

Network Optimisation - A Statistical Physics Perspective

K. Y. Michael Wong
Department of Physics
Hong Kong University
of Science and Technology
Hong Kong, China
Email: phkywong@ust.hk

David Saad
Aston University
Neural Computing Research Group
Birmingham, B4 7ET, United Kingdom
Email: d.saad@aston.ac.uk

C. H. Yeung
Department of Physics
Hong Kong University
of Science and Technology
Hong Kong, China
Email: phbill@ust.hk

Abstract—Inference and optimisation of real-value edge variables in sparse graphs are studied using the tree based Bethe approximation optimisation algorithms. Equilibrium states of general energy functions involving a large set of real edge-variables that interact at the network nodes are obtained for networks in various cases. These include different cost functions, connectivity values, constraints on the edge bandwidth and the case of multiclass optimisation.

I. INTRODUCTION

The links between statistical physics models and a variety of inference and optimisation problems have been significantly strengthened over the last decade [1]. Two aspects of these similarities have been exploited. Macroscopically, using the statistical physics framework, one describes typical properties of the problem and provides valuable insight into its generic characteristics. Microscopically, established techniques of statistical physics such as the cavity method and Bethe approximation have been used for devising efficient inference algorithms, some of which have been independently discovered and used in other disciplines [2], [3], [4], [5]. These links have become more transparent with the increasing use of graphical models and iterative approximation techniques such as *belief propagation* (BP) to a variety of probabilistic information processing problems. By mapping the probabilistic dependencies between parameters onto edges in graphs, the problem could be approximately solved (exactly on polytrees) by iterating a set of message-passing equations which couple the conditional probabilities on neighbouring nodes [2], [3].

Most studies so far have focused on cases of discrete variables, while networks of continuous variables were much less explored. One of the main reasons for this limited activity is the difficulty in applying message passing approximation algorithms [2], [3] in this case as the discrete messages passed between variables become *functions* of real variables. Applied message passing for systems of real variables typically relies on modelling the functions using a reduced number of parameters [6]. We have recently studied systems with real variables that can be mapped onto a sparse graph and suggested an efficient message-passing approximation method for inference and optimisation, especially in the context of resource allocation on sparse random systems [7], [8] chosen

as an exemplar of practical relevance.

Traditionally, network resource allocation and routing problems have been solved by *global* optimisation techniques, such as linear or quadratic programming [9]. However, with the increasing network size and/or the evolving configuration of wireless networks, centralised control becomes increasingly costly and infeasible. *Distributed control* in networks consists of a group of independent controllers which make locally optimal decisions. Compared with the traditional centralised approach, this has the advantages of a lower computational load and communication overhead, and robustness against network breakdown; BT's *Dynamic Alternative Routing* was an early successful example [10]. Also, in computer science, many algorithmic solutions have been proposed to distribute computational load between computers connected in a network, but they tend to be more heuristic; some aim to optimise the benefit to an individual node or task, without considering the impact it makes to the remainder of the network [11].

The optimisation problems addressed here include the resource allocation problem under various cost functions and connectivity profiles as well as the cases of restricted bandwidth and multiclass optimisation. Emphasis will be given to results obtained in the various cases; details of the analysis and algorithms derived for the various cases will be only briefly mentioned and can be found in [8], [12], [13] and [14].

II. THE BASIC MODEL

We consider the following resource allocation task on a sparse graph of N nodes, labelled $i = 1, \dots, N$. Each node i is randomly connected to c other nodes¹, and has a capacity Λ_i , randomly drawn from a distribution $\rho(\Lambda_i)$. The objective is to migrate tasks between nodes such that each node will be capable of carrying out its tasks. The *current* $y_{ij} \equiv -y_{ji}$ drawn from node j to i is aimed at satisfying the constraint

$$\sum_j \mathcal{A}_{ij} y_{ij} + \Lambda_i \geq 0, \quad (1)$$

¹Although we focus here on graphs of fixed connectivity, one can easily accommodate any connectivity profile within the same framework; the algorithms presented later are completely general.

