holds for an arbitrary orthonormal system $\{\boldsymbol{u}_i\}$. From $(\boldsymbol{x}, \sum_{i=1}^n \boldsymbol{u}_i\boldsymbol{u}_i^\top \boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{I}\boldsymbol{x})$, we obtain the following identity for an arbitrary vector and an arbitrary orthonormal system $\{\boldsymbol{u}_i\}$:

$$\sum_{i=1}^n (\boldsymbol{u}_i, \boldsymbol{x})^2 = \|\boldsymbol{x}\|^2. \tag{2.64}$$

Let $\{\lambda_i\}$ be the eigenvalues of (nn)-matrix $\boldsymbol{A}$, and $\{\boldsymbol{u}_i\}$ the corresponding eigensystem. Since $\{\boldsymbol{u}_i\}$ is an orthonormal system, the matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n)$ is orthogonal. Eq. (2.62) is equivalent to

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top, \tag{2.65}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix with diagonal elements λ_1, λ_2, ..., λ_n in that order; we write

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n). \tag{2.66}$$

From eq. (2.65), we obtain

$$\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}, \tag{2.67}$$

which is called the *diagonalization* of $\boldsymbol{A}$. Applying the fourth of eqs. (2.2) and eq. (2.12) to eq. (2.65), we obtain the following identities:

$$\mathrm{tr}\boldsymbol{A} = \sum_{i=1}^n \lambda_i, \qquad \det \boldsymbol{A} = \prod_{i=1}^n \lambda_i. \tag{2.68}$$

From the spectral decomposition (2.62), the kth power $\boldsymbol{A}^k$ for an arbitrary integer $k > 0$ is given by

$$\boldsymbol{A}^k = \sum_{i=1}^n \lambda_i{}^k \boldsymbol{u}_i\boldsymbol{u}_i^\top. \tag{2.69}$$

This can be extended to an arbitrary polynomial $p(x)$:

$$p(\boldsymbol{A}) = \sum_{i=1}^n p(\lambda_i)\boldsymbol{u}_i\boldsymbol{u}_i^\top. \tag{2.70}$$

If $\boldsymbol{A}$ is of full rank, its inverse $\boldsymbol{A}^{-1}$ is given by

$$\boldsymbol{A}^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i}\boldsymbol{u}_i\boldsymbol{u}_i^\top. \tag{2.71}$$

This can be extended to an arbitrary negative power of $\boldsymbol{A}$ (see the third of eqs. (2.21)):

$$\boldsymbol{A}^{-k} = \sum_{i=1}^n \frac{1}{\lambda_i{}^k}\boldsymbol{u}_i\boldsymbol{u}_i^\top. \tag{2.72}$$

2.2.2 *Generalized inverse*

An (nn)-matrix $\boldsymbol{A}$ is *positive definite* if its eigenvalues are all positive, and is *positive semi-definite* if its eigenvalues are all nonnegative; it is *negative definite* if its eigenvalues are all negative, and is *negative semi-definite* if its eigenvalues are all nonpositive.

For a positive semi-definite (nn)-matrix $\boldsymbol{A}$, eq. (2.69) can be extended to arbitrary non-integer powers $\boldsymbol{A}^q$, $q > 0$. In particular, the "square root" $\sqrt{\boldsymbol{A}}$ of $\boldsymbol{A}$ is defined by

$$\sqrt{\boldsymbol{A}} = \sum_{i=1}^{n} \sqrt{\lambda_i} \boldsymbol{u}_i \boldsymbol{u}_i^\top. \tag{2.73}$$

It is easy to see that $(\sqrt{\boldsymbol{A}})^2 = \boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, eq. (2.69) can be extended to arbitrary negative non-integer powers such as $\boldsymbol{A}^{-2/3}$.

Let $\{\boldsymbol{r}_1, ..., \boldsymbol{r}_l\}_L$ denote the linear subspace *spanned* (or *generated*) by $\boldsymbol{r}_1$, ..., $\boldsymbol{r}_l$, i.e., the set of all vectors that can be expressed as a linear combination $\sum_{i=1}^{l} c_i \boldsymbol{r}_i$ for some real numbers c_1, ..., c_l. A positive semi-definite (nn)-matrix of rank r $(\leq n)$ has the following spectral decomposition:

$$\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top, \qquad \lambda_i > 0, \quad i = 1, ..., r. \tag{2.74}$$

