\beta}{2}x^2) = \frac{\int \exp(-\frac{\beta}{2}\phi^2 \pm \beta\phi x)\mathrm{d}\phi}{\int \exp(-\frac{\beta}{2}\phi^2)\mathrm{d}\phi}. \tag{69}$$

This trick can easily be derived by elaborating the integral of the numerator of the right-hand side. Applying it with

$$xy = \left(\frac{x+y}{2}\right)^2 - \left(\frac{x-y}{2}\right)^2 \tag{70}$$

we can write:

$$\begin{aligned} \exp(\tfrac{\beta}{2}xy) &= \frac{\int \exp[-\frac{\beta}{2}(\phi^2 - \psi^2) \pm \frac{\beta}{2}(\phi x + \phi y - \psi x + \psi y)]\mathrm{d}\phi\mathrm{d}\psi}{\int \exp[-\frac{\beta}{2}(\phi^2 - \psi^2)]\mathrm{d}\phi\mathrm{d}\psi} \\ &= \frac{\int \exp[-\frac{\beta}{2}\tilde{\phi}\tilde{\psi} \pm \frac{\beta}{2}(\tilde{\phi}x + \tilde{\psi}y)]\mathrm{d}\tilde{\phi}\mathrm{d}\tilde{\psi}}{\int \exp[-\frac{\beta}{2}\tilde{\phi}\tilde{\psi}]\mathrm{d}\tilde{\phi}\mathrm{d}\tilde{\psi}}, \end{aligned} \tag{71}$$

where $\tilde{\phi} = \phi - \psi$ and $\tilde{\psi} = \phi + \psi$. We can generalize this result to

$$\exp(\tfrac{\beta}{2}\mathbf{x}^T Ay) = \frac{\int \exp[-\frac{\beta}{2}\boldsymbol{\phi}^T\boldsymbol{\psi} \pm \frac{\beta}{2}(\boldsymbol{\phi}^T\mathbf{x} + \boldsymbol{\psi}^T A\mathbf{y})]\mathrm{d}\boldsymbol{\phi}\mathrm{d}\boldsymbol{\psi}}{\int \exp[-\frac{\beta}{2}\boldsymbol{\phi}^T\boldsymbol{\psi}]\mathrm{d}\boldsymbol{\phi}\mathrm{d}\boldsymbol{\psi}}. \tag{72}$$

Supposing the matrix A is symmetrical and non-singular, we can substitute $\psi \to A^{-1}\psi$ (implying that $\psi^T Ay \to (A^{-1}\psi)^T Ay = \psi^T(A^{-1})^T Ay = \psi^T y$) yielding:

$$\exp(\tfrac{\beta}{2}\mathbf{x}^T Ay) = \frac{\int \exp[-\frac{\beta}{2}\boldsymbol{\phi}^T A^{-1}\boldsymbol{\psi} \pm \frac{\beta}{2}(\boldsymbol{\phi}^T\mathbf{x} + \boldsymbol{\psi}^T\mathbf{y})]\mathrm{d}\boldsymbol{\phi}\mathrm{d}\boldsymbol{\psi}}{\int \exp[-\frac{\beta}{2}\boldsymbol{\phi}^T A^{-1}\boldsymbol{\psi}]\mathrm{d}\boldsymbol{\phi}\mathrm{d}\boldsymbol{\psi}}. \tag{73}$$

Now, by substituting $y \to x$ and by writing $\mathrm{d}\boldsymbol{\phi} = \prod_\mathrm{i} \mathrm{d}\phi_\mathrm{i}$ the theorem is found. □

Lemma 2. *Expansion around the saddle-point of the numerator and the denumerator in the right-hand side makes the following approximation exact*

$$\exp(\tfrac{1}{2}wx^2) = \lim_{a\to\infty} \frac{\int_0^a \exp(-\frac{y^2}{2w} + xy)\mathrm{dy}}{\int_0^a \exp(-\frac{y^2}{2w})\mathrm{dy}}. \tag{74}$$

Proof. Taking $f(y) = \exp(-\frac{y^2}{2w} + xy)$, the saddle-point $\hat{y} = wx$ is found by solving $df(y)/dy = 0$. Application of a Taylor series expansion around the saddle point yields

$$\begin{aligned} f(y) &= f(\hat{y}) + \tfrac{f''(\hat{y})}{2}(y-\hat{y})^2 + \cdots \\ &= f(wx) - \tfrac{f(wx)}{2\beta}(y - wx)^2 + \cdots \end{aligned} \tag{75}$$

