SIAM J. Matrix Anal. Appl.
Vol. ?, No. ?, pp. ?–?

MAJORIZATION FOR CHANGES IN ANGLES BETWEEN SUBSPACES, RITZ VALUES, AND GRAPH LAPLACIAN SPECTRA *

ANDREW V. KNYAZEV† AND MERICO E. ARGENTATI‡

Abstract. Many inequality relations between real vector quantities can be succinctly expressed as "weak (sub)majorization" relations using the symbol $\prec_w$. We explain these ideas and apply them in several areas: angles between subspaces, Ritz values, and graph Laplacian spectra, which we show are all surprisingly related.

Let $\Theta(\mathcal{X},\mathcal{Y})$ be the vector of principal angles in nondecreasing order between subspaces $\mathcal{X}$ and $\mathcal{Y}$ of a finite dimensional space $\mathcal{H}$ with a scalar product. We consider the change in principal angles between subspaces $\mathcal{X}$ and $\mathcal{Z}$, where we let $\mathcal{X}$ be perturbed to give $\mathcal{Y}$. We measure the change using the weak majorization. We prove that $|\cos^2\Theta(\mathcal{X},\mathcal{Z})-\cos^2\Theta(\mathcal{Y},\mathcal{Z})|\prec_w \sin\Theta(\mathcal{X},\mathcal{Y})$, and give similar results for differences of cosines, i.e. $|\cos\Theta(\mathcal{X},\mathcal{Z})-\cos\Theta(\mathcal{Y},\mathcal{Z})|\prec_w \sin\Theta(\mathcal{X},\mathcal{Y})$, and of sines, and of sines squared, assuming $\dim\mathcal{X}=\dim\mathcal{Y}$.

We observe that $\cos^2\Theta(\mathcal{X},\mathcal{Z})$ can be interpreted as a vector of Ritz values, where the Rayleigh-Ritz method is applied to the orthogonal projector on $\mathcal{Z}$ using $\mathcal{X}$ as a trial subspace. Thus, our result for the squares of cosines can be viewed as a bound on the change in the Ritz values of an orthogonal projector. We then extend it to prove a general result for Ritz values for an arbitrary Hermitian operator A, not necessarily a projector: let $\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right)$ be the vector of Ritz values in nonincreasing order for A on a trial subspace $\mathcal{X}$, which is perturbed to give another trial subspace $\mathcal{Y}$, then $|\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right)-\Lambda\left(P_{\mathcal{Y}}A|_{\mathcal{Y}}\right)|\prec_w (\lambda_{\max}-\lambda_{\min})\ \sin\Theta(\mathcal{X},\mathcal{Y})$, where the constant is the difference between the largest and the smallest eigenvalues of A. This establishes our conjecture that the root two factor in our earlier estimate may be eliminated. Our present proof is based on a classical but rarely used technique of extending a Hermitian operator in $\mathcal{H}$ to an orthogonal projector in the "double" space $\mathcal{H}^2$.

An application of our Ritz values weak majorization result for Laplacian graph spectra comparison is suggested, based on the possibility to interpret eigenvalues of the edge Laplacian of a given graph as Ritz values of the edge Laplacian of the complete graph. We prove that $\sum_k|\lambda_k^1-\lambda_k^2|\leq nl$, where λ_k^1 and λ_k^2 are all ordered elements of the Laplacian spectra of two graphs with the same n vertices and with l equal to the number of differing edges.

Key words. Majorization, principal angles, canonical angles, canonical correlations, subspace, orthogonal projection, perturbation analysis, Ritz values, Rayleigh–Ritz method, graph spectrum, graph vertex Laplacian, graph edge Laplacian.

AMS subject classification. 15A42, 15A60, 65F35, 05C50

1. Introduction.

Many inequality relations between real vector quantities can be succinctly expressed as "weak (sub)majorization" relations using the symbol $\prec_w$ that we now introduce. For a real vector $x=[x_1,\cdots,x_n]$ let $x^{\downarrow}$ be the vector obtained by rearranging the entries of x in an algebraically non-increasing order. Vector y weakly majorizes vector x, i.e. $x\prec_w y$, if $\sum_{i=1}^k x_i^{\downarrow}\leq\sum_{i=1}^k y_i^{\downarrow},\ k=1,\ldots,n$. The importance of weak majorization can be seen from the classical statement that the

*Received by the editors August 31, 2005; accepted for publication (in revised form) February 22, 2006; published electronically March 5, 2006. Preliminary results of this paper were presented at the 12th ILAS Conference in Canada in June 2005 at the mini-symposium "Canonical Angles Between Subspaces: Theory and Applications."

†Department of Mathematical Sciences University of Colorado at Denver and Health Sciences Center, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364 (andrew.knyazev[at]cudenver.edu, http://math.cudenver.edu/~aknyazev/). This material is based upon work supported by the National Science Foundation award DMS 0208773 and the Intelligence Technology Innovation Center through the joint "Approaches to Combat Terrorism" NSF Program Solicitation NSF 03-569.

‡Department of Mathematical Sciences University of Colorado at Denver and Health Sciences Center, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364 (rargenta[at]math.cudenver.edu).

following two conditions are equivalent: $x \prec_w y$ and $\sum_{i=1}^n \phi(x_i) \le \sum_{i=1}^n \phi(y_i)$ for all nondecreasing convex functions ϕ. Thus, a single weak majorization result implies a great variety of inequalities. We explain these ideas and apply them in several areas: angles between subspaces, Ritz values, and graph Laplacian spectra, which we show are all surprisingly related.

The concept of principal angles, also referred to as canonical angles between subspaces is one of the classical mathematical ideas originated from Jordan [15] with many applications. In functional analysis, the gap between subspaces, which is related to the sine of the largest principal angle, bounds the perturbation of a closed linear operator by measuring the change in its graph, while the smallest nontrivial principal angle between two subspaces determines if the sum of the subspaces is closed. In numerical analysis, principal angles appear naturally to estimate how close an approximate eigenspace is to the true eigenspace. The chordal distance, the Frobenius norm of the sine of the principal angles, on the Grassmannian space of finite dimensional subspaces is used, e.g., for subspace packing with applications in control theory. In statistics, the cosines of principal angles are called canonical correlations and have applications in information retrieval and data visualization.

Let $\mathcal{H}$ be a real or complex $n < \infty$ dimensional vector space equipped with an inner product (x, y) and a vector norm $\|x\| = (x, x)^{1/2}$. The acute angle between two non–zero vectors x and y is defined as

$$\theta(x, y) = \arccos \frac{|(x, y)|}{\|x\|\|y\|} \in [0, \pi/2].$$

For three nonzero vectors x, y, z, we have bounds on the change in the angle

$$|\theta(x, z) - \theta(y, z)| \le \theta(x, y), \tag{1.1}$$

in the sine

$$|\sin(\theta(x, z)) - \sin(\theta(y, z))| \le \sin(\theta(x, y)), \tag{1.2}$$

in the cosine

$$|\cos(\theta(x, z)) - \cos(\theta(y, z))| \le \sin(\theta(x, y)), \tag{1.3}$$

and a more subtle bound on the change in the sine or cosine squared

$$\left|\cos^2(\theta(x, z)) - \cos^2(\theta(y, z))\right| = \left|\sin^2(\theta(x, z)) - \sin^2(\theta(y, z))\right| \le \sin(\theta(x, y)). \tag{1.4}$$

Let us note that we can project the space $\mathcal{H}$ into the span$\{x, y, z\}$ without changing the angles, i.e. the inequalities above present essentially the case of a $3D$ space.

Inequality (1.1) is proved in Qiu et al. [33]. We note that (1.2) follows from (1.1), since the sine function is increasing and subadditive, see [33].

It is instructive to provide a simple proof of the sine inequality (1.2) using orthogonal projectors. Let $P_{\mathcal{X}}$, $P_{\mathcal{Y}}$, and $P_{\mathcal{Z}}$ be, respectively, the orthogonal projectors onto the subspaces spanned by the vectors x, y, and z, and let $\|\cdot\|$ also denote the induced operator norm. When we are dealing with $1D$ subspaces, we have the following elementary formula $\sin(\theta(x, y)) = \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|$ (indeed, $P_x - P_y$ has rank at most two, so it has at most two non-zero singular values, but $(P_x - P_y)^2 x = (1 - |(x, y)|^2)x$ and $(P_x - P_y)^2 y = (1 - |(x, y)|^2)y$ for unit vectors x and y, so $1 - |(x, y)|^2 = \sin^2(\theta\{x, y\})$ is a double eigenvalue of $(P_x - P_y)^2$). Then the sine (1.2) inequality is equivalent to the triangle inequality $|\ \|P_{\mathcal{X}} - P_{\mathcal{Z}}\| - \|P_{\mathcal{Y}} - P_{\mathcal{Z}}\|\ | \le \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|$.

In this paper, we replace $1D$ subspaces spanned by the vectors x, y and z, with multi–dimensional subspaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, and we use the concept of principal angles between subspaces. Principal angles are very well studied in the literature, however, some important gaps still remain. Here, we are interested in generalizing inequalities (1.2)-(1.4) above to multi–dimensional subspaces to include all principal angles, using weak majorization.

Let us denote by $\Theta(\mathcal{X},\mathcal{Y})$ the vector of principal angles in nondecreasing order between subspaces $\mathcal{X}$ and $\mathcal{Y}$. Let $\dim\mathcal{X} = \dim\mathcal{Y}$, and let another subspace $\mathcal{Z}$ be given. We prove that $|\cos^2\Theta(\mathcal{X},\mathcal{Z}) - \cos^2\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y})$, and give similar results for differences of cosines, i.e. $|\cos\Theta(\mathcal{X},\mathcal{Z}) - \cos\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y})$, and of sines, and of sines squared. This is the first main result of the present paper, see Section 3. The proof of weak majorization for sines is a direct generalization of the 1D proof above. Our proofs of weak majorization for cosines and sines or cosines squared do not have such simple 1D analogs.

