An analysis of various elastic net algorithms

Jan van den Berg and Jock H. Geselschap

Department of Computer Science, Faculty of Economics, Erasmus University, Room H4–13, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
E-mail: berg,bigesel@cs.few.eur.nl

ABSTRACT

The Elastic Net Algorithm (ENA) for solving the Traveling Salesman Problem is analyzed applying statistical mechanics. Using some general properties of the free energy function of stochastic Hopfield Neural Networks, we argue why Simic's derivation of the ENA from a Hopfield network is incorrect. However, like the Hopfield-Lagrange method, the ENA may be considered a specific *dynamic penalty method*, where, in this case, the weights of the various penalty terms decrease during execution of the algorithm. This view on the ENA corresponds to the view resulting from the theory on 'deformable templates', where the term *stochastic penalty method* seems to be most appropriate.
Next, the ENA is analyzed both on the level of the energy function as well as on the level of the motion equations. It will be proven and shown experimentally, why a non-feasible solution is sometimes found. It can be caused either by a too rapid lowering of the temperature parameter (which is avoidable), or by a peculiar property of the algorithm, namely, that of *adhering to equidistance* of the elastic net points.
Thereupon, an alternative, *Non-equidistant* Elastic Net Algorithm (NENA) is presented and analyzed. It has a correct distance measure and it is hoped to guarantee feasibility in a more natural way. For small problem instances, this conjecture is confirmed experimentally. However, trying larger problem instances, the pictures changes: our experimental results show that the elastic net points appear to become 'lumpy' which may cause non-feasibility again. Moreover, in cases both algorithms yield a feasible solution, the quality of the solution found by the NENA is often slightly worse than the one found by the original ENA. This motivated us to try an *Hybrid* Elastic Net Algorithm (HENA), which starts using the ENA and, after having found a good approximate solution, switches to the NENA in order to guarantee feasibility too. In practice, the ENA and HENA perform more or less the same. Up till now, we did not find parameters of the HENA, which invariably guarantee the desired feasibility of solutions.

1. Motivation and results

Artificial Neural Networks (ANN's) are sometimes used to find good solutions to combinatorial optimization problems: for a survey we refer to [6,3]. The approach differs substantially from other search, mostly heuristic methods. The ANN approach often has a special statistical mechanics interpretation using *mean field* theory. The dynamics of the neural network should be stable in such way that the final, equilibrium state corresponds to a good approximate solution of the optimization problem. The ANN approach is attractive because of the potentiality of straightforward hardware implementations. Many solution methods using neural nets are modifications of one of the Hopfield models [7,8]. The most widely used approach concerns the 'penalty' method, where penalty terms are added to the original energy function [6,9,21]. These terms penalize violation of constraints. This type of constraints enforcement is sometimes indicated as the 'soft' one. 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 [17]. Then, the constrained optimization problem is converted into an unconstrained extremization one. The correct values of the multipliers are determined automatically using a gradient ascent. Still another way to deal with the constraints consists of changing the properties of the neural net [20,12,18,19]. 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'. In physics, these models are called Potts glass models. The type of constraints enforcement is sometimes indicated as the 'strong' one. We shall use this type of networks in our analysis.

An alternative, very specific ANN for solving the classical Travelling Salesman Problem (TSP), the so-termed Elastic neural Network Algorithm (ENA), was introduced by Durbin and Willshaw [4]. Like a 'Feature Map' ANN [6], the ENA tries to find a *topology preserving map* between two spaces, in this case between a plane and a line. Durbin and Willshaw derived their ENA from a hypothetical 'tea trade model' [10]. The ENA has an important *scaling* property: the number of variables (2-dimensional units) needed is linear relative to the number of cities, while in the case of the Hopfield model the number of neurons needed is quadratic relative to that number. Some researchers [15,2] have proposed small modifications of the original algorithm. Analysis shows that, contrary to the conclusions of the various designers, these modified algorithms achieve less [5]. In 1990, Simic [14] introduced an interesting idea of using statistical mechanics as the underlying theory of both the Hopfield ANN approach (and modifications) and the ENA approach. Recently, we came across another unifying approach called 'deformable templates' [13] which applies statistical mechanics to another energy function. The overall approach of this method is that a part of the constraints is enforced in the strong way, while the remaining part is supplied with noise keeping the original cost function unattached.

In this paper, we first argue that Simic's derivation of the ENA (from a certain constrained Hopfield model) is incorrect. In our view, the ENA should be considered as a specific *dynamic penalty model*, where the weights of the various penalty terms decrease during the execution of the algorithm. If, finally, a feasible solution is reached, the energy function is mainly determined by the original cost function. The other, at that time less important term is a sum of small penalty weights causing small pits in the energy landscape of the original cost function. It is hoped that the system will have settled down in a small pit close to the global constrained minimum. The new view – the ENA being a dynamic penalty method – opens up new avenues for search for formulations of the TSP using other dynamic penalty terms. The 'dynamic penalty' view corresponds to the approach as applied in the theory on deformable templates. There, the elastic net algorithm is derived from a different energy function composed of a cost function to be minimized as well as a sum of penalty terms of which each one enforces a match between a fixed city and one of the variable net points. Since stochastic noise is here only added to the penalty terms leaving the cost function unattached, we introduce the terminology of 'stochastic penalty terms'.

Secondly, using the analysis method as introduced in [5], the original ENA will be discussed both on the level of the energy landscape and on the level of the updating rule. It will be proven and shown experimentally, why a non-feasible solution is sometimes found by the ENA. This is caused by either a too rapid lowering of the temperature parameter (which can easily be avoided at the cost of a certain amount of additional computation time), or by a peculiar property of the algorithm namely, that of adhering to *equidistance* of the elastic net points. Initially, when the elastic net is stretching out, this property of equidistance may be useful to achieve a good 'first order approximation' of the final tour of all cities. Eventually, the equidistance property is unnecessary and may even be the cause of persisting in a non-feasible solution. The tendency towards equidistance has to do with with the fact that the tour length is expressed as a sum of square distances between the succeeding elastic net points instead of as a sum of linear distances.

Knowing this, we searched for an alternative, non-equidistant algorithm (NENA) which has a correct distance measure and which guarantees feasibility in a better way. We shall present our first idea and give the experimental results with this new algorithm. This time, the elastic net points appear to become 'lumpy' due to the fact that the mutual attracting forces are strongly reduced. For small problem instances, the NENA works fine and better than the original ENA. However, trying larger problem instances, the new algorithm may lead to non-feasibility too. And, the larger the problem instance, the easier the NENA appear to yield a non-valid solution. Moreover, in cases both the ENA and the NENA yield a feasible solution, the quality of the solution found by the last one is usually slightly worse than the solution found by the original one. Therefore, we decided to try a hybrid algorithm (HENA), which was hoped to combine the good properties of the old and new ENA-algorithm. The HENA starts using the ENA in order to achieve a steady stretching out of the elastic net (corresponding to solutions of high quality) and, thereafter, switches to the NENA in order to guarantee feasibility at the end as well. Unfortunately, up to now, we did not find a larger problem instance (a 100-city problem), where the HENA performed better than the original ENA.

