Basics: Complex Matrices

Marianna Bolla and Tamás Szabados, BME Math. Inst.

September 14, 2020

We consider finite dimensional complex Euclidean spaces that are also Hilbert spaces. Linear operations between them can be described by matrices of complex entries. Vectors are treated as column-vectors and denoted by bold-face, lower-case letters. The inner product of the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ is therefore written with matrix multiplication as $\mathbf{x}^*\mathbf{y}$, where * stands for the conjugate transpose of a complex vector, hence $\mathbf{x}^*$ is a row-vector. Matrices will be denoted by bold-face upper-case letters. An $m \times n$ matrix $\boldsymbol{A} = [a_{ij}]$ of complex entries a_{ij}'s corresponds to a $\mathbb{C}^n \to \mathbb{C}^m$ linear transformation (operator). Its adjoint, $\boldsymbol{A}^*$, is an $n \times m$ matrix, the entries of which are $a^*_{ij} = \overline{a_{ji}}$. An $n \times n$ matrix is called quadratic (or square) and it maps $\mathbb{C}^n$ into itself. The identity matrix is denoted by $\boldsymbol{I}$ or $\boldsymbol{I}_n$ if we want to refer to its size.

Definition 1. *Some special kinds of matrices:*

- *The $n \times n$ complex matrix $\boldsymbol{A}$ is* self-adjoint (Hermitian) *if $\boldsymbol{A} = \boldsymbol{A}^*$. In paricular, a real matrix with $\boldsymbol{A}^T = \boldsymbol{A}$ is called* symmetric.
- *The $n \times n$ real matrix $\boldsymbol{A}$ is* anty-symmetric *if $\boldsymbol{A}^T = -\boldsymbol{A}$.*
- *The $n \times n$ complex matrix $\boldsymbol{A}$ is* unitary *if $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}_n$. It means that both its rows and columns constitute a complete orthonormal system in $\mathbb{C}^n$.*
- *The $n \times r$ complex matrix ($r \leq n$) is* sub-unitary *if its columns constitute a (usually not complete) orthonormal system in $\mathbb{C}^n$; consequently, $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}_r$, whereas $\boldsymbol{A}\boldsymbol{A}^*$ is the matrix of the orthogonal projection onto the subspace spanned by the column vectors of $\boldsymbol{A}$.*

- *The $n \times n$ complex matrix $\boldsymbol{P}$ is* Hermitian projector *if it is self-adjoint and idempotent, i.e., $\boldsymbol{P}^2 = \boldsymbol{P}$.*

- *The complex matrix $\boldsymbol{A}$ is called* normal *if $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.*

The $n \times n$ matrix $\boldsymbol{A}$ has an inverse if and only if its determinant, $\det \boldsymbol{A} = |\boldsymbol{A}| \neq 0$, and its inverse is denoted by $\boldsymbol{A}^{-1}$. In this case, the linear transformation corresponding to $\boldsymbol{A}^{-1}$ undoes the effect of the $\mathbb{C}^n \to \mathbb{C}^n$ transformation corresponding to $\boldsymbol{A}$, i.e. $\boldsymbol{A}^{-1}\mathbf{y} = \mathbf{x}$ if and only if $\boldsymbol{A}\mathbf{x} = \mathbf{y}$ for any $\mathbf{y} \in \mathbb{C}^n$; equivalently, $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$. A unitary matrix $\boldsymbol{A}$ is always invertible and $\boldsymbol{A}^{-1} = \boldsymbol{A}^*$.

It is important that in case of an invertible (*regular*) matrix $\boldsymbol{A}$, the *range* (or image space) of $\boldsymbol{A}$ – denoted by $\mathcal{R}(\boldsymbol{A})$ – is the whole $\boldsymbol{C}^n$, and in exchange, the *kernel* of $\boldsymbol{A}$ (the subspace of vectors that are mapped into the zero vector by $\boldsymbol{A}$) consists of only the vector $\mathbf{0}$. Note that for an $m \times n$ matrix $\boldsymbol{A}$, its range is

$$\mathcal{R}(\boldsymbol{A}) = \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\},$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the column vectors of $\boldsymbol{A}$ for which fact the notation $\boldsymbol{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ will be used. The *rank* of $\boldsymbol{A}$ is the dimension of its range:

$$\mathrm{rank}(\boldsymbol{A}) = \dim \mathcal{R}(\boldsymbol{A}),$$

and it also equals the maximum number of linearly independent rows, or equivalently, the maximum number of linearly independent columns of $\boldsymbol{A}$, or the maximal size of a nonzero minor (subdeterminant) of $\boldsymbol{A}$. Trivially, $\mathrm{rank}(\boldsymbol{A}) \leq \min\{m, n\}$; if equality is attained, we say that $\boldsymbol{A}$ has *full rank*. In case of $m = n$, $\boldsymbol{A}$ is regular if and only if $\mathrm{rank}(\boldsymbol{A}) = n$, and *singular* otherwise.

Eigenvalues and eigenvectors tell "everything" about a quadratic matrix. The complex number λ is an eigenvalue of the $n \times n$ complex matrix $\boldsymbol{A}$ with corresponding eigenvector $\mathbf{u} \neq \mathbf{0}$ if $\boldsymbol{A}\mathbf{u} = \lambda\mathbf{u}$. If $\mathbf{u}$ is an eigenvector of $\boldsymbol{A}$, it is easy to see that for $c \neq 0$, $c\mathbf{u}$ is also an eigenvector with the same eigenvalue. Therefore, it is better to speak about *eigen-direction*s instead of eigenvectors; or else, we will consider specially normalized, e.g. unit-norm eigenvectors, when only the orientation is divalent. It is well known that an $n \times n$ matrix $\boldsymbol{A}$ has exactly n eigenvalues (with multiplicities) which are (complex) roots of the characteristic polynomial $|\boldsymbol{A} - \lambda\boldsymbol{I}|$. Knowing the eigenvalues, the corresponding eigenvectors are obtained by solving the system of linear equations $(\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u} = \mathbf{0}$ which must have a non-trivial

solution due to the choice of λ. In fact, there are infinitely many solutions (in case of a single eigenvalue, they are constant multiples of each other).