Pioneering results using angles between subspaces in the framework of unitarily invariant norms and symmetric gauge functions, equivalent to majorization, appear in Davis and Kahan [6], which introduces many of the tools that we use here. The main goal of [6] is however entirely different — analyzing the perturbations of eigenvalues and eigenspaces, while in the present paper we are concerned with sensitivity of angles and Ritz values with respect to changes in subspaces.

Our second main result, see Section 4, bounds the change in the Ritz values with the change of the trial subspace. We attack the problem by discovering a simple, but deep, connection between the principal angles and the Rayleigh–Ritz method.

We first give a brief definition of Ritz values. Let $A : \mathcal{H} \to \mathcal{H}$ be a Hermitian operator and let $\mathcal{X}$ be a (so-called "trial") subspace of $\mathcal{H}$. We define an operator $P_{\mathcal{X}}A|_{\mathcal{X}}$ on $\mathcal{X}$, where $P_{\mathcal{X}}$ is the orthogonal projector onto $\mathcal{X}$ and $P_{\mathcal{X}}A|_{\mathcal{X}}$ denotes the restriction of operator $P_{\mathcal{X}}A$ to its invariant subspace $\mathcal{X}$, as discussed, e.g., in Parlett [31]. The eigenvalues $\Lambda(P_{\mathcal{X}}A|_{\mathcal{X}})$ are called Ritz values of the operator A with respect to the trial subspace $\mathcal{X}$.

We observe that the cosines squared $\cos^2\Theta(\mathcal{X},\mathcal{Z})$ of principal angles between subspaces $\mathcal{X}$ and $\mathcal{Z}$ can be interpreted as a vector of Ritz values, where the Rayleigh-Ritz method is applied to the orthogonal projector $P_{\mathcal{Z}}$ onto $\mathcal{Z}$ using $\mathcal{X}$ as a trial subspace. Let us illustrate this connection for one–dimensional $\mathcal{X} = \operatorname{span}\{x\}$ and $\mathcal{Z} = \operatorname{span}\{z\}$, where it becomes trivial:

$$\cos^2(\theta(x,z)) = \frac{(x, P_{\mathcal{Z}}x)}{(x,x)}.$$

The ratio on the right is the Rayleigh quotient for $P_{\mathcal{Z}}$ — the one dimensional analog of the Ritz value. In this notation, estimate (1.4) turns into

$$\left|\frac{(x, P_z x)}{(x,x)} - \frac{(y, P_z y)}{(y,y)}\right| \le \sin(\theta(x,y)), \tag{1.5}$$

which clearly now is a particular case of a general estimate for the Rayleigh quotient, cf. Knyazev and Argentati [18],

$$\left|\frac{(x, Ax)}{(x,x)} - \frac{(y, Ay)}{(y,y)}\right| \le \left(\lambda_{\max} - \lambda_{\min}\right)\sin(\theta(x,y)), \tag{1.6}$$

where A is a Hermitian operator and $\lambda_{\max} - \lambda_{\min}$ is the spread of its spectrum.

We show that the multi–dimensional analog of (1.5) can be interpreted as a bound on the change in the Ritz values with the change of the trial subspace, in the particular case where the Rayleigh-Ritz method is applied to an orthogonal projector. We then extend it to prove a general result for Ritz values for an arbitrary Hermitian operator A, not necessarily a projector: let $\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right)$ be the vector of Ritz values in nonincreasing order for the operator A on a trial subspace $\mathcal{X}$, which is perturbed to give another trial subspace $\mathcal{Y}$, then $|\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right) - \Lambda\left(P_{\mathcal{Y}}A|_{\mathcal{Y}}\right)| \prec_w (\lambda_{\max} - \lambda_{\min}) \sin\Theta(\mathcal{X},\mathcal{Y})$, which is a multi–dimensional analog of (1.6). Our present proof is based on a classical but rarely used idea of extending a Hermitian operator in $\mathcal{H}$ to an orthogonal projector in the "double" space $\mathcal{H}^2$ preserving its Ritz values.

An application of our Ritz values weak majorization result for Laplacian graph spectra comparison is suggested in Section 5, based on the possibility to interpret eigenvalues of the edge Laplacian of a given graph as Ritz values of the edge Laplacian of the complete graph. We prove that $\sum_k |\lambda_k^1 - \lambda_k^2| \leq nl$, where λ_k^1 and λ_k^2 are all ordered elements of the Laplacian spectra of two graphs with the same n vertices and with l equal to the number of differing edges.

The rest of the paper is organized as follows. In Section 2, we provide some background, definitions and several statements concerning weak majorization, principal angles between subspaces, and extensions of Hermitian operators to projectors. In Section 3, we prove in Theorems 3.2 and 3.3 that the absolute value of the change in (the squares of) the sines and cosines is weakly majorized by the sines of the angles between the original and perturbed subspaces. In Section 4, we prove in Theorem 4.3 that a change in the Ritz values in the Rayleigh-Ritz method with respect to the change in the trial subspaces is weakly majorized by the sines of the principal angles between the original and perturbed trial subspaces times a constant. In Section 5 we apply our Ritz values weak majorization result to Laplacian graphs spectra comparison.

This paper is related to several different subjects: majorization, principal angles, Rayleigh-Ritz method, and Laplacian graph spectra. In most cases, whenever possible, we cite books rather than the original works in order to keep our already quite long list of references within a reasonable size.

2. Definitions and Preliminaries.

In this section we introduce some definitions, basic concepts and mostly known results for later use.

2.1. Weak Majorization.

Majorization is a well known, e.g., Hardy et al. [13], Marshall and Olkin [23], important mathematical concept with numerous applications.

For a real vector $x = [x_1, \cdots, x_n]$ let $x^{\downarrow}$ be the vector obtained by rearranging the entries of x in an algebraically non-increasing order, $x_1^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$. We denote $[|x_1|, \cdots, |x_n|]$ by $|x|$. We say that vector y weakly majorizes vector x and we use the notation $[x_1, \cdots, x_n] \prec_w [y_1, \cdots, y_n]$ or $x \prec_w y$ if $\sum_{i=1}^k x_i^{\downarrow} \leq \sum_{i=1}^k y_i^{\downarrow}$, $k = 1, \ldots, n$. If in addition the sums above for $k = n$ are equal, y (strongly) majorizes vector x, but we do not use this type of majorization in the present paper. Two vectors of different lengths may be compared by simply appending zeroes to increase the size of the smaller vector to make the vectors the same length.

Weak majorization is a powerful tool for estimates involving eigenvalues and singular values and is covered, e.g., in Gohberg and Kreĭn [9], Marshall and Olkin [23], Bhatia [1] and Horn and Johnson [14], which we follow here and refer the reader to for references to the original works and all necessary proofs. In the present paper, we use several well known statements that we formulate for operators $\mathcal{H} \to \mathcal{H}$ and overview briefly below.

Let $S(A)$ denote the vector of all singular values of $A : \mathcal{H} \to \mathcal{H}$ in nonincreasing order, i.e. $S(A) = S^{\downarrow}(A)$, while individual singular values of A enumerated in nonincreasing order are denoted by $s_i(A)$. For Hermitian A let $\Lambda(A)$ denote the vector of all eigenvalues of A in nonincreasing order, i.e. $\Lambda(A) = \Lambda^{\downarrow}(A)$, while individual eigenvalues of A enumerated in nonincreasing order are denoted by $\lambda_i(A)$.

The starting point for weak majorization results we use in this paper is

THEOREM 2.1. *[e.g., Th. 9.G.1, p. 241 [23]]* $\Lambda(A + B) \prec_w \Lambda(A) + \Lambda(B)$ *for Hermitian A and B,* which follows easily from Ky Fan's trace maximum principle [e.g. Th. 20.A.2 [23]] and the fact that the maximum of a sum is bounded from above by the sum of the maxima. For general $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{H} \to \mathcal{H}$, it follows from Theorem 2.1, since the top half of the spectrum of the Hermitian 2-by-2 block operator $\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ is nothing but $S(A)$, that

COROLLARY 2.2. *[e.g., Cor. 3.4.3, p. 196 [14]]* $S(A \pm B) \prec_w S(A) + S(B)$.

A more delicate and stronger result is the following Lidskii theorem [e.g., Th. III.4.1 [1]], which can be proved using the Wielandt maximum principle [e.g., Th. III.3.5 [1]],

THEOREM 2.3. *For Hermitian A and B and any set of indices* $1 \le i_1 < \ldots < i_k \le n = \dim\mathcal{H}$, *we have* $\sum_{j=1}^{k} \lambda_{i_j}(A+B) \le \sum_{j=1}^{k} \lambda_{i_j}(A) + \sum_{j=1}^{k} \lambda_j(B)$, $k = 1, \ldots, n$.
By choosing an appropriate set of indices, Theorem 2.3 for Hermitian A and B immedeately gives $\Lambda(A) - \Lambda(B) \prec_w \Lambda(A - B)$, which for singular values of arbitraty $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{H} \to \mathcal{H}$ is equivalent [e.g., Sec. IV.3, pp.98–101 [1]] to

COROLLARY 2.4. *[e.g., Th.3.4.5, p. 198 [14] or Th. IV.3.4, p. 100 [1]]* $|S(A) - S(B)| \prec_w S(A - B)$. Applying Corollary 2.4 to properly shifted Hermitian operators, we get

COROLLARY 2.5. $|\Lambda(A) - \Lambda(B)| \prec_w S(A - B)$ *for Hermitian A and B.*

We finally need the so called "pinching" inequality,

THEOREM 2.6. *[e.g., Th. II.5.1 [9] or (II.38), p. 50 [1]] If P is an orthogonal projector then* $S(PAP \pm (I - P)A(I - P)) \prec_w S(A)$.