where $\mathcal{A}_{ij} = 1/0$ for connected/unconnected node pairs i and j , respectively. We consider the load balancing task of minimising the energy function (cost) $E = \sum_{(ij)} \mathcal{A}_{ij} \phi(y_{ij})$, where the summation (ij) runs over all pairs of nodes, subject to the constraints (1); $\phi(y)$ is a general function of the current y . For load balancing tasks, $\phi(y)$ is typically a convex function, which will be assumed in our study. The analysis of the graph is done by introducing the free energy $F = -T \ln \mathcal{Z}_y$ for a temperature $T \equiv \beta^{-1}$, where \mathcal{Z}_y is the partition function

$$\mathcal{Z}_y = \prod_{(ij)} \int dy_{ij} \prod_i \Theta \left(\sum_j \mathcal{A}_{ij} y_{ij} + \Lambda_i \right) \exp \left[-\beta \sum_{(ij)} \mathcal{A}_{ij} \phi(y_{ij}) \right].$$

The Θ function returns 1 for non-negative arguments and 0 otherwise.

The system can be analysed using various methods such as the replica method [1], [15] and the Bethe approximation. We will briefly outline here the tree based Bethe approximation while full details of the derivation as well as the complete replica based analysis and the identification of links between these approaches and BP can be found in [8].

When the connectivity c is low, the probability of finding a loop of finite length on the graph is low, and the Bethe approximation well describes the local environment of a node. In the approximation, a node is connected to c branches in a tree structure and correlations among the tree branches can be neglected. In each branch, nodes are arranged in generations. A node is connected to an ancestor node and $c-1$ descendent nodes of the previous and generations, respectively.

Consider a vertex $V(\mathbf{T})$ of capacity $\Lambda_V(\mathbf{T})$, and a current y is drawn from the vertex. One can write an expression for the free energy $F(y|\mathbf{T})$ as a function of the free energies $F(y_k|\mathbf{T}_k)$ of its descendants, that branch out from this vertex

$$F(y|\mathbf{T}) = -T \ln \left\{ \prod_{k=1}^{c-1} \left(\int dy_k \right) \Theta \left(\sum_{k=1}^{c-1} y_k - y + \Lambda_V \right) \times \exp \left[-\beta \sum_{k=1}^{c-1} (F(y_k|\mathbf{T}_k) + \phi(y_k)) \right] \right\}, \quad (2)$$

where \mathbf{T}_k represents the tree terminated at the k^{th} descendent of the vertex. The free energy can be considered as the sum of two parts, $F(y|\mathbf{T}) = N_{\mathbf{T}} F_{\text{av}} + F_V(y|\mathbf{T})$, where $N_{\mathbf{T}}$ is the number of nodes in the tree \mathbf{T} , F_{av} is the average free energy per node, and $F_V(y|\mathbf{T})$ is referred to as the *vertex free energy*². Note that when a vertex is added to a tree, there is a change in the free energy due to the added vertex. Since the number of nodes increases by 1, the vertex free energy is obtained by subtracting the free energy change by the average free energy.

²This term is marginalised over all inputs to the current vertex, leaving y as its sole argument, hence the terminology used.

This allows us to obtain the recursion relation

$$F_V(y|\mathbf{T}) = -T \ln \left\{ \prod_{k=1}^{c-1} \left(\int dy_k \right) \Theta \left(\sum_{k=1}^{c-1} y_k - y + \Lambda_V(\mathbf{T}) \right) \times \exp \left[-\beta \sum_{k=1}^{c-1} (F_V(y_k|\mathbf{T}_k) + \phi(y_k)) \right] \right\} - F_{\text{av}}, \quad (3)$$

and the average free energy per node is given by

$$F_{\text{av}} = -T \left\langle \ln \left\{ \prod_{k=1}^c \left(\int dy_k \right) \Theta \left(\sum_{k=1}^c y_k + \Lambda_V \right) \times \exp \left[-\beta \sum_{k=1}^c (F_V(y_k|\mathbf{T}_k) + \phi(y_k)) \right] \right\} \right\rangle_{\Lambda}, \quad (4)$$

where Λ_V is the capacity of the vertex V fed by c trees $\mathbf{T}_1, \dots, \mathbf{T}_c$, and $\langle \cdot \rangle_{\Lambda}$ represents the average over the distribution $\rho(\Lambda)$. In the zero temperature limit, Eq. (3) reduces to

$$F_V(y|\mathbf{T}) = \min_{\{y_k | \sum_{k=1}^{c-1} y_k - y + \Lambda_V(\mathbf{T}) \geq 0\}} \left[\sum_{k=1}^{c-1} (F_V(y_k|\mathbf{T}_k) + \phi(y_k)) \right] - F_{\text{av}}. \quad (5)$$