Normal matrices have the following important spectral property: to their eigenvalues there corresponds a complete orthonormal set of eigenvectors; choosing this as a new basis, the matrix becomes *diagonal* (all the off-diagonal entries are zeros). Here we only state the analogous version for Hermitian matrices.

Theorem 1 (Hilbert–Schmidt theorem)**.** *The $n \times n$ self-adjoint complex matrix $\boldsymbol{A}$ has real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ (with multiplicities), and the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ can be chosen so that they constitute a complete orthonormal set in $\mathbb{C}^n$.*

Theorem 1 implies the following *Spectral Decomposition (SD)* of the $n \times n$ self-adjoint matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^* = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^*, \tag{1}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalues – called *spectrum* – in its main diagonal, while $\boldsymbol{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is the unitary matrix containing the corresponding eigenvectors of $\boldsymbol{A}$ in its columns in the order of the eigenvalues. Of course, permuting the eigenvalues in the main diagonal of $\boldsymbol{\Lambda}$, and the columns of $\boldsymbol{U}$ accordingly, will lead to the same SD, however – if not otherwise stated – we will enumerate the real eigenvalues in non-increasing order. About the uniqueness of the above SD we can state the following: the unit-norm eigenvector corresponding to a single eigenvalue is unique (up to orientation), whereas to an eigenvalue with multiplicity m there corresponds a unique m-dimensional so-called *eigen-subspace* within which any orthonormal set can be chosen for the corresponding eigenvectors.

Definition 2. *The parsimonious SD of the $n \times n$ self-adjoint matrix $\boldsymbol{A}$ of rank r is*

$$\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{u}_i^* = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{U}}^*, \tag{2}$$

where $\tilde{\boldsymbol{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$ is $n \times r$ sub-unitary matrix and $\tilde{\boldsymbol{\Lambda}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ is $r \times r$ diagonal matrix, where $\lambda_1 \geq \cdots \geq \lambda_r > 0$.

The quadratic form $\mathbf{x}^*\boldsymbol{A}\mathbf{x}$ with the SD of the self-adjoint $\boldsymbol{A}$ is

$$\mathbf{x}^*\boldsymbol{A}\mathbf{x} = \sum_{i=1}^{n} \lambda_i (\mathbf{x}^*\mathbf{u}_i)\overline{(\mathbf{x}^*\mathbf{u}_i)} = \sum_{i=1}^{n} \lambda_i |\mathbf{x}^*\mathbf{u}_i|^2$$

that is a real number. Some properties of the self-adjoint matrix $\boldsymbol{A}$ and of the quadratic forms generated by it follow:

- $\boldsymbol{A}$ is singular if and only if it has a 0 eigenvalue, and

  $$r = \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{\Lambda}) = |\{i \,:\, \lambda_i \neq 0\}|;$$

  moreover, $\mathcal{R}(\boldsymbol{A}) = \operatorname{Span}\{\mathbf{u}_i \,:\, \lambda_i \neq 0\}$. Therefore, the SD of $\boldsymbol{A}$ simplifies to $\sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{u}_i^*$.

- $\boldsymbol{A}$ is *positive (negative) definite* if $\mathbf{x}^*\boldsymbol{A}\mathbf{x} > 0$ ($\mathbf{x}^*\boldsymbol{A}\mathbf{x} < 0$), $\forall \mathbf{x} \neq \mathbf{0}$; equivalently, all the eigenvalues of $\boldsymbol{A}$ are positive (negative).

- $\boldsymbol{A}$ is *positive (negative) semidefinite* if $\mathbf{x}^*\boldsymbol{A}\mathbf{x} \geq 0$ ($\mathbf{x}^*\boldsymbol{A}\mathbf{x} \leq 0$), $\forall \mathbf{x} \in \mathbb{C}^n$; equivalently, all the eigenvalues of $\boldsymbol{A}$ are non-negative (non-positive).

- Note that the notion *non-negative (non-positive) definite* can be used instead of positive (negative) semidefinite. In some literature if $\boldsymbol{A}$ is called positive (negative) semidefinite, then it is understood that $\mathbf{x}^*\boldsymbol{A}\mathbf{x} = \mathbf{0}$ for at least one $\mathbf{x} \neq \mathbf{0}$; and so the spectrum of $\boldsymbol{A}$ contains the zero eigenvalue too.

- $\boldsymbol{A}$ is indefinite if $\mathbf{x}^*\boldsymbol{A}\mathbf{x}$ takes both positive and negative values (with different, non-zero $\mathbf{x}$'s); equivalently, the spectrum of $\boldsymbol{A}$ contains at least one positive and one negative eigenvalue.

- $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i$ and $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$.

A canonical decomposition for a rectangular matrix is a useful tool.