Let the symbol $\mathcal{R}^n$ denote the n-dimensional space of all n-vectors. The r-dimensional subspace

$$\mathcal{R}_{\boldsymbol{A}} = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_r\}_L \subset \mathcal{R}^n \tag{2.75}$$

is called the *range* (or *image space*) of $\boldsymbol{A}$, for which the set $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_r\}$ is an orthonormal basis. The $(n-r)$-dimensional subspace

$$\mathcal{N}_{\boldsymbol{A}} = \{\boldsymbol{u}_{r+1}, ..., \boldsymbol{u}_n\}_L \subset \mathcal{R}^n \tag{2.76}$$

is called the *null space* of $\boldsymbol{A}$, for which the set $\{\boldsymbol{u}_{r+1}, ..., \boldsymbol{u}_n\}$ is an orthonormal basis. The n-dimensional space is the direct sum of $\mathcal{R}_{\boldsymbol{A}}$ and $\mathcal{N}_{\boldsymbol{A}}$, each being the orthogonal complement of the other:

$$\mathcal{R}^n = \mathcal{R}_{\boldsymbol{A}} \oplus \mathcal{N}_{\boldsymbol{A}}, \qquad \mathcal{R}_{\boldsymbol{A}} \perp \mathcal{N}_{\boldsymbol{A}}. \tag{2.77}$$

This definition implies

$$\boldsymbol{P}_{\mathcal{N}_A} \boldsymbol{A} = \boldsymbol{A} \boldsymbol{P}_{\mathcal{N}_A} = \boldsymbol{A}. \tag{2.78}$$

The *(Moore-Penrose) generalized* (or *pseudo*) *inverse*[10] $\boldsymbol{A}^-$ of $\boldsymbol{A}$ is defined

[10]The Moore-Penrose generalized inverse is often denoted by $\mathbf{A}^+$ in order to distinguish it from the generalized inverse in general, which is defined as the matrix $\mathbf{X}$ that satisfies $\mathbf{AXA} = \mathbf{A}$ and denoted by $\mathbf{A}^-$. The generalized inverse we use throughout this book is always the Moore-Penrose type, so we adopt the generic symbol $\mathbf{A}^-$. The symbol $\mathbf{A}^+$ will be given another meaning (see Section 2.2.6).

by

pseudo-inverse rank r

$$A^- = \sum_{i=1}^{r} \frac{1}{\lambda_i} u_i u_i^\top .$$

with SVD, not much harder for general ***A*** (2.121)

Evidently, the generalized inverse A^- coincides with the inverse A^{-1} if A is of full rank. From this definition, the following relationships are obtained (see eqs. (2.50) and (2.63)):

$$(A^-)^- = A, \qquad P_{\mathcal{N}_A} A^- = A^- P_{\mathcal{N}_A} = A^-,$$
$$A^- A = A A^- = P_{\mathcal{N}_A}. \tag{2.80}$$

From eqs. (2.78) and (2.80), we obtain

$$A A^- A = A, \qquad A^- A A^- = A^-. \tag{2.81}$$

The rank and the generalized inverse of a matrix are well defined concepts in a mathematical sense only; it rarely occurs in finite precision numerical computation that some eigenvalues are precisely zero. In computing the generalized inverse numerically, the rank of the matrix should be predicted by a theoretical analysis first. Then, the matrix should be modified so that it has the desired rank. Let A be a positive semi-definite (nn)-matrix of rank r; let $A = \sum_{i=1}^{r} \lambda_i u_i u_i^\top$, $\lambda_1 \geq \cdots \geq \lambda_r > 0$, be its spectral decomposition. Its *rank-constrained generalized inverse* $(A)_{r'}^-$ of rank r' $(\leq r)$ is defined by

$$(A)_{r'}^- = \sum_{i=1}^{r'} \frac{1}{\lambda_i} u_i u_i^\top . \tag{2.82}$$

From this definition, the following identities are obtained:

$$(A)_{r'}^- A = A (A)_{r'}^- = P_{\mathcal{N}_{(A)_{r'}^-}}, \qquad (A)_{r'}^- A (A)_{r'}^- = (A)_{r'}^- . \tag{2.83}$$

Let A be an (nn)-matrix, and B an (mm)-matrix. Let S and T be nm-matrices. Even if A and B are not of full rank, the matrix inversion formula (2.22) holds in the form

$$(A + P_{\mathcal{N}_A} S B T^\top P_{\mathcal{N}_A})^- = A^- - A^- S (B^- + P_{\mathcal{N}_B} T^\top A^- S P_{\mathcal{N}_B})^- T^\top A^-, \tag{2.84}$$

provided that matrix $A + P_{\mathcal{N}_A} S B T^\top P_{\mathcal{N}_A}$ has the same rank as A and matrix $B^- + P_{\mathcal{N}_B} T^\top A^- S P_{\mathcal{N}_B}$ has the same rank as B^-. We call eq. (2.84) the *generalized matrix inversion formula*.