In the final section, we evaluate our work calling to mind that static as well as stochastic penalty methods always have to be tuned carefully. Because elastic networks appear to behave as a (stochastic) penalty method, they inherently have a parameter tuning problem. We further conclude, that the quadratic distance measure of the ENA is an essential ingredient of it.

2. Hopfield and elastic networks

In this section, we touch upon the background theories of this article. For more details about the general aspects, we refer to [6,14,16]. More specific references are mentioned in the text. We start considering binary Hopfield networks. Secondly, we dwell on 'elastic networks'.

2.1. Statistical mechanics of Hopfield networks

In 1982, Hopfield introduced the idea of an 'energy function' into neural network theory using an asynchronous updating rule and binary units [7]. Like Simic, we shall use Hopfield's energy expression multiplied by minus one, that is:

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

where $\mathbf{S} \in \{0,1\}^n$ is the state vector $(S_1, \ldots, S_n)$ of the neural network, S_i the output value and I_i the external input of neuron i and where w_{ij} represents the interconnection strength from neuron j to neuron i. $E(\mathbf{S})$ equals the energy or 'cost function' to be minimized. In this paper, we suppose that $\forall i, j : w_{ij} \geq 0$.

Making the units stochastic, supposing random fluctuations, the stochastic neural net can be analyzed applying statistical mechanics. The starting point of statistical mechanics is always an energy function often called the Hamiltonian denoted by H_α, in this paper denoted by $E(\mathbf{S})$ [a]. Next, the central expression to calculate is the so-called partition function Z_β defined by

$$Z_\beta = \sum_{\mathbf{S}} \exp(-\beta E(\mathbf{S})), \tag{2}$$

where the summation takes place over the set of states $\mathbf{S}$ of the system. Related to the partition function is the thermodynamic free energy defined by $F_\beta = -T\ln(Z_\beta)$, where $T = 1/\beta$ is the 'temperature' of the system. It can also be written as

$$F = \langle E(\mathbf{S})\rangle - T\mathcal{S}^{eq} = \sum_{\mathbf{S}} P^{eq}(\mathbf{S})E(\mathbf{S}) + T\sum_{\mathbf{S}} P^{eq}(\mathbf{S})\ln P^{eq}(\mathbf{S}). \tag{3}$$

$\langle E(\mathbf{S})\rangle$ represents the average energy of the system at thermal equilibrium, $\mathcal{S}^{eq}$ is the so-called entropy at thermal equilibrium, and $P^{eq}(\mathbf{S})$ is the probability of finding the system in state $\mathbf{S}$ at thermal equilibrium. The importance of the free energy notion comes from a variational formulation [11] on it, called the principle of minimal free energy $F(P)$. In this case, F is considered as a function of an *arbitrary* probability distribution P over the states of the system:

$$F(P) = E(P) - T\mathcal{S}(P) = \sum_{\mathbf{S}} P(\mathbf{S})E(\mathbf{S}) + T\sum_{\mathbf{S}} P(\mathbf{S})\ln P(\mathbf{S}) \tag{4}$$

The principle states that a minimum of the free energy $F(P)$ corresponds to a (dynamic) equilibrium state of the thermodynamic system. Therefore, the free energy can be used to find such a state. Moreover, at high temperatures, the surface of the free energy function appears to be much smoother than the energy surface of the original energy function $E(\mathbf{S})$: this is caused by a high level of the thermal energy under these circumstances. On lowering the temperature, the smoothing effect diminishes gradually and fine details of the energy landscape appear, while the free energy goes over to the original energy function. At the end, it is hoped that the system will not

[a] In statistical mechanics, the probability distribution *in equilibrium* is supposed to be proportional to the number of different ways the 'particles' can be divided over the various energy levels. In these considerations, the driving mechanism by which the articles of the system – on account of their mutual interaction – are divided over the available energy levels is ignored. However, one can construct various dynamics [11] which have the property of leading to 'thermal equilibrium'.

be caught in a local minimum, but that it has reached its ground state, i.e., the state with the lowest cost. In fact, these properties are the background of *simulated* and *mean field annealing*.

In [16], a thorough analysis is presented of both unconstrained and some constrained binary Hopfield neural networks. Among other things, the following theorems are proved:

Theorem 1. *In mean field approximation, the free energy of unconstrained stochastic binary Hopfield networks equals*

$$F_u(\mathbf{V}) \quad = \quad -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \tfrac{1}{\beta}\sum_i \ln[1+\exp(-\beta(\sum_j w_{ij}V_j + I_i))], \tag{5}$$

where $\forall i : V_i = \mathrm{P}(S_i = 1)$. *The stationary points of* F_u *are found at points of the state space where*

$$\forall i : V_i = \frac{1}{1+\exp(\beta(\sum_j w_{ij}V_j + I_i))}. \tag{6}$$

Theorem 2. *In mean field approximation, the free energy of constrained stochastic binary Hopfield networks, submitted to the constraint*

$$\sum_i S_i = 1 \tag{7}$$

equals

$$F_c(\mathbf{V}) = -\tfrac{1}{2}\sum_{ij} w_{ij}V_iV_j - \tfrac{1}{\beta}\ln[\sum_i \exp(-\beta(\sum_j w_{ij}V_j + I_i))], \tag{8}$$

where $\forall i : V_i = \mathrm{P}(S_i = 1 \ \wedge \ \forall j \neq i : S_j = 0)$. *The stationary points of* F_c *are found at points of the state space where*

$$\forall i : V_i = \frac{\exp(-\beta(\sum_j w_{ij}V_j + I_i))}{\sum_l \exp(-\beta(\sum_j w_{lj}V_j + I_l))}. \tag{9}$$

The first theorem can be used to solve the TSP using the 'soft' approach with penalty terms. A generalization of the second theorem can be used to solve the TSP using a combination of the 'strong' and the 'soft' approach: if S_p^i denotes whether the salesman 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 Hamiltonian may be formulated as [14]:

$$E(\mathbf{S}) = \tfrac{1}{4}\sum_i \sum_{pq} d_{pq}^2 S_p^i (S_q^{i+1} + S_q^{i-1}) + \tfrac{\alpha}{4}\sum_i \sum_{pq} d_{pq}^2 S_p^i S_q^i. \tag{10}$$

The first term represents the sum of distance-squares between visited cities, while the second term is a penalty term which penalizes the simultaneous presence of the salesman at more than one position. Note, that in this formulation the dimension of the energy surface is quadratic in the number of cities, because there are N^2 neurons S_p^i (the index p as well as the index i ranges from 1 to N). The other constraints, which should guarantee that every city is visited once and only once, can be 'strongly' fulfilled. Using the cost function (10) and applying a generalization of theorem 2, the following expression of the free energy can be obtained [14,16]:

$$\begin{aligned} F_{tsp1}(\mathbf{V}) &= -\tfrac{1}{4}\sum_i \sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) - \tfrac{\alpha}{4}\sum_i \sum_{pq} d_{pq}^2 V_p^i V_q^i - \\ & \tfrac{1}{\beta}\sum_p \ln\left[\sum_i \exp(-\tfrac{\beta}{2}\sum_q d_{pq}^2(\alpha V_q^i + V_q^{i+1} + V_q^{i-1}))\right]. \quad (11) \end{aligned}$$

This energy expression has been used by Simic [14] to derive the elastic net algorithm. However, we think his derivation is not correct (section 3).