Proof. Indeed, $A = PAP + (I - P)A(I - P) + PA(I - P) + (I - P)AP$ so let $B = PAP + (I - P)A(I - P) - PA(I - P) - (I - P)AP$ then $(2P - I)A(2P - I) = B$, where $2P - I$ is unitary Hermitian, so A^*A and B^*B are similar and $S(A) = S(B)$. Evidently, $PAP + (I - P)A(I - P) = (A + B)/2$ so the pinching result with the plus follows from Corollary 2.2. The pinching result with the minus is equivalent to the pinching result with the plus since the sign does not change the singular values on the left-hand side: $S(PAP \pm (I - P)A(I - P)) = S(PAP) \cup S((I - P)A(I - P))$, since the ranges of PAP and $(I - P)A(I - P)$ are disjoint. □

2.2. Principal Angles Between Subspaces. Let $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ be orthogonal projectors onto the subspaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, of the space $\mathcal{H}$. We define the set of cosines of principal angles between subspaces $\mathcal{X}$ and $\mathcal{Y}$ by

$$\cos\Theta(\mathcal{X}, \mathcal{Y}) = [s_1(P_{\mathcal{X}}P_{\mathcal{Y}}), \ldots, s_m(P_{\mathcal{X}}P_{\mathcal{Y}})],\ m = \min\{\dim\mathcal{X}; \dim\mathcal{Y}\}\,. \tag{2.1}$$

Our definition (2.1) is evidently symmetric: $\Theta(\mathcal{X}, \mathcal{Y}) = \Theta(\mathcal{Y}, \mathcal{X})$. By definition, the cosines are arranged in nonincreasing order, i.e. $\cos(\Theta(\mathcal{X}, \mathcal{Y})) = (\cos(\Theta(\mathcal{X}, \mathcal{Y})))^{\downarrow}$, while the angles $\theta_i(\mathcal{X}, \mathcal{Y}) \in [0, \pi/2], i = 1, \ldots, m$ and their sines are in nondecreasing order.

The concept of principal angles is closely connected to cosine–sine (CS) decompositions of unitary operators; and we refer the reader to the books Bhatia [1], Stewart and Sun [36], Stewart [37] for the history and references to the original publications on the principal angles and the CS decomposition. We need several simple but important statements about the angles provided below. In the particular case $\dim\mathcal{X} = \dim\mathcal{Y}$, the standard CS decomposition can be used and the statements are easy to derive. For the general case $\dim\mathcal{X} \neq \dim\mathcal{Y}$ that is necessary for us here, they can be obtained using the general (rectangular) form of the CS decomposition described, e.g., in Paige and Saunders [29], Paige and Wei [30]. We cannot find exact references for the facts that we need in the general case, so for completeness we provide the proofs here using ideas from Davis and Kahan [6], Halmos [12], preparing our work to be more easily extended to infinite dimensional Hilbert spaces.

Theorem 2.7. *When one of the two subspaces is replaced with its orthogonal complement, the corresponding pairs of angles sum up to $\pi/2$, specifically:*

$$\left[\frac{\pi}{2},\dots,\frac{\pi}{2},(\Theta(\mathcal{X},\mathcal{Y}))^{\downarrow}\right] = \left[\frac{\pi}{2} - \Theta(\mathcal{X},\mathcal{Y}^{\perp}),0,\dots,0\right], \tag{2.2}$$

where there are $\max(\dim\mathcal{X} - \dim\mathcal{Y}, 0)$ values $\pi/2$ on the left, and possibly extra zeros on the right to match the sizes.

The angles between subspaces and between their orthogonal complements are essentially the same,

$$[(\Theta(\mathcal{X},\mathcal{Y}))^{\downarrow},0,\dots,0] = [(\Theta(\mathcal{X}^{\perp},\mathcal{Y}^{\perp}))^{\downarrow},0,\dots,0], \tag{2.3}$$

where extra 0s at the end may need to be added on either side to match the sizes.

Proof. Let $\mathfrak{M}_{00} = \mathcal{X}\cap\mathcal{Y}$, $\mathfrak{M}_{01} = \mathcal{X}\cap\mathcal{Y}^{\perp}$, $\mathfrak{M}_{10} = \mathcal{X}^{\perp}\cap\mathcal{Y}$, $\mathfrak{M}_{11} = \mathcal{X}^{\perp}\cap\mathcal{Y}^{\perp}$, as suggested in Halmos [12]. Each of the subspaces is invariant with respect to orthoprojectors $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ and their products, and so each of the subspaces contributes independently to the set of singular values of $P_{\mathcal{X}}P_{\mathcal{Y}}$ in (2.1). Specifically, there are $\dim\mathfrak{M}_{00}$ ones, $\dim\mathfrak{M}_0$ singular values in the interval $(0,1)$ equal to $\cos\Theta(\mathfrak{M}_0,\mathcal{Y})$, where $\mathfrak{M}_0 = \mathcal{X}\cap(\mathfrak{M}_{00}\oplus\mathfrak{M}_{01})^{\perp}$, and all other singular values are zeros; thus,

$$\Theta(\mathcal{X},\mathcal{Y})^{\downarrow} = \left[\frac{\pi}{2},\dots,\frac{\pi}{2},(\Theta(\mathfrak{M}_0,\mathcal{Y}))^{\downarrow},0,\dots,0\right], \tag{2.4}$$

where there are $\min\{\dim(\mathfrak{M}_{01});\dim(\mathfrak{M}_{10})\}$ values $\pi/2$ and $\dim(\mathfrak{M}_{00})$ zeros.

The subspace $\mathfrak{M}_0$ does not change if we substitute $\mathcal{Y}^{\perp}$ for $\mathcal{Y}$ in (2.4), so we have

$$\Theta(\mathcal{X},\mathcal{Y}^{\perp})^{\downarrow} = \left[\frac{\pi}{2},\dots,\frac{\pi}{2},(\Theta(\mathfrak{M}_0,\mathcal{Y}^{\perp}))^{\downarrow},0,\dots,0\right],$$

where there are $\min\{\dim(\mathfrak{M}_{00});\dim(\mathfrak{M}_{11})\}$ values $\pi/2$ and $\dim(\mathfrak{M}_{01})$ zeros. Since λ is an eigenvalue of $(P_{\mathcal{X}}P_{\mathcal{Y}})|_{\mathfrak{M}_0}$ if and only if $1-\lambda$ is an eigenvalue of $(P_{\mathcal{X}}P_{\mathcal{Y}^{\perp}})|_{\mathfrak{M}_0}$, we have $\frac{\pi}{2} - \Theta(\mathfrak{M}_0,\mathcal{Y}^{\perp}) = (\Theta(\mathfrak{M}_0,\mathcal{Y}))^{\downarrow}$, and the latter equality turns into

$$\frac{\pi}{2} - \Theta(\mathcal{X},\mathcal{Y}^{\perp}) = \left[\frac{\pi}{2},\dots,\frac{\pi}{2},(\Theta(\mathfrak{M}_0,\mathcal{Y}))^{\downarrow},0,\dots,0\right], \tag{2.5}$$

where there are $\dim(\mathfrak{M}_{01})$ values $\pi/2$, and $\min\{\dim(\mathfrak{M}_{00});\dim(\mathfrak{M}_{11})\}$ zeros. To obtain (2.2), we make (2.4) and (2.5) equal by adding $\max\{\dim\mathfrak{M}_{01} - \dim\mathfrak{M}_{10};0\}$ values $\pi/2$ to (2.4) and $\max\{\dim\mathfrak{M}_{00} - \dim\mathfrak{M}_{11};0\}$ zeros to (2.5), and noting that, since $\dim(\mathcal{X}\cap\mathcal{Y}^{\perp}) = \dim\mathcal{X} + \dim\mathcal{Y}^{\perp} - \dim(\mathcal{X}+\mathcal{Y}^{\perp})$, $\mathcal{X}^{\perp}\cap\mathcal{Y} = (\mathcal{X}+\mathcal{Y}^{\perp})^{\perp}$, and $\dim\mathcal{Y}^{\perp} - \dim((\mathcal{X}+\mathcal{Y}^{\perp})^{\perp}) = \dim(\mathcal{X}+\mathcal{Y}^{\perp}) - \dim\mathcal{Y}$, we have $\dim\mathfrak{M}_{01} - \dim\mathfrak{M}_{10} = \dim(\mathcal{X}\cap\mathcal{Y}^{\perp}) - \dim(\mathcal{X}^{\perp}\cap\mathcal{Y}) = \dim\mathcal{X} + \dim\mathcal{Y}^{\perp} - \dim(\mathcal{X}+\mathcal{Y}^{\perp}) - \dim((\mathcal{X}+\mathcal{Y}^{\perp})^{\perp}) = \dim\mathcal{X} - \dim\mathcal{Y}$.