2.2.3 Rayleigh quotient and quadratic form

For an (nn)-matrix A, the expression $(u, Au)/\|u\|^2$ is called the *Rayleigh quotient* of vector u for A. Let $\lambda_{\min}$ and $\lambda_{\max}$ be, respectively, the largest and the smallest eigenvalues of $\boldsymbol{A}$. The following inequality holds for an arbitrary nonzero vector $\boldsymbol{u}$:

$$\lambda_{\min} \leq \frac{(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})}{\|\boldsymbol{u}\|^2} \leq \lambda_{\max}. \tag{2.85}$$

The left equality holds if $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ for eigenvalue $\lambda_{\min}$; the right equality holds if $\boldsymbol{u}$ is an eigenvector for eigenvalue $\lambda_{\max}$.

The Rayleigh quotient $(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})/\|\boldsymbol{u}\|^2$ is invariant to multiplication of $\boldsymbol{u}$ by a constant and hence is a function of the orientation of $\boldsymbol{u}$: if we put $\boldsymbol{n} = N[\boldsymbol{u}]$, then $(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})/\|\boldsymbol{u}\|^2 = (\boldsymbol{n}, \boldsymbol{A}\boldsymbol{n})$, which is called the *quadratic form* in $\boldsymbol{n}$ for $\boldsymbol{A}$. Eq. (2.85) implies

$$\min_{\|\mathbf{n}\|=1} (\boldsymbol{n}, \boldsymbol{A}\boldsymbol{n}) = \lambda_{\min}, \qquad \max_{\|\mathbf{n}\|=1} (\boldsymbol{n}, \boldsymbol{A}\boldsymbol{n}) = \lambda_{\max}. \tag{2.86}$$

The minimum is attained by any unit eigenvector $\boldsymbol{n}$ of $\boldsymbol{A}$ for eigenvalue $\lambda_{\min}$; the maximum is attained by any unit eigenvector $\boldsymbol{n}$ for eigenvalue $\lambda_{\max}$. It follows that an (nn)-matrix $\boldsymbol{A}$ is positive definite if and only if $(\boldsymbol{r}, \boldsymbol{A}\boldsymbol{r}) > 0$ for an arbitrary nonzero vector $\boldsymbol{r}$; it is positive semi-definite if and only if $(\boldsymbol{r}, \boldsymbol{A}\boldsymbol{r}) \geq 0$ for an arbitrary n-vector $\boldsymbol{r}$.

For an arbitrary mn-matrix $\boldsymbol{B}$, the matrix $\boldsymbol{B}^\top\boldsymbol{B}$ is symmetric (see the second of eq. (2.2)). It is also positive semi-definite since $(\boldsymbol{r}, \boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{r}) = \|\boldsymbol{B}\boldsymbol{r}\|^2 \geq 0$ for an arbitrary n-vector $\boldsymbol{r}$. If $\boldsymbol{B}$ is an nn-matrix of full rank, equality holds if and only if $\boldsymbol{r} = \boldsymbol{0}$. For an (nn)-matrix $\boldsymbol{A}$, its square root $\sqrt{\boldsymbol{A}}$ is also symmetric (see eq. (2.73)). We can also write $\boldsymbol{A} = \sqrt{\boldsymbol{A}}^\top\sqrt{\boldsymbol{A}}$. From these observations, we conclude the following:

- Matrix $\boldsymbol{A}$ is positive semi-definite if and only if there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}^\top\boldsymbol{B}$.
- Matrix $\boldsymbol{A}$ is positive definite if and only if there exists a nonsingular matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}^\top\boldsymbol{B}$.
- If $\boldsymbol{A}$ is a positive semi-definite (nn)-matrix, matrix $\boldsymbol{B}^\top\boldsymbol{A}\boldsymbol{B}$ is a positive semi-definite (mm)-matrix for any nm-matrix $\boldsymbol{B}$.

2.2.4 Nonsingular generalized eigenvalue problem

Let $\boldsymbol{A}$ be an (nn)-matrix, and $\boldsymbol{G}$ a positive semi-definite (nn)-matrix. If there exists a nonzero vector $\boldsymbol{u}$ and a scalar λ such that

$$\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{G}\boldsymbol{u}, \tag{2.87}$$

the scalar λ is called the *generalized eigenvalue* of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$; the vector $\boldsymbol{u}$ is called the corresponding *generalized eigenvector*. The problem of computing such $\boldsymbol{u}$ and λ is said to be *nonsingular* if $\boldsymbol{G}$ is of full rank, and *singular* otherwise.

Consider the nonsingular generalized eigenvalue problem. Eq. (2.87) can be rewritten as $(\lambda\boldsymbol{G} - \boldsymbol{A})\boldsymbol{u} = \boldsymbol{0}$, which has a nonzero solution $\boldsymbol{u}$ if and only if function

$$\phi_{\boldsymbol{A},\boldsymbol{G}}(\lambda) = |\lambda\boldsymbol{G} - \boldsymbol{A}| \tag{2.88}$$

has a zero: $\phi_{\boldsymbol{A},\boldsymbol{G}}(\lambda) = 0$. The function $\phi_{\boldsymbol{A},\boldsymbol{G}}(\lambda)$ is an nth degree polynomial in λ and is called the *generalized characteristic polynomial* of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$. The equation $\phi_{\boldsymbol{A},\boldsymbol{G}}(\lambda) = 0$ is called the *generalized characteristic equation* of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$ and has n roots $\{\lambda_i\}$ (with multiplicities counted). The generalized eigenvalue problem with respect to $\boldsymbol{I}$ reduces to the usual eigenvalue problem.