We note, that a more 'natural' mean field approximation of the free energy may be obtained as follows:

$$F_{tsp2}(\mathbf{V}) = \tfrac{1}{4}\sum_i \sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) + \tfrac{\alpha}{4}\sum_i \sum_{pq} d_{pq}^2 V_p^i V_q^i + \tfrac{1}{\beta}\sum_{pi} V_i^p \ln V_i^p, \quad (12)$$

which has the structure of equation (3), where the entropy equals $\mathcal{S} = -\sum_{pi} V_i^p \ln V_i^p$. However, this free energy expression is not used in the rest of this paper.

2.2. The original elastic net algorithm

The term 'elastic algorithm' has been introduced by Durbin and Willshaw [4] and deals with a specific type of neural network for solving the TSP. The elastic net algorithm was derived from a hypothetical tea trade model. Here, we merely mention the main results. The energy function [b] to be minimized of the elastic net equals:

$$F_{en}(\mathbf{x}) = \tfrac{\alpha_2}{2}\sum_i \mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid^2 - \tfrac{\alpha_1}{\beta}\sum_p \ln \sum_j \exp(\tfrac{-\beta^2}{2} \mid \mathbf{x}_p - \mathbf{x}^j \mid^2). \quad (13)$$

Here, $\mathbf{x}^i$ represents the i-th elastic net point (the succeeding M elastic net points form a ring) and $\mathbf{x}_p$ represents the location of city p. Application of the gradient descent method on equation (13) yields the updating rule:

$$\Delta \mathbf{x}^i = \tfrac{\alpha_2}{\beta}(\mathbf{x}^{i+1} - 2\mathbf{x}^i + \mathbf{x}^{i-1}) + \alpha_1 \sum_p \Lambda^p(i)(\mathbf{x}_p - \mathbf{x}^i), \quad (14)$$

where the time-step $\Delta t = 1/\beta$ equals the current temperature T and where

$$\Lambda^p(i) = \frac{\exp(-\frac{\beta^2}{2} \mid \mathbf{x}_p - \mathbf{x}^i \mid^2)}{\sum_l \exp(-\frac{\beta^2}{2} \mid \mathbf{x}_p - \mathbf{x}^l \mid^2)}. \quad (15)$$

In practice, all $\mathbf{x}_p$ should be normalized to points in the unit square. In that case, the following parameter values appear to be efficient [4]: $\alpha_1 = 2.0$ and $\alpha_2 = 0.2$. The initial value of the temperature $T = 1/\beta$ is set to 0.2, and is reduced by 1% every n iterations to a final value in the range 0.01-0.02. As will be shown, the general effect of this lowering is that large-scale, global adjustments occur early on, resulting in a general

[b] In the original paper [4], this energy function is denoted by an E. Since Simic conjectured that the energy was a type of free energy, he chose the notation F, which we adopt here too.

stretching out of the elastic net. Later on, smaller refinements occur corresponding to a more local adaption of the elastic net towards city points.

3. Why the ENA is a dynamic penalty method

In this section, we shall argue why Simic's derivation of ENA from a certain constrained stochastic Hopfield model is invalid. However, the analysis yields another view on the relationship: the ENA should be considered as a dynamic penalty method.

In order to simplify the formulas somewhat, we provisionally set the parameter values $\alpha_1 = \alpha_2 = 1$. We start by briefly recapitulating 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. Part of the constraints is enforced 'strongly' by summing only over those configurations which guarantee 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. The corresponding energy function or Hamiltonian equals (10). We already mentioned in the previous section that, in mean field approximation, the free energy may be stated as equation (11) [c]. From now on, we continue to sketch Simic's derivation (eventually resulting in the ENA) as well as to formulate our objections against it.

Objection 1. In order to derive an energy expression in the standard form of the free energy ($F = \langle E(\mathbf{S})\rangle - TS$), Simic applies a Taylor series expansion on the last term of equation (11). We tried to do the same. Taking

$$f(\mathbf{x}) = \tfrac{1}{\beta}\sum_p \ln\,[\sum_i \exp(x_p^i)], \tag{16}$$

$$a_p^i = -\beta\tfrac{\alpha}{2}\sum_q d_{pq}^2 V_q^i, \quad \text{and} \tag{17}$$

$$h_p^i = -\beta\tfrac{1}{2}\sum_q d_{pq}^2 (V_q^{i+1} + V_q^{i-1}), \tag{18}$$

we found using the mean field equation (9):

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= \tfrac{1}{\beta}\sum_p \ln\,[\sum_i \exp(a_p^i)] + \tfrac{1}{\beta}\sum_{ip} h_p^i \frac{\partial f}{\partial x_p^i}(a_p^i) + \mathcal{O}(\mathbf{h}^2) & (19)\\ &\approx \tfrac{1}{\beta}\sum_p \ln \sum_i \exp\big(-\beta\tfrac{\alpha}{2}\sum_q d_{pq}^2 V_q^i\big) - \tfrac{1}{2}\sum_i\sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}). & (20)\end{aligned}$$

[c] It is interesting to note Simic's observation that expression (11) has the 'wrong' sign: indeed, the structure of that equation suggests, that stationary points in that case correspond to *maxima*, while those of the ENA are *minima*. Especially this phenomenon aroused our suspicions regarding the derivation.