The proof above shows that there are $\dim\mathfrak{M}_{00} = \dim(\mathcal{X}\cap\mathcal{Y})$ zeros on the right in (2.2). To prove (2.3), we substitute in (2.2) $\mathcal{X}^\perp$ for $\mathcal{X}$ to get $\left[\frac{\pi}{2},\ldots,\frac{\pi}{2},(\Theta(\mathcal{X}^\perp,\mathcal{Y}))^\downarrow\right] = \left[\frac{\pi}{2}-\Theta(\mathcal{X}^\perp,\mathcal{Y}^\perp),0,\ldots,0\right]$ with $\dim(\mathcal{X}^\perp\cap\mathcal{Y})$ zeros on the right on the one hand and exchange $\Theta(\mathcal{X},\mathcal{Y}) = \Theta(\mathcal{Y},\mathcal{X})$ in (2.2) and then substitute $\mathcal{X}^\perp$ for $\mathcal{X}$ to obtain $\left[\frac{\pi}{2},\ldots,\frac{\pi}{2},(\Theta(\mathcal{Y},\mathcal{X}^\perp))^\downarrow\right] = \left[\frac{\pi}{2}-\Theta(\mathcal{Y},\mathcal{X}),0,\ldots,0\right]$ with $\dim(\mathcal{Y}\cap\mathcal{X}^\perp)$ zeros on the right on the other hand. We have the equal number of zeros on the right in both equalities and $\Theta(\mathcal{X}^\perp,\mathcal{Y}) = \Theta(\mathcal{Y},\mathcal{X}^\perp)$ by the symmetry of our definition (2.1), so subtracting both equalities from $\pi/2$ leads to (2.3). □

We also use the following trivial, but crucial, statement.

LEMMA 2.8. $\Lambda\left((P_\mathcal{X}P_\mathcal{Y})|_\mathcal{X}\right) = [\cos^2\Theta(\mathcal{X},\mathcal{Y}),0,\ldots,0]$, *with* $\max\{\dim\mathcal{X}-\dim\mathcal{Y},0\}$ *extra* 0*s.*

Proof. The operator $(P_\mathcal{X}P_\mathcal{Y})|_\mathcal{X} = \left((P_\mathcal{X}P_\mathcal{Y})(P_\mathcal{X}P_\mathcal{Y})^\star\right)|_\mathcal{X}$ is Hermitian nonnegative definite, and its spectrum can be represented using the definition of angles (2.1). The number of extra 0s is exactly the difference between the number $\dim\mathcal{X}$ of Ritz values and the number $\min\{\dim\mathcal{X},\dim\mathcal{Y}\}$ of principal angles . □

Finally, we need the following characterization of singular values of the difference of projectors:

THEOREM 2.9.

$$[S(P_\mathcal{X}-P_\mathcal{Y}),0,\ldots,0] = [1,\ldots,1,(\sin\Theta(\mathcal{X},\mathcal{Y}),\sin\Theta(\mathcal{X},\mathcal{Y}))^\downarrow,0,\ldots,0],$$

where there are $|\dim\mathcal{X}-\dim\mathcal{Y}|$ *extra* 1*s upfront, the set* $\sin\Theta(\mathcal{X},\mathcal{Y})$ *is repeated twice and ordered, and extra* 0*s at the end may need to be added on either side to match the sizes.*

Proof. The projectors $P_\mathcal{X}$ and $P_\mathcal{Y}$ are idempotent, which implies on the one hand

$$(P_\mathcal{X}-P_\mathcal{Y})^2 = P_\mathcal{X}(I-P_\mathcal{Y}) + P_\mathcal{Y}(I-P_\mathcal{X}) = P_\mathcal{X}P_{\mathcal{Y}^\perp} + P_\mathcal{Y}P_{\mathcal{X}^\perp},$$

so the subspace $\mathcal{X}$ is invariant under $(P_\mathcal{X}-P_\mathcal{Y})^2$. On the other hand,

$$(P_\mathcal{X}-P_\mathcal{Y})^2 = (I-P_\mathcal{X})P_\mathcal{Y} + (I-P_\mathcal{Y})P_\mathcal{X} = P_{\mathcal{X}^\perp}P_\mathcal{Y} + P_{\mathcal{Y}^\perp}P_\mathcal{X},$$

so the subspace $\mathcal{X}^\perp$ is also invariant under $(P_\mathcal{X}-P_\mathcal{Y})^2$. The projectors $P_\mathcal{X}$ and $P_\mathcal{Y}$ are orthogonal, thus the operator $(P_\mathcal{X}-P_\mathcal{Y})^2$ is Hermitian, and its spectrum can be represented as a union (counting the multiplicities) of the spectra of its restrictions to the complementary invariant subspaces $\mathcal{X}$ and $\mathcal{X}^\perp$:

$$\Lambda\left((P_\mathcal{X}-P_\mathcal{Y})^2\right) = [\Lambda((P_\mathcal{X}P_{\mathcal{Y}^\perp})|_\mathcal{X}),\Lambda((P_{\mathcal{X}^\perp}P_\mathcal{Y}))|_{\mathcal{X}^\perp}]^\downarrow.$$

Using Lemma 2.8 and statement (2.2) of Theorem 2.7,

$$\begin{aligned}[\Lambda((P_\mathcal{X}P_{\mathcal{Y}^\perp})|_\mathcal{X}),0,\ldots,0] &= [\cos^2\Theta(\mathcal{X},\mathcal{Y}^\perp),0,\ldots,0]\\ &= [1,\ldots,1,(\sin^2\Theta(\mathcal{X},\mathcal{Y}))^\downarrow,0,\ldots,0],\end{aligned}$$

where there are $\max(\dim\mathcal{X}-\dim\mathcal{Y},0)$ leading 1s and possibly extra zeros to match the sizes, and

$$\begin{aligned}[\Lambda((P_{\mathcal{X}^\perp}P_\mathcal{Y}))|_{\mathcal{X}^\perp}),0,\ldots,0] &= [\cos^2\Theta(\mathcal{X}^\perp,\mathcal{Y}),0,\ldots,0]\\ &= [1,\ldots,1,(\sin^2\Theta(\mathcal{X},\mathcal{Y}))^\downarrow,0,\ldots,0],\end{aligned}$$

where there are $\max(\dim\mathcal{Y}-\dim\mathcal{X},0)$ leading 1s and possibly extra zeros to match the sizes. Combining these two relations and taking the square root, completes the proof. □

2.3. Extending Operators to Isometries and Projectors. In this subsection we present a simple and known technique, e.g., Halmos [11] and Riesz and Sz.-Nagy [34], p. 461, for extending a Hermitian operator to a projector. We give an alternative proof based on extending an arbitrary normalized operator B to an isometry $\hat{B}$ (in matrix terms, a matrix with orthonormal columns). Glazman and Ljubič [8] Problem X.1.26 and Bhatia [1] Exersize I.3.6, p. 11 extend B to a block 2-by-2 unitary operator. Our technique is similar and results in a 2-by-1 isometry operator $\hat{B}$ that coincides with the first column of the 2-by-2 unitary extension.

Lemma 2.10. *Given an operator $B : \mathcal{H} \to \mathcal{H}$ with singular values less than or equal to one, there exists a block 2-by-1 isometry operator $\hat{B} : \mathcal{H} \to \mathcal{H}^2$, such that the upper block of $\hat{B}$ coincides with B.*

Proof. B^*B is Hermitian nonnegative definite, and all its eigenvalues are bounded by one, since all singular values of B are bounded by one. Therefore, $I - B^*B$ is Hermitian and nonnegative definite, and thus possesses a Hermitian nonnegative square root. Let

$$\hat{B} = \begin{bmatrix} B \\ \sqrt{I - B^*B} \end{bmatrix}.$$

By direct calculation, $\hat{B}^*\hat{B} = B^*B + \sqrt{I - B^*B}\sqrt{I - B^*B} = I$, i.e. $\hat{B}$ is an isometry. □

Now we use Lemma 2.10 to extend, in a similar sense, a shifted and normalized Hermitian operator to an orthogonal projector.

Theorem 2.11 ([11] and [34], p. 461). *Given a Hermitian operator $A : \mathcal{H} \to \mathcal{H}$ with eigenvalues enclosed in the segment $[0, 1]$, there exists a block 2-by-2 orthogonal projector $\hat{A} : \mathcal{H}^2 \to \mathcal{H}^2$, such that its upper left block is equal to A.*

Proof. There exists $\sqrt{A}$, which is also Hermitian and has its eigenvalues enclosed in $[0, 1]$. Applying Lemma 2.10 to $B = \sqrt{A}$, we construct the isometry $\hat{B}$ and set

$$\hat{A} = \hat{B}\hat{B}^* = \begin{bmatrix} \sqrt{A} \\ \sqrt{I - A} \end{bmatrix} \begin{bmatrix} \sqrt{A} & \sqrt{I - A} \end{bmatrix} = \begin{bmatrix} A & \sqrt{A(I - A)} \\ \sqrt{A(I - A)} & I - A \end{bmatrix}.$$

We see that indeed the upper left block is equal to A. We can use the fact that $\hat{B}$ is an isometry to show that $\hat{A}$ is an orthogonal projector, or that can be checked directly by calculating $\hat{A}^2 = \hat{A}$ and noticing that $\hat{A}$ is Hermitian by construction. □

Introducing $S = \sqrt{A}$ and $C = \sqrt{I - A}$, we obtain

$$\hat{A} = \begin{bmatrix} S^2 & SC \\ SC & C^2 \end{bmatrix},$$

which is a well known, e.g., Davis [5], Halmos [12], block form of an orthogonal projector that can alternatively be derived using the CS decomposition of unitary operators, e.g., Bhatia [1], Stewart and Sun [36], Stewart [37].