The generalized eigenvalues $\{\lambda_i\}$ of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$ are all real. The corresponding generalized eigenvectors $\{\boldsymbol{u}_i\}$ can be chosen so that

$$(\boldsymbol{u}_i, \boldsymbol{G}\boldsymbol{u}_j) = \delta_{ij}, \tag{2.89}$$

which implies

$$(\boldsymbol{u}_i, \boldsymbol{A}\boldsymbol{u}_j) = \lambda_j\delta_{ij}. \tag{2.90}$$

Let us call the set $\{\boldsymbol{u}_i\}$ so defined the *generalized eigensystem* of the (nn)-matrix with respect to the positive definite (nn)-matrix $\boldsymbol{G}$. Let $\boldsymbol{U} = (\boldsymbol{u}_1, ..., \boldsymbol{u}_n)$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_n)$, respectively. Eqs. (2.89) and (2.90) can be rewritten as

$$\boldsymbol{U}^\top\boldsymbol{G}\boldsymbol{U} = \boldsymbol{I}, \qquad \boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}. \tag{2.91}$$

By multiplying the first equation by $\boldsymbol{G}\boldsymbol{U}$ from the left and $\boldsymbol{U}^\top\boldsymbol{G}$ from the right, the following *generalized spectral decomposition* is obtained:

$$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top\boldsymbol{G} = \sum_{i=1}^{n} \lambda_i(\boldsymbol{G}\boldsymbol{u}_i)(\boldsymbol{G}\boldsymbol{u}_i)^\top. \tag{2.92}$$

The number of nonzero generalized eigenvalues is equal to the rank of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, $\{\lambda_i\}$ are all positive; if $\boldsymbol{A}$ is positive semi-definite, $\{\lambda_i\}$ are all nonnegative.

The generalized eigenvalue problem $\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{G}\boldsymbol{u}$ reduces to an ordinary eigenvalue problem as follows. Let $\boldsymbol{C} = \boldsymbol{G}^{-1/2}$ and $\tilde{\boldsymbol{u}} = \boldsymbol{C}^{-1}\boldsymbol{u}$ (see eqs. (2.71) and (2.73)). It is easy to see that eq. (2.87) can be written as

$$\tilde{\boldsymbol{A}}\tilde{\boldsymbol{u}} = \lambda\tilde{\boldsymbol{u}}, \qquad \tilde{\boldsymbol{A}} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{C}. \tag{2.93}$$

If an eigenvector $\tilde{\boldsymbol{u}}$ of $\tilde{\boldsymbol{A}}$ is computed, the corresponding generalized eigenvector is given by

$$\boldsymbol{u} = \boldsymbol{C}\tilde{\boldsymbol{u}}. \tag{2.94}$$

The expression $(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})/(\boldsymbol{u}, \boldsymbol{G}\boldsymbol{u})$ for an (nn)-matrix $\boldsymbol{A}$ and a positive definite (nn)-matrix $\boldsymbol{G}$ is called the *generalized Rayleigh quotient* of $\boldsymbol{u}$. It satisfies

$$\lambda_{\min} \leq \frac{(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{G}\boldsymbol{u})} \leq \lambda_{\max}, \tag{2.95}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the largest and the smallest generalized eigenvalues of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$. The left equality holds if $\boldsymbol{u}$ is a generalized eigenvector of $\boldsymbol{A}$ for the generalized eigenvalue $\lambda_{\min}$; the right equality holds if $\boldsymbol{u}$ is a generalized eigenvector for the generalized eigenvalue $\lambda_{\max}$.

2.2.5 Singular generalized eigenvalue problem

save for when we need it

Consider the singular generalized eigenvalue problem of an (nn)-matrix $\boldsymbol{A}$ with respect to a positive semi-definite (nn)-matrix $\boldsymbol{G}$ of rank m $(< n)$. Let $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_m\}$ be an orthonormal basis of the range $\mathcal{R}_{\boldsymbol{G}}$ of $\boldsymbol{G}$, and $\{\boldsymbol{v}_{m+1}, ..., \boldsymbol{v}_n\}$ an orthonormal basis of its null space $\mathcal{N}_{\boldsymbol{G}}$. Define an nm-matrix $\boldsymbol{P}_1$ and an $n(n-m)$-matrix $\boldsymbol{P}_0$ by

$$\boldsymbol{P}_1 = (\boldsymbol{v}_1, ..., \boldsymbol{v}_m), \qquad \boldsymbol{P}_0 = (\boldsymbol{v}_{m+1}, ..., \boldsymbol{v}_n). \tag{2.96}$$

