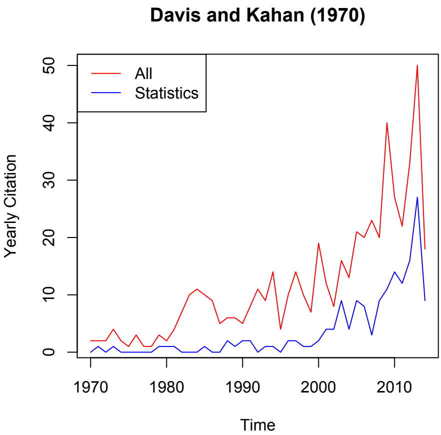
A useful variant of the Davis–Kahan theorem for statisticians

Joint work with Yi Yu and Richard J. Samworth

Tengyao Wang (t.wang@statslab.cam.ac.uk)
Statistical Laboratory, University of Cambridge

Davis, C. & Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7**, 1-46.



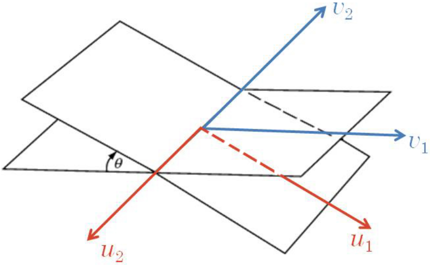
These citations include...

- Sun, T. & Zhang, C.-H. (2012). Calibrated elastic regularization in matrix completion. *Adv. Neural Inf. Proc. Sys.* **25**.
- Cai, T. T., Ma, Z. & Wu, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41**, 3074–3110.
- Fan, J. & Han, X. (2013). Estimation of false discovery proportion with unknown dependence. *arXiv preprint*, axXiv:1305.7007.
- Fan, J., Liao, Y. & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. Roy. Statist. Soc., Ser. B* **75**, 603–680.
- Gao, C., Ma, Z., Ren, Z. & Zhou, H. H. (2014). Minimax estimation in sparse canonical correlation analysis. *arXiv preprint*, arXiv:1405.1595.
- Yu, Y., Feng, Y. & Samworth, R. J. (2014). Fused community detection. *manuscript*.
- Wang, T., Berthet, Q. & Samworth, R. J. (2015). Computational and statistical trade-offs in estimation of sparse principal components. *arXiv preprint*, arXiv:1408.5369.

Our results could be applied directly to allow these authors to assume more natural conditions, to simplify proofs, and in some cases, to improve bounds.

If $U, V \in \mathbb{R}^{p \times d}$ with $p > d$ are matrices with orthonormal columns, then the principal angles between them are given by $\cos^{-1} \sigma_1, \ldots, \cos^{-1} \sigma_d$, where $\sigma_1 \leq \cdots \leq \sigma_d$ are the singular values of $U^\top V$.

Let $\Theta(U, V)$ denote the $d \times d$ diagonal matrix whose jth diagonal entry is the jth principal angle, and let $\sin \Theta(U, V)$ be defined entrywise.



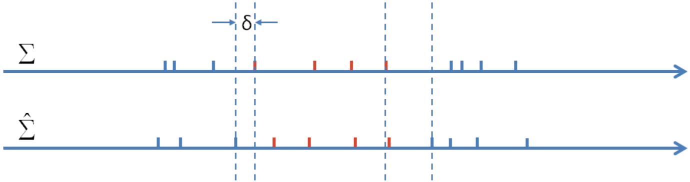
Theorem (Davis–Kahan $\sin\theta$ theorem). Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p\times p}$ be symmetric, with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$ respectively. Fix $1 \leq r \leq s \leq p$ and set $d = s - r + 1$. Let $V = (v_r, \ldots, v_s) \in \mathbb{R}^{p\times d}$ and $\hat{V} = (\hat{v}_r, \ldots, \hat{v}_s) \in \mathbb{R}^{p\times d}$ have orthonormal columns satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$ for $j = r, r+1, \ldots, s$. If