Theorem 2. *Let $\boldsymbol{A}$ be an $m \times n$ rectangular matrix of complex entries,* $\operatorname{rank}(\boldsymbol{A}) = r \leq \min\{m, n\}$. *Then there exist an orthonormal set* $(\mathbf{v}_1, \ldots, \mathbf{v}_r) \subset \mathbb{C}^m$ *and* $(\mathbf{u}_1, \ldots, \mathbf{u}_r) \subset \mathbb{C}^n$ *together with the positive real numbers* $s_1 \geq s_2 \geq \cdots \geq s_r > 0$ *such that*

$$\boldsymbol{A}\mathbf{u}_j = s_j \mathbf{v}_j, \quad \boldsymbol{A}^*\mathbf{v}_j = s_j \mathbf{u}_j, \quad j = 1, 2, \ldots, r. \tag{3}$$

The elements $\mathbf{v}_j \in \mathbb{C}^m$ and $\mathbf{u}_j \in \mathbb{C}^n$ $(j = 1, \ldots, r)$ in (3) are called *relevant singular vector pairs* (or *left and right singular vectors*) corresponding to the *singular value* s_j $(j = 1, 2, \ldots, r)$. The transformations in (3) give a one-to one mapping between $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{R}(\boldsymbol{A}^*)$, all the other vectors of $\mathbb{C}^n$ and $\mathbb{C}^m$ are mapped into the zero vector of $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively. However, the left and right singular vectors can appropriately be completed into a complete orthonormal set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathbb{C}^m$ and $\{\mathbf{u}_1, \ldots \mathbf{u}_n\} \subset \mathbb{C}^n$, respectively, such that, the so introduced extra vectors in the kernel subspaces in $\mathbb{C}^m$ and $\mathbb{C}^n$ are mapped into the zero vector of $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. With the unitary matrices $\boldsymbol{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ and $\boldsymbol{U} = (\mathbf{u}_1, \ldots \mathbf{u}_n)$, the following *singular value decomposition (SVD)* of $\boldsymbol{A}$ and $\boldsymbol{A}^*$ holds:

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^* = \sum_{i=1}^{r} s_i \mathbf{v}_i \mathbf{u}_i^* \quad \text{and} \quad \boldsymbol{A}^* = \boldsymbol{U}\boldsymbol{S}^*\boldsymbol{V}^* = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^*, \tag{4}$$

where $\boldsymbol{S}$ is an $m \times n$ so-called *generalized diagonal matrix* which contains the singular values $s_1, \ldots, s_r$ in the first r positions of its main diagonal (starting from the upper left corner) and zeros otherwise. We remark that there are other equivalent forms of the above SVD depending on, whether $m < n$ or $m \geq n$. For example, in the $m < n$ case, $\boldsymbol{V}$ can be an $m \times m$ unitary, $\boldsymbol{S}$ an $m \times m$ diagonal, and $\boldsymbol{U}$ an $n \times m$ sub-unitary matrix with the same relevant entries. About the uniqueness of the SVD the following can be stated: to a single positive singular value there corresponds a unique singular vector pair (of course, the orientation of the left and right singular vectors can be changed at the same time). To a positive singular value of multiplicity $k > 1$ a k-dimensional left and right so-called *isotropic subspace* corresponds, within which, any k-element orthonormal sets can embody the left and right singular vectors with orientation such that the requirements in (3) are met.

We also remark that the singular values of a self-adjoint matrix are the absolute values of its real eigenvalues. In case of a positive eigenvalue, the left and right singular vectors are the same (they coincide with the corresponding eigenvector with any, but the same orientation). In case of a negative eigenvalue, the left and right side singular vectors are opposite (any of them is the corresponding eigenvector which have a divalent orientation). In case of a zero singular value the orientation is immaterial, as it does not contribute to the SVD of the underlying matrix.

Assume that the $m \times n$ complex matrix $\boldsymbol{A}$ of rank r has SVD (4). It is easy

to see that the matrices $\boldsymbol{A}\boldsymbol{A}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ are self-adjoint, positive semidefinite matrices of rank r, and their SD is

$$\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{V}(\boldsymbol{S}\boldsymbol{S}^*)\boldsymbol{V}^* = \sum_{j=1}^{r} s_j^2 \mathbf{v}_j \mathbf{v}_j^* \quad \text{and} \quad \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{U}(\boldsymbol{S}^*\boldsymbol{S})\boldsymbol{U}^* = \sum_{j=1}^{r} s_j^2 \mathbf{u}_j \mathbf{u}_j^*,$$

where the diagonal matrices $\boldsymbol{S}\boldsymbol{S}^*$ and $\boldsymbol{S}^*\boldsymbol{S}$ both contain the numbers $s_1^2, \ldots, s_r^2$ in the leading positions of their main diagonals as non-zero eigenvalues.

These facts together also imply that the only positive singular value of a sub-unitary matrix is the 1 with multiplicity of its rank.

By means of SD and SVD we are able to define so-called *generalized inverses* of singular quadratic or rectangular matrices.

Definition 3. *The $m \times n$ complex matrix $\boldsymbol{X}$ is a generalized inverse of the $n \times m$ complex matrix $\boldsymbol{A}$ if $\boldsymbol{A}\boldsymbol{X}\boldsymbol{A} = \boldsymbol{A}$.*

A generalized inverse $\boldsymbol{X}$ satisfying $\boldsymbol{A}\boldsymbol{X}\boldsymbol{A} = \boldsymbol{A}$ is denoted by $\boldsymbol{A}^-$. In fact, any matrix that undoes the effect of the underlying linear transformation between the ranges of $\boldsymbol{A}^*$ and $\boldsymbol{A}$ will do. A generalized inverse is far not unique as any transformation operating on the kernels can be added. However, the following *pseudoinverse (Moore–Penrose inverse)* is unique and, in case of a quadratic matrix, it coincides with the usual inverse if exists.