The current distribution and the average free energy per link can be derived by integrating the current y' in a link from one vertex to another, fed by the trees \mathbf{T}_1 and \mathbf{T}_2 , respectively; the obtained expressions are $P(y) = \langle \delta(y - y') \rangle_{\star}$ and $\langle E \rangle = \langle \phi(y') \rangle_{\star}$ where $\langle \bullet \rangle_{\star}$ represents

$$\left\langle \frac{\int dy' \exp [-\beta (F_V(y'|\mathbf{T}_1) + F_V(-y'|\mathbf{T}_2) + \phi(y'))]}{\int dy' \exp [-\beta (F_V(y'|\mathbf{T}_1) + F_V(-y'|\mathbf{T}_2) + \phi(y'))]} \right\rangle_{\Lambda}.$$

Equations (3) and (5) are at the core of our analysis; they are being solved iteratively and facilitate the derivation of the free energy per link, internal energy and current distribution.

We have outlined only the main steps of the derivation to provide a rough idea for the methods used to obtain the numerical results presented later on. Slight modifications will be applied in the cases of different costs [8] and bandwidth limited network [12].

III. DISTRIBUTED ALGORITHMS

The local nature of the recursion relation Eq. (5) points to the possibility that the network optimisation can be solved by message passing approaches, which have been successful in problems such as error-correcting codes [2] and probabilistic inference [3]. The major advantage of message passing is its potential to solve a global optimisation problem via local updates, thereby reducing the computational complexity. For example, the computational complexity of quadratic programming for the load balancing task typically scales as N^3 , whereas capitalising on the network topology underlying the connectivity of the variables, message passing scales as N . An even more important advantage, relevant to practical implementation, is its distributive nature; it does not require a global optimiser, and is particularly suitable for distributive control in evolving networks.

However, in contrast to other message passing algorithms which pass conditional probability estimates of *discrete variables* to neighbouring nodes, the messages in the present context are more complex, since they are *functions* $F_V(y|\mathbf{T})$ of the current y . We simplify the message to two parameters, namely, the first and second derivatives of the vertex free energies. For the quadratic load balancing task, it can be shown that a self-consistent solution of the recursion relation, Eq. (5), consists of vertex free energies which are piecewise quadratic with continuous slopes. This makes the 2-parameter message a very precise approximation.

Let $(A_{ij}, B_{ij}) \equiv (\partial F_V(y_{ij}|\mathbf{T}_j)/\partial y_{ij}, \partial^2 F_V(y_{ij}|\mathbf{T}_j)/\partial y_{ij}^2)$ be the message passed from node j to i ; using Eq.(5), the recursion relation of the messages become

$$A_{ij} \leftarrow -\mu_{ij}, \quad B_{ij} \leftarrow \Theta(-\mu_{ij}) \left[\sum_{k \neq i} \mathcal{A}_{jk} (\phi'_{jk} + B_{jk})^{-1} \right]^{-1}, \quad (6)$$

where,

$$\mu_{ij} = \min \left[0, \frac{\sum_{k \neq i} \mathcal{A}_{jk} [y_{jk} - (\phi'_{jk} + A_{jk})(\phi'_{jk} + B_{jk})^{-1}] + \Lambda_j - y_{ij}}{\sum_{k \neq i} \mathcal{A}_{jk} (\phi'_{jk} + B_{jk})^{-1}} \right]$$

with ϕ'_{jk} and ϕ''_{jk} representing the first and second derivatives of $\phi(y)$ at $y = y_{jk}$ respectively. The forward passing of the message from node j to i is followed by a backward message from node j to k for updating the currents y_{jk} according to

$$y_{jk} \leftarrow y_{jk} - \frac{\phi'_{jk} + A_{jk} + \mu_{ij}}{\phi''_{jk} + B_{jk}}; \quad (7)$$

updated nodes are selected randomly and sequentially.