Then,

$$\boldsymbol{P}_1^\top \boldsymbol{P}_1 = \boldsymbol{I}, \qquad \boldsymbol{P}_1^\top \boldsymbol{P}_0 = \boldsymbol{O}, \qquad \boldsymbol{P}_0^\top \boldsymbol{P}_0 = \boldsymbol{I}. \tag{2.97}$$

Here, we only consider the case where $\boldsymbol{P}_0^\top \boldsymbol{A} \boldsymbol{P}_0$ is nonsingular[11]. Since $\mathcal{R}^n = \mathcal{R}_{\boldsymbol{G}} \oplus \mathcal{N}_{\boldsymbol{G}}$, an arbitrary n-vector can be uniquely written in the form

$$\boldsymbol{u} = \boldsymbol{P}_1 \boldsymbol{x} + \boldsymbol{P}_0 \boldsymbol{y}, \tag{2.98}$$

where $\boldsymbol{x}$ is an m-vector and $\boldsymbol{y}$ is an $(n-m)$-vector. Eqs. (2.97) imply that $\boldsymbol{x}$ and $\boldsymbol{y}$ are respectively given by

$$\boldsymbol{x} = \boldsymbol{P}_1^\top \boldsymbol{u}, \qquad \boldsymbol{y} = \boldsymbol{P}_0^\top \boldsymbol{u}. \tag{2.99}$$

Substituting eq. (2.98) into eq. (2.87) and noting the identities $\boldsymbol{G}\boldsymbol{P}_0 = \boldsymbol{O}$ and $\boldsymbol{P}_0^\top \boldsymbol{G} = \boldsymbol{O}$, we can split eq. (2.87) into the following two equations:

$$\boldsymbol{A}^* \boldsymbol{x} = \lambda \boldsymbol{G}^* \boldsymbol{x}, \qquad \boldsymbol{y} = \boldsymbol{B}^* \boldsymbol{x}. \tag{2.100}$$

Here, $\boldsymbol{A}^*$ and $\boldsymbol{G}^*$ are (mm)-matrices; $\boldsymbol{B}^*$ is an $(n-m)m$-matrix. They are defined by

$$\boldsymbol{A}^* = \boldsymbol{P}_1^\top \boldsymbol{A} \boldsymbol{P}_1 - \boldsymbol{P}_1^\top \boldsymbol{A} \boldsymbol{P}_0 \boldsymbol{C}^{*-1} \boldsymbol{P}_0^\top \boldsymbol{A} \boldsymbol{P}_1,$$
$$\boldsymbol{G}^* = \boldsymbol{P}_1^\top \boldsymbol{G} \boldsymbol{P}_1, \qquad \boldsymbol{B}^* = -\boldsymbol{C}^{*-1} \boldsymbol{P}_0^\top \boldsymbol{A} \boldsymbol{P}_1, \tag{2.101}$$

where $\boldsymbol{C}^*$ is an $(n-m)(n-m)$-matrix defined by

$$\boldsymbol{C}^* = \boldsymbol{P}_0^\top \boldsymbol{A} \boldsymbol{P}_0. \tag{2.102}$$

[11] This is always true if $\mathbf{A}$ is positive definite or negative definite.

The definition of the matrix $\boldsymbol{P}_0$ implies that the matrix $\boldsymbol{G}^*$ is positive definite. Hence, the first of eqs. (2.100) is a nonsingular generalized eigenvalue problem.

The generalized Rayleigh quotient of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$ for $\boldsymbol{u} \notin \mathcal{N}_{\boldsymbol{G}}$ (i.e., $\boldsymbol{x} \neq \boldsymbol{0}$) can be written as follows:

$$\frac{(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{G}\boldsymbol{u})} = \frac{(\boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{x}) + (\boldsymbol{y} - \boldsymbol{B}^*\boldsymbol{x}, \boldsymbol{C}^*(\boldsymbol{y} - \boldsymbol{B}^*\boldsymbol{x}))}{(\boldsymbol{x}, \boldsymbol{G}^*\boldsymbol{x})}. \tag{2.103}$$

If $\boldsymbol{C}^*$ is positive definite[12], we observe that

$$\frac{(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{G}\boldsymbol{u})} \geq \frac{(\boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{G}^*\boldsymbol{x})} \geq \lambda_{\min}, \tag{2.104}$$

where $\lambda_{\min}$ is the smallest generalized eigenvalue of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$ (see eqs. (2.100)). Equality holds if $\boldsymbol{u}$ is the corresponding generalized eigenvector. If $\boldsymbol{C}^*$ is negative definite[13], we observe that

$$\frac{(\boldsymbol{u}, \boldsymbol{A}\boldsymbol{u})}{(\boldsymbol{u}, \boldsymbol{G}\boldsymbol{u})} \leq \frac{(\boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{G}^*\boldsymbol{x})} \leq \lambda_{\max}, \tag{2.105}$$

where $\lambda_{\max}$ is the smallest generalized eigenvalue of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$. Equality holds if $\boldsymbol{u}$ is the corresponding generalized eigenvector.