Substitution of this result in (11) eventually yields:

$$F_{ap}(\mathbf{V}) \;=\; \tfrac{1}{4}\sum_i\sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) - \tfrac{\alpha}{4}\sum_i\sum_{pq} d_{pq}^2 V_p^i V_q^i - \tfrac{1}{\beta}\sum_p \ln \sum_i \exp\big(-\beta\tfrac{\alpha}{2}\sum_q d_{pq}^2 V_q^i\big). \qquad (21)$$

Simic found a slightly different expression with the weight value $\frac{\alpha}{2}$ instead of the value $-\frac{\alpha}{4}$. He simply ignores this term arriving at the following expression for the free energy:

$$F_{sim}(\mathbf{V}) = \tfrac{1}{4}\sum_i\sum_{pq} d_{pq}^2 V_p^i (V_q^{i+1} + V_q^{i-1}) - \tfrac{1}{\beta}\sum_p \ln \sum_i \exp\big(-\beta\tfrac{\alpha}{2}\sum_q d_{pq}^2 V_q^i\big). \qquad (22)$$

However, inspection of equation (19) reveals that the chosen Taylor-approximation does not hold for low values of the temperature, i.e., for high values of β. This is a fundamental objection because during the execution of the ENA, the parameter β is increased step by step until the end, when it has reached a relatively high value. □

Objection 2. In order to transform the Hopfield network formulation of the TSP into an elastic net, Simic performs a 'decomposition of the particle trajectory':

$$\mathbf{x}^i = \langle\mathbf{x}(i)\rangle = \sum_p \mathbf{x}_p \langle S_p^i\rangle = \sum_p \mathbf{x}_p V_p^i. \qquad (23)$$

Here, $\mathbf{x}(i)$ is the (stochastic) position of the particle at time i, $\mathbf{x}_p$ is the vector denoting the position of city point p, and $\mathbf{x}^i$ denotes the *average* (or expected) position of the particle at time i. Using the decomposition by, among other things, writing

$$\sum_q d_{pq}^2 V_q^i = \mid \mathbf{x}_p - \mathbf{x}^i \mid^2, \qquad (24)$$

he is able to make both a notable as well as a crucial transformation from a linear function in V_p^i into a quadratic one in $\mathbf{x}^i$. Using this result, he obtains the free energy expression (13) of the ENA.

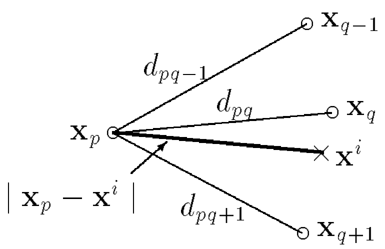

Figure 1: An elucidation of inequality (25).

However, careful analysis shows that in general

$$\sum_q d_{pq}^2 V_q^i = \sum_q (\mathbf{x}_p - \mathbf{x}_q)^2 V_q^i \neq \mid \mathbf{x}_p - \mathbf{x}^i \mid^2 . \qquad (25)$$

The left-hand side of this inequality represents the expected sum of distance squares between city point p and the particle position at time i, while the right-hand side represents the square of the distance between city point p and the expected particle position at time i. Under special conditions (e.g., if the constraints are fulfilled), the inequality sign must be replaced by the equality sign, but in general, the inequality holds (see also figure 1). □

Objection 3. The free energy (11) and (13), the latter of which is in our opinion incorrectly derived from the former, appear to have very different properties. The free energy (11) is a special case of the general energy function:

$$F_{cg}(\mathbf{V}) = -\tfrac{1}{2}\sum_{ij}\sum_{pq} w_{pq}^{ij} V_p^i V_q^j - \tfrac{1}{\beta}\sum_p \ln\,[\sum_i \exp(-\beta \sum_{jq} w_{pq}^{ij} V_q^j)] \tag{26}$$

with stationary points

$$\forall i, p : V_p^i = \frac{\exp(-\beta \sum_{jq} w_{pq}^{ij} V_q^j)}{\sum_l \exp(-\beta \sum_{jq} w_{pq}^{lj} V_q^j)}. \tag{27}$$

As can be concluded from (27), the free energy expression (26) *as a whole* has the peculiar property that – *whatever the value of the temperature parameter* – the stationary points are found at states where on average all strongly submitted constraints are fulfilled [19]. In case of the present TSP formulation, this is mathematically expressed by

$$\forall p : \sum_i V_p^i = 1, \tag{28}$$

meaning that, on average, any city p will be visited once. Moreover, the stationary points of (26) are often maxima.
However, inspection of the free energy (13) yields a very different view: an analysis of that expression (see the next section) clarifies that each term *on its own* creates a set of local minima, the first one trying to minimize the tour length, the second trying to force a valid solution. *The current value of the temperature*, which is a weight factor of the second term, *determines the overall effect* of summation over all these local minima , e.g. which of the two types will predominate. So, a competition takes place between the two types of minima (likewise, after having applied the above mentioned transformation of (22) into (13), the two modified terms still compete with each other). This phenomenon is remarkable, since the competition is similar to the one found by applying the classical penalty method (with fixed weights). A difference from that classical method is that in the present case – like in the Hopfield-Lagrange model [17] – the weights of the penalty terms change dynamically: in the case of the ENA, the weights ($T = 1/\beta$) decrease during updating of the motion equations, while in case of the Hopfield-Lagrange model the weights (the multipliers) often increase. This view on the ENA explains why we consider it a *dynamic penalty method.* □

We think the last observation corresponds to the theory of 'deformable templates' [13,22]. In that approach, the elastic net is considered as a 'template trajectory' (corresponding to Simic's particle trajectory), whose correct parameters should be determined.

These parameters are the 'template coordinates' (the elastic net points) and the binary Potts spins S_{pj} (where $\forall p : \sum_j S_{pj} = 1$). We note that $S_{pj} = 1$ has the meaning that net point j is assigned to template coordinate p. The corresponding Hamiltonian equals

$$E_{dt}(\mathbf{S}, \mathbf{x}) = \frac{\alpha_2}{2} \sum_i \mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid^2 + \sum_{pj} S_{pj} \mid \mathbf{x}_p - \mathbf{x}^j \mid^2 . \tag{29}$$

Thus, the energy E_{dt} is a function of both binary decision functions S_{pj} and of continuous template coordinates x^i. The first term in (29) equals the first term in the elastic net energy expression (13) and minimizes the tour length. The second term enforces a match between each city and one of the elastic net points. In other words, the energy (29) describes a penalty method! A statistical analysis of E_{DT} using the fact that the binary spins S_{pj} are stochastic, yields the free energy expression (13) of the elastic net: the derivation is straightforward [13,22], among other things because E_{DT} is a linear function in the Potts spins. By inspection of both (13) and (29) we conclude that the first energy expression is derived from the second, by adding stochastic noise exclusively to the penalty terms of (29). Therefore, one might say that the deformable template method applies *stochastic penalty terms*, which may be considered as a specific type of dynamic penalty terms.

4. An analysis of the ENA

We start by presenting a thorough analysis of the forces of the ENA and the corresponding energy landscapes. We will visualize things as much as possible. It will become clear, why the ENA may come up with a non-feasible solution. We also scrutinize the quadratic distance measure of the ENA.

4.1. Energy landscapes and elastic net forces

We can analyze the ENA at two levels, namely at the level of the energy equation (13) by inspection of the energy surface, and at the level of the updating rule (14) by analysis of the various forces acting on every net point. Afterwards, we shall deal – in a direct mathematical way – with the properties of the energy equation on lowering the temperature. From now on, we adopt the parameter values of the algorithm as given in subsection 2.2.