Definition 4. *The $m \times n$ complex matrix $\boldsymbol{X}$ is the pseudoinverse (in other words, the Moore–Penrose inverse) of the $n \times m$ complex matrix $\boldsymbol{A}$ if it satisfies all of the following conditions:*

$$\begin{aligned} \boldsymbol{A}\boldsymbol{X}\boldsymbol{A} &= \boldsymbol{A}, \\ \mathbf{X}\boldsymbol{A}\boldsymbol{X} &= \boldsymbol{X}, \\ (\boldsymbol{A}\boldsymbol{X})^* &= \boldsymbol{A}\boldsymbol{X}, \\ (\boldsymbol{X}\boldsymbol{A})^* &= \boldsymbol{X}\boldsymbol{A}. \end{aligned}$$

It can be proven that there uniquely exists a pseudoinverse satisfying the conditions in the above definition, and it is denoted by $\boldsymbol{A}^+$. Actually, it can be obtained from the SVD (4) of $\boldsymbol{A}$ as follows:

$$\boldsymbol{A}^+ = \boldsymbol{U}\boldsymbol{S}^+\boldsymbol{V}^* = \sum_{j=1}^{r} \frac{1}{s_j} \mathbf{u}_j \mathbf{v}_j^*, \tag{5}$$

where $\boldsymbol{S}^+$ is the $m \times n$ generalized diagonal matrix containing the reciprocals of the non-zero singular values, otherwise zeros, in its main diagonal.

In particular, the Moore–Penrose inverse of the $n \times n$ self-adjoint matrix with SD (1) is

$$\boldsymbol{A}^+ = \sum_{j=1}^{r} \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^* = \boldsymbol{U} \boldsymbol{\Lambda}^+ \boldsymbol{U}^*,$$

where $\boldsymbol{\Lambda}^+ = \operatorname{diag}(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_r}, 0, \ldots, 0)$ is the diagonal matrix containing the reciprocals of the non-zero eigenvalues, otherwise zeros, in its main diagonal.

Note that any analytic function f of the self-adjoint matrix $\boldsymbol{A}$ can be defined by its SD, $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^*$, in the following way:

$$f(\boldsymbol{A}) := \boldsymbol{U} f(\boldsymbol{\Lambda}) \boldsymbol{U}^* \tag{6}$$

where $f(\boldsymbol{\Lambda}) = \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n))$, of course, only if every eigenvalue is in the domain of f. In this way, for a positive semidefinite $\boldsymbol{A}$, its square root is

$$\boldsymbol{A}^{1/2} := \boldsymbol{U} \boldsymbol{\Lambda}^{1/2} \boldsymbol{U}^*, \tag{7}$$

and for a regular $\boldsymbol{A}$ its inverse is obtained by applying the $f(x) = x^{-1}$ function to it:

$$\boldsymbol{A}^{-1} = \boldsymbol{U} \boldsymbol{\Lambda}^{-1} \boldsymbol{U}^*.$$

For a singular $\boldsymbol{A}$, the Moore–Penrose inverse is obtained by using $\boldsymbol{\Lambda}^+$ instead of $\boldsymbol{\Lambda}^{-1}$. Accordingly, for a positive definite matrix, its $-1/2$ power is defined as the square root of $\boldsymbol{A}^{-1}$.

Now a special type of matrices is introduced.

Definition 5. *We say that the $n \times n$ self-adjoint complex matrix $\boldsymbol{G} = (g_{ij})$ is a* Gram-matrix (Gramian) *if its entries are inner products; i.e., there is a dimension $d > 0$ and vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^d$ such that*

$$g_{jk} = \mathbf{x}_j^* \mathbf{x}_k, \quad j, k = 1, \ldots n.$$

Proposition 1. *The self-adjoint matrix $\boldsymbol{G}$ is a Gramian if and only if it is positive semidefinite.*

Proof. If $\boldsymbol{G}$ is a Gram-matrix, then it can be decomposed as $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{A}^*$, where $\boldsymbol{A}^* = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ with its generating vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^d$. Therefore,

$$\mathbf{x}^* \boldsymbol{G} \mathbf{x} = \mathbf{x}^* \boldsymbol{A} \boldsymbol{A}^* \mathbf{x} = (\boldsymbol{A}^* \mathbf{x})^* (\boldsymbol{A}^* \mathbf{x}) = \|\boldsymbol{A}^* \mathbf{x}\|^2 \geq 0, \quad \forall \mathbf{x} \in \mathbb{C}^n.$$

Conversely, if $\boldsymbol{G}$ is positive semidefinite with rank $r \leq n$, then its SD – using (1) – can be written as $\boldsymbol{G} = \sum_{j=1}^{r} \lambda_j \mathbf{u}_j \mathbf{u}_j^*$ with its positive real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_r > 0$. Let the $n \times r$ matrix $\boldsymbol{A}$ be defined as follows:

$$\boldsymbol{A} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Lambda}}^{1/2} = [\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_r}\mathbf{u}_r], \tag{8}$$

where $\tilde{\boldsymbol{U}}$ and $\tilde{\boldsymbol{\Lambda}}$ are as defined in Definition 2.

Then the row vectors of the matrix $\boldsymbol{A}$ will be the vectors $\mathbf{x}_j \in \mathbb{C}^r$ ($j = 1, \ldots n$) reproducing $\boldsymbol{G}$. Of course, the decomposition $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{A}^*$ is far not unique: first of all, instead of $\boldsymbol{A}$ the matrix $\boldsymbol{A}\boldsymbol{Q}$ will also do, where $\boldsymbol{Q}$ is an arbitrary $r \times r$ unitary matrix (obviously, $\mathbf{x}_j$'s can be rotated); and $\mathbf{x}_j$'s can also be put in a higher ($d > r$) dimension with attaching any (but the same) number of zero coordinates to them. □

Now matrix norms are summarized.

