# Graded Rescaling in Hopfield Networks

Xinchuan Zeng, Tony R. Martinez
Computer Science Department, Brigham Young University, Provo, Utah 84602
e-mail: zengx@axon.cs.byu.edu, martinez@cs.byu.edu

**Abstract**

In this work we propose a method with the capability of improving the performance of the *Hopfield network* for solving optimization problems by using a *graded rescaling* scheme on the distance matrix of the energy function. This method controls the magnitude of rescaling by adjusting a parameter (scaling factor) in order to explore the optimal range for performance. We have evaluated different scaling factors through 20,000 simulations, based on 200 randomly generated city distributions of the 10-city *traveling salesman problem*. The results show that the graded rescaling can improve the performance significantly for a wide range of scaling factors. It increases the percentage of valid tours by 72.2%, reduces the error rate of tour length by 10.2%, and increases the chance of finding optimal tours by 39.0%, as compared to the original Hopfield network without rescaling.

## 1 Introduction

Hopfield and Tank [1] proposed a neural network approach to find approximate solutions to combinatorial optimization problems, such as the *traveling salesman problem* (*TSP*). Much research work has focused on analyzing the model and improving its performance.

Wilson and Pawley [2] showed that the Hopfield network often failed to converge to valid solutions, and many solutions were far from optimal. Brandt et al. [3], Aiyer et al. [4], and Protzel et al. [5] showed that they could improve the convergence of the network by modifying constraints in the energy function. Li [6] combined the Hopfield network with the "augmented Lagrange multipliers" algorithm from optimization theory. Catania et al. [7] applied a fuzzy scheme to tune the parameters in the Hopfield network. Although these approaches have demonstrated some improvement, most were tested on only a single or a small number of TSP city distributions, and general performance on a large number of city distributions were not reported.

Despite the improvement of its performance over the past decade, the Hopfield network still has some basic problems [8, 9] that remain unsolved. One problem is the inconsistent performance, which is usually better for simple city distributions, but poor for those with complex topologies and multiple clusters. Another problem is that the performance is sensitive to the choice of parameters in the energy function. In previous work, we proposed a new activation function to reduce effects of noise [10], a new relaxation procedure to reach a smoother relaxation process [11], and a simple rescaling scheme to reduce the effects of clustering [12]. These approaches have been shown capable of significantly improving performance based on a large number of simulations.

In this work we propose an approach which uses *graded rescaling* of the distance matrix in the energy function to reduce the negative impacts of distribution clusterings on the performance. The proposed graded rescaling scheme has the capability to improve the balance of the distance matrix. This approach has been tested with a large number $(20,000)$ of simulations based on 200 randomly generated city distributions of the 10-city *TSP*. The results indicate that it is capable of improving the performance significantly: 72.2% increase in the percentage of valid tours, 10.2% decrease in the error rate, and 39.0% increase in the chance of finding optimal tours.

## 2 Graded Rescaling of Distance Matrix

Hopfield's original energy function for an $N$-city *TSP* is given by [1]:

$$E = \frac{A}{2}\sum_{X=1}^{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} V_{Xi}V_{Xj}$$

$$+\frac{B}{2}\sum_{i=1}^{N}\sum_{X=1}^{N}\sum_{Y=1,Y\neq X}^{N} V_{Xi}V_{Yi} + \frac{C}{2}(\sum_{X=1}^{N}\sum_{i=1}^{N} V_{Xi} - N_0)^2$$

$$+\frac{D}{2}\sum_{X=1}^{N}\sum_{i=1}^{N}\sum_{Y=1,Y\neq X}^{N} d_{XY}V_{Xi}(V_{Y,i+1}+V_{Y,i-1}) \qquad (1)$$

where $X$, $Y$ are row indices, and $i$, $j$ are column indices, $V_{Xi}$ is the activation for neuron $(X, i)$, and $d_{XY}$ is the distance between cities $X$ and $Y$. The first three terms enforce the constraints for a valid tour, and the last term represents the cost function for obtaining a short tour. The value of each parameter ($A$, $B$, $C$, and $D$) measures the importance of the corresponding term. The connecting weight between neuron $(X, i)$ and $(Y, j)$ is set according to:

$$W_{Xi,Yj} = -A\delta_{XY}(1-\delta_{ij}) - B\delta_{ij}(1-\delta_{XY}) - C \\ -Dd_{XY}(\delta_{j,i+1} + \delta_{j,i-1}) \tag{2}$$