A second type of algorithm we have introduced, termed *price iteration*, is based on the optimisation of the cost given a set of constraints (e.g., non-negative resources $\sum_j \mathcal{A}_{ij} y_{ij} + \Lambda_i$) enforced by a set of Lagrange multipliers, for instance, of the form $L = \sum_{(ij)} \mathcal{A}_{ij} y_{ij}^2/2 + \sum_i \mu_i (\sum_j \mathcal{A}_{ij} y_{ij} + \Lambda_i)$. The optimisation task can be transformed to an adjoint optimisation problem of the Lagrange multipliers $\{\mu_i\}$, also termed *chemical potentials*. The extremum condition yields $y_{ij} = \phi'^{-1}(\mu_j - \mu_i)$, and using the K uhn-Tucker condition, μ_i can be solved in terms of μ_j of its neighbours, namely,

$$\mu_i = \min(g_i^{-1}(0), 0); \quad g_i(x) = \sum_j \mathcal{A}_{ij} \phi'^{-1}(\mu_j - x) + \Lambda_i.$$

This provides a local iterative optimisation method. We may interpret this algorithm as a *price iteration* scheme, by noting that the Lagrangian can be written as $L = \sum_{(ij)} \mathcal{A}_{ij} L_{ij} + \text{constant}$, where $L_{ij} = \phi(y_{ij}) + (\mu_i - \mu_j)y_{ij}$. Therefore, the problem can be decomposed into independent optimisation subproblems, each for a current on a link. μ_i is the storage price at node i , and each subproblem involves balancing the transportation cost on the link, and the storage cost at node i less that at node j , yielding the optimal solution. This provides

a pricing scheme for the individual links to optimise, which simultaneously optimise the global performance [10].

IV. QUADRATIC COST

Arguably the most illuminating example is that of quadratic cost. The solution of Eq. (5) is obtained numerically. Since the vertex free energy of a node depends on its own capacity and the disordered configuration of its descendants, we generate 1000 nodes at each iteration of Eq. (5), with capacities randomly drawn from the distribution $\rho(\Lambda)$, each being fed by $c-1$ nodes randomly drawn from the previous iteration. The free and internal energies and current distribution are computed numerically from the vertex free-energy distribution.

Convergence -To study the convergence rate of the iterations, we fit the average energy at iteration step t using $\langle E(t) - E(\infty) \rangle \sim \exp(-\gamma t)$ in the asymptotic regime. As shown in the inset of Fig. 1(a), the relaxation rate γ increases with the average capacity. It is interesting to note that a cusp exists at the average capacity of about 0.45. Below that value, convergence of the iteration is slow, since the average energy curve starts to develop a plateau before the final convergence. On the other hand, the plateau disappears and the convergence is fast above the cusp. The slowdown of convergence below the cusp is probably due to the appearance of increasingly large clusters of nonzero currents on the network, since clusters of nodes with negative capacities become increasingly extensive, and need to draw currents from increasingly extensive regions of nodes with excess capacities to satisfy the demand.

Dependence on the connectivity - We have observed the scaling laws of the currents [7] (roughly as c^{-1}), and average energy $c^2 \langle \phi \rangle$. A more accurate scaling in the high connectivity limit has been derived in [8] both analytically, in the large c limit and **numerically**; it depends on an empirical scaling factor $s = \sqrt{\frac{\lim_{c \rightarrow \infty} c^2 \langle \phi \rangle}{\langle \phi \rangle}} \approx 1.02c - 0.43$, converging to c in the high connectivity limit, as shown in Fig. 1(b).

V. OTHER COSTS

Two other representative costs have been studied are:

- (a) The *anharmonic* cost function is used to model the effects of costs rising faster than quadratic $\phi(y) = y^2/2 + u|y|^3/3$.
- (b) The *frictional* cost function is used to model the effects of frictional forces, which add an extra cost per unit current in a link irrespective of direction. Hence it is also useful in networks with multiple classes of traffic sharing the same links. The cost function takes the form $\phi(y) = y^2/2 + v|y|$.

Note that these cost functions represent two distinct cases. The former has well defined first and second derivatives for all of its arguments. In the latter case, the frictional cost function does not have a second derivative at $y = 0$. There is a kink in the cost function at the point of zero current, thus increasing the preference for idle links, or equivalently the reluctance to switch on a link. Results in both cases have been obtained using the message passing and price iteration algorithms.