The importance of Theorem 2.11 can be better seen if we reformulate it as

Theorem 2.12. *Given a Hermitian operator $A : \mathcal{H} \to \mathcal{H}$ with eigenvalues enclosed in a segment $[0, 1]$, there exist subspaces $\mathcal{X}$ and $\mathcal{Y}$ in $\mathcal{H}^2$ such that A is unitarily equivalent to $(P_\mathcal{X} P_\mathcal{Y})|_\mathcal{X}$, where $P_\mathcal{X}$ and $P_\mathcal{Y}$ are the corresponding orthogonal projectors in $\mathcal{H}^2$ and $|_\mathcal{X}$ denotes a restriction to the invariant subspace $\mathcal{X}$.*

Proof. We use Theorem 2.11 and take $P_\mathcal{Y} = \hat{A}$ and $P_\mathcal{X} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. □

Similar to Theorem 2.12, Lemma 2.10 implicitly states that an arbitrary normalized operator B is unitary equivalent to a product of the partial isometry $\hat{B}$ in $\mathcal{H}^2$ and the orthogonal projector in $\mathcal{H}^2$ that selects the upper block in $\hat{B}$ (called $P_{\mathcal{X}}$ in the proof of Theorem 2.12). It is instructive to compare this product to the classical polar decomposition of B that is a product of a partial isometry and a Hermitian nonnegative operator in $\mathcal{H}$. In $\mathcal{H}^2$, we can choose the second factor to be an orthogonal projector! This statement together with Theorem 2.12 can provide interesting canonical decompositions is $\mathcal{H}^2$ that apparently are not used at present, but in our opinion deserve attention.

We take advantage of Theorem 2.12 in the present paper. Using Lemma 2.8 with (2.1), Theorem 2.12 implies that *the spectrum of an arbitrary Hermitian operator after a proper shift and scaling is nothing but a set of cosines squared of principal angles between some pair of subspaces.* This surprising idea appears to be very powerful. It allows us, in Section 4, to obtain a novel result on sensitivity of Ritz values with respect to the trial subspace by reducing the investigation of the Rayleigh-Ritz method to the analysis of the principal angles between subspaces that we provide in the next section.

3. Majorization for Angles. In this section we prove the main results of the present paper involving sines and cosines and their squares of principal angles, but we start with a known statement that involves the principal angles themselves:

THEOREM 3.1 (Theorem 2 Qiu et al. [33]). *Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be subspaces of the same dimension. Then*

$$|\Theta(\mathcal{X},\mathcal{Z}) - \Theta(\mathcal{Y},\mathcal{Z})| \prec_w \Theta(\mathcal{X},\mathcal{Y}). \tag{3.1}$$

Theorem 3.1 deals with the principal angles themselves, and the obvious question is: are there similar results for a function of these angles, in particular for sines and cosines and their squares? For one dimensional subspaces, estimate (3.1) turns into (1.1) that, as discussed in the Introduction, implies the estimate (1.2) for the sine. According to an anonymous referee, it appears to be known to some specialists that the same inference can be made for tuples of angles, but there is no good reference for this at present. Below we give easy direct proofs in a unified way for the sines and cosines and their squares.

We first prove the estimates for sine and cosine, which are straightforward generalizations of the 1D sine (1.2) and cosine (1.3) inequalities from the Introduction.

THEOREM 3.2. *Let* $\dim\mathcal{X} = \dim\mathcal{Y}$ *then*

$$|\sin\Theta(\mathcal{X},\mathcal{Z}) - \sin\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y}), \tag{3.2}$$

$$|\cos\Theta(\mathcal{X},\mathcal{Z}) - \cos\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y}). \tag{3.3}$$

Proof. Let $P_{\mathcal{X}}$, $P_{\mathcal{Y}}$ and $P_{\mathcal{Z}}$ be the corresponding orthogonal projectors onto the subspaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively. We prove the sine estimate (3.2), using the idea of Qiu and Zhang [32]. Starting with $(P_{\mathcal{X}} - P_{\mathcal{Z}}) - (P_{\mathcal{Y}} - P_{\mathcal{Z}}) = P_{\mathcal{X}} - P_{\mathcal{Y}}$, as in the proof of the 1D sine estimate (1.2), we use Corollary 2.4 to obtain

$$|S(P_{\mathcal{X}} - P_{\mathcal{Z}}) - S(P_{\mathcal{Y}} - P_{\mathcal{Z}})| \prec_w S(P_{\mathcal{X}} - P_{\mathcal{Y}}).$$

The singular values of the difference of two orthoprojectors are described by Theorem 2.9. Since $\dim\mathcal{X} = \dim\mathcal{Y}$ we have the same number of extra 1s upfront in $S(P_\mathcal{X} - P_\mathcal{Z})$ and in $S(P_\mathcal{Y} - P_\mathcal{Z})$ so that the extra 1's are canceled and the set of nonzero entries of $|S(P_\mathcal{X} - P_\mathcal{Z}) - S(P_\mathcal{Y} - P_\mathcal{Z})|$ consists of nonzero entries of $|\sin\Theta(\mathcal{X},\mathcal{Z}) - \sin\Theta(\mathcal{Y},\mathcal{Z})|$ repeated twice. The nonzero entries of $S(P_\mathcal{X} - P_\mathcal{Y})$ are by Theorem 2.9 the nonzero entries of $\sin\Theta(\mathcal{X},\mathcal{Y})$ also repeated twice, thus we come to (3.2).

The cosine estimate (3.3) follows directly from the sine estimate (3.2) with $\mathcal{Z}^\perp$ instead of $\mathcal{Z}$ because of (2.2) utilizing the assumption $\dim\mathcal{X} = \dim\mathcal{Y}$. ☐

In our earlier paper, Knyazev and Argentati [17], in Lemmas 5.1 and 5.2 we obtained a particular case of Theorem 3.2, only for the largest change in the sine and the cosine, but with improved constants. We are not presently able, however, to modify the proofs of [17] using weak majorization, in order to improve the estimates of Theorem 3.2 by introducing the same constants as in [17].

Our last, but not least, result in this series is the weak majorization estimate for the sines or cosines squared, which provides the foundation for the rest of the paper.

Theorem 3.3. *Let* $\dim\mathcal{X} = \dim\mathcal{Y}$*, then*

$$|\cos^2\Theta(\mathcal{X},\mathcal{Z}) - \cos^2\Theta(\mathcal{Y},\mathcal{Z})| = |\sin^2\Theta(\mathcal{X},\mathcal{Z}) - \sin^2\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y}).$$

Proof. The equality is evident. To prove the majorization result for the sines squared, we start with the useful pinching identity

$$(P_\mathcal{X} - P_\mathcal{Z})^2 - (P_\mathcal{Y} - P_\mathcal{Z})^2 = P_{\mathcal{Z}^\perp}(P_\mathcal{X} - P_\mathcal{Y})P_{\mathcal{Z}^\perp} - P_\mathcal{Z}(P_\mathcal{X} - P_\mathcal{Y})P_\mathcal{Z}.$$

Applying Corollary 2.4 we obtain

$$\left|S\left((P_\mathcal{X} - P_\mathcal{Z})^2\right) - S\left((P_\mathcal{Y} - P_\mathcal{Z})^2\right)\right| \prec_w S\left(P_{\mathcal{Z}^\perp}(P_\mathcal{X} - P_\mathcal{Y})P_{\mathcal{Z}^\perp} - P_\mathcal{Z}(P_\mathcal{X} - P_\mathcal{Y})P_\mathcal{Z}\right).$$

For the left-hand side we use Theorem 2.9 as in the proof of Theorem 3.2, except that we are now working with the squares. For the right-hand side, the pinching Theorem 2.6 gives

$$S\left(P_{\mathcal{Z}^\perp}(P_\mathcal{X} - P_\mathcal{Y})P_{\mathcal{Z}^\perp} - P_\mathcal{Z}(P_\mathcal{X} - P_\mathcal{Y})P_\mathcal{Z}\right) \prec_w S\left(P_\mathcal{X} - P_\mathcal{Y}\right)$$

and we use Theorem 2.9 again to characterize $S\left(P_\mathcal{X} - P_\mathcal{Y}\right)$ the same way as in the proof of Theorem 3.2. ☐

4. Changes in the Trial Subspace in the Rayleigh–Ritz Method. In this section, we explore a simple, but deep, connection between the principal angles and the Rayleigh–Ritz method that we discuss in the Introduction. We demonstrate that the analysis of the influence of changes in a trial subspace in the Rayleigh–Ritz method is a natural extension of the theory concerning principal angles and the proximity of two subspaces developed in the previous section.

For the reader's convenience, let us repeat here the definition of Ritz values from the Introduction: Let $A : \mathcal{H} \to \mathcal{H}$ be a Hermitian operator and $P_\mathcal{X}$ be an orthogonal projector to a subspace $\mathcal{X}$ of $\mathcal{H}$. The eigenvalues $\Lambda(P_\mathcal{X} A|_\mathcal{X})$ are the Ritz values of operator A with respect to $\mathcal{X}$, which is called the trial subspace.

Let $\mathcal{X}$ and $\mathcal{Y}$ both be subspaces of $\mathcal{H}$ and $\dim\mathcal{X} = \dim\mathcal{Y}$. The goal of this section is to analyze sensitivity of the Ritz values with respect to the trial subspaces, specifically, to bound the change $|\Lambda\left(P_\mathcal{X} A|_\mathcal{X}\right) - \Lambda\left(P_\mathcal{Y} A|_\mathcal{Y}\right)|$ in terms of $\sin\Theta(\mathcal{X},\mathcal{Y})$ using weak majorization. Such an estimate is already obtained in Theorem 10 of our earlier paper Knyazev and Argentati [18] by applying Corollary 2.5 to the matrices of $P_{\mathcal{X}}A|_{\mathcal{X}}$ and $P_{\mathcal{Y}}A|_{\mathcal{Y}}$. This approach, however, leads to an extra factor $\sqrt{2}$ on the right-hand side, which is conjectured in [18] to be artificial.

We remove this $\sqrt{2}$ factor in our new Theorem 4.3 by using the entirely different and novel approach: we connect the Ritz values with extension Theorem 2.11 on the one hand and with the cosine squared of principal angles on the other hand. We have shown in Theorem 2.11 that a Hermitian nonnegative definite contraction operator can be extended to an orthogonal projector in a larger space. The extension has an extra nice property: it preserves the Ritz values.