During relaxation, the input value $U_{Xi}^{(n+1)}$ for neuron $(X, i)$ at iteration step *(n+1)* is the sum of its value $U_{Xi}^{(n)}$ at step $n$ and $\Delta U_{Xi}$. The value of $\Delta U_{Xi}$ is given by the following equation:

$$\Delta U_{Xi} = (-\frac{U_{Xi}}{\tau} + \sum_{Y=1}^{N}\sum_{j=1}^{N} W_{Xi,Yj}V_{Yj} + I_{Xi})^{(n)}\Delta t \tag{3}$$

where $\tau (= RC)$ is the time constant of an RC circuit and $I_{Xi} (= CN_0)$ is an external input current connected to each neuron $(X, i)$.

The distance matrix $d_{XY}$ in the energy function has the function of controlling the quality of solutions. From Eq. (2) and (3), we can see that the second term in Eq. (3) for the activation change $\Delta U_{Xi}$ is proportional to the value: $-D\sum_{Y=1}^{N} d_{XY}V_{XY}$. Since each $V_{XY}$ has a similar average level, that term is approximately proportional to the value $-D\sum_{Y=1}^{N} d_{XY}$.

Now consider a case of two clusters $C_1$ and $C_2$, where $N_1$ and $N_2$ are the number of cities in $C_1$ and $C_2$ respectively. Assume that cluster $C_1$ is much larger than $C_2$, i.e. $N_1 >> N_2$. For a city $X_1$ in $C_1$ and $X_2$ in $C_2$, $d_{X_1Y}$ would be smaller than $d_{X_2Y}$ for $(N_1 - 1)$ cities in $C_1$ while larger for $N_2$ cities in $C_2$. Since $N_1 >> N_2$, the value $\sum_{Y=1}^{N} d_{X_1Y}$ would be smaller than $\sum_{Y=1}^{N} d_{X_2Y}$. The net effect is that the activations for the cities in cluster $C_1$ have overall higher values than those in $C_2$ (a smaller $\sum_{Y=1}^{N} d_{X_2Y}$ leads to a larger $-D\sum_{Y=1}^{N} d_{XY}$). A potential consequence is the formation of an invalid tour – a closed loop inside cluster $C_1$ without inclusion of those cities in $C_2$. Even when a valid tour is formed, its quality may be poor since each cluster constructs its own local configuration with less consideration of the other cluster because of the imbalance in their overall activations. We have observed such phenomenon for some city distributions where one or two cities are isolated from the main cluster.

Based on this analysis we propose the following approach, called *graded rescaling*, to rescale the distance matrix so that activations in a smaller cluster become closer to those in a larger cluster. For each row $X$ $(X = 1, 2, ..., N)$, we first calculate the summation $S_X = \sum_{Y=1}^{N} d_{XY}$, and then calculate the average of $S_X$ over all $N$ rows using $\overline{S} = (\sum_{X=1}^{N} S_X)/N$. Next we rescale each original matrix element $d_{XY}$ to be a new element $d'_{XY}$ using the formula:

$$d'_{XY} = \frac{((1-\alpha)S_X + \alpha\overline{S})d_{XY}}{S_X} \tag{4}$$

where $\alpha$ is a parameter, called *scaling factor*, which controls the amount of rescaling and is in the range of $0 \leq \alpha \leq 1$. For the limit $\alpha = 0$, there is no rescaling and the matrix keeps its original form: $d'_{XY} = d_{XY}$. For the limit $\alpha = 1$, we have $d'_{XY} = (\overline{S}/S_X)d_{XY}$. Summation $S'_X$ over row $X$ gives $S'_X = \sum_{Y=1}^{N} d'_{XY} = (\overline{S}/S_X)\sum_{Y=1}^{N} d_{XY} = (\overline{S}/S_X)S_X = \overline{S}$, i.e. all rows are normalized to the same constant $\overline{S}$. In this case, the two terms $\sum_{Y=1}^{N} d'_{X_1Y}V_{XY}$ and $\sum_{Y=1}^{N} d'_{X_2Y}V_{XY}$ would have a similar overall level even though they are located in two unbalanced clusters. When $0 < \alpha < 1$, however, it is easy to show that the range for $S'_X$ is between $S_X$ and $\overline{S}$ ($S_X < S'_X < \overline{S}$ if $S_X < \overline{S}$, and $\overline{S} < S'_X < S_X$ if $S_X > \overline{S}$). The choice of $\alpha$ is important for the performance. A small $\alpha$ value has little influence on balancing activations. However, too large an $\alpha$ value can also have a negative effect since it causes too much distortion of the distance matrix, which can reduce the chance of finding optimal tours, and also degrades the percentage and quality of valid solutions. We have found empirically that an $\alpha$ value of 0.5 is both robust and balances the above concerns.

## 3 Simulation Results

The performance of the proposed graded rescaling approach has been evaluated through a large number of simulations based on 200 randomly generated 10-city *TSP* city distributions (including wide varieties of topologies). Many previous studies evaluated the performance of their algorithms based on only one (e.g., [1]) or a small number of city distributions (e.g., 10 distributions in [2]). In this simulation, 100 runs are conducted for each of 200 city distributions. Different random noise is added to the initial neuron values for each of the 100 runs of a given city distribution. The quantities that measure the performance are evaluated by first averaging over 100 runs for each city distribution, and then averaging over the entire 200 city distributions. Thus 20,000 runs are used to calculate each simulation data point shown in the following figures. We have experimented with different sets of 20,000 runs for a fixed set of parameters. The results show that the estimated quantities are fairly stable and their values vary within a range of about 1% among different sets of 20,000 runs. This demonstrates the robustness of the evaluation for the scaling factor $\alpha$. The original energy function of the Hopfield network was used in the simulation, and the parameters in the energy function are those used by Hopfield and Tank [1].

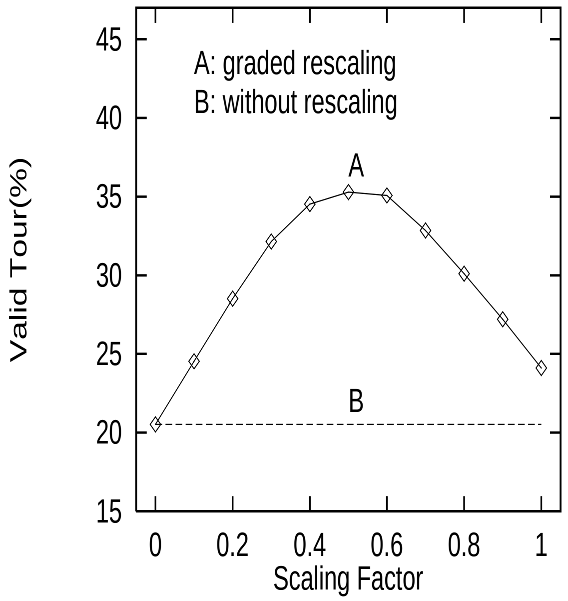


**Fig. 1.** The percentage of valid tours of different scaling factors using the graded rescaling procedure compared to the original Hopfield network without rescaling.