$$\delta = \inf\{|\hat{\lambda} - \lambda| : \lambda \in [\lambda_s, \lambda_r], \hat{\lambda} \in (-\infty, \hat{\lambda}_{s-1}] \cup [\hat{\lambda}_{r+1}, \infty)\} > 0,$$

where $\hat{\lambda}_0 = -\infty$ and $\hat{\lambda}_{p+1} = \infty$, then

$$\|\sin\Theta(\hat{V}, V)\|_{\mathrm{F}} \leq \frac{\|\hat{\Sigma} - \Sigma\|_{\mathrm{F}}}{\delta}.$$

Remark Both occurrences of the Frobenius norm can be replaced with the operator norm $\|\cdot\|_{\mathrm{op}}$, or any other orthogonally invariant norm.



Frequently in applications, we have $r = s = j$, in which case we can conclude that

$$\sin \Theta(\hat{v}_j, v_j) \leq \frac{\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}}{\min(|\hat{\lambda}_{j-1} - \lambda_j|, |\hat{\lambda}_{j+1} - \lambda_j|)}.$$

Since we may reverse the sign of $\hat{v}_j$ if necessary, there is a choice of orientation of $\hat{v}_j$ for which $\hat{v}_j^T v_j \geq 0$. For this choice, we can also deduce that

$$\|\hat{v}_j - v_j\| \leq \sqrt{2} \sin \Theta(\hat{v}_j, v_j).$$

How to use

S1. Argue $\hat{\Sigma}$ is close to Σ.

S2. Argue, e.g. using Weyl's inequality, that with high probability,

$$|\hat{\lambda}_{j-1} - \lambda_j| \geq (\lambda_{j-1} - \lambda_j)/2 \text{ and } |\hat{\lambda}_{j+1} - \lambda_j| \geq (\lambda_j - \lambda_{j+1})/2.$$

Theorem 1. Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p\times p}$ be symmetric, with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$ respectively. Fix $1 \leq r \leq s \leq p$ and assume that $\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$, where $\lambda_0 = \infty$ and $\lambda_{p+1} = -\infty$. Let $d = s - r + 1$, and let $V = (v_r, v_{r+1}, \ldots, v_s) \in \mathbb{R}^{p\times d}$ and $\hat{V} = (\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s) \in \mathbb{R}^{p\times d}$ have orthonormal columns satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$ for $j = r, r+1, \ldots, s$. Then

$$\|\sin\Theta(\hat{V}, V)\|_{\mathrm{F}} \leq \frac{2\min(d^{1/2}\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}, \|\hat{\Sigma} - \Sigma\|_{\mathrm{F}})}{\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})}.$$

Moreover, there exists an orthogonal matrix $\hat{O} \in \mathbb{R}^{d\times d}$ such that

$$\|\hat{V}\hat{O} - V\|_{\mathrm{F}} \leq \frac{2^{3/2}\min(d^{1/2}\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}, \|\hat{\Sigma} - \Sigma\|_{\mathrm{F}})}{\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})}.$$

Example

$$\Sigma = \begin{pmatrix} 3 & \\ & 1 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} \sqrt{1-\epsilon^2} & -\epsilon \\ \epsilon & \sqrt{1-\epsilon^2} \end{pmatrix} \begin{pmatrix} 3 & \\ & 1 \end{pmatrix} \begin{pmatrix} \sqrt{1-\epsilon^2} & -\epsilon \\ \epsilon & \sqrt{1-\epsilon^2} \end{pmatrix}^\top$$

Leading eigenvectors of Σ and $\hat{\Sigma}$ are $v = (1,0)^T$ and $\hat{v} = \big((1-\epsilon^2)^{1/2}, -\epsilon\big)^T$ respectively. Then

$$\sin\Theta(\hat{v}, v) = \epsilon, \quad \|\hat{v} - v\|^2 = 2 - 2(1-\epsilon^2)^{1/2}, \quad \text{and} \quad \frac{2\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}}{3-1} = 2\epsilon.$$

Wedin, P.-Å. (1972) proved a generalisation of Davis–Kahan theorem for the singular values of perturbed matrices.