Let us start regarding the first, so-called 'elastic ring term' of (13). It is composed of a sum of M (the number of elastic net points) quadratic position differences. Of course, this term is minimized if all points coincide at some place. However, if the elastic net has a given length, this term is minimized whenever all space-points are equidistant. In figure 2 and 3, the 2-dimensional energy landscape of one net point $\mathbf{x}^i = (x, y)$ is shown at two different positions, one time using $d = 0.02$ as the mutual distance between neighbouring net points, the other time taking $d = 0.2$. The shapes of the two landscapes do not differ very much: in both cases, the variable point is forced to the middle of the other two (temporally fixed) points: in figure 2, these points are (0.49;0.5) and (0.51;0.5), in figure 3 they are (0.5;0.5) and (0.7;0.5). At the

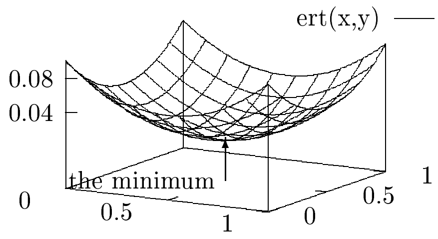

Fig. 2. The elastic ring term for one point with d=0.02

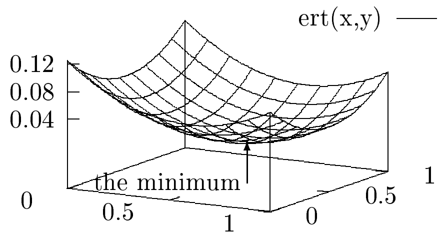

Fig. 3. The elastic ring term for another point with d=0.2

level of the motion equation (14), we see by writing

$$\mathbf{x}^{i+1} - 2\mathbf{x}^{i} + \mathbf{x}^{i-1} = (\mathbf{x}^{i+1} - \mathbf{x}^{i}) + (\mathbf{x}^{i-1} - \mathbf{x}^{i}), \tag{30}$$

that every $\mathbf{x}^i$ is forced to the *midpoint* between $\mathbf{x}^{i-1}$ and $\mathbf{x}^{i+1}$. Summarizing, if the 'elastic ring term' would be the only one, the ring points would become equidistant and, eventually, would coincide at one position, somewhere in state space.

But the second so-called 'mapping term' term of (13), makes its influence felt too. It is composed of a sum of N (the number of cities to be visited) logarithms, each logarithm having a sum of M exponentials as its argument. Every exponential is a Gaussian function with one local maximum, namely in the position where $\mathbf{x}_p$ coincides with $\mathbf{x}^j$. We may conclude, that the total mapping term (with the minus sign) corresponds to a set of 'pits' in the energy landscape. The width and depth of these pits depend on two things, namely the temperature and the distance between a city and the most nearby elastic net point. Initially, when the temperature T is relatively high, the attraction of elastic net points by every city is more or less uniformly distributed: this corresponds to a wide and shallow pit in the energy landscape around every city. The resulting, total energy landscape shelves slightly and is lowest in

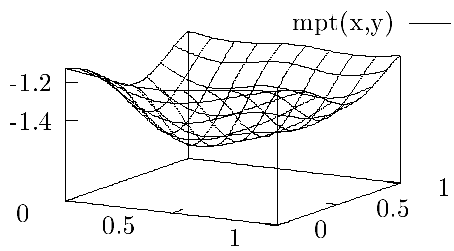

Fig. 4. The mapping term, initially at high temperature.

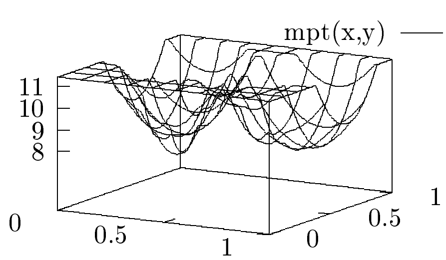

Fig. 5. The mapping term, in case of non-feasibility.

regions with a high city density. This phenomenon is quite independent of the (initial) position of the elastic net points in the unit square. A simple example is given in figure 4: again, the energy landscape of one of five elastic net points is shown, while the positions of the city points are (0.2;0.63), (0.8;0.63), (0.65;0.37), (0.37;0.37) and (0.33;0.37). The city positions will be kept the same in the next examples and can be found in figure 7. As can be seen in figure 4, the lowest part of the energy landscape of the mapping term is found around the last two, closely situated, cities. Experiments show, that the positions of the other four elastic net points do not matter very much, i.e., whatever these positions are, in all cases approximately the same energy surface is found provided that the initially high temperature $T = 0.2$ is used.

On lowering the temperature, a city will attract nearby net points more and more and distant net points less and less because, in general, the pit in the energy landscape around a city becomes narrower. However, a second parameter plays an important part. If a city remains without a nearby elastic net point, the width of the pit shrinks only slowly and the depth even grows: apparently, the city persists in trying to catch a not too remote elastic net point. In figure 5, an example is given at $T = 0.027$, which is an almost final temperature of the algorithm. The four net points are still chosen around the center of the unit square, far away from any city. The basins of attraction around every city are clearly present.

If, on the other hand, a city has been able to (almost) catch a net point, the surrounding pit in the energy landscape will become very narrow and shallow. In figure 6, an example is given with, once more, four (temporally) fixed net points. Again, $T = 0.027$. The position of one net point coincides exactly with a city, the position of a second one is chosen close to a city, a third net point is situated on a somewhat larger distance from another city, and the position of the fourth net point is precisely in the middle between two close city points. The city point and net point positions are shown in figure 7. The energy landscape in figure 6 shows narrow and shallow pits around cities: the smaller the distance of the most neighbouring elastic net point is, the narrower and shallower the pit. The figure also demonstrates an unpleasant phenomenon concerning the elastic net point in the middle of the two close cities. Both cities seem to consider themselves owner of that elastic net point. Consequently, the surrounding energy landscape of the two cities will generally not be able to catch another elastic net point, so, in those circumstances, the system persists in non-feasibility!

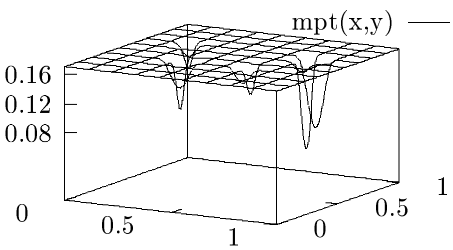

Fig. 6. The mapping term, in case of an almost feasible solution.

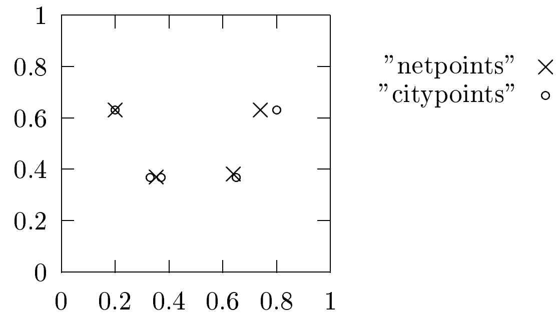

Fig. 7. The positions of net and city points.