Definition 6. *The* spectral norm *(or* operator norm*) of an $m \times n$ complex matrix $\boldsymbol{A}$ of rank r, with singular values $s_1 \geq \cdots \geq s_r > 0$, is*

$$\|\boldsymbol{A}\| := \max_{|\mathbf{x}|=1} |\boldsymbol{A}\mathbf{x}| = s_1,$$

where $|\cdot|$ is the Euclidean (L^2) norm. Then for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we have $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|$.

The Frobenius norm*, denoted by $\|.\|_F$, is*

$$\|\boldsymbol{A}\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^*)} = \sqrt{\operatorname{tr}(\boldsymbol{A}^*\boldsymbol{A})} = \left(\sum_{i=1}^{r} s_i^2 \right)^{1/2}.$$

For a self-adjoint matrix $\boldsymbol{A}$,

$$\|\boldsymbol{A}\| = \max_{|\mathbf{x}|=1} |\boldsymbol{A}\mathbf{x}| = \max_i |\lambda_i| \quad \text{and} \quad \|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{r} \lambda_i^2 \right)^{1/2}.$$

Obviously, for a matrix $\boldsymbol{A}$ of rank r,

$$\|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_F \leq \sqrt{r}\|\boldsymbol{A}\|. \tag{9}$$

More generally, a matrix norm is called *unitary invariant* if

$$\|\boldsymbol{A}\|_{\text{un}} = \|\boldsymbol{Q}\boldsymbol{A}\boldsymbol{R}\|_{\text{un}}$$

with any $m \times m$ and $n \times n$ unitary matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively. It is easy to see that a unitary invariant norm of a matrix merely depends on its singular values (or eigenvalues if it is self-adjoint). For example, the spectral and Frobenius norms are such.

Next, the spectral radius of a quadratic matrix is defined and related to its so-called natural norms.

Definition 7. *The* spectrum $\sigma(\boldsymbol{A})$ *of the matrix* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *is the set of all its eigenvalues* λ_j, $j = 1, \ldots, n$. *The* spectral radius *of the matrix* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *is*

$$\rho(\boldsymbol{A}) = \max\{|\lambda| : \lambda \in \sigma(\boldsymbol{A})\}.$$

Note that for a self-adjoint matrix $\rho(\boldsymbol{A}) = \|\boldsymbol{A}\|$, *where* $\|\boldsymbol{A}\|$ *is the spectral norm of* $\boldsymbol{A}$.

Definition 8. *A* natural matrix norm *(or* matrix norm induced by a vector norm*) of a matrix* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *is defined as*

$$\|\boldsymbol{A}\|_p = \max_{\|\mathbf{x}\|_p=1} \|\boldsymbol{A}\mathbf{x}\|_p,$$

where $\|\mathbf{x}\|_p$ *can be any* L^p *vector norm in* $\mathbb{C}^n$, $1 \le p \le \infty$.

Note that $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\|$ is the previous spectral norm.

Lemma 1 (Rózsa Pál)**.** *Between the spectral radius and any natural norm of the quadratic, complex matrix* $\boldsymbol{A}$, *the relation*

$$\rho(\boldsymbol{A}) \le \|\boldsymbol{A}\|_p$$

holds.

Proof. Let $\lambda^* := \rho(\boldsymbol{A})$, and let $\mathbf{x}^*$ denote a unit-norm eigenvector corresponding to λ^*. Then

$$\|\boldsymbol{A}\|_p = \max_{\|\mathbf{x}\|_p=1} \|\boldsymbol{A}\mathbf{x}\|_p \ge \|\boldsymbol{A}\mathbf{x}^*\|_p = \|\lambda^*\mathbf{x}^*\|_p = \rho(\boldsymbol{A})\,\|\mathbf{x}^*\|_p = \rho(\boldsymbol{A}),$$

and that proves the lemma. □

Here we quote some important facts about spectral radius. These can be proved using the Jordan normal form of $\boldsymbol{A}$, but here we omit the proof.

Lemma 2. *For any matrix* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *and its spectral norm* $\|\boldsymbol{A}\| = \|\boldsymbol{A}\|_2$ *we have*

1. $\rho(\boldsymbol{A}) = \lim_{k\to\infty} \|\boldsymbol{A}^k\|^{1/k} = \inf_{k\geq 1} \|\boldsymbol{A}^k\|^{1/k}$;
2. *if* $\rho(\boldsymbol{A}) < 1$, *then for any* c, $\rho(\boldsymbol{A}) < c < 1$, *there exists a constant* K *such that*
$$\|\boldsymbol{A}^j\| \leq Kc^j \quad (j \geq 0);$$
3. $\rho(\boldsymbol{A}) < 1 \quad \Leftrightarrow \quad \lim_{k\to\infty} \boldsymbol{A}^k = 0$;
4. $\rho(\boldsymbol{A}) > 1 \quad \Leftrightarrow \quad \lim_{k\to\infty} \|\boldsymbol{A}^k\| = \infty$;
5. $\rho(\boldsymbol{A}) = 1 \quad \Rightarrow \quad \|\boldsymbol{A}^k\| \geq 1$ *for any* $k \geq 1$.

It is obvious that all the complex eigenvalues of an $n \times n$ complex matrix $\boldsymbol{A}$ are within the closed circle of radius $\rho(\boldsymbol{A})$ around the origin of the complex plane. However, the subsequent Gersgorin disc theorem gives a finer allocation of them.