Fig. 1 shows the percentage of valid tours with different scaling factors using the proposed graded rescaling procedure and compares them to that using the original procedure without rescaling. For each city distribution, valid tour percentage is evaluated based on number of valid tours among 100 runs. Each data point shown in the figure is the average of valid tour percentages over 200 city distributions. The result shows that the rescaled distance matrix significantly improves the percentage of valid tours over the original one. The largest improvement is at the middle range for $\alpha$ from 0.4 to 0.6. At $\alpha = 0.5$, the percentage of valid tours (35.3%) with rescaling is increased by 72.2% compared to that (20.5%) without rescaling.

Fig. 2 shows the error rate of different scaling factors using the proposed graded rescaling procedure. For city distribution $i$, let $d_{i,optimal}$ be its optimal (shortest) tour length, and let $d_{i,j}$ be tour length of a valid tour $j$. Then the error rate of valid tour $j$

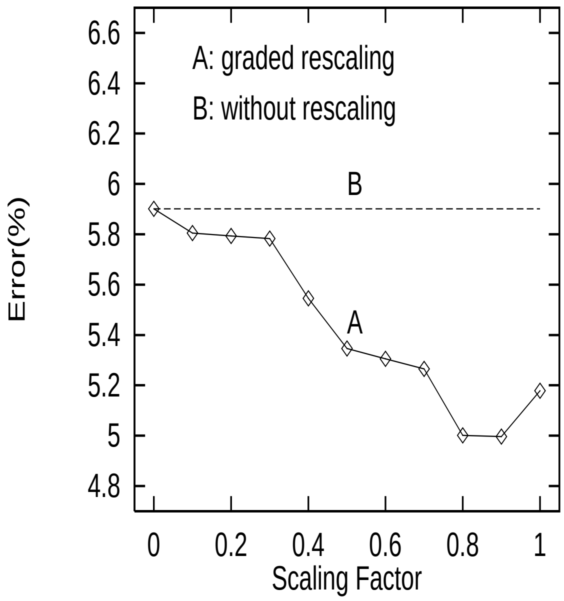


**Fig. 2.** The errors (%) of different scaling factors using the graded rescaling procedure compared to the original Hopfield network without rescaling.

is defined as $(d_{i,j} - d_{i,optimal})/d_{i,optimal}$. The error for each city distribution is averaged over all valid tours for that distribution. Each data point shown in the figure is the average error of valid tours in 200 city distributions and is weighted by the percentage of valid tours for each city distribution. From the result we can see that all the errors with graded scaling are smaller than that without rescaling. The best improvement is at $\alpha = 0.9$, with a decrease of 15.3% in error (5.0% vs 5.9%). At $\alpha = 0.5$, the decrease in error is 10.2% (5.3% vs 5.9%).

Fig. 3 compares the graded rescaling and original scheme in terms of the percentage of city distributions for which the optimal tour is by at least one of the 100 runs, which measures the chance of finding optimal tours by the network. The result shows that all scaling factors have better chances to find optimal tours. At $\alpha = 0.5$, the improvement is 39.0% (82.0% vs 59.0%).

The above results demonstrate the advantage of using an adjustable graded scaling factor to achieve an optimal range of rescaling. A proper choice of scaling factor can lead to significant improvement in performance over that without rescaling ($\alpha = 0$), and also has an overall better performance than simple rescaling ($\alpha = 1$) [12].

## 4 Summary

In this paper, we have proposed a graded rescaling approach to rescale the distance matrix in the Hop-

