On Phase Transitions in Learning Sparse Networks

Goele Hollanders[a] Geert Jan Bex[a] Marc Gyssens[a] Ronald L. Westra[b]

Karl Tuyls[b]

[a] *Department of Mathematics, Physics, and Computer Science,*
Hasselt University and Transnational University of Limburg,
Hasselt, Belgium
E-mail to: goele.hollanders@uhasselt.be
[b] *Department of Mathematics and Computer Science,*
Maastricht University and Transnational University of Limburg,
Maastricht, the Netherlands

In this paper [1] we study the identification of sparse interaction networks, from a given set of observations, as a machine learning problem. An example of such a network is a sparse gene-protein interaction network, for more details see [2]. Sparsity means that we are provided with a small data set and a high number of unknown components of the system, most of which are zero. Under these circumstances, a model needs to be learned that fits the underlying system, capable of generalization. This corresponds to the student-teacher setting in machine learning.

Problem setting In some engineering applications, the number of measurements M available for system identification and model validation is much smaller than the system order N, which represents the number of components. This substantial lack of data can give rise to an identifiability problem, in which case a larger subset of the model class is entirely consistent with the observed data so that no unique model can be proposed. Since conventional techniques for system identification are not well suited to deal with such situations, it thus becomes important to work around this by exploiting as much additional information as possible about the underlying system. In particular, we are interested in the relation between the number of measurements and the number of components, the sparsity of the network and the influence of noise.

Definitions and Algorithm In the first part of this paper we introduce a learning algorithm, based on L_1-minimization, to identify interaction networks from poor data and analyze its dynamics with respect to phase transitions.

Therefore we assume that a *training set* of M input/output pairs $\chi_{\text{tr}} = \{(x_m, \dot{x}_m) \mid m : 1, \ldots, M\}$ is given, where $x_m, \dot{x}_m \in \mathbb{R}^N$. The components of the input vectors x_m are independently and identically distributed so that they are linearly independent. Since the data is assumed to be generated by some interaction network, this network will be denoted by $T = (A^T, B^T)$ where $A^T \in \mathbb{R}^{N\times N}$ and $B^T \in \mathbb{R}^N$. In this context, we refer to T as the unknown *teacher* and its output $\dot{x}$ on some input x is computed as follows: $\dot{x} = T(x) \equiv A^T \cdot x + B^T$.

The learning algorithm should return a network $S = (A^S, B^S)$, referred to as the *student*, with $A^S \in \mathbb{R}^{N\times N}$ and $B^S \in \mathbb{R}^N$ and with as many zeros as possible. This student should be able to reproduce the training set χ_{tr}: $\dot{x}_m = A^S \cdot x_m + B^S$ for $m : 1, \ldots, M$ and generalize beyond. The efficiency of the algorithm is measured by the generalization error ϵ_{gen}, which represents the probability that the student is a good fit to the teacher. Our findings illustrate that the quality of the fit depends on several factors such as the ratio of the training set size to the system size and on the sparsity of the network. The transition towards generalization is quite abrupt at an α-value of 0.28, indicative of a phase transition. From this training set size onwards the student will be a good fit to the teacher, see Fig. 1. Concerning the sparsity of the network we observed that when the number of non-zero components increases, the efficiency of the process will gradually increase. This

phenomenon is shown in Fig. 2.

Note that in this setting, it is natural to link network identification to feature selection. Only very few components influence the expression level of any given component, so one can restate the problem as selecting exactly those few among the large amount of components under consideration. Hence the results presented here will not only be applicable to network identification, but more generally to feature selection as well. As for network identification, we can define the generalization error for feature selection $\varepsilon^{\text{fs}}_{\text{gen}}$.