Figure 2 shows results obtained for the chemical potentials distribution for the case of anharmonic and friction costs, obtained by price iteration. While results for the anharmonic

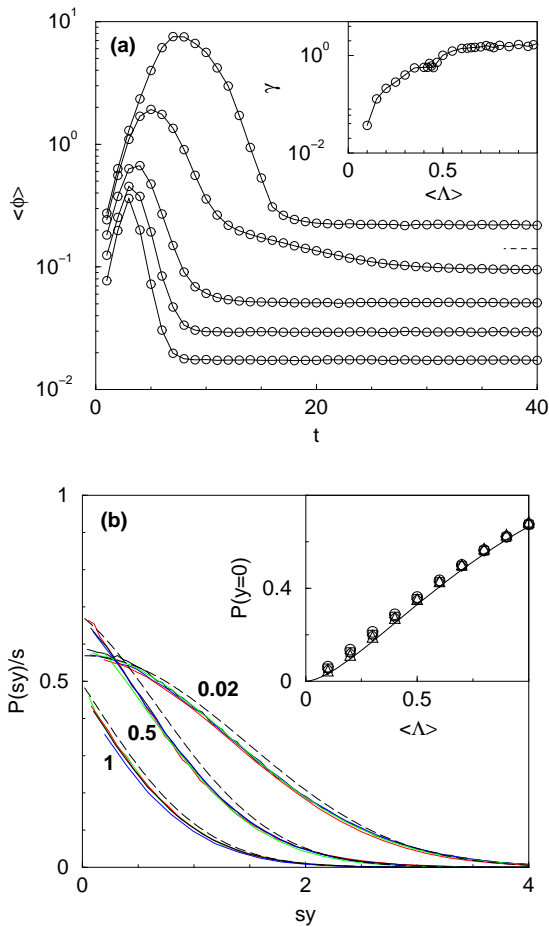


Fig. 1. Results for system size $N=1000$ and $\phi(y) = y^2/2$. (a) $\langle\phi\rangle$ obtained by iterating Eq. (5) as a function of t for $\langle\Lambda\rangle = 0.1, 0.2, 0.4, 0.6, 0.8$ (top to bottom) and $c=3$. Dashed line: The asymptotic $\langle E\rangle$ for $\langle\Lambda\rangle = 0.1$. Inset: γ as a function of $\langle\Lambda\rangle$. (b) The continuous component of the current distribution $P(sy)/s$ for $\langle\Lambda\rangle = 0.02, 0.5, 1$. Lines: $c=3$ (solid), 4 (dotted), 5 (dashed), 10 (dot-dashed), high c (long dashed). Inset: $P(y=0)$ as a function of $\langle\Lambda\rangle$. Symbols: $c=3$ (\circ), 4 (\square), 5 (\diamond), 10 (\triangle), high c (line).

case are not qualitatively different than for the quadratic cost, results in the friction cost case show a distinctly different behaviour - a much lower value between the zero chemical potential and some negative value, a threshold phenomena that can be explained by the need for a sufficiently high chemical potential difference to get over the friction term (Fig. 2(b)).

VI. BANDWIDTH LIMITED SYSTEMS

Another aspect of the problem, which is arguably more realistic, includes a bandwidth restriction on the network edges; currents are bounded by a predetermined coefficient W . For links with small bandwidth W , or nodes with negative average capacity, there exist nodes which violate the capacity constraint. In these unsatisfiable cases, it is expedient to relax the constraints and search for optimal solutions which limit