Theorem 2. Let $A, \hat{A} \in \mathbb{R}^{p\times q}$ have singular values $\sigma_1 \geq \ldots \geq \sigma_{\min(p,q)}$ and $\hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_{\min(p,q)}$ respectively. Fix $1 \leq r \leq s \leq \mathrm{rank}(A)$ and assume that $\min(\sigma_{r-1}^2 - \sigma_r^2, \sigma_s^2 - \sigma_{s+1}^2) > 0$, where $\sigma_0^2 = \infty$ and $\sigma_{\mathrm{rank}(A)+1}^2 = -\infty$. Let $d = s - r + 1$, and let $V = (v_r, v_{r+1}, \ldots, v_s) \in \mathbb{R}^{q\times d}$ and $\hat{V} = (\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s) \in \mathbb{R}^{q\times d}$ have orthonormal columns satisfying $Av_j = \sigma_j u_j$ and $\hat{A}\hat{v}_j = \hat{\sigma}_j \hat{u}_j$ for $j = r, r+1, \ldots, s$. Then

$$\|\sin\Theta(\hat{V}, V)\|_{\mathrm{F}} \leq \frac{2(2\sigma_1 + \|\hat{A} - A\|_{\mathrm{op}})\min(d^{1/2}\|\hat{A} - A\|_{\mathrm{op}}, \|\hat{A} - A\|_{\mathrm{F}})}{\min(\sigma_{r-1}^2 - \sigma_r^2, \sigma_s^2 - \sigma_{s+1}^2)}.$$

Moreover, there exists an orthogonal matrix $\hat{O} \in \mathbb{R}^{d\times d}$ such that

$$\|\hat{V}\hat{O} - V\|_{\mathrm{F}} \leq \frac{2^{3/2}(2\sigma_1 + \|\hat{A} - A\|_{\mathrm{op}})\min(d^{1/2}\|\hat{A} - A\|_{\mathrm{op}}, \|\hat{A} - A\|_{\mathrm{F}})}{\min(\sigma_{r-1}^2 - \sigma_r^2, \sigma_s^2 - \sigma_{s+1}^2)}.$$

Theorem 1 bound the distance between matching eigenspaces by the ratio of the matrix distance and the eigengap $\delta := \min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})$:

$$\|\sin\Theta(\hat{V}, V)\|_{\mathrm{F}} \le \frac{2\min(d^{1/2}\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}, \|\hat{\Sigma} - \Sigma\|_{\mathrm{F}})}{\delta}.$$

Proof outline of Theorem 1.

1. Use the definition of principal angles to rewrite
$$\|\sin\Theta(\hat{V}, V)\|_{\mathrm{F}} = \textstyle\sum_{j=r}^{s} \|(I_p - VV^\top)\hat{v}_j\|^2.$$
2. As $(I_p - VV^\top)\hat{v}_j$ is orthogonal to $\mathrm{span}(V)$, when transformed by $\lambda_j I_p - \Sigma$ it satisfies
$$\delta\|(I_p - VV^\top)\hat{v}_j\| \le \|(\lambda_j I_p - \Sigma)(I_p - VV^\top)\hat{v}_j\| \le \|(\lambda_j I_p - \Sigma)\hat{v}_j\|.$$
3. Split $(\lambda_j I_p - \Sigma)\hat{v}_j = (\hat{\Sigma} - \Sigma)\hat{v}_j - (\hat{\lambda}_j - \lambda_j)\hat{v}_j$ and bound everything in terms of $\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}$ using Weyl's inequality.

References

- Davis, C. & Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7**, 1-46.
- Wedin, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics* **12**, 99–111.
- Yu, Y., Wang, T. & Samworth, R. J. (2015) A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, **102**, 315–323.

Thank you!