In the second part of this paper we show that from a system with a specific system size value the generalization error of other system sizes can be estimated. Suppose we have a curve for $\varepsilon^{\text{fs}}_{\text{gen}}$ versus α for N_0, then the curve for system size N can be obtained by scaling $\alpha(N) = \alpha^{\text{fs}}_{\text{gen}} + \sqrt{N_0/N}\,(\alpha(N_0) - \alpha^{\text{fs}}_{\text{gen}})$, with $\alpha^{\text{fs}}_{\text{gen}}$ the training set size for $\varepsilon^{\text{fs}}_{\text{gen}} = 1/2$. A comparison with a set of simulation experiments, Fig. 4, shows a very good fit.

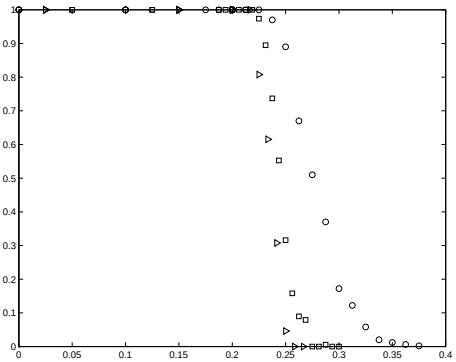

Figure 1: The generalization error ε_{gen} as a function of the training set size α for $N = 100$ (○), $N = 160$ (□) and $N = 300$ (▷) for $\kappa_T \approx 0.03$.

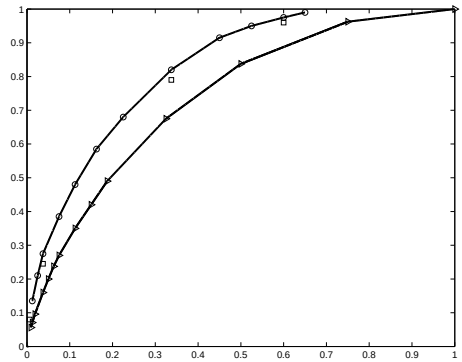

Figure 2: The generalization threshold α_{gen} as a function of the sparsity κ_T for $N = 80$ (○) and —for comparison— a few values for $N = 160$ (□). The generalization threshold for feature selection $\alpha^{\text{fs}}_{\text{gen}}$ (▷) is given as reference.

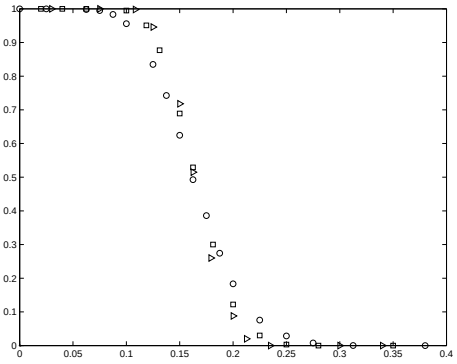

Figure 3: The generalization error for feature selection $\varepsilon^{\text{fs}}_{\text{gen}}$ as a function of the training set size α for $N = 100$ (○), $N = 160$ (□) and $N = 300$ (▷) for $\kappa_T \approx 0.03$.

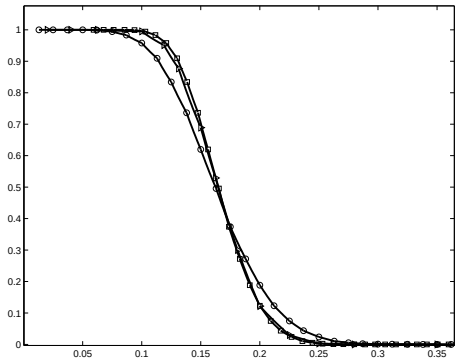

Figure 4: Fig. 3 using the scaling-equation for $N = 100$ (○), $N = 160$ (□) computed, $N = 160$ observed (▷), $\kappa_T \approx 0.03$.

References

[1] G. Hollanders, G. Bex, M. Gyssens, R. L. Westra and K. Tuyls. On Phase Transitions in Learning Sparse Networks. *Accepted for European Conference on Machine Learning (ECML)*, 2007.

[2] R. L. Westra, G. Hollanders, G. Bex, M. Gyssens and K. Tuyls. The identification of dynamic gene-protein networks. *LNBI*, 4366:157–170, 2007.