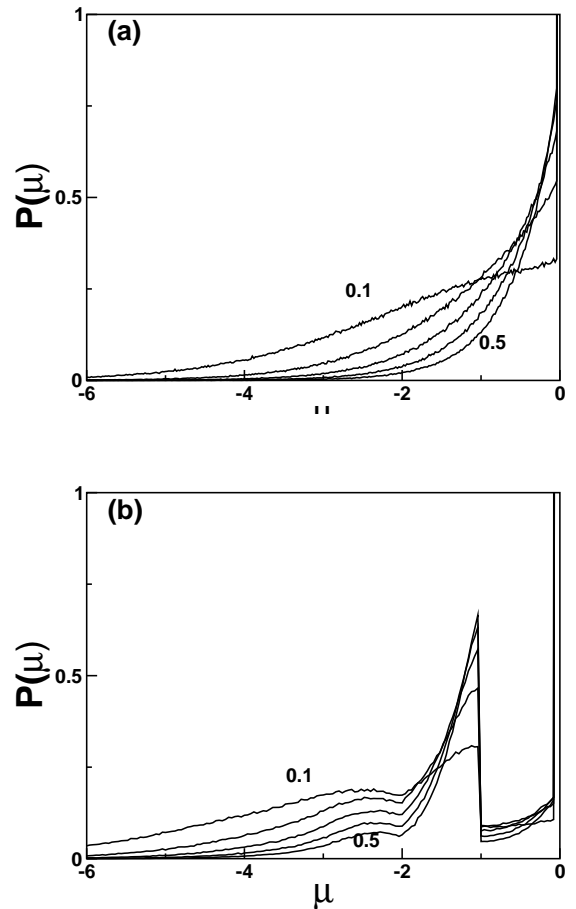


Fig. 2. Results for $N=1000$, $c=3$ and 1000 samples - chemical potential distribution $P(\mu)$ for $\langle\Lambda\rangle = 0.1, 0.2, 0.3, 0.4, 0.5$ (top to bottom). (a) Anharmonic cost function with $u=1$; (b) Frictional cost with $v=1$.

the violations. Hence we consider the cost

$$E = \frac{1}{2} \sum_i \Theta \left[- \sum_j \mathcal{A}_{ij} y_{ij} - \Lambda_i \right] \left[\sum_j \mathcal{A}_{ij} y_{ij} + \Lambda_i \right]^2 + \frac{R}{2} \sum_{(ij)} y_{ij}^2. \quad (8)$$

The first term represents the cost of unsatisfaction while the second term represents the transportation cost; R serves as the weight of transportation cost in the global cost. Here we study the case when the unsatisfaction cost dominates, i.e. R is small. The message-passing can be derived in a similar manner to the unlimited bandwidth case [12].

Figure 3(a) shows results for various $\langle\Lambda\rangle$ values and fixed bandwidth $W=1$. The average cost increases rapidly when $\langle\Lambda\rangle$ enters the unsatisfiable regime, and the results obtained by the theory, the message-passing and price iteration algorithms show excellent agreement. There are three types of links in the network: idle ($|y_{ij}|=0$), unsaturated ($|y_{ij}| < W$) and saturated ($|y_{ij}|=W$). When $\langle\Lambda\rangle$ enters the unsatisfiable regime, the fraction of idle links vanishes rapidly, while that of saturated links increases to a steady level, implying that more resources

are transported in the links in response to the networkwide demand on resources (Fig. 3(a) inset).

Figure 3(b) shows the simulation results when W varies. For large values of W , the average cost is effectively constant, since the link bandwidth constraints become irrelevant. On the other hand, when W decreases, the average cost increases rapidly, since the links become increasingly ineffective in allocating resources in the network. The inset shows how the fraction of saturated links increases when W decreases. It is interesting to note that the fraction of idle links *increases* when W decreases, contrary to the expectation that more links are involved in resource provision. This can be explained in terms of the *bottleneck effect* as follows; if the links in the network were unconstrained, nodes with sufficiently large violations would have drawn currents from distant neighbours, causing currents to flow through many intermediate nodes. However, when W is small, the currents drawn by nodes with violations from their nearest neighbours may have already saturated the links, and there is no use to draw currents from further neighbours. In the limit of vanishing W , the links are exclusively either idle or saturated. In this limit, a link is idle only when both nodes at its ends have positive Λ . Hence the fraction of idle links is $f_{\text{idle}} = 1 - f_{\text{sat}} = [P(\Lambda > 0)]^2$. Since the transportation cost is negligible in this limit, the contribution to the average cost only comes from the violated nodes, given by $\langle \phi \rangle / N = \langle \Theta(-\Lambda) \Lambda^2 / 2 \rangle_{\Lambda}$. The case of bandwidth limited network optimisation has been further investigated in [13].

VII. MULTICLASS TRANSMISSION

In the case of multiclass optimisation one considers the allocation of D classes of resources on the same sparse network. The capacity of class ν at node i is defined as Λ_i^ν and the currents y_{ij}^ν satisfy the resource constraints

$$\sum_j \mathcal{A}_{ij} y_{ij}^\nu + \Lambda_i^\nu \geq 0. \quad (9)$$

The links are duplex, carrying currents flowing in both directions. The total current of link (ij) is given by $\sum_\nu |y_{ij}^\nu|$. One aims to optimise the cost function

$$E = \sum_{(ij)} \mathcal{A}_{ij} \phi(\{y_{ij}^\nu\}), \quad \phi(\{y_{ij}^\nu\}) = \frac{1}{2} \left(\sum_\nu |y_{ij}^\nu| \right)^2. \quad (10)$$

A price iteration and message passing algorithms can be derived in a similar manner for this case [14]. The price iteration algorithm uses gradient descent with a learning rate η and takes the form

$$\mu_i^\nu = \mu_i^\nu + \eta \{ \min([g_i^\nu]^{-1}(0), 0) - \mu_i^\nu \},$$

using the notation of the quadratic case.

