

Perturbation for Sparse PCA: Project Overview

Daniel Ringwalt

April 1, 2023

1 Background

Berk et. al [1] created a modular, best-first search design for combinatorial Sparse PCA. Nodes in the search tree have a lower bound (generally stochastic method for a k -sparse eigenvalue approximation), as well as an upper bound (which is more challenging to bound). We work in this combinatorial framework, and try to improve performance of the lower and upper bounds.

We adapt Berk et. al's treatment of the Hermitian eigenproblem (PCA on the covariance matrix) to a singular value analysis, without changing the actual implementation. Sparse PCA is treated on a matrix of m observations and n variables: M , and considering different n -by- k matrices, D , where the columns are taken to be standard basis vectors (subsetting the columns of the identity matrix). Then $\max_D \|MD\|_2^2$ provides us with the maximum of Hermitian eigenproblem taking k rows and columns: $\max_D \|D^T \Sigma D\|$, where $\Sigma = M^T M$ is a sample covariance matrix.

Berk et. al perform a branch-and-bound [5] tree search, where the root node has D with no columns, the next level has $D_1 \in M_{n,1}(\mathbb{R})$, and has a terminal level where $D \in M_{n,k}(\mathbb{R})$. The singular value treatment immediately lends a rank-one update computational tactic to each graph edge [3], which was not pursued by Berk et. al. This tactic does not improve performance on its own, and analysis of the partial matrix MD_i and its perturbation needs to be performed in this project.

We will refer to the “trace bound”, which is a basic bound on the objective from the squared column norms of M (with equality if M is rank-one). If M is unit-normed, then the trace bound is a constant $\|MD\|_2^2 \leq k$. For more general data, then the trace bound admits a greedy tactic for variable selection, rather than Sparse PCA, by retaining only high-variance variables. We will display results from Berk et. al showing that although unit-normed data does not admit a greedy tactic, the covariance matrix's spectrum was actually more amenable to analysis (an observation from their program's performance).

2 Lower bounds on Sparse PCA

We will leverage: $\|D^T M^T u\|_2 \leq \|MD\|_2$. A good u will come, at a graph node that is not trivially shallow, from the first singular vector of MD_i at this node. The lower bound on the system (from projecting the system onto a subspace spanned by this vector) allows us to elide all of the stochastic, or power iteration, methods pursued by Berk et. al, but will not prove impactful on the provided code and data sets.

3 Upper bound from Frobenius norm

We will review Berk et. al's Gershgorin circle theorem bound, which appears as a useful tactic for bounding many matrices whose rows and columns are drawn from a discrete distribution (the set of full rows and columns of the covariance matrix). We will generalize this algorithm, instead of retaining one matrix row, to retain k matrix rows (after elementwise squaring) in order to bound the squared Frobenius norm $\|D^T \Sigma D\|_F^2$. We will place this bound into an ordering, compared to the Gershgorin bound and to the trace bound.

4 Perturbation upper bound from Davis-Kahan

We will finally treat the singular values of MD as a perturbation from MD_i . For this section, we will move to the Hermitian eigenproblem for $D^T \Sigma D$. We will justify the first singular vectors of MD_i and MD being similar, but the dimensions of MD_i possibly being very small. Instead, we will project all variables onto a one-dimensional subspace and its complement. The subspace is spanned by the first singular vector u of MD_i . We have $A = D^T M^T u u^T M D$, and $E = \Sigma - A$.

E is compared to Σ , and the Davis-Kahan $\sin \theta$ theorem is an extremely good fit to analyze the shrinking of a rank- k Hermitian matrix to the rank- $k - 1$ Hermitian matrix after projection. This produces an inequality formula with $\|E\|_2$ appearing on the left-hand side and $\|E\|^2$ (any unitarily invariant norm) appearing on the right-hand side. The right-hand occurrences of $\|E\|^2$ are substituted with a linear expression relaxing this norm to the squared Frobenius upper-bound, which is linear in the entries of D . This expression is added to the lower bound from projection ($\|A\|$), producing $\|A\| + \|E\|$ (bound on the Sparse PCA objective, from Weyl's inequalities). Compared to the trace bound, we have shrunken the value of some of the variables according to our perturbation analysis.

We started with one tactic which was linear in the entries of D (the trace bound). We produced a projected rank-one system (lower bound), a concave Frobenius norm bound (square root of the linear expression), and a linearized (Weyl) upper bound (using squared Frobenius norm bound), all of which have a simple greedy solver. As a tactic, we will introduce comparing the lower bound

and upper bound of variables' contributions (now that the contributions are linearized), as a novel variable selection step.

5 Future Directions

Berk's collaborator, Bertsimas, has produced follow-ups using mixed-linear solvers and branch-and-cut [2]. Our Davis-Kahan analysis produces interesting linear cuts in the variables, and may be best suited to branch-and-cut rather than branch-and-bound.

References

- [1] Berk, Lauren, & Dimitris Bertsimas. "Certifiably optimal sparse principal component analysis." *Mathematical Programming Computation* 11.3 (2019): 381-420.
- [2] Bertsimas, D., Cory-Wright, R., & Pauphilet, J. (2022). Solving Large-Scale Sparse PCA to Certifiable (Near) Optimality. *J. Mach. Learn. Res.*, 23(13), 1-35.
- [3] Bunch, James R., & Christopher P. Nielsen. "Updating the singular value decomposition." *Numerische Mathematik* 31.2 (1978): 111-129.
- [4] Davis, C., & Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1), 1-46.
- [5] Lawler, E. L., & Wood, D. E. (1966). Branch-and-bound methods: A survey. *Operations research*, 14(4), 699-719.