It follows that

$$\begin{aligned} \frac{\int_0^a \exp(-\frac{y^2}{2w} + xy)\mathrm{dy}}{\int_0^a \exp(-\frac{y^2}{2w})\mathrm{dy}} &\approx \frac{\int_0^a f(wx)(1 - \frac{(y-wx)^2}{2\beta})\mathrm{dy}}{\int_0^a f(0)(1 - \frac{y^2}{2\beta})\mathrm{dy}} \\ &= \exp(\tfrac{1}{2}wx^2) \times \frac{[y - \frac{(y-wx)^3}{6\beta}]_0^a}{[y - \frac{y^3}{6\beta}]_0^a} \\ &= \exp(\tfrac{1}{2}wx^2) \times \frac{a - \frac{(a-wx)^3}{6\beta} - \frac{(wx)^3}{6\beta}}{a - \frac{a^3}{6\beta}} \\ &= \exp(\tfrac{1}{2}wx^2) \times \left(1 + \frac{\frac{3awx}{6\beta} - \frac{3w^2x^2}{6\beta}}{1 - \frac{a^2}{6\beta}}\right). \end{aligned}$$

Taking the limit with $a \to \infty$, lemma 2 is found. Also, if a larger expansion around the saddle-point is chosen the same result will be found. This completes the proof. □

Lemma 3.

$$V_i = -\langle\frac{\partial E(\boldsymbol{\phi})}{\partial I_i}\rangle \approx -\frac{\partial E(\tilde{\phi})}{\partial I_i}, \tag{76}$$

where

$$\langle\frac{\partial E(\boldsymbol{\phi})}{\partial I_i}\rangle = \frac{\int \frac{\partial E(\boldsymbol{\phi})}{\partial I_i}\exp(-\beta E(\phi))\mathrm{d}\boldsymbol{\phi}}{\int \exp(-\beta E(\phi))\mathrm{d}\boldsymbol{\phi}}, \tag{77}$$

and for the partition function the following equation holds:

$$Z_\beta[\mathbf{I}] = \frac{\int \exp(-\beta E(\boldsymbol{\phi}))\mathrm{d}\boldsymbol{\phi}}{\int \exp(-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j)\mathrm{d}\boldsymbol{\phi}} = \frac{\int \exp(-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j + g(\boldsymbol{\phi},\mathbf{I}))\mathrm{d}\boldsymbol{\phi}}{\int \exp(-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j)\mathrm{d}\boldsymbol{\phi}}, \tag{78}$$

where $g(\boldsymbol{\phi},\mathbf{I})$ is a certain differentiable function of $\boldsymbol{\phi}$ and $\mathbf{I}$.

Proof. Using (10), (78) and (77) we can write:

$$V_i = \tfrac{1}{\beta}\frac{\partial \ln Z_\beta[\mathbf{I}]}{\partial I_i} = -\frac{1}{Z}.\frac{\int \frac{\partial E(\boldsymbol{\phi})}{\partial I_i}\exp(-\beta E(\phi))\mathrm{d}\boldsymbol{\phi}}{\int \exp(-\frac{\beta}{2}\sum_{ij}\phi_i w_{ij}^{-1}\phi_j)\mathrm{d}\boldsymbol{\phi}} \tag{79}$$

$$= -\langle\frac{\partial E(\boldsymbol{\phi})}{\partial I_i}\rangle. \tag{80}$$

This is the proof of the first part of the lemma.

For the proof of the second part we use the 'saddle-point method' [5]. Then, $E(\boldsymbol{\phi})$ is approximated by $E(\tilde{\phi})$, where $\tilde{\phi}$ equals the saddle-point, so we may write $E(\boldsymbol{\phi}) \approx E(\tilde{\phi})$. Using this approximation, we find that

$$Z_\beta[\mathbf{I}] \approx \frac{\int \exp(-\beta E(\tilde{\phi}))\mathrm{d}\boldsymbol{\phi}}{\int \exp(-\beta E(0))\mathrm{d}\boldsymbol{\phi}} = \exp(-\beta E(\tilde{\phi}). \tag{81}$$

Substituting this result in (10), we find:

$$V_i = \tfrac{1}{\beta}\frac{\partial \ln Z_\beta[\mathbf{I}]}{\partial I_i} \approx -\frac{\partial E(\tilde{\phi})}{\partial I_i}. \tag{82}$$