4.2. The total energy landscape

Of course, we should analyze the combined effect of the elastic ring and the mapping term. For that purpose, we selected some, more or less representative examples starting with an initial elastic net situated around the center of the unit square at $T = 0.2$. Then, the energy landscape appears to resemble that of figure 4 (as expected): the mapping term predominates, pushing the elastic net to regions of high city density (in practice, there may exist more than one such region, resulting in a stretching out of the net). In the background, the elastic net term keeps the net more or less together. On lowering the temperature a little, until $T = 0.15$, the mapping term becomes more important as long as feasibility has not been reached. In figure 8, the energy landscape of the free elastic net point is shown under the assumption that the initial configuration of all other points would have remained the same. It is clear that the landscape has become somewhat steeper. Thus, the system is trying to reach

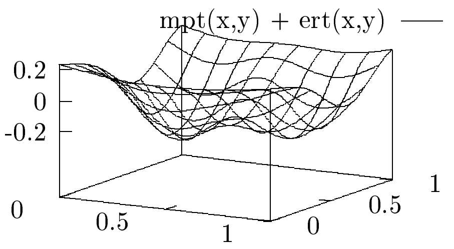

Fig. 8. The total energy landscape, initially at $T = 0.15$.

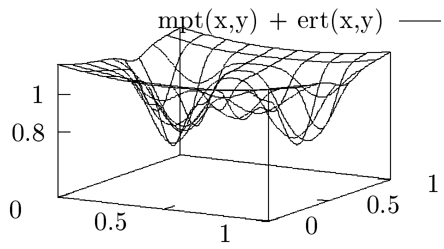

Fig. 9. The total energy landscape, an intermediate state at $T = 0.08$.

feasibility with a bit more strength. Now supposing the more realistic scenario that the elastic net has stretched out somewhat (with elastic net positions (0.57;0.44), (0.43;0.44), (0.35;0.56), (0.65;0.56), while the 'free' elastic net point is supposed to be somewhere between the last two given positions), then more details in the energy landscape are apparent: in figure 9, the energy landscape is shown at $T = 0.08$.

Next, we show two potential, nearly final states. In figure 10, a solution is shown,

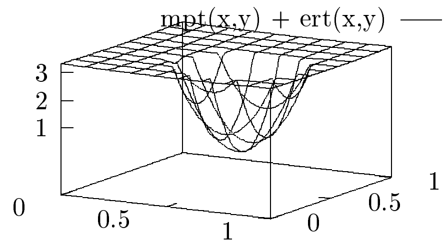

Fig. 10. The total energy landscape, a non-feasible state at $T = 0.027$.

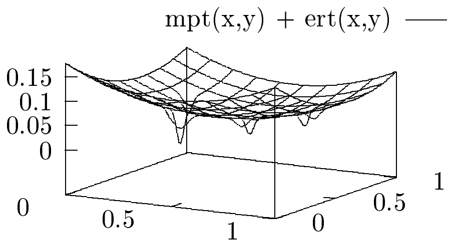

Fig. 11. The total energy landscape, an almost feasible state at $T = 0.027$.

where all cities except one have caught an elastic net point. If the remaining net point is not too far away from the non-visited city, it can still be attracted by it, otherwise this city will never be visited. This shows, that a too rapid lowering of the temperature may lead to a non-valid solution, because a further lowering of the temperature will lead to a further narrowing of the energy pit of figure 10. Note too, that in this case the pits corresponding to the elastic ring term are not visible: comparatively, they are too small. In figure 11, an almost feasible solution is shown, where the positions of three net points coincide with the position of a city, while a fourth elastic net point is precisely in the middle between the two close cities. Because an almost feasible solution has been reached, the mapping term becomes relatively small (corresponding to some small pits), and the remaining elastic net point is forced to the middle of its neighbors. The final state will be equidistant, but not feasible! The example shows clearly that in case of (almost) feasibility the influence of the mapping term becomes small and, at the same time, is capable of maintaining feasibility. Under these conditions, the algorithm tries to realize equidistance.

4.3. Non-feasibility

The analysis of the previous subsection reveals that it is possible to end up in a non-feasible solution for at least two reasons:

- The parameter T may be lowered too rapidly yielding a non-feasible solution, where one or more cities have not 'caught' an elastic point.
- Two close cities may have received the same elastic net point as the nearest one.

The determination of the optimal schedule for decreasing T is often mentioned in literature and is also associated with 'optimal simulated annealing'. We want to emphasize here, that the similarity is less than would appear. In simulated annealing [1], the temperature should be decreased carefully in order to *escape* from local minima. Here, this lowering should be done carefully in order get and keep a valid solution, in other words, to *end up* in a local (constrained) minimum!

Like any other penalty method, the ENA tries to fulfill two competing requirements: in this case these are minimal *equidistance* and *feasibility* (a tour through all city points). To be able to fulfill both requirements, it is generally necessary to use more elastic net points than city points (see figure 12). It should be clear that the more diversity exists in the shortest distances between cities, the more elastic net points are needed [d]. Using a large number of elastic net points gives rise to the additional drawback of increasing computation time. Finally, we note that the property of equidistance – which is a consequence of the quadratic distance measure of the ENA – is not at all a necessary qualification of the final solution.

The above mentioned observations that (a) a non-feasible solution might be found and (b) the ENA pursues equidistance, motivated us to investigate alternative elastic net algorithms (section 5 and 6). For reasons of clarity, experimental results of the ENA are given together with those of the NENA and the HENA in section 7.

5. A non-equidistant elastic net algorithm

In order to get rid of the equidistance property, we only need to change the first term of the original energy expression (13). Here, a *linear* distance function is

[d] Durbin and Willshaw choose as relation between the number M of elastic net points and N, the number of city points: $M = 2.5N$.