COROLLARY 4.1. *Under the assumptions of Theorem 2.11, the Ritz values of operator $A : \mathcal{H} \to \mathcal{H}$ in the trial subspace $\mathcal{X} \subset H$ are the same as the Ritz values of operator $\hat{A} : \mathcal{H}^2 \to \mathcal{H}^2$ in the trial subspace*

$$\hat{\mathcal{X}} = \begin{bmatrix} \mathcal{X} \\ 0 \end{bmatrix} \subset \hat{\mathcal{H}} = \begin{bmatrix} \mathcal{H} \\ 0 \end{bmatrix} \subset \mathcal{H}^2.$$

Proof. Let $P_{\hat{\mathcal{H}}} : \mathcal{H}^2 \to \mathcal{H}^2$ be an orthogonal projector on the subspace $\hat{\mathcal{H}}$ and $P_{\hat{\mathcal{X}}} : \mathcal{H}^2 \to \mathcal{H}^2$ be an orthogonal projector on the subspace $\hat{\mathcal{X}}$. We use the equality sign to denote the trivial isomorphism between $\mathcal{H}$ and $\hat{\mathcal{H}}$, i.e. we simply write $\mathcal{H} = \hat{\mathcal{H}}$ and $\mathcal{X} = \hat{\mathcal{X}}$.

In this notation, we first observe that $A = P_{\hat{\mathcal{H}}}\hat{A}|_{\hat{\mathcal{H}}}$, i.e. the operator A itself can be viewed as a result of the Rayleigh–Ritz method applied to the operator $\hat{A}$ in the trial subspace $\hat{\mathcal{H}}$. Second, we use the fact that a recursive application of the Rayleigh–Ritz method on a system of enclosed trial subspaces is equivalent to a direct single application of the Rayleigh–Ritz method to the smallest trial subspace, indeed, in our notation, $P_{\hat{\mathcal{H}}}P_{\hat{\mathcal{X}}} = P_{\hat{\mathcal{X}}}P_{\hat{\mathcal{H}}} = P_{\hat{\mathcal{X}}}$, since $\hat{\mathcal{X}} \subset \hat{\mathcal{H}}$, thus

$$P_{\mathcal{X}}A|_{\mathcal{X}} = \left(P_{\hat{\mathcal{X}}}P_{\hat{\mathcal{H}}}\hat{A}|_{\hat{\mathcal{H}}}\right)\Big|_{\hat{\mathcal{X}}} = P_{\hat{\mathcal{X}}}\hat{A}|_{\hat{\mathcal{X}}}.$$

□

Next we note that Lemma 2.8 states that the Rayleigh–Ritz method applied to an orthogonal projector produces Ritz values, which are essentially the cosines squared of the principal angles between the range of the projector and the trial subspace. For the reader's convenience we reformulate Lemma 2.8 here:

LEMMA 4.2. *Let the Rayleigh–Ritz method be applied to $A = P_{\mathcal{Z}}$, where $P_{\mathcal{Z}}$ is an orthogonal projector onto a subspace $\mathcal{Z}$, and let $\mathcal{X}$ be the trial subspace in the Rayleigh–Ritz method. Then the set of the Ritz values is*

$$\Lambda\left(P_{\mathcal{X}}P_{\mathcal{Z}}|_{\mathcal{X}}\right) = [\cos^2\Theta(\mathcal{X},\mathcal{Z}), 0, \ldots, 0],$$

where there are $\max\{\dim\mathcal{X} - \dim\mathcal{Z}, 0\}$ *extra* 0*s.*

Now we are ready to direct our attention to the main topic of this section: the influence of changes in a trial subspace in the Rayleigh–Ritz method on the Ritz values.

THEOREM 4.3. *Let $A : \mathcal{H} \to \mathcal{H}$ be Hermitian and let $\mathcal{X}$ and $\mathcal{Y}$ both be subspaces of $\mathcal{H}$ and* $\dim\mathcal{X} = \dim\mathcal{Y}$. *Then*

$$|\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right) - \Lambda\left(P_{\mathcal{Y}}A|_{\mathcal{Y}}\right)| \prec_w (\lambda_{\max} - \lambda_{\min})\ \sin\Theta(\mathcal{X},\mathcal{Y}), \tag{4.1}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of A, respectively.

Proof. We prove Theorem 4.3 in two steps. First we show that we can assume that A is a nonnegative definite contraction without losing generality. Second, under these assumptions, we extend the operator A to an orthogonal projector by Theorem 2.11 and use the facts that such an extension does not affect the Ritz values by Corollary 4.1 and that the Ritz values of an orthogonal projector can be interpreted as the cosines squared of principal angles between subspaces by Lemma 4.2, thus reducing the problem to the already established result on weak majorization of the cosine squared Theorem 3.3.

We observe that the statement of the theorem is invariant with respect to a shift and a scaling, indeed, for real α and β if the operator A is replaced with $\beta(A-\alpha)$ and $\lambda_{\min}$ and $\lambda_{\max}$ are correspondingly updated, both sides of (4.1) are just multiplied by β and (4.1) is thus invariant with respect to α and β. Choosing $\alpha = \lambda_{\min}$ and $\beta = 1/(\lambda_{\max} - \lambda_{\min})$, the transformed operator $(A - \lambda_{\min})/(\lambda_{\max} - \lambda_{\min})$ is Hermitian with its eigenvalues enclosed in a segment $[0, 1]$, thus the statement (4.1) of the theorem can be equivalently rewritten as

$$|\Lambda\left(P_{\mathcal{X}}A|_{\mathcal{X}}\right) - \Lambda\left(P_{\mathcal{Y}}A|_{\mathcal{Y}}\right)| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y}), \tag{4.2}$$

where we from now on assume that A is a nonnegative definite contraction without losing generality.

The second step of the proof is to recast the problem into an equivalent problem for an orthogonal projector with the same Ritz values and principal angles. By Theorem 2.11 we can extend the nonnegative definite contraction A to an orthogonal projector $P_{\hat{\mathcal{Z}}}$, where $\hat{\mathcal{Z}}$ is a subspace of $\mathcal{H}^2$. $P_{\hat{\mathcal{Z}}}$ has by Corollary 4.1 the same Ritz values with respect to trial subspaces

$$\hat{\mathcal{X}} = \begin{bmatrix} \mathcal{X} \\ 0 \end{bmatrix} \subset \hat{\mathcal{H}} = \begin{bmatrix} \mathcal{H} \\ 0 \end{bmatrix} \subset \mathcal{H}^2 \text{ and } \hat{\mathcal{Y}} = \begin{bmatrix} \mathcal{Y} \\ 0 \end{bmatrix} \subset \hat{\mathcal{H}} = \begin{bmatrix} \mathcal{H} \\ 0 \end{bmatrix} \subset \mathcal{H}^2$$

as A has with respect to the trial subspaces $\mathcal{X}$ and $\mathcal{Y}$. By Lemma 4.2, these Ritz values are equal to the cosines squared of the principal angles between $\hat{\mathcal{Z}}$ and the trial subspace $\hat{\mathcal{X}}$ or $\hat{\mathcal{Y}}$ possibly with the same number of 0s being added. Moreover, the principal angles between $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ in $\mathcal{H}^2$ are clearly the same as those between $\mathcal{X}$ and $\mathcal{Y}$ in $\mathcal{H}$ and $\dim\hat{\mathcal{X}} = \dim\mathcal{X} = \dim\mathcal{Y} = \dim\hat{\mathcal{Y}}$. Thus, (4.2) can be equivalently reformulated as

$$|\cos^2\Theta(\hat{\mathcal{X}},\hat{\mathcal{Z}}) - \cos^2\Theta(\hat{\mathcal{Y}},\hat{\mathcal{Z}}))| \prec_w \sin\Theta(\hat{\mathcal{X}},\hat{\mathcal{Y}}). \tag{4.3}$$

Finally, we notice that (4.3) is already proved in Theorem 3.3. ◻

REMARK 4.1. *As in Remark 7 of Knyazev and Argentati [18], the constant* $\lambda_{\max} - \lambda_{\min}$ *in Theorem 4.3 can be replaced with*

$$\max_{x\in\mathcal{X}+\mathcal{Y},\,\|x\|=1}(x,Ax) - \min_{x\in\mathcal{X}+\mathcal{Y},\,\|x\|=1}(x,Ax),$$

which for some subspaces $\mathcal{X}$ and $\mathcal{Y}$ can provide a significant improvement.