Since the total current $\sum_\nu |y_{ij}^\nu|$ in a link should scale as D , it is convenient to rescale the energy per link $\langle \phi \rangle$ by D^2 . As shown in Fig. 4, $\langle \phi \rangle / D^2$ decreases with the number of classes. This shows that the network is able to increase its efficiency due to the reduced fluctuations of the total currents

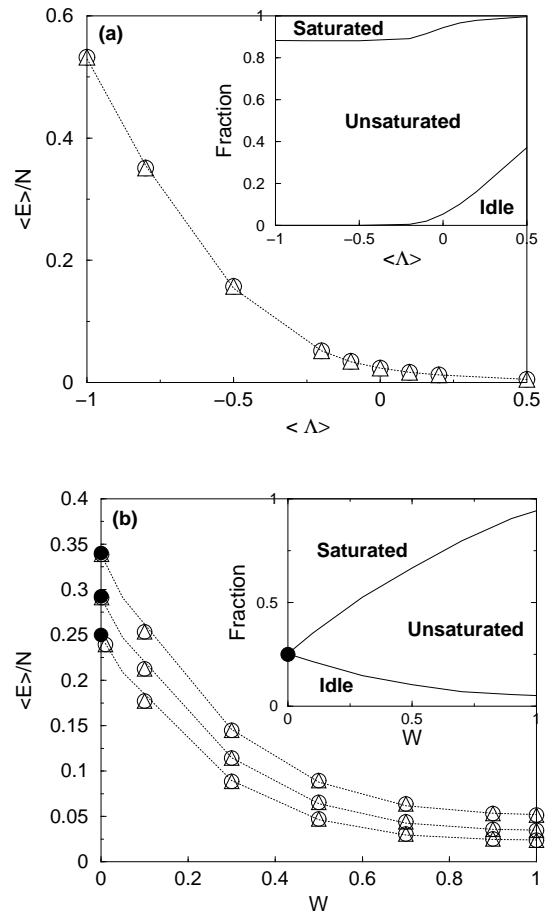


Fig. 3. Results for $N=1000$, $c=3$, $R=0.1$ and 100 samples. (a) $\langle E \rangle / N$ as a function of $\langle \Lambda \rangle$ for $W = 1$. Symbols: Bethe approximation (dotted-line), message-passing (\triangle), price iteration (\circ). Inset: the fraction of idle, unsaturated and saturated links as a function of $\langle \Lambda \rangle$ for $W = 1$; the vertical height of each region for a given $\langle \Lambda \rangle$ corresponds to the respective fraction. (b) $\langle E \rangle / N$ as a function of W for $\langle \Lambda \rangle = 0$. Symbols: Bethe approximation (dotted-line), message-passing (\triangle), price iteration (\circ), $W \rightarrow 0$ theoretical limit (\bullet). $\langle \Lambda \rangle = -0.2, -0.1, 0$ from top to bottom line. Inset: the fraction of idle, unsaturated and saturated links as a function of W for $\langle \Lambda \rangle = 0$. Symbol: $W \rightarrow 0$ theoretical limit of the fraction of idle links (\bullet).

in the links. In the engineering literature, this is known as *statistical multiplexing*.

Figure 4 also shows that the fraction of idle links increases with D , presumably due to friction-like effects resulting from the interaction between the different classes on the same edges.

VIII. SUMMARY

We have studied the problem of network optimisation in the case of sparsely connected graphs. The problem has been investigated using various methods of statistical physics; in this proceedings we outline the derivation based on the Bethe approximation and some of the distributed optimisation methods devised for solving the problem; the presentation is brief due to lack of space and we refer the reader to the corresponding papers. We also briefly describe some of the key results obtained focusing on the resulting internal energy per node, and current and chemical potential distributions for