2.2.6 Perturbation theorem

Let $\boldsymbol{A}$ and $\boldsymbol{D}$ be (nn)-matrices. Let $\{\lambda_i\}$ be the eigenvalues of $\boldsymbol{A}$, and $\{\boldsymbol{u}_i\}$ the corresponding eigensystem:

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \qquad (\boldsymbol{u}_i, \boldsymbol{u}_j) = \delta_{ij}. \tag{2.106}$$

Consider a perturbed matrix

$$\boldsymbol{A}' = \boldsymbol{A} + \epsilon \boldsymbol{D} \tag{2.107}$$

for a small ϵ. Let $\{\lambda_i'\}$ and $\{\boldsymbol{u}_i'\}$ be, respectively, the eigenvalues and the eigensystem of $\boldsymbol{A}'$ corresponding to $\{\lambda_i\}$ and $\{\boldsymbol{u}_i\}$. The following relations hold (the *perturbation theorem*):

$$\lambda_i' = \lambda_i + \epsilon(\boldsymbol{u}_i, \boldsymbol{D}\boldsymbol{u}_i) + O(\epsilon^2), \tag{2.108}$$

$$\boldsymbol{u}_i' = \boldsymbol{u}_i + \epsilon \sum_{j \neq i} \frac{(\boldsymbol{u}_j, \boldsymbol{D}\boldsymbol{u}_i)\boldsymbol{u}_j}{\lambda_i - \lambda_j} + O(\epsilon^2). \tag{2.109}$$

Let $\boldsymbol{u}_n$ be the unit eigenvector of $\boldsymbol{A}$ for the smallest eigenvalue λ_n, which is assumed to be a simple root. Let $\{\boldsymbol{u}_i\}$ be the eigensystem of $\boldsymbol{A}$ defined so

[12] This is always true if $\mathbf{A}$ is positive definite.

[13] This is always true if $\mathbf{A}$ is negative definite.

that the corresponding eigenvalues are $\lambda_1 \geq \cdots \geq \lambda_{n-1} > \lambda_n$. Define matrix $\boldsymbol{A}^+$ by

$$\boldsymbol{A}^+ = \sum_{i=1}^{n-1} \frac{\boldsymbol{u}_i \boldsymbol{u}_i^\top}{\lambda_i - \lambda_n} \tag{2.110}$$

This is a positive semi-definite matrix having eigenvalues $\{1/(\lambda_i - \lambda_n)\}$ for the same eigensystem $\{\boldsymbol{u}_i\}$. If $\lambda_n = 0$, the matrix $\boldsymbol{A}^+$ coincides with the generalized inverse $\boldsymbol{A}^-$. Eq. (2.109) can be rewritten as

$$\boldsymbol{u}'_n = \boldsymbol{u}_n - \epsilon \boldsymbol{A}^+ \boldsymbol{D} \boldsymbol{u}_n + O(\epsilon^2). \tag{2.111}$$

Let $\boldsymbol{A}$ and $\boldsymbol{D}$ be (nn)-matrices, and $\boldsymbol{G}$ a positive definite (nn)-matrix. Let $\{\lambda_i\}$ be the generalized eigenvalues of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$, and $\{\boldsymbol{u}_i\}$ the corresponding generalized eigensystem:

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{G}\boldsymbol{u}_i, \qquad (\boldsymbol{u}_i, \boldsymbol{G}\boldsymbol{u}_j) = \delta_{ij}. \tag{2.112}$$

If $\boldsymbol{A}$ is perturbed in the form of eq. (2.107), the perturbation theorem holds in the same form. Eq. (2.111) also holds if $\{\lambda_i\}$ in eq. (2.110) are interpreted as generalized eigenvalues of $\boldsymbol{A}$ with respect to $\boldsymbol{G}$.

2.3 Linear Systems and Optimization

2.3.1 *Singular value decomposition and generalized inverse*

If $\boldsymbol{A}$ is an mn-matrix, $\boldsymbol{A}^\top \boldsymbol{A}$ is a positive semi-definite (nn)-matrix, and $\boldsymbol{A}\boldsymbol{A}^\top$ is a positive semi-definite (mm)-matrix. They share the same nonzero eigen- ... $\geq \sigma_r$ (> 0), $r \leq \min(m,n)$. The number r is called the ... $\sqrt{\sigma_i}$, $i = 1, ..., r$, and $\lambda_i = 0$, $i = r+1, ..., \max(m,n)$. It can be shown that orthonormal systems $\{\boldsymbol{u}_i\}$, $i = 1, ..., n$, and $\{\boldsymbol{v}_i\}$, $i = 1, ..., m$, exist such that

no restrictions on ***A*** !

