Optimizing Eigen-Gaps and Spectral Functions using Iterated SDP

Tony Jebara, Blake Shaw and Andrew Howard
Columbia University
New York, NY 10027
jebara@cs.columbia.edu

Semidefinite programming is widely used in machine learning and yields improvements for sparse PCA, robust SVMs, kernel selection, maximum margin matrix factorization, and nonlinear dimensionality reduction. In most formulations, the trace of a matrix (the sum of its eigenvalues) is maximized or minimized. The trace is convenient to implement but is usually only *loosely* related to the learning task at hand. We show how SDPs can readily accomodate more general spectral functions of the eigenvalues beyond the sum. For instance, one can optimize the eigengap of a matrix such that it has a few large eigenvalues while the bottom eigenvalues are aggressively driven towards zero. We revisit SDP learning formulations and extend solvers beyond the trace to novel spectral cost functions such as monotonic and eigengap spectral functions. These are easily solvable by interleaving SDP and SVD computations. These new costs improve machine learning SDP approaches producing lower dimensionality embeddings, lower rank matrix factorizations and improved VC-dimension estimates for SVMs.

First note that most SDP solvers (SeDuMi, CSDP and YALMIP) set up an optimization over matrices $K \in \Re^{D \times D}$ that are positive semidefinite $K \succeq 0$. The user further tailors the problem with extra linear inequalities $tr(B_i^T K) \geq \beta_i$ for $i = 1 \ldots N$ to restrict the solution to a subset of the PSD cone, denoted $\mathcal{K}$ for short. Finally, the user selects a cost to minimize which is either $\min_{K \in \mathcal{K}} tr(K)$, $\min_{K \in \mathcal{K}} tr(B^T K)$ or $\min_{K \in \mathcal{K}} - \log |K|$. While many machine learning problems can be squeezed into such a formula, it is restrictive and may lead us away from our desired cost function. Instead, consider the cost $\min_{K \in \mathcal{K}} \sum_{d=1}^{D} \alpha_d \lambda_d(K)$ where λ_d are the eigenvalues of K such that $\lambda_d \geq \lambda_{d+1}$. Clearly, if α_d is constant for all d we get back the regular $\min_{K \in \mathcal{K}} tr(K)$ problem. However, if α_d varies, the following holds:

$$\min_{K \in \mathcal{K}} \sum_{d=1}^{D} \alpha_d \lambda_d(K) = \min_{K \in \mathcal{K}} \min_{\substack{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_D \\ v_i^T v_j = \delta_{ij}}} \alpha_d tr(K v_d v_d^T)$$

In other words, the eigenvalues of the matrix K can be rewritten as an inner minimization over orthonormal vectors. This variatonal optimization now permits us to iteratively optimize the cost by solving for the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_D$ using SVD and then updating K using SDP. Convergence is monotonic and requires about a dozen outer loop iterations. Furthermore, when $\alpha_d \geq \alpha_{d+1}$ this is a convex optimization with a single global optimum. In experiments we see direct improvements over maximum variance unfolding (MVU) and maximum margin matrix factorization (MMMF) as well as other max trace problems.