REMARK 4.2. *The implications of the weak majorization inequality in Theorem 4.3 may not be obvious to every reader. To clarify, let $m = \dim\mathcal{X} = \dim\mathcal{Y}$ and let $\alpha_1 \geq \cdots \geq \alpha_m$ be the Ritz values of A with respect to $\mathcal{X}$ and $\beta_1 \geq \cdots \geq \beta_m$ be the Ritz values of A with respect $\mathcal{Y}$. The weak majorization inequality in Theorem 4.3 directly implies*

$$\sum_{i=1}^{k} |\alpha_i - \beta_i|^{\downarrow} \leq (\lambda_{\max} - \lambda_{\min}) \sum_{i=1}^{k} \sin(\Theta_i(\mathcal{X},\mathcal{Y}))^{\downarrow}, \quad k = 1, \ldots, m,$$

e.g., for $k = m$ we obtain

$$\sum_{i=1}^{m} |\alpha_i - \beta_i| \le (\lambda_{\max} - \lambda_{\min}) \sum_{i=1}^{m} \sin(\Theta_i(\mathcal{X}, \mathcal{Y})), \tag{4.4}$$

and for $k = 1$ we have

$$\max_{j=1,\dots,m} |\alpha_j - \beta_j| \le (\lambda_{\max} - \lambda_{\min}) \operatorname{gap}(\mathcal{X}, \mathcal{Y}), \tag{4.5}$$

where the gap $\operatorname{gap}(\mathcal{X}, \mathcal{Y})$ *between equidimensional subspaces $\mathcal{X}$ and $\mathcal{Y}$ is the sine of the largest angle between $\mathcal{X}$ and $\mathcal{Y}$. Inequality (4.5) is proved in Knyazev and Argentati [18].*

For real vectors x and y the weak majorization $x \prec_w y$ is equivalent to the inequality $\sum_{i=1}^{n} \phi(x_i) \le \sum_{i=1}^{n} \phi(y_i)$ for any continuous nondecreasing convex real valued function ϕ, e.g., Marshall and Olkin [23], Statement 4.B.2. Taking, e.g., $\phi(t) = t^p$ with $p \ge 1$, Theorem 4.3 also implies

$$\Big(\sum_{i=1}^{m} |\alpha_i - \beta_i|^p \Big)^{\frac{1}{p}} \le (\lambda_{\max} - \lambda_{\min}) \Big(\sum_{i=1}^{m} \sin(\Theta_i(\mathcal{X}, \mathcal{Y}))^p \Big)^{\frac{1}{p}}, \quad 1 \le p < \infty.$$

We finally note that the results of Theorem 4.3 are not intended for the case where one of the subspaces $\mathcal{X}$ or $\mathcal{Y}$ is invariant with respect to operator A. In such a case, it is natural to expect a much better bound that involves the square of the $\sin \Theta(\mathcal{X}, \mathcal{Y})$. Majorization results of this kind are not apparently known in the literature. Without majorization, estimates just for the largest change in the Ritz values are available, e.g., Knyazev and Argentati [18], Knyazev and Osborn [20].

5. Application of the Majorization Results to Graph Spectra Comparison. In this section, we show that our majorization results can be applied to compare graph spectra. The graph spectra comparison can be used for graph matching and has applications in data mining, cf. Kosinov and Caelli [22].

The section is divided into three subsections. In Subsection 5.1, we give all necessary definitions and basic facts concerning Laplacian graph spectra. In Subsection 5.2, we connect the Laplacian graph spectrum and Ritz values, by introducing the graph edge Laplacian. Finally, in Subsection 5.3, we prove our main result on the Laplacian graph spectra comparison.

5.1. Incidence Matrices and Graph Laplacians. Here, we give mostly well known relevant definitions, e.g., Chung [3], Cvetković et al. [4], Merris [25, 26, 27], Mohar [28], just slightly tailored for our specific needs.

Let V be a finite ordered set (of vertices), with an individual element (vertex) denoted by $v_i \in V$. Let E_c be the finite ordered set (of all edges), with an individual element (edge) denoted by $e_k \in E_c$ such that every $e_k = [v_i, v_j]$ for all possible $i > j$. E_c can be viewed as the set of edges of a complete simple graph with vertices V (without self-loops and/or multiple edges). The results of the present paper are invariant with respect to specific ordering of vertices and edges.

Let $w_c : E_c \to \mathbf{R}$ be a function describing edge weights, i.e. $w_c(e_k) \in \mathbf{R}$. If for some edge e_k the weight is positive, $w_c(e_k) > 0$, we call this edge present, if $w_c(e_k) = 0$ we say that the edge is absent. In this paper we do not allow negative edge weights. For a given weight function w_c, we define $E \subseteq E_c$ such that $e_k \in E$ if $w_c(e_k) \neq 0$ and we define w to be the restriction of w_c on all present edges E, i.e. w is made of all nonzero values of w_c. A pair of sets of vertices V and present edges E with weights w is called a graph (V, E) or a weighted graph (V, E, w).

The vertex–edge incidence matrix Q_c of a complete graph (V, E_c) is a matrix which has a row for each vertex and a column for each edge, with column-wise entries determined as $q_{ik} = 1$, $q_{jk} = -1$ for every edge $e_k = [v_i, v_j]$, $i > j$ in E_c and with all other entries of Q_c equal to zero. The vertex–edge incidence matrix Q of a graph (V, E) is determined in the same way, but only for the edges present in E. The vertex–edge incidence matrix can be viewed as a matrix representation of a graph analog of the divergence operator from partial differential equations (PDE).

Extending the analogy with PDE, the matrix $L = QQ^*$ is called the graph Laplacian. In the PDE context, this definition corresponds to the negative Laplacian with the natural boundary conditions, cf. McDonald and Meyers [24]. Let us note that in the graph theory literature such a definitions of the graph Laplacian is usually attributed to directed graphs, even though changing any edge direction into the opposite does not affect the graph Laplacian.

If we want to take into account the weights, we can work with the matrix $Q \operatorname{diag}(w(E))Q^*$, which is an analog of an isotropic diffusion operator, or we can introduce a more general edge matrix W and work with QWQ^*, which corresponds to a general anisotropic diffusion. It is interesting to notice the equality

$$Q_c \operatorname{diag}(w_c(E_c))Q_c^* = Q \operatorname{diag}(w(E))Q^*, \tag{5.1}$$

which shows two alternative equivalent formulas for the graph diffusion operator.

For simplicity of presentation, we assume in the rest of the paper that the weights w_c take only the values zero and one. Under this assumption, we introduce matrix $P = \operatorname{diag}(w_c(E_c))$ and notice that P is the matrix of an orthogonal projector on a subspace spanned by coordinate vectors with indices corresponding to the indices of edges present in E and that equality (5.1) turns into

$$Q_c P Q_c^* = QQ^*. \tag{5.2}$$

Let us note that our results can be easily extended to a more general case of arbitrary nonnegative weights, or even to the case of the edge matrix W, assuming that it is nonnegative definite, $W \geq 0$.

Fiedler's pioneering work [7] on using the eigenpairs of the graph Laplacian to determine some structural properties of the graph has attracted much attention in the past. Recent advances in large-scale eigenvalue computations using multilevel preconditioning, e.g., Knyazev [16], Knyazev and Neymeyr [19], Koren et al. [21], suggest novel efficient numerical methods to compute the Fiedler vector and may rejuvenate this classical approach, e.g., for graph partitioning. In this paper, we concentrate on the whole set of eigenvalues of L, which is called the Laplacian graph spectrum.

It is known that the Laplacian graph spectrum does not determine the graph uniquely, i.e. that there exist isospectral graphs, see, e.g., van Dam and Haemers [38] and references there. However, intuition suggests that a small change in a large graph should not change the Laplacian graph spectrum very much; and attempts have been made to use the closeness of Laplacian graph spectra to judge the closeness of the graphs in applications; for alternative approaches, see Blondel et al. [2]. The goal of this section is to backup this intuition with rigorous estimates for proximity of the Laplacian graph spectra.

5.2. Laplacian graph spectrum and Ritz values. In the previous section, we obtain in Theorem 4.3 the weak majorization bound on for changes in the Ritz values depending on a change in the trial subspace, which we would like to apply to analyze the graph spectrum. In this subsection, we present an approach that allows us to interpret the Laplacian graph spectrum as a set of Ritz values obtained by the Rayleigh–Ritz method applied to the complete graph.

A graph (V, E) can evidently be obtained from the complete graph (V, E_c) by removing edges, moreover, as we already discussed, we can construct the (V, E) graph Laplacian by either of the terms in equality (5.2). The problem is that such a construction cannot be recast as an application of the Rayleigh–Ritz method, since the multiplication by the projector P takes place inside of the product, not outside, as required by the Rayleigh–Ritz method.

To resolve this difficulty, we use the matrix $K = Q^*Q$ that is sometimes called the matrix of the graph *edge* Laplacian, instead of the matrix of the graph *vertex* Laplacian $L = QQ^*$, as both matrices K and L share the same nonzero eigenvalues. The advantage of the edge Laplacian K is that it can be obtained from the edge Laplacian of the complete graph $Q_c^*Q_c$ simply by removing the rows and columns that correspond to missing edges. Mathematically, this procedure can be viewed as an instance of the classical Rayleigh–Ritz method:

LEMMA 5.1. *Let us remind the reader that the weights w_c take only the values zero and one and that $P = \operatorname{diag}(w_c(E_c))$ is a matrix of an orthogonal projector on a subspace spanned by coordinate vectors with indices corresponding to the indices of edges present in E. Then $Q^*Q = (PQ_c^*Q_c)|_{\mathrm{Range}(P)}$, in other words, the matrix Q^*Q is the result of the Rayleigh–Ritz method applied to the matrix $Q_c^*Q_c$ on the trial subspace* $\mathrm{Range}(P)$. The application of the Rayleigh–Ritz method in this case is reduced to simply crossing out rows and columns of the matrix $Q_c^*Q_c$ corresponding to absent edges, since P projects onto a span of coordinate vectors with the indices of the present edges.

Lemma 5.1 is a standard tool in the spectral graph theory, e.g., Haemers [10], to prove the eigenvalues interlacing; however, the procedure is not apparently recognized in the spectral graph community as an instance of the classical Rayleigh–Ritz method. Lemma 5.1 provides us with the missing link in order to apply our Theorem 4.3 to Laplacian graph spectra comparison.

5.3. Majorization of Ritz Values for Laplacian Graph Spectra Comparison. Using the tools that we have presented in the previous subsections, we now can apply our weak majorization result of Section 4 to analyze the change in the graph spectrum when several edges are added to or removed from the graph.

THEOREM 5.2. *Let (V, E^1) and (V, E^2) be two graphs with the same set of n vertices V, with the same number of edges E^1 and E^2, and with the number of differing edges in E^1 and E^2 equal to l. Then*

$$\sum_k |\lambda_k^1 - \lambda_k^2| \le nl, \tag{5.3}$$

where λ_k^1 and λ_k^2 are all elements of the Laplacian spectra of the graphs (V, E^1) and (V, E^2) in nonincreasing order.