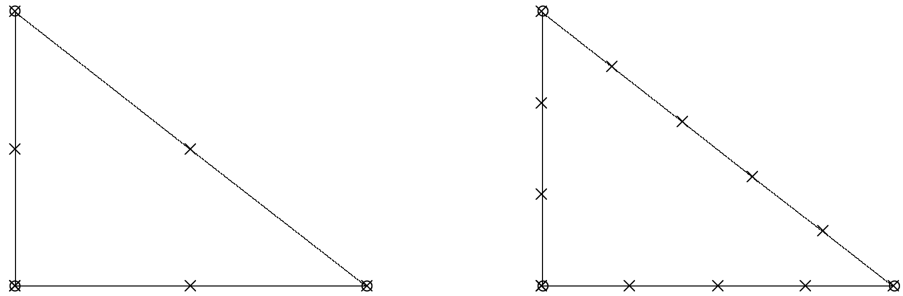

Figure 12: To realize both feasibility and equidistance many net points are needed.

chosen[e], whose minimal constrained length equals, by definition, the global minimal tour length. The new energy function is:

$$F_{lin}(\mathbf{x}) = \alpha_2 \sum_i \mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid - \tfrac{\alpha_1}{\beta} \sum_p \ln \sum_j \exp(\tfrac{-\beta^2}{2} \mid \mathbf{x}_p - \mathbf{x}^j \mid^2). \tag{31}$$

Applying gradient descent, the corresponding motion equations are found [5]:

$$\Delta \mathbf{x}^i = \tfrac{\alpha_2}{\beta} \left(\frac{\mathbf{x}^{i+1} - \mathbf{x}^i}{\mid \mathbf{x}^{i+1} - \mathbf{x}^i \mid} \right) + \left(\frac{\mathbf{x}^{i-1} - \mathbf{x}^i}{\mid \mathbf{x}^{i-1} - \mathbf{x}^i \mid} \right) + \alpha_1 \sum_p \Lambda^p(i)(\mathbf{x}_p - \mathbf{x}^i), \tag{32}$$

where again, the time-step Δt equals the current temperature. We notice that all elastic net forces are normalized now. Moreover, if $\exists i : x^{i+1} = x^i$, we get into trouble [f]. A self-evident analysis [5] shows that, as in the original ENA, the elastic net forces try to push elastic net points onto a straight line. There is however another important difference: once a net point is situated at *any* point on the straight line between its neighbouring net points, it no longer feels any elastic net force (this is simply caused by the normalization of the elastic net forces: see figure 13). This means, that equidistance is no longer pursued. Consequently, elastic net points will

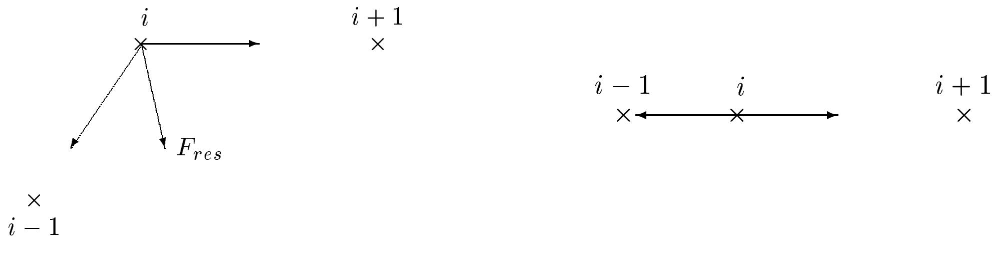

Figure 13: The new elastic net force: general case (left), 3 points in line (right).

have more freedom in moving towards cities. It is therefore hoped that application of the NENA (a) will nearly always yield feasible solutions (of high quality), if the same number of elastic net points is used as in the original ENA and-or (b) will often

[e]At some places in the literature [14,22], a linear distance measure is suggested, but nowhere did we find an elaborated implementation of this idea.

[f]In practice, this fortunately never occurred.

yield feasible solutions too, if a smaller number of elastic net points is chosen [g]. Or stated in more general terms, we hope the new algorithm will yield valid solutions more easily than the original ENA.

Since the elastic net forces are normalized by the new algorithm (those of the old one are not), a tuning problem arises. To solve this problem, the following simple approach is chosen: in the motion equations (32), all elastic net forces will be multiplied by the same factor

$$A(\mathbf{x}) = \tfrac{1}{M} \sum_{1}^{M} \mid x^{i+1} - x^{i} \mid , \tag{33}$$

which represents the average distance between two elastic net points. Thus, the average elastic net force is roughly equal to the average in the original algorithm, and the final updating rule becomes:

$$\Delta \mathbf{x}^{i} = \tfrac{\alpha_2}{\beta} A(\mathbf{x}) \left(\frac{\mathbf{x}^{i+1} - \mathbf{x}^{i}}{\mid \mathbf{x}^{i+1} - \mathbf{x}^{i} \mid} + \frac{\mathbf{x}^{i-1} - \mathbf{x}^{i}}{\mid \mathbf{x}^{i-1} - \mathbf{x}^{i} \mid} \right) + \alpha_1 \sum_{p} \Lambda^{p}(i)(\mathbf{x}_p - \mathbf{x}^{i}), \tag{34}$$

where the values α_1, α_2 and β are chosen conform the original ENA. Experimental results with this updating rule are described in section 7.

6. The Hybrid Approach

A fundamental problem of ENA is, that it might lead to non-feasible solutions due to adhering to equidistance of the elastic net points. Moreover, equidistance is not required for the final solution of the elastic net, although it might be very useful in the initial phase of the algorithm in order to realize a smooth stretching out of the elastic ring. A fundamental problem of NENA is, that net points may become too lumpy, which, at least for larger problem instances, leads to non-feasibility and a lower quality of the subsequent solutions.

Contemplating these considerations we tried to merge the two algorithms into a hybrid one retaining the best properties of both. The approach of the Hybrid Elastic neural Net Algorithm (HENA) is simple: the algorithm starts using ENA and, after a certain number of iterations, switches to NENA. The first phase is used in order to get a balanced stretching of the elastic net which is hoped to lead to solutions of high quality, the second phase is used in order to try to guarantee feasibility at the end. A consequence of this hybrid approach is the introduction of two new parameters. First, we have to decide at what time the switch should take place, and then, we have to choose the starting temperature after the switch. The experimental results using the HENA are presented in the next section.

7. Experiments

We now describe some of the results as obtained with the NENA and the HENA,

[g] In [14], it is even conjectured that, using a linear distance measure, the number of elastic net points could be equal to the number of cities.

and compare them with results found using the original ENA.

7.1. A small problem instance

We start by using the configuration of cities as described in the theoretical analysis of section 4 (the 5 cities are situated as given in figure 7). In all cases, we used the following initialization of the elastic net: elastic net points are put in a small ring in the center of the state space, where the position of every net point is slightly randomized.

Using 5, 7, 10 or 12 elastic net points, the ENA produced only non-feasible solutions: in all experiments, one elastic net point is found in the middle between the two closely situated city points. The other 3 cities are always visited, while all other net points are more or less spread equidistantly. However, using 15 elastic net points, the optimal and feasible solution is always found: apparently, the number of elastic net points is now large enough to guarantee both feasibility and optimality.

Using 5 elastic net points, the NENA nearly always produced a non-feasible solution, but sometimes the optimal, feasible one. A gradual increase of the number of elastic net points results into a rise of the percentage of optimal solutions found. Using only 10 elastic net points yields a 100% score. An inspection of the final results reveals that the elastic net points become lumpy: they appear to come together around a city, which is, in all probability, a consequence of their increased freedom. The number of net points per city depends the initialization as well as the location of the city.