- $\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{v}_i$, $i = 1, ..., \min(m,n)$.
- $\{\boldsymbol{u}_i\}$, $i = 1, ..., n$, is the eigensystem of $\boldsymbol{A}^\top \boldsymbol{A}$ for eigenvalues $\{\lambda_i^2\}$, $i = 1, ..., n$.
- $\{\boldsymbol{v}_i\}$, $i = 1, ..., m$, is the eigensystem of $\boldsymbol{A}\boldsymbol{A}^\top$ for eigenvalues $\{\lambda_i^2\}$, $i = 1, ..., m$.

Matrix $\boldsymbol{A}$ is expressed in terms of $\{\boldsymbol{u}_i\}$, $\{\boldsymbol{v}_i\}$, and $\{\lambda_i\}$ in the form

$$\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i \boldsymbol{u}_i^\top. \tag{2.113}$$

This is called the ***singular value decomposition*** of $\boldsymbol{A}$; the values $\{\lambda_i\}$, $i = 1, ..., \min(m,n)$, are called the *singular values* of $\boldsymbol{A}$. Let us call $\{\boldsymbol{u}_i\}$, $i = 1, ..., n$,

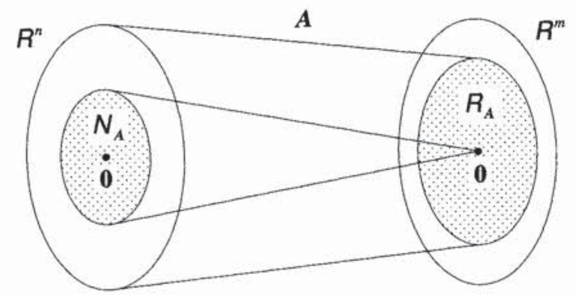

Fig. 2.4. The range $\mathcal{R}_{\mathbf{A}}$ and the null space $\mathcal{N}_{\mathbf{A}}$ of linear mapping $\mathbf{A}$.

and $\{\boldsymbol{v}_i\}$, $i = 1, ..., m$, the *right orthonormal system* and the *left orthonormal system* of $\boldsymbol{A}$, respectively.

If we define orthogonal matrices $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n)$ and $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m)$, eq. (2.113) can be rewritten in the form

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^\top, \tag{2.114}$$

where $\boldsymbol{\Lambda}$ is an mn matrix whose first r diagonal elements are $\lambda_1, ..., \lambda_r$ in that order and whose other elements are all zero. If $m = n$, matrix $\boldsymbol{\Lambda}$ is diagonal.

The r-dimensional linear subspace

$$\mathcal{R}_{\boldsymbol{A}} = \{\boldsymbol{v}_1, ..., \boldsymbol{v}_r\}_L \subset \mathcal{R}^m \tag{2.115}$$

is called the *range* (or *image space*) of $\boldsymbol{A}$: for any m-vector $\boldsymbol{y} \in \mathcal{R}_{\boldsymbol{A}}$, there exists an n-vector $\boldsymbol{x}$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ (Fig. 2.4). The $(n-r)$-dimensional linear subspace

$$\mathcal{N}_{\boldsymbol{A}} = \{\boldsymbol{u}_{r+1}, ..., \boldsymbol{u}_n\}_L \subset \mathcal{R}^n \tag{2.116}$$

is called the *null space* (or *kernel*) of $\boldsymbol{A}$: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for any n-vector $\boldsymbol{x} \in \mathcal{N}_{\boldsymbol{A}}$ (Fig. 2.4). If $\boldsymbol{A}$ is symmetric, its right and left orthonormal systems coincide with its eigensystem, and its singular value decomposition coincides with its spectral decomposition (see eq. (2.62)).

Since $\{\boldsymbol{u}_i\}$ is an orthonormal system, eq. (2.64) holds for an arbitrary n-vector $\boldsymbol{x}$. Let $\lambda_{\max}$ be the maximum singular value. Since $\{\boldsymbol{v}_i\}$ is also an orthonormal system, we see from eq. (2.113) that

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = \|\sum_{i=1}^{r} \lambda_i (\boldsymbol{u}_i, \boldsymbol{x}) \boldsymbol{v}_i\|^2 = \sum_{i=1}^{r} \lambda_i^2 (\boldsymbol{u}_i, \boldsymbol{x})^2 \leq \sum_{i=1}^{r} \lambda_{\max}^2 (\boldsymbol{u}_i, \boldsymbol{x})^2 = \lambda_{\max}^2 \|\boldsymbol{x}\|^2. \tag{2.117}$$

Hence, if we define the *spectral norm* (or the *natural norm*) of $\boldsymbol{A}$ by

$$\|\boldsymbol{A}\|_s = \lambda_{\max}, \tag{2.118}$$

eq. (2.117) implies the following inequality:

$$\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|_s \|\boldsymbol{x}\| \tag{2.119}$$

Equality holds for

$$\boldsymbol{x} \propto \boldsymbol{u}_{\max} + \mathcal{N}_{\boldsymbol{A}}. \tag{2.120}$$

The right-hand side means the first term plus any element of $\mathcal{N}_{\boldsymbol{A}}$ (and such a form only), and $\boldsymbol{u}_{\max}$ is the vector $\boldsymbol{u}_i$ corresponding to the singular value $\lambda_{\max}$.