Proof. The spectra of the graph vertex and edge Laplacians QQ^* and Q^*Q are the same apart from zero, which does not affect the statement of the theorem, so we redefine λ_k^1 and λ_k^2 as elements of the spectra, counting the multiplicities, of the edge Laplacians of the graphs (V, E^1) and (V, E^2). Then, by Theorem 5.1, λ_k^1 and λ_k^2

are the Ritz values of the edge Laplacian matrix $A = Q_c^* Q_c$ of the complete graph, corresponding to the trial subspaces $\mathcal{X} = \text{Range}(P_1)$ and $\mathcal{Y} = \text{Range}(P_2)$ spanned by coordinate vectors with indices of the edges present in E^1 and E^2, respectively.

Let us apply Theorem 4.3, taking the sum over all available nonzero values in the weak majorization statement as in (4.4). This already gives us the left-hand side of (5.3). To obtain the right-hand side of (5.3) from Theorem 4.3, we now show in our case that, first, $\lambda_{\max} - \lambda_{\min} = n$ and, second, the sum of sines of all angles between the trial subspaces $\mathcal{X}$ and $\mathcal{Y}$ is equal to l.

The first claim follows from the fact, which is easy to check by direct calculation, that the spectrum of the vertex (and thus the edge) Laplacian of the complete graph with n vertices consists of only two eigenvalues $\lambda_{\max} = n$ and $\lambda_{\min} = 0$. Let us make a side note that we can interpret the Laplacian of the complete graph as a scaled projector, i.e. in this case we could have applied Theorem 3.3 directly, rather than Theorem 4.3, which would still result in (5.3).

The second claim, on the sum of sines of all angles, follows from the definition of $\mathcal{X}$ and $\mathcal{Y}$ and the assumption that the number of differing edges in E^1 and E^2 is equal to l. Indeed, $\mathcal{X}$ and $\mathcal{Y}$ are spanned by coordinate vectors with indices of the edges present in E^1 and E^2. The edges that are present both in E^1 and E^2 contribute zero angles into $\Theta(\mathcal{X}, \mathcal{Y})$, while the l edges that are different in E^1 and E^2 contribute l right angles into $\Theta(\mathcal{X}, \mathcal{Y})$, so that the sum of all terms in $\sin \Theta(\mathcal{X}, \mathcal{Y})$ is equal to l. $\square$

Remark 4.1 is also applicable for Theorem 5.2 — while the min term is always zero, since all graph Laplacians are degenerate, the max term can be made smaller by replacing n with the largest eigenvalue of the Laplacian of the graph $(V, E^1 \cup E^2)$.

It is clear from the proof that we do not use the full force of our weak majorization results in Theorem 5.2, because it concerns angles which are zero or $\pi/2$. Nevertheless, the results of Theorem 5.2 appear to be novel in graph theory. We note that these results can be easily extended on k-partite graphs, and possibly to mixed graphs.

Let us finally mention an alternative approach to compare Laplacian graph spectra, which we do not cover in the present paper, by applying Corollary 2.5 directly to graph Laplacians and estimating the right-hand side using the fact that the changes in l edges represents a low–rank perturbation of the graph Laplacian, cf. [35].

Conclusions. We use majorization to investigate the sensitivity of angles between subspaces and Ritz values with respect to subspaces, and to analyze changes in graph Laplacian spectra where edges are added and removed. We discover that these seemingly different areas are all surprisingly related. We establish in a unified way new results on weak majorization of the changes in the sine/cosine (squared) and in the Ritz values. The main strength of the paper in our opinion is, however, not so much in the results themselves but rather in a novel and elegant proof technique that is based on a classical but rarely used idea of extending Hermitian operators to orthogonal projectors in a larger space. We believe that such a technique is very powerful and should be known to a wider audience.

Acknowledgments. We are indebted to Li Qiu and Yanxia Zhang, who have suggested to us the idea of the proof of the sine estimate in Theorem 3.2 that has allowed us also to simplify our original proof of Theorem 3.3. The authors thank our Ph.D. students Ilya Lashuk for contributing to Section 2.3 and Abram Jujunashvili for many helpful comments. Finally, we thank anonymous referees for their numerous useful and illuminating suggestions that have dramatically improved the paper.

References.

[1] R. Bhatia. *Matrix Analysis*. Springer-Verlag, Berlin, 1997. ISBN 0-387-94846-5.

[2] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren. A measure of similarity between graph vertices: applications to synonym extraction and web searching. *SIAM Rev.*, 46(4):647–666 (electronic), 2004.

[3] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997. ISBN 0-8218-0315-8.

[4] D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of graphs*. Johann Ambrosius Barth, Heidelberg, third edition, 1995. ISBN 3-335-00407-8. Theory and applications.

[5] C. Davis. Separation of two linear subspaces. *Acta Sci. Math. Szeged*, 19:172–187, 1958.

[6] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[7] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Math. J.*, 23(98): 298–305, 1973.

[8] I. M. Glazman and Ju. I. Ljubič. *Finite-dimensional linear analysis: a systematic presentation in problem form*. Translated from the Russian and edited by G. P. Barker and G. Kuerti. The M.I.T. Press, Cambridge, Mass.-London, 1974.

[9] I. C. Gohberg and M. G. Kreĭn. *Introduction to the theory of linear nonselfadjoint operators*. Translated from the Russian by A. Feinstein. Translations of Mathematical Monographs, Vol. 18. American Mathematical Society, Providence, R.I., 1969.

[10] W. H. Haemers. Interlacing eigenvalues and graphs. *Linear Algebra Appl.*, 226/228:593–616, 1995.

[11] P. R. Halmos. Normal dilations and extensions of operators. *Summa Brasil. Math.*, 2:125–134, 1950.

[12] P. R. Halmos. Two subspaces. *Trans. Amer. Math. Soc.*, 144:381–389, 1969.

[13] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1959.

[14] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, NY, 1999.

[15] C. Jordan. Essai sur la géométrie à n dimensions. *Bull. Soc. Math. France*, 3: 103–174, 1875.

[16] A. V. Knyazev. Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. *SIAM Journal on Scientific Computing*, 23(2):517–541, 2001.

[17] A. V. Knyazev and M. E. Argentati. Principal angles between subspaces in an A-based scalar product: Algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2009–2041, 2002.

[18] A. V. Knyazev and M. E. Argentati. On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces. *Linear Algebra and its Applications*, 2006. In press (available online).

[19] A. V. Knyazev and K. Neymeyr. Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method. *Electron. Trans. Numer. Anal.*, 15:38–55 (electronic), 2003. Tenth Copper Mountain Conference on Multigrid Methods (Copper Mountain, CO, 2001).

[20] A. V. Knyazev and J. Osborn. New a priori FEM error estimates for eigenval-

ues. *SIAM Journal on Numerical Analysis*, 43(6):2647–2667, 2006 An extendent version published as a technical report UCD-CCM 215, 2004 (http://www-math.cudenver.edu/ccm/reports/rep215.pdf) at the Center for Computational Mathematics, University of Colorado at Denver.

[21] Y. Koren, L. Carmel, and D. Harel. Drawing huge graphs by algebraic multigrid optimization. *Multiscale Model. Simul.*, 1(4):645–673 (electronic), 2003.

[22] S. Kosinov and T. Caelli. Inexact multisubgraph matching using graph eigenspace and clustering models. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, pages 133–142, London, UK, 2002. Springer-Verlag. ISBN 3-540-44011-9.

[23] A. W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979. ISBN 0-12-473750-1.

[24] P. McDonald and R. Meyers. Diffusions on graphs, Poisson problems and spectral geometry. *Trans. Amer. Math. Soc.*, 354(12):5111–5136 (electronic), 2002.

[25] R. Merris. Laplacian matrices of graphs: a survey. *Linear Algebra Appl.*, 197/198: 143–176, 1994. ISSN 0024-3795. Second Conference of the International Linear Algebra Society (ILAS) (Lisbon, 1992).

[26] R. Merris. A survey of graph Laplacians. *Linear and Multilinear Algebra*, 39 (1-2):19–31, 1995.

[27] R. Merris. Laplacian graph eigenvectors. *Linear Algebra Appl.*, 278(1-3):221–236, 1998.

[28] B. Mohar. Some applications of Laplace eigenvalues of graphs. In *Graph symmetry (Montreal, PQ, 1996)*, volume 497 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 225–275. Kluwer Acad. Publ., Dordrecht, 1997.

[29] C. C. Paige and M. A. Saunders. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.*, 18(3):398–405, 1981.

[30] C. C. Paige and M. Wei. History and generality of the CS decomposition. *Linear Algebra Appl.*, 208/209:303–326, 1994.

[31] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. ISBN 0-89871-402-8. Corrected reprint of the 1980 original.

[32] L. Qiu and Y. Zhang. Private communication, 2005.

[33] L. Qiu, Y. Zhang, and C.-K. Li. Unitarily invariant metrics on the Grassmann space. *SIAM Journal on Matrix Analysis and Applications*, 27 (2): 507–531, 2005.

[34] F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Dover Publications, Inc., New York, NY, 1990.

[35] W. So. Rank one perturbation and its application to the Laplacian spectrum of a graph. *Linear and Multilinear Algebra*, 46(3):193–198, 1999.

[36] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Academic Press Inc., Boston, MA, 1990. ISBN 0-12-670230-6.

[37] G. W. Stewart. *Matrix Algorithms Volume II: Eigensystems*. SIAM, Philadelphia, PA, 2001.

[38] E. R. van Dam and W. H. Haemers. Which graphs are determined by their spectrum? *Linear Algebra Appl.*, 373:241–272, 2003.