This completes the proof. □

Lemma 4. *If*

$$m = \frac{1}{1+\exp(-a)}, \tag{83}$$

then

$$\ln(1+\exp(a)) = -m\ln m - (1-m)\ln(1-m) + ma. \tag{84}$$

Proof. Equation (83) implies that

$$1 - m = \frac{1}{1+\exp(a)}. \tag{85}$$

Using (83) and (85), we can proof the lemma directly:

$$\begin{aligned}
&-m\ln m - (1-m)\ln(1-m) + ma = \\
&= \frac{-1}{1+\exp(-a)}\ln(\frac{-1}{1+\exp(-a)}) - \frac{-1}{1+\exp(a)}\ln(\frac{-1}{1+\exp(a)}) + \frac{a}{1+\exp(-a)} \\
&= \frac{\ln(1+\exp(-a)) + \ln\exp(a)}{1+\exp(-a)} + \frac{\ln(1+\exp(a))}{1+\exp(a)} \\
&= \ln(1+\exp(a))(m+1-m) = \ln(1+\exp(a)).
\end{aligned} \tag{86}$$

□

Lemma 5. *If*

$$V_i = \frac{\exp(\pm\beta U_i)}{\sum_l \exp(\pm\beta U_l)}, \tag{87}$$

then

$$\ln \sum_i \exp(\pm\beta U_i) = -\sum_i V_i \ln V_i \pm \sum_i \beta U_i V_i. \tag{88}$$

Proof. From equation (87) it follows that

$$\pm\beta U_i = \ln\,(V_i \sum_l \exp(\pm\beta U_l)). \tag{89}$$

Using this result and the fact that $\sum_i V_i = 1$ we can write

$$\pm \sum_i \beta U_i V_i = \sum_i \ln\,(V_i \sum_l \exp(\pm\beta U_l)) V_i \tag{90}$$

$$= \sum_i V_i \ln V_i + \sum_i V_i \ln(\sum_l \exp\,(\,\pm\,\beta U_l)) \tag{91}$$

$$= \sum_i V_i \ln V_i + \ln\,(\sum_l \exp(\pm\beta U_l)). \tag{92}$$

By rewriting this equation, the lemma is found immediately. □

Lemma 6. *If* (87) *holds, if* $l \geq 2$, *and if* $l \neq i$, *then*

$$\frac{\partial V_i}{\partial U_i} = \beta V_i (1 - V_i) > 0 \quad \text{and} \quad \frac{\partial V_i}{\partial U_l} = -\beta V_i V_l < 0. \tag{93}$$

Proof.

$$\frac{\partial V_i}{\partial U_i} = \frac{\sum_l \exp\,(\beta U_l).\exp\,(\beta U_i).\beta - \exp\,(\beta U_i).\exp\,(\beta U_i).\beta}{(\sum_l \exp\,(\beta U_l))^2} \tag{94}$$

$$= \frac{\beta \exp\,(\beta U_i).(\sum_l \exp\,(\beta U_l) - \exp\,(\beta U_i)}{(\sum_l \exp\,(\beta U_l))^2} \tag{95}$$

$$= \frac{\beta \exp\,(\beta U_i).\sum_{l\neq i} \exp\,(\beta U_l)}{(\sum_l \exp\,(\beta U_l))^2} = \beta V_i (1 - V_i) > 0. \tag{96}$$

The second result is found in the same way. Taking $l \neq i$ we find

$$\frac{\partial V_i}{\partial U_l} = \frac{0 - \exp\,(\beta U_i).\exp\,(\beta U_l).\beta}{(\sum_l \exp\,(\beta U_l))^2} - \beta V_i V_l < 0. \tag{97}$$

B Appendix

Let us start by briefly recaputilating Simic's approach [14]. In order to solve the classical Travelling Salesman Problem (TSP), a 'statistical mechanics' is defined regarding 'particle trajectories' as an 'ensemble', where the paths of legal trajectories must obey the global constraints of the TSP: the particle (salesman) cannot visit two space-points (cities) at the same time and it (he) visits all the points (cities) once and only once. The legal trajectory with the shortest path length equals the optimal tour for the travelling salesman and that is the solution we are trying to find.