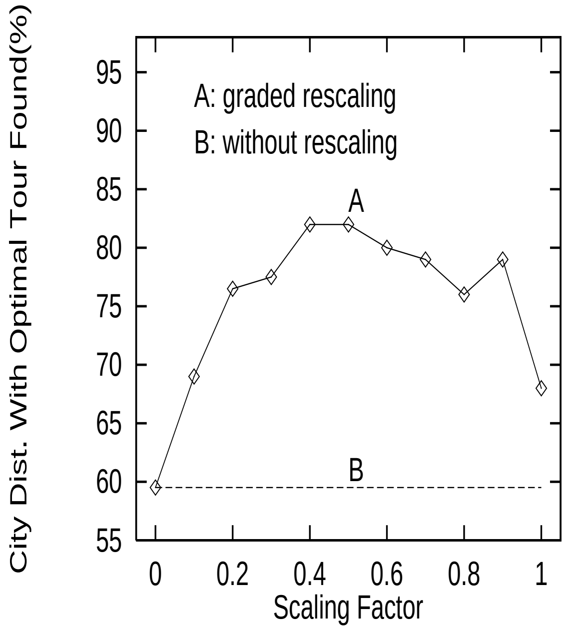


**Fig. 3.** The percentage of city distributions which optimal tour is found by at least one of 100 runs, showing results of different scaling factors using the graded rescaling procedure compared to the original Hopfield network without rescaling.

field network for solving *TSP*. A parameter (scaling factor $\alpha$) is applied to adjust the amount of rescaling in order to achieve better performance. The proposed approach has capability of reducing activation imbalances in the network caused by distribution clustering so that more valid tours with higher qualities can be obtained. We have evaluated its performance for different scaling factors based on a large number (20,000) of simulations using 200 randomly generated city distributions of the 10-city $TSP$. The results show that a significant improvement can be achieved by this approach. With scaling factor of 0.5, the network obtains a 72.2% increase in the percentage of valid tours, a 10.2% decrease in the error rate, and a 39.0% increase in the chance of finding optimal tours, compared to those without rescaling.

## 5 Acknowledgments


This research is funded in part by a grant from *fonix* Corp.


## References


[1] J. J. Hopfield and D. W. Tank, "Neural Computations of Decisions in Optimization Problems", *Biological Cybernetics*, vol. 52, pp. 141–152, 1985.

[2] G. V. Wilson and G. S. Pawley, "On the Stability of the Traveling Salesman Problem Algorithm of Hopfield and Tank", *Biological Cybernetics*, vol. 58, pp. 63–70, 1988.

[3] R. D. Brandt, Y. Wang, A. J. Laub and S. K. Mitra, "Alternative Networks for Solving the Traveling Salesman Problem and the List-Matching Problem", *Proceedings of IEEE International Conference on Neural Networks*, San Diego, CA, vol. 2: pp. 333–340, 1988.

[4] S. V. B. Aiyer, M. Niranjan and F. Fallside, "A Theoretical Investigation into the Performance of the Hopfield Model", *IEEE Transactions on Neural Networks*, vol. 1, no. 2, pp. 204-215, 1990.

[5] P. W. Protzel, D. L. Palumbo and M. K. Arras, "Performance and Fault-Tolerance of Neural Networks for Optimization", *IEEE Transactions on Neural Networks*, Vol. 4, pp. 600-614, 1993.

[6] S. Z. Li, "Improving Convergence and Solution Quality of Hopfield-Type Neural Networks with Augmented Lagrange Multipliers", *IEEE Transactions On Neural Networks*, vol. 7, no. 6, pp. 1507-1516, 1996.

[7] V. Catania, S. Cavalieri and M. Russo, "Tuning Hopfield Neural Network by a Fuzzy Approach", *Proceedings of IEEE International Conference on Neural Networks*, pp. 1067-1072, 1996.

[8] B. S. Cooper, "Higher Order Neural Networks-Can they help us Optimise?", *Proceedings of the Sixth Australian Conference on Neural Networks (ACNN'95)*, pp. 29-32, 1995.

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

[10] X. Zeng and T. R. Martinez, "A New Activation Function in the Hopfield Network for Solving Optimization Problems", *Fourth International Conference on Artificial Neural Networks and Genetic Algorithms*, pp. 67-72, 1999.

[11] X. Zeng and T. R. Martinez, "A New Relaxation Procedure in the Hopfield Network for Solving Optimization Problems", *Neural Processing Letters*, vol. 10, pp. 1-12, 1999.

[12] X. Zeng and T. R. Martinez, "Rescaling the Energy Function in Hopfield Networks", *Proceedings of the IEEE International Joint Conference on Neural Networks* vol. 6, pp. 498-504, 2000.