Repeated application of Zhu's rank-one update bounds

Daniel Ringwalt

December 4, 2022

1 Background

Rank-one update of a symmetric matrix has efficient implementations, and now has a bound that behaves reasonably as expected for a single perturbed eigenvalue in $O(n)$ (or possibly $O(1)$ time). Successive rank-one updates, which are random variables drawn from a pool of potential candidates, are likely to produce poor bounds. However, the known upper bounds also have notably poor performance (putting the updates in columns of a matrix, and using Gershgorin's circle theorem for a subset of columns maximizing any eigenvalue).

2 Algorithm

Each rank-one update (add a variable to the sparse PCA) is drawn from a pool of variables, without replacement. After each update, we store a lower and upper bound of every eigenvalue. The bounds might start to overlap (especially, the small eigenvalues will no longer be useful), but Zhu computes the sequence of eigenvalues placed in order, so the difference $\lambda_{i-1} - \lambda_i$ of descending eigenvalues is capped at a lower bound of 0. The upper bound for the gap is computed using the lower bound of the smaller eigenvalue, and the upper bound of the larger eigenvalue.

When perturbing the eigenvalues, the eigenvectors are also perturbed. Rank-one updates produce an eigenvector (not to scale) $(\Lambda - \lambda_i' I)^{-1} v$ in the diagonalized system, which is rotated using the transposed eigenvector matrix of the original system. The uncertainty in the dominant eigenvector is stored as an angle $|\theta|$, based on the arc cosine of the first component of the eigenvector which we have given, and the value in radians is added after each successive update. The remaining components of the rank-one update vector v can be normed, and either $|v_{2:k}| \sin \theta$ is rotated to grow the size of the first component of the update vector (upper bound), or is rotated away from it.

Lower bounding the eigenvalue would become inaccurate with significant θ. When θ reaches $\frac{\pi}{2}$, then it should be saturated, and further updates will update

only the dominant eigenvalue upper bound, by adding the trace of vv^T to the upper bound.

We should probably also capture the radial uncertainty in the second eigenvector by multiplying a 3x3 rotation matrix, instead of using an angle. The upper bound of the first eigenvalue and lower bound of the second eigenvalue form a system (Zhu's formulae use the difference, or gap, between eigenvalues). When we have further eigenvectors which are never updated, then the second eigenvalue upper bound might no longer be able to rely on either bound of the third eigenvalue. We might use simple bounds on successive eigenvalues (Gershgorin upper bound, system before update as the lower bound). Zhu's formulae, applied to the dominant two eigenvalues only, might be strictly better for performance than Gershgorin upper bound for all eigenvalues.

3 Searching for extrema of Zhu's lower and upper bounds

We need to search for critical points, in case Zhu's formulae must be applied to the system using any matrix entries, other than those extremes of the variables which we are tracking.

3.1 Lower bound (case 1)

This is the lower bound on the dominant eigenvalue. Let the unit-norm rank-one update vector (rotated along the eigenvectors) be v, and let $w_i = v_i^2$ (the diagonals of the rank-one update). Let $g_i = \lambda_{i-1} - \lambda_i, 2 \leq i \leq n$ (matching Zhu's terminology for the *gap*).

$$\ell_1(w_1, w_2, g_2) = \frac{1}{2} \left(w_1 + w_2 - g_2 + \sqrt{(g_2 + w_1 + w_2)^2 - 4gw_2} \right)$$

$$\frac{\delta \ell_1}{\delta g_2} = \frac{1}{2} \left(-1 + \frac{g_2 + w_1 - w_2}{\sqrt{(g_2 + w_1 + w_2)^2 - 4g_2 w_2}} \right) = 0$$

$$(g_2 + w_1 - w_2)^2 = (g_2 + w_1)^2 - 2w_2(g_2 + w_1) + w_2^2 = (g_2 + w_1 + w_2)^2 - 4g_2 w_2$$

$$4w_2(g_2 + w_1) - 4g_2 w_2 = 0$$

$$g_2 + w_1 = g_2$$

We should never have $w_1 = 0$ (covariance matrix should be sanitized to not contain duplicates with 1 correlation, or data with 0 correlation). Normally there would be a deflation procedure, but that has implications for the sparsity level of the result. This could be a fatal error instead.

2

$$\frac{\delta \ell_2}{\delta w_1} = \frac{1}{2}\left(1 + \frac{g_2 + w_1 + w_2}{\sqrt{(g_2 + w_1 + w_2)^2 - 4g_2 w_2}}\right) = 0$$

$$(g_2 + w_1 + w_2)^2 = (g_2 + w_1 + w_2)^2 - 4g_2 w_2$$

$$\frac{\delta \ell_2}{\delta w_2} = \frac{1}{2}\left(1 + \frac{-g_2 + w_1 + w_2}{\sqrt{(g_2 + w_1 + w_2)^2 - 4g_2 w_2}}\right) = 0$$

By symmetry with $\frac{\delta \ell_1}{\delta g_2}$ expansion, we think that $\frac{\delta \ell_1}{\delta w_2}$ will not have a critical point of interest either.

3.1.1 Analysis

The saddle points occur where one of the variables is 0. This would represent a non-distinct eigenvalue or rank-deficient covariance matrix (variable is linearly dependent with the eigenvector produced so far), which should be fatal errors in the system. Zhu's other 3 bounds (initial or final lower/upper bound, other lower/upper bound) ought to produce the same result.

4 Conclusions

We have three variables when solving for a single eigenvalue lower and upper bound. These variables themselves lie in an interval with a lower and upper bound, produced previously. Therefore, for successive application of Zhu's bounds, we would apply the function 4+ times and take the min/max depending on lower/upper bound function. We would use the sign and lack of poles (except at zero) of the gradient to attempt to evaluate the function on less than 8 inputs, if we can determine which application would have a larger/smaller value.

References

[1] Zhu, L., Peng, X. & Liu, H. Rank-one perturbation bounds for singular values of arbitrary matrices. J Inequal Appl 2019, 138 (2019).