A part of the constraints is enforced 'strongly' by summing only over those configurations which obey that part of the constraints guaranteeing that all space points (cities) are visited once and only once. The other part of the constraints is enforced 'softly' by adding a penalty term in order to guarantee that at any time, one and only one city is visited. If S_p^i denotes whether the particle at time i occupies space-point p ($S_p^i = 1$) or not ($S_p^i = 0$), and if d_{pq} is the distance between points p and q, then the corresponding energy function (1) of the particle trajectory equals

$$E(\mathbf{S}) = \tfrac{1}{4} \sum_i \sum_{pq} d_{pq}^2 S_p^i (S_q^{i+1} + S_q^{i-1}) + \tfrac{\alpha}{4} \sum_i \sum_{pq} d_{pq}^2 S_p^i S_q^i. \tag{98}$$

Here, the first term represents the sum of distance-squares of the particle. The second term is the penalty term which penalizes the simultaneous presence of a particle at more than one position. Now, the statistical mechanics approach of the constrained model of section 3.2 can be applied. Using the cost function (98), the following expression of the free energy is obtained:

$$\begin{aligned} F(\mathbf{V}) \;=\; & -\tfrac{1}{4} \sum_i \sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) - \tfrac{\alpha}{4} \sum_i \sum_{pq} d_{pq}^2 V_p^i V_q^i - \\ & \tfrac{1}{\beta} \sum_p \ln \left[\sum_i \exp(-\tfrac{\beta}{2} \sum_q d_{pq}^2 (\alpha V_q^i + V_q^{i+1} + V_q^{i-1})) \right]. \end{aligned} \tag{99}$$

This free energy expression can be seen as a special case of the general energy function:

$$F_C(\mathbf{V}) = -\tfrac{1}{2} \sum_{ij} \sum_{pq} w_{pq}^{ij} V_p^i V_q^j - \tfrac{1}{\beta} \sum_p \ln \left[\sum_i \exp(-\beta \sum_{jq} w_{pq}^{ij} V_q^j) \right]. \tag{100}$$

In order to derive an energy expression in the standard form (24), Simic applies a Taylor series expansion on the last term of equation (99). Taking

$$f(\mathbf{V}) \;=\; \sum_p \ln[\sum_i \exp(V_p^i)], \tag{101}$$

$$a_p^i \;=\; -\beta \tfrac{\alpha}{2} \sum_q d_{pq}^2 V_q^i, \quad \text{and} \tag{102}$$

$$h_p^i \;=\; -\beta \tfrac{1}{2} \sum_q d_{pq}^2 (V_q^{i+1} + V_q^{i-1}), \tag{103}$$

he obtains

$$F(\mathbf{V}) \;\approx\; \sum_p \ln \left[\sum_i \exp(a_p^i) \right] + \sum_{ip} h_p^i \frac{\partial f}{\partial V_p^i}(a_p^i) \tag{104}$$

$$\begin{aligned} \;=\; & \tfrac{1}{4} \sum_i \sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) + \tfrac{\alpha}{2} \sum_i \sum_{pq} d_{pq}^2 V_p^i V_q^i - \\ & \tfrac{1}{\beta} \sum_p \ln \sum_i \exp\left(- \beta \tfrac{\alpha}{2} \sum_q d_{pq}^2 V_q^i \right). \end{aligned} \tag{105}$$

In our derivation, we found a slightly different expression with the weight value $-\frac{\alpha}{4}$ instead of the value $+\frac{\alpha}{2}$. Moreover, inspection of equation (103) reveals, that the chosen Taylor-approximation does not hold for low values of the temperature, i.e., high values of β. This underpins the first objection of subsection 3.4.

In order to transform the Hopfield network formulation of the TSP into an elastic net, Simic performs a 'decomposition of the particle trajectory':

$$\mathbf{x}^i = <\mathbf{x}(i)> = \sum_p \mathbf{x}_p <S_p^i> = \sum_p \mathbf{x}_p V_p^i. \tag{106}$$

Here, $\mathbf{x}(i)$ is the position of the particle at time i, $\mathbf{x}_p$ is the vector denoting the position of space-point p, and $\mathbf{x}^i$ denotes the *average* position of the particle at time i. Using the decomposition, he obtains the free energy expression of the elastic net algorithm

$$F(\mathbf{x}) = \sum_i \frac{1}{2} \mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid^2 - \frac{1}{\beta} \sum_p \ln\left[\sum_j \exp(-\beta \frac{\alpha}{2} \mid \mathbf{x}_p - \mathbf{x}^j \mid^2)\right]. \tag{107}$$

However, careful analysis shows that in general

$$\sum_q d_{pq}^2 V_q^i = \sum_q (\mathbf{x}_p - \mathbf{x}_q)^2 V_q^i \neq \mid \mathbf{x}_p - \mathbf{x}^i \mid^2 . \tag{108}$$

If the constraints are fulfilled, the inequality sign must be replaced by the equality sign, but in general the inequality holds. This motivates our second objection against Simic's result.