We conclude that for this small problem instance the NENA produces better results than the ENA, or, stated more precisely, using a smaller number of elastic net points the NENA finds the the same optimal solution as the original ENA. The described experimental results are completely consonant with the theoretical conjectures of section 4.

7.2. Larger problem instances

Using a 15-city-problem, we had the similar experiences: it is easier to arrive at a feasible solution using the NENA. E.g., using 30 elastic net points, the NENA always yielded the same solution (namely the best solution found with both the ENA and the NENA), while the ENA sometimes yielded that solution, and sometimes a non-valid one.

However, the picture starts to change, if 30-city problem instances are chosen. As a rule, both algorithms are equally disposed to finding a valid solution, but another phenomenon turns up: the quality of the solutions found by the original ENA was generally better. Inspection of the solutions found by the NENA, demonstrated a strong lumping effect. The lumping can be so strong that sometimes a city is left out completely. Especially cities which are situated at a point, where the final tour bends strongly, may be overlooked. Apparently, by disregarding the property of equidistance, a new problem has originated. Re-evaluating, we conclude that the equidistance property of the ENA has an important contribution towards finding solutions of high quality, i.e., short tours.

At this point, the hybrid approach of HENA comes to mind. Because for small problem instances the NENA works better than the ENA, we only tried larger problem instances. Unfortunately, in our experiments the HENA appears to be slightly worse than the original ENA both in relation to the quality of the solution and in relation to feasibility. E.g., taking a 100-city problem, the ENA usually yielded a solution where 99 of the 100 cities are visited, while in case if the HENA, on average 98 to 99 cities are visited. Moreover, the encountered tour length using the ENA is, on average, slightly better than the tour length found by the HENA. Trying larger problem instances, we were unable to find parameters of the HENA, which yield better solutions than the original ENA or which guarantee feasibility of solutions.

8. Discussion and outlook

A fundamental conclusion of the analysis of elastic neural networks as given in this paper is that, in principle, they are penalty methods (using a type of *dynamic* penalty terms). A consequence is, that any application using such a method is always confronted with a *tuning* problem, that should be resolved in practice. This is, of course, a fundamental drawback of this type of networks.

From the angle of the theory on deformable templates, elastic net algorithms can be considered *stochastic penalty methods*, where, contrary to simulated annealing and to what is mentioned in the literature, the network should *end up* in a local, i.e., a constrained minimum.

Considering the original ENA, we conclude that it is relatively well-tuned: for small problem instances, it generally yields a valid solution of high quality. A non-valid solution may come up, if two cities are very close to each other. For larger problem instances, up to 100 cities, solutions are often almost-feasible or can be made so, by enlarging the number of elastic net points or the number of iterations of the algorithm. Of course, in practice it is relatively easy to transform (such) a non-feasible solution in a valid one, by taking up the non-visited city (cities) at a logical place on the ring.

In order to fundamentally improve the ENA (especially, to guarantee feasibility), we proposed a new algorithm, named the NENA, having a *linear* distance measure. This measure seems to be more natural from a theoretical point of view. However, the success of this algorithm is limited to small problem instances as experiments have shown. Apparently, the quadratic distance measure is an essential ingredient of the original ENA!

Thereupon, we proposed the HENA which would combine the good properties of the ENA and the NENA. We are still trying to find a better tuning of the parameters, so a final judgement is difficult. But up to now, the old ENA performed slightly better than the new HENA, if larger problem instances were tried.

In future research, an alternative for HENA could be considered by realizing a gradual switch from the ENA to the NENA in order to enlarge the 'freedom of the elastic net points' little by little. This could be implemented by applying a gradual normalization of the elastic net forces. It should be noted that this approach may introduce new tuning difficulties. Another possible would be try other penalty terms: some explorations have been done [5], but in all cases, parameter tuning seems to be a tough job.

9. Acknowledgements

We thank Cor Bioch for reading this report and, more specifically, for his advice on the mathematical underpinning of the analysis of section 3. We thank Marc Leipoldt for reading this report and suggesting some changes to the English.

References

1. E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines, A Stochastic Approach to Combinatorial Optimization and Neural Computing.* John Wiley & Sons, Inc., 1989.
2. D.J. Burr. An Improved Elastic Net Method for the Traveling Salesman Problem. *IEEE International Conference on Neural Networks*, vol. 1, 69–76, New York: IEEE, 1988.
3. A. Cichocki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* John Wiley & Suns, 1993
4. R. Durbin and D. Willshaw. An Analogue Approach of the Travelling Salesman Problem Using an Elastic Net Method. *Nature*, 326:689–691, 1987.
5. 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.
6. J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company, The advanced Book Program, 1991.
7. 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.
8. 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.
9. J.J. Hopfield and D.W. Tank. "Neural" Computation of Decisions in Optimization Problems. *Biological Cybernetics*, 52:141–152, 1986.
10. Ch. von der Malsburg and D.J. Willshaw How to label nerve cells so that they can interconnect in an ordered fashion. *Proceedings in the National Academy of Science*, USA 74-11:5176–5178, 1977.
11. G. Parisi. *Statistical Field Theory.* Addison-Wesley Publishing Company, The advanced Book Program, 1988.
12. C. Peterson and B. Söderberg. A New Method for Mapping Optimization Problems onto Neural Networks. *International Journal of Neural Systems*, 1:3–22, 1989.
13. 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.
14. P.D. Simic. Statistical Mechanics as the Underlying Theory of 'Elastic' and 'Neural' Optimisations. *Network*, 1:88–103, 1990.
15. J.V. Stone. The Optimal Elastic Net: Finding Solutions for the Traveling Salesman. *Artificial Neural Networks*, 2 vol. 2, 1077–1080, Amsterdam: North-Holland, 1992.

16. J. van den Berg and J.C. Bioch. On the Statistical Mechanics of (Un)Constrained Stochastic Hopfield and 'Elastic' Neural Networks. Technical Report EUR-CS-94-08, Erasmus University Rotterdam, Comp. Sc. Dept. (AI Section), Faculty of Economics, 1994.
17. 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.
18. J. van den Berg and J. C. Bioch. On the (Free) Energy of Hopfield Networks. To appear in: World Scientific, *Proceedings of the Workshop on the Theory of Neural Networks: The Statistical Mechanics Perspective*, Pohang, Korea, February 1995.
19. J. van den Berg and J.C. Bioch. Some Theorems Concerning the Free Energy of (Un)Constrained Stochastic Hopfield Neural Networks. *Lecture Notes in Artificial Intelligence* 904: 298–312, Second European Conference on Computational Learning Theory, EuroCOLT'95, Barcelona, March 1995.
20. 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.
21. 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.
22. A.L. Yuille. Generalized Deformable Models, Statistical Physics, and Matching Problems. *Neural Computation*, 2: 1–24, 1990.