Theorem 3 (Gersgorin disc theorem)**.** *Let* $\boldsymbol{A}$ *be an* $n \times n$ *matrix of entries* $a_{ij} \in \mathbb{C}$. *The Gersgorin disks of* $\boldsymbol{A}$ *are the following regions of the complex plane:*
$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j\neq i} |a_{ij}|\}, \quad i = 1, \ldots n.$$
Let $\lambda_1, \ldots, \lambda_n$ *denote the eigenvalues of* $\boldsymbol{A}$. *Then*
$$\{\lambda_1, \ldots, \lambda_n\} \subset \cup_{i=1}^n D_i.$$
Furthermore, any connected component of the set $\cup_{i=1}^n D_i$ *contains as many eigenvalues of* $\boldsymbol{A}$ *as the number of discs that form this component.*

Theorem 4 (Cayley–Hamilton theorem)**.** *For any* $n \times n$ *complex matrix* $\boldsymbol{A}$, $p_n(\boldsymbol{A}) = \boldsymbol{O}$ *(the* $n \times n$ *zero matrix), where* p_n *is the characteristic polynomial of* $\boldsymbol{A}$, *i.e.,* $p_n(z) = |\boldsymbol{A} - z\boldsymbol{I}|$ *(nth degree polynomial of* z*).*

Now some perturbation results follow for self-adjoint matrices,

Theorem 5 (Weyl perturbation theorem)**.** *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be* $n \times n$ *self-adjoint matrices. Then*
$$|\lambda_j(\boldsymbol{A}) - \lambda_j(\boldsymbol{B})| \leq \|\boldsymbol{A} - \boldsymbol{B}\|, \quad j = 1, \ldots, n$$
in spectral norm, where the eigenvalues of $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are enumerated in non-increasing order.*

A theorem for the perturbation of spectral subspaces (sometimes called Davis–Kahan theorem) is stated here for self-adjoint matrices.

Theorem 6. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be self-adjoint matrices; S_1 and S_2 are subsets of $\mathbb{R}$ such that $dist(S_1, S_2) = \delta > 0$. Let $\boldsymbol{P_A}(S_1)$ and $\boldsymbol{P_B}(S_2)$ be orthogonal projections onto the subspace spanned by the eigenvectors of the matrix corresponding to the eigenvalues within the subset in the lower index. Then with any unitary invariant norm:*

$$\|\boldsymbol{P_A}(S_1)\boldsymbol{P_B}(S_2)\| \leq \frac{c}{\delta}\|\boldsymbol{A} - \boldsymbol{B}\|$$

where c is a constant.

The statement is true for any unitary invariant norm. In case of the Frobenius norm, $c = 1$ will always do.

We also need the following simple lemma.

Lemma 3. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are self-adjoint, positive semidefinite quadratic matrices of the same size, then $\boldsymbol{AB}$ has real nonnegative eigenvalues.*

Proof. Though $\boldsymbol{AB}$ is usually not self-adjoint, it is still diagonalizable as follows. The eigenvalue–eigenvector equation for the matrix $\boldsymbol{AB}$ is:

$$\boldsymbol{AB}\mathbf{x} = \lambda\mathbf{x}$$

that is equivalent to

$$(\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2})(\boldsymbol{A}^{-1/2}\mathbf{x}) = \lambda(\boldsymbol{A}^{-1/2}\mathbf{x}),$$

where $\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2} = (\boldsymbol{A}^{1/2}\boldsymbol{B}^{1/2})(\boldsymbol{A}^{1/2}\boldsymbol{B}^{1/2})^*$ is a Gram matrix (Definition 5), so positive semidefinite. Each of its nonnegative real eigenvalue λ is also an eigenvalue of $\boldsymbol{AB}$ with eigenvector that is obtained with premultiplying its eigenvector with $\boldsymbol{A}^{1/2}$. □

Definition 9. *The quadratic matrix $\boldsymbol{A} = [a_{jk}]$ is of Toeplitz type if it has the same entries along its main diagonal and along all lines parallel to the main diagonal. In other words, the value of the entry a_{jk} depends only on $|j - k|$, $\forall j, k$.*

Definition 10. *The quadratic matrix $\boldsymbol{A} = [a_{jk}]$ is of Hankel type if it has the same entries along its anti-diagonal and along all lines parallel to its anti-diagonal. In other words, the value of the entry a_{jk} depends only on $j + k$, $\forall j, k$.*

Block Toeplitz and block Hankel matrices are defined analogously: they are of Toeplitz and of Hankel type in terms of their blocks considered as entries, respectively.

Without proofs, we enlist some notable matrix decompositions, (see Golub-=Van Loan).

- The *Gram-decomposition* of the self-adjoint, positive semidefinite, $n \times n$ matrix $\boldsymbol{G}$ is the decomposition $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{A}^*$ in the proof of Proposition 1. As we saw, it is not unique, and the minimal size of $\boldsymbol{A}$ is $n \times r$, where $r = \operatorname{rank}(\boldsymbol{A})$.

  The Gram-decomposition $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{A}^*$ with the $\boldsymbol{A}$ of equation (8) is called *parsimonious Gram-decomposition.*

- The *QR-decomposition* of a complex $m \times n$ matrix $\boldsymbol{A}$ is

  $$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R},$$

  where the matrix $\boldsymbol{Q}$ is $m \times m$ unitary, whereas $\boldsymbol{R}$ is $m \times n$ generalized upper triangular matrix (there are 0 entries below its main diagonal, starting at its upper left corner). This (not necessarily unique) decomposition always exists, and can be derived by applying the Gram–Schmidt orthogonalization procedure to the column vectors of $\boldsymbol{A}$.