Thirdly, calculation of the stationary points of equation (105) (or, equation (107)) yields that the stationary points of the energy function do not correspond automatically to constraints of the form (37). Therefore, the effect of the $\sum \ln[\sum \exp()]$-term differs from the effect in the statistical mechanics approach. This explains our third objection against Simic's derivations and conclusions.

References

[1] A.R. Bizzarri. Convergence Properties of a Modified Hopfield-Tank Model. *Biological Cybernetics*, 64:293–300, 1991.

[2] D.E. Van den Bout and T.K. Miller. Improving the Performance of the Hopfield-Tank Neural Network Through Normalization and Annealing. *Biological Cybernetics*, 62:129–139, 1989.

[3] R. Durbin and D. Willshaw. An Analogue Approach of the Travelling Salesman Problem Using an Elastic Net Method. *Nature*, 326:689–691, 1987.

[4] J.H. Geselschap. Een Verbeterd 'Elastic Net' Algoritme (An Improved Elastic Net Algorithm). Master's thesis, Erasmus University Rotterdam, Comp. Sc. Dept. (AI Section), Faculty of Economics, 1994.

[5] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, The Advanced Book Program, 1991.

[6] J.J. Hopfield. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, USA 79:2554–2558, 1982.

[7] J.J. Hopfield. Neurons with Graded Responses Have Collective Computational Properties Like Those of Two-State Neurons. *Proceedings of the National Academy of Sciences*, USA 81:3088–3092, 1984.

[8] J.J. Hopfield and D.W. Tank. “Neural” Computation of Decisions in Optimization Problems. *Biological Cybernetics*, 52:141–152, 1986.

[9] A. Joppe, H.R.A. Cardon, and J.C. Bioch. A Neural Network for Solving the Travelling Salesman Problem on the Basis of City Adjacency in the Tour. In *Proceedings of the International Joint Conference on Neural Networks*, pages 961–964, San Diego, june 1990.

[10] C. Peterson and B. Söderberg. A New Method for Mapping Optimization Problems onto Neural Networks. *International Journal of Neural Systems*, 1:3–22, 1989.

[11] C. Peterson and B. Söderberg. Artificial Neural Networks and Combinatorial Optimization Problems. To appear in E.H.L. Aarts and J.K. Lenstra, editors, *Local Search in Combinatorial Optimization*, John Wiley & Sons, New York.

[12] J.C. Platt and A.H. Barr. Constrained Differential Optimization. *Proceedings of the IEEE 1987 NIPS Conference*, 1988.

[13] D. Sherrington. Neural networks: The Spin Glass Approach. In J.G. Taylor, editor, *Mathematical Approaches to Neural Networks*, pages 261–292. North-Holland, 1993.

[14] P.D. Simic. Statistical Mechanics as the Underlying Theory of ‘Elastic’ and ‘Neural’ Optimisations. *Network*, 1:88–103, 1990.

[15] J. van den Berg and J.C. Bioch. Capabilities of the Symbiosis between the Hopfield Model and Lagrange Multipliers in Resolving Constrained Optimization Problems. *Proceedings of Computing Science in the Netherlands 1994*, pages 54–65, Utrecht, 1994.

[16] J. van den Berg and J.C. Bioch. Constrained Optimization With the Hopfield-Lagrange Model. In *Proceedings of the 14th IMACS World Congress*, pages 470–473, Atlanta, GA 30332 USA, 1994.

[17] E. Wacholder, J. Han, and R.C. Mann. A Neural Network Algorithm for the Traveling Salesman Problem. *Biological Cybernetics*, 61:11–19, 1989.

[18] G.V. Wilson and G.S. Pawley. On the Stabilitity of the Travelling Salesman Problem Algorithm of Hopfield and Tank. *Biological Cybernetics*, 58:63–70, 1988.