Weighted Low Rank Approximations

Nathan Srebro & Tommi Jaakkola

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: We study the problem of approximating a target matrix with a matrix of low rank, with respect to a weighted Frobenius norm.

Motivation: Low-rank matrix approximation with respect to the squared or Frobenius norm has wide applicability in estimation and can be easily solved with singular value decomposition. For many application, however, the deviation between the observed matrix and the low-rank approximation has to be measured relative to a weighted-norm.

Weighted-norms can arise in several situations. A zero/one weighted-norm, for example, arises when some of the entries in the matrix are not observed. External estimates of the noise variance associated with each measurement may be available (e.g. gene expression analysis) and using weights inversely proportional to the noise variance can lead to better reconstruction of the underlying structure. In other applications, entries in the target matrix represent aggregates of many samples. When using *un*weighted low-rank approximations (e.g. for separating style and content [4]), we assume a uniform number of samples for each entry. By incorporating weights, we can account for varying numbers of samples in such situations. Low-rank approximations are also used in the design of two-dimensional digital filters, in which case weights might arise from constraints of varying importance [3].

We also consider other measures of deviation, beyond weighted-Frobenius norms. Such measures arise, for example, when the noise model associated with matrix elements is known, but is not Gaussian. Classification, rather than regression, also gives rise to different measures of deviation. Classification tasks over matrices arise, for example, in the context of collaborative filtering. To predict the unobserved entries, one can fit a partially observed binary matrix using a logistic model with an underlying low-rank representation (input matrix).

Approach: We develop two alternate methods for solving the problem. We first study the problem from a numeric optimization perspective. We study in detail the structure of the problem and the differences between weighted and unweighted low-rank approximation which are responsible for the gap in their hardness. We then show a closed-form solution for part of the parameters of the problem, and calculate the derivatives with respect to the other parameters, enabling us to use standard numerical optimization techniques to solve the problem.

We also apply an expectation-maximization (EM) approach in which the problem is viewed as a missing-values problem. To do so, we actually relax the problem to a slightly more general problem that can be viewed as a missing-values problem. In fact, we consider a family of such problems, with increasingly more missing values, that approach a weighted low-rank approximation when the number of missing values goes to infinity. Each such problem can be solved using EM, and the solution is independent of the number of parameters, enabling us to apply the EM approach also for the weighted low-rank approximation problem. This results in an extremely simple iterative algorithm that can implemented in a few lines of code and is effective in finding weighted low-rank approximations in many cases.

Having developed methods for solving weighted low-rank approximation problems, we show how they can be applied as a subroutine, using iterative variational methods, for solving more general low-rank problems.

Impact: Low-rank matrix approximation is one of the most commonly used tools in estimation, control, machine learning, computer vision and in other areas of engineering and statistics. Currently, low-rank matrix approximation is almost invariantly performed with respect to the Frobenius norm. In many situations the use of the Frobenius norm is well founded, and is the appropriate loss function, corresponding to a maximum likelihood estimation in the presence of white Gaussian noise. However, in many other cases, the Frobenius norm is used not because it

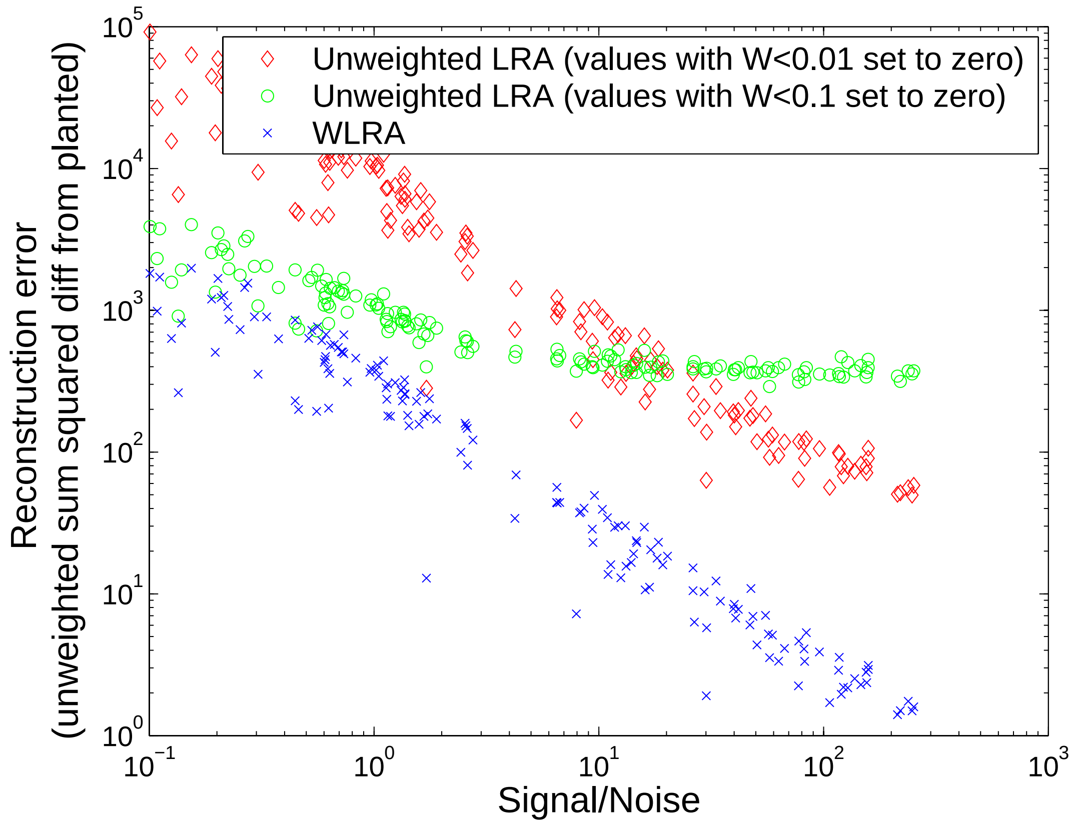

Figure 1: The improvement in reconstruction attained by introducing weights. Reconstruction error as a function of noise in reconstructing a planted low-rank matrix with simulated noise of varying variance. Weighted low-rank approximation is compared to a simpler approach of "clipping" entries with extreme noise variance.

is the *correct* loss function, but because it is the *convenient* loss function, for which optimization tools are readily available. Our work opens the door to other measures of deviation, and enables researchers and engineers to use loss functions which are appropriate for the task at hand.

Future Work: We are continuing to study the weighted low-rank approximation problem, understanding the sensitivity of EM and gradient methods to the weight distribution. We are applying these techniques to problems involving factor-gene binding arrays and robust collaborative filtering. We are also studying further loss functions, such as those in which the noise model is viewed as a nuisance parameter, and those arising in Sufficient Dimensionality Reductions [1].

Research Support: This work was supported in part by Nippon Telegraph and Telephone Corporation, by ARO MURI grant DAAD19-00-1-0466, and by NSF ITR grant #IIS-0085836.

References:

[1] Amir Globerson and Naftali Tishby. Sufficient dimensionality reduction - a novel analysis principle. In *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002.

[2] Michal Irani and P Anandan. Factorization with uncertainty. In *European Conference on Computer Vision*, June 2000.

[3] W.-S. Lu, S.-C. Pei, and P.-H. Wang. Weighted low-rank approximation of general complex matrices and its application in the design of 2-D digital filters. *IEEE Transactions on Circuits and Systems—I*, 44(7):650–655, July 1997.

[4] Joshua B. Tenenbaum and William T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12(6):1247–1283, 2000.