  The related QR-transformation (Francis, 1961) uses the QR-decomposition for an iteration converging to $\boldsymbol{\Lambda}$ of the SD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ of the self-adjoint matrix $\boldsymbol{A}$. The iteration is as follows:

  $$\boldsymbol{A}_0 = \boldsymbol{A}, \quad \boldsymbol{Q}_0 = \boldsymbol{Q}, \quad \boldsymbol{R}_0 = \boldsymbol{R},$$

  and for $t = 1, 2, \ldots$, if $\boldsymbol{A}_{t-1} = \boldsymbol{Q}_{t-1}\boldsymbol{R}_{t-1}$, then $\boldsymbol{A}_t := \boldsymbol{R}_{t-1}\boldsymbol{Q}_{t-1}$. Then $\lim_{t\to\infty} \boldsymbol{A}_t = \boldsymbol{\Lambda}$ in L^2-norm.

- The *LDL-decomposition* of the complex $n \times n$ self-adjoint matrix $\boldsymbol{A}$ is

  $$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^*,$$

  where $\boldsymbol{L}$ is $n \times n$ lower triangular with 1s along its main diagonal, and $\boldsymbol{D}$ is $n \times n$ diagonal matrix (with nonnegative diagonal entries).

  Moreover, the LDL-decomposition is nested in the following sense: if $\boldsymbol{A}_k$, $\boldsymbol{L}_k$, and $\boldsymbol{D}_k$ denote the $k \times k$ submatrices of the underlying matrices formed by their first k rows and columns, then

  $$\boldsymbol{A}_k = \boldsymbol{L}_k\boldsymbol{D}_k\boldsymbol{L}_k^*$$

is also LDL-decomposition for $k = 1, 2, \ldots, n$.

- The *LU-decomposition* of the complex $n \times n$ matrix $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U},$$

where $\boldsymbol{L}$ is $n \times n$ lower, and $\boldsymbol{U}$ is $n \times n$ upper triangular matrix. It can be arranged that each diagonal entry of $\boldsymbol{L}$ is 1. This decomposition is sometimes called *Cholesky decomposition.*

If $\boldsymbol{A}$ is self-adjoint, then the LU-decomposition can be obtained from the LDL-decomposition via manipulations with the nonnegative entries of $\boldsymbol{D}$.

The related LR-transformation (Rutishauser, 1958) uses the LU-decomposition for an iteration converging to $\boldsymbol{\Lambda}$ of the SD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ of the self-adjoint matrix $\boldsymbol{A}$. Here LR denotes left-right, which is the same as LU (lower-upper). The iteration is as follows:

$$\boldsymbol{A}_0 = \boldsymbol{A} = \boldsymbol{L}\boldsymbol{R}, \quad \boldsymbol{L}_0 = \boldsymbol{L}, \quad \boldsymbol{R}_0 = \boldsymbol{R},$$

and for $t = 1, 2, \ldots$, if $\boldsymbol{A}_{t-1} = \boldsymbol{L}_{t-1}\boldsymbol{R}_{t-1}$, then $\boldsymbol{A}_t := \boldsymbol{R}_{t-1}\boldsymbol{L}_{t-1}$. Eventually, $\lim_{t\to\infty} \boldsymbol{R}_t = \lim_{t\to\infty} \boldsymbol{A}_t = \boldsymbol{\Lambda}$ and $\lim_{t\to\infty} \boldsymbol{L}_t = \boldsymbol{I}_n$ in L^2-norm.

Note that the above matrix decompositions work for block-matrices similarly. The computational complexity is increased with the understanding that here matrix multiplications are substituted for entry-wise multiplications.

Block-matrices sometimes arise as Kronecker-products.

Definition 11. *Let $\boldsymbol{A}$ be $p \times n$ and $\boldsymbol{B}$ be $q \times m$ complex matrix. Their Kronecker-product, denoted by $\boldsymbol{A} \otimes \boldsymbol{B}$, is the following $pq \times nm$ block-matrix: it has p block rows and n block columns; each block is a $q \times m$ matrix such that the block indexed by (j, k) is the matrix $a_{jk}\boldsymbol{B}$ $(j = 1, \ldots, p;\ k = 1, \ldots, n)$.*

This product is associative, for the addition distributive, but usually not commutative. If $\boldsymbol{A}$ is $n \times n$ and $\boldsymbol{B}$ is $m \times m$ quadratic matrix, then

$$\det(\boldsymbol{A} \otimes \boldsymbol{B}) = (\det \boldsymbol{A})^m \cdot (\det \boldsymbol{B})^n;$$

further, if both are regular, then so is their Kronecker-product. Namely,

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}.$$

It is also useful to know that – provided $\boldsymbol{A}$ and $\boldsymbol{B}$ are self-adjoint – the spectrum of $\boldsymbol{A} \otimes \boldsymbol{B}$ consists of the real numbers

$$\alpha_j \beta_k \quad (j = 1, \ldots, n;\, k = 1, \ldots, m),$$

where α_j's and β_k's are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Definition 11 naturally extends to vectors: the Kronecker-product of vectors $\mathbf{a} \in \mathbb{C}^n$ and $\mathbf{b} \in \mathbb{C}^m$ is a vector $\mathbf{a} \otimes \mathbf{b} \in \mathbb{C}^{nm}$.

The eigenvalues of other types of block-matrices are characterized in the following theorem.

Theorem 7 (Theorem 5.3.1 in Rózsa)**.** *Let $\boldsymbol{A}$ be a $d \times d$ self-adjoint matrix with spectral decomposition*

$$\boldsymbol{A} = \sum_{k=1}^{d} a_k \mathbf{u}_k \mathbf{u}_k^*;$$

the analytic functions $g_{ij}(z)$ for $i, j = 1, \ldots, n$ satisfy

$$g_{ij}(z) = \overline{g_{ji}(z)},$$

and the eigenvalues $a_1, \ldots, a_k$ are within the convergence region of every $g_{ij}(z)$.