Let eq. (2.113) be the singular value decomposition of matrix $\boldsymbol{A}$. Its *(Moore-Penrose) generalized inverse* is defined by

$$\boldsymbol{A}^- = \sum_{i=1}^{r} \frac{\boldsymbol{u}_i \boldsymbol{v}_i^\top}{\lambda_i}. \tag{2.121}$$

Evidently, the generalized inverse $\boldsymbol{A}^-$ coincides with the inverse $\boldsymbol{A}^{-1}$ if $\boldsymbol{A}$ is nonsingular. In correspondence with eq. (2.78) and eqs. (2.80), the following relationships hold:

$$(\boldsymbol{A}^-)^- = \boldsymbol{A}, \quad \boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{P}_{\mathcal{N}_A}, \quad \boldsymbol{A}\boldsymbol{A}^- = \boldsymbol{P}^{\mathcal{R}_A},$$
$$\boldsymbol{P}^{\mathcal{R}_A}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{P}_{\mathcal{N}_A} = \boldsymbol{A}, \quad \boldsymbol{P}_{\mathcal{N}_A}\boldsymbol{A}^- = \boldsymbol{A}^-\boldsymbol{P}^{\mathcal{R}_A} = \boldsymbol{A}^-. \tag{2.122}$$

Here, $\boldsymbol{P}^{\mathcal{R}_A}$ $(= \boldsymbol{P}_{\mathcal{R}_A^\perp})$ and $\boldsymbol{P}_{\mathcal{N}_A}$ $(= \boldsymbol{P}^{\mathcal{N}_A^\perp})$ are the projection matrices onto $\mathcal{R}_{\boldsymbol{A}}$ and $\mathcal{N}_{\boldsymbol{A}}^\perp$, respectively. From the above equations, we obtain

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}, \qquad \boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^- = \boldsymbol{A}^-. \tag{2.123}$$

The *rank-constrained generalized inverse* $(\boldsymbol{A})_{r'}^-$ of rank r' $(\leq r)$ is defined by

$$(\boldsymbol{A})_{r'}^- = \sum_{i=1}^{r'} \frac{\boldsymbol{u}_i \boldsymbol{v}_i^\top}{\lambda_i}, \tag{2.124}$$

and the following relations hold:

$$(\boldsymbol{A})_{r'}^-\boldsymbol{A} = \boldsymbol{P}^{\mathcal{R}_{(A)_{r'}^-}}, \quad \boldsymbol{A}(\boldsymbol{A})_{r'}^- = \boldsymbol{P}_{\mathcal{N}_{(A)_{r'}^-}},$$
$$(\boldsymbol{A})_{r'}^-\boldsymbol{A}(\boldsymbol{A})_{r'}^- = (\boldsymbol{A})_{r'}^-. \tag{2.125}$$

2.3.2 Linear equations

Let $\boldsymbol{A}$ be an mn-matrix, and $\boldsymbol{b}$ an m-vector. Consider the following linear equation for n-vector $\boldsymbol{x}$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{2.126}$$

The following is the fundamental theorem for linear equations:

- The solution exists if and only if $\boldsymbol{b} \in \mathcal{R}_{\boldsymbol{A}}$ (or $\boldsymbol{P}_{\mathcal{R}_{\boldsymbol{A}}}\boldsymbol{b} = \boldsymbol{0}$).
- If the solution exists, it is unique if and only if $\mathcal{N}_{\boldsymbol{A}} = \{\boldsymbol{0}\}$.

The problem (2.126) is said to be *consistent* (or *solvable*) when $\boldsymbol{b} \in \mathcal{R}_{\boldsymbol{A}}$, and *inconsistent* (or *unsolvable*) otherwise; if it is consistent, it is said to be *determinate* when $\mathcal{N}_{\boldsymbol{A}} = \{\boldsymbol{0}\}$, and *indeterminate* otherwise.

If eq. (2.126) is solvable, the solution can be explicitly written in the following form:

$$\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{b} + \mathcal{N}_{\boldsymbol{A}}. \tag{2.127}$$

If $\boldsymbol{A}$ is nonsingular, the solution is given by

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} = \frac{\boldsymbol{A}^{\dagger}\boldsymbol{b}}{\det