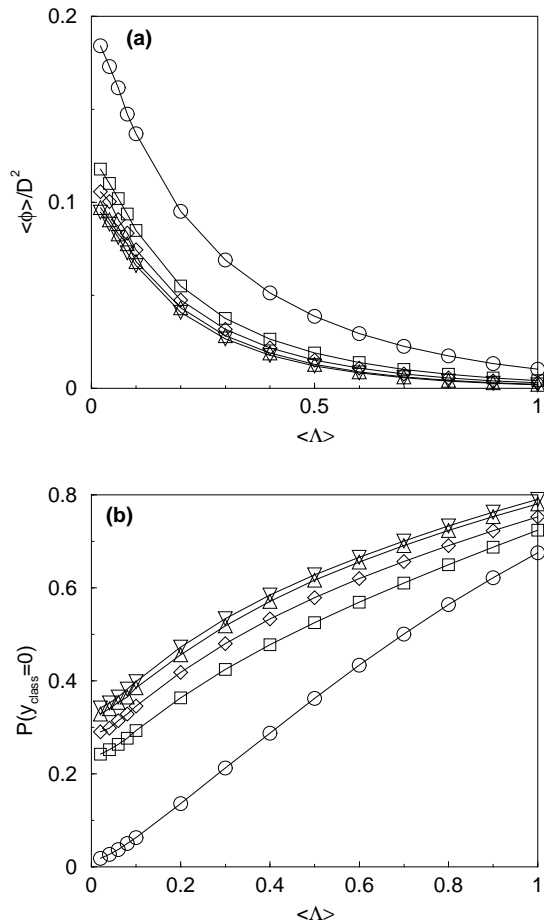


Fig. 4. (a) The dependence of the asymptotic $\langle \phi \rangle / D^2$ on $\langle \Lambda \rangle$ for given values of D . (b) The dependence of the fraction of zero class current on $\langle \Lambda \rangle$ for given values of D . Symbols: $D=1$ (\circ), 3 (\square), 5 (\diamond), 10 (\triangle), 15 (∇). Parameters: $N = 1000$, $c = 3$. Ranges of other parameters: from $\eta = 1$, 500 steps and 1000 samples at $D = 1$ to $\eta = 0.7$, 5000 steps and 200 samples at $D = 15$.

the various costs considered. We also reported briefly results on the bandwidth limited case where edge currents are limited and the case of multiclass optimisation.

The main contribution to the field of network optimisation is twofold. Firstly, using methods of statistical mechanics one can better analyse the properties of typical networks and gain generic insight on the behaviour of the network. Secondly, based on this insight one can devise novel optimisation algorithms. One can identify potential contributions also in the area of labelled data transmission and in the limit of a large number of different classes; research activities in both directions are under way.

Acknowledgements: Support from Evergrow, EU-IP 1935, EPSRC E049516 and Research Grant Council of Hong Kong (HKUST603606 and HKUST603607) is acknowledged.

REFERENCES

[1] H. Nishimori *Statistical Physics of Spin Glasses and Information Processing*, Oxford University Press, Oxford UK (2001)
 [2] M. Opper and D. Saad *Advanced Mean Field Methods: Theory and Practice*, MIT Press, Cambridge MA (2001)

[3] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press (2003)
 [4] J. S. Yedidia, W. T. Freeman and Y. Weiss, *NIPS* **13** 689 (2000)
 [5] M. Mézard, G. Parisi and R. Zecchina, *Science* **297** 812 (2002)
 [6] S. L. Lauritzen, *J. of the American Stat. Assoc.* **87** 1098 (1992)
 [7] K. Y. M. Wong, D. Saad, and Z. Gao, *Advances in Neural Information Processing Systems* **18**, 1529, Y. Weiss, B. Schölkopf and J. Platt eds, MIT Press, Cambridge (2006)
 [8] K. Y. M. Wong and D. Saad, *Phys. Rev. E* **74**, 010104(R) (2006); *Phys. Rev. E* **76**, 011115 (2007)
 [9] D. Bertsekas, *Linear Network Optimization*, MIT Press, Cambridge MA (1991)
 [10] F. P. Kelly, *Phil. Trans. R. Soc. Lond. A* **337**, 343 (1991)
 [11] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, NJ (1992)
 [12] K. Y. M. Wong, C. H. Yeung, and D. Saad, *Lecture Notes Comput. Sci.* **4233** Part II, 754, I. King *et al* eds, Springer, Berlin (2006)
 [13] C. H. Yeung and K. Y. M. Wong, in preparation (2008)
 [14] K. Y. M. Wong, in preparation (2008)
 [15] M. Mezard and G. Parisi and M.A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore (1987)