Denoting the spectral decomposition of the self-adjoint matrix $[g_{ij}(a_k)]_{i,j=1}^n$ with

$$[g_{ij}(a_k)]_{i,j=1}^n = \sum_{\ell=1}^{n} \lambda_\ell^{(k)} \mathbf{v}_\ell^{(k)} \mathbf{v}_\ell^{(k)*}, \quad k = 1, \ldots, d,$$

the spectral decomposition of the $nd \times nd$ block matrix $[\boldsymbol{A}_{ij}]_{i,j=1}^n = [g_{ij}(\boldsymbol{A})]_{i,j=1}^n$ is

$$[\boldsymbol{A}_{ij}]_{i,j=1}^n = \sum_{\ell=1}^{n} \sum_{k=1}^{d} \lambda_\ell^{(k)} (\mathbf{u}_k \otimes \mathbf{v}_\ell^{(k)})(\mathbf{u}_k \otimes \mathbf{v}_\ell^{(k)})^*.$$

Note that if $\boldsymbol{A}$ is normal and the matrices $[g_{ij}(a_k)]_{i,j=1}^n$ are as well all normal, the statement also holds irrespective whether g_{ij}s are analytic.

Next we discuss low rank approximations of a matrix.

Theorem 8. *Let $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ with SVD $\boldsymbol{A} = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^*$, where r is the rank of $\boldsymbol{A}$ and $s_1 \geq \cdots \geq s_r > 0$. Then for any $1 \leq k < r$ such that $s_k > s_{k+1}$ we have*

$$\min \|\boldsymbol{A} - \boldsymbol{B}\| = s_{k+1} \quad \textit{and} \quad \min \|\boldsymbol{A} - \boldsymbol{B}\|_F = \left(\sum_{i=k+1}^{r} s_i^2 \right)^{1/2},$$

where the minima are taken for all matrices $\boldsymbol{B} \in \mathbb{C}^{m\times n}$ of rank k. Both minima are attained with the matrix $\boldsymbol{B} = \boldsymbol{A}_k := \sum_{i=1}^{k} s_i \mathbf{u}_i \mathbf{v}_i^*$.

Note that $\boldsymbol{A}_k$ is called the *best rank k approximation* of $\boldsymbol{A}$, and the aforementioned theorem guarantees that it is the best approximation both in spectral and Frobenius norm. In fact, it is true for any unitary invariant norm:

$$\min_{\substack{\boldsymbol{B}\,\text{is}\,m\times n \\ \operatorname{rank}(\boldsymbol{B})=k}} \|\boldsymbol{A} - \boldsymbol{B}\|_{\text{un}} = \|\boldsymbol{A} - \boldsymbol{A}_k\|_{\text{un}}.$$

Corollary 1. *Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ be a self-adjoint, positive semidefinite matrix with SD $\boldsymbol{A} = \sum_{j=1}^{r} \lambda_j \mathbf{u}_j \mathbf{u}_j^*$, where r is the rank of $\boldsymbol{A}$ and the eigenvalues are $\lambda_1 \geq \cdots \geq \lambda_r > 0$. Then for any $1 \leq k < r$ such that $\lambda_k > \lambda_{k+1}$ we have*

$$\min \|\boldsymbol{A} - \boldsymbol{B}\| = \lambda_{k+1} \quad \textit{and} \quad \min \|\boldsymbol{A} - \boldsymbol{B}\|_F = \left(\sum_{i=k+1}^{r} \lambda_i^2 \right)^{1/2},$$

where the minima are taken for all self-adjoint, positive semidefinite matrices $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ of rank k. Both minima are attained with the best rank k approximation

$$\boldsymbol{A}_k = \sum_{j=1}^{k} \lambda_j \mathbf{u}_j \mathbf{u}_j^* = \tilde{\boldsymbol{U}}_k \tilde{\boldsymbol{\Lambda}}_k \tilde{\boldsymbol{U}}_k^*$$

of $\boldsymbol{A}$, where $\tilde{\boldsymbol{U}}_k = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ is $n \times k$ sub-unitary and $\tilde{\boldsymbol{\Lambda}}_k = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ is $k \times k$ diagonal matrix.

In particular,

$$\boldsymbol{A} = \boldsymbol{A}_r = \tilde{\boldsymbol{U}}_r \tilde{\boldsymbol{\Lambda}}_r \tilde{\boldsymbol{U}}_r^*,$$

of which we can take the square-root:

$$\boldsymbol{A}^{1/2} = \tilde{\boldsymbol{U}}_r \tilde{\boldsymbol{\Lambda}}_r^{1/2} \tilde{\boldsymbol{U}}_r^*,$$

where $\tilde{\mathbf{\Lambda}}_r^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r})$. However, with any matrix $\boldsymbol{M} = \tilde{\boldsymbol{U}}_r \tilde{\mathbf{\Lambda}}_r^{1/2} \boldsymbol{Q}$, where $\boldsymbol{Q}$ is $r \times r$ unitary, the decomposition $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{M}^*$ holds. In particular it holds with $\boldsymbol{M} = \tilde{\boldsymbol{U}}_r \tilde{\mathbf{\Lambda}}_r^{1/2} = [\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_r}\mathbf{u}_r]$, see the parsimonious Gram-decomposition (8).

Theorem 8 is proved for self-adjoint matrices by C. R. Rao with the Frobenius norm, but it can be easily extended to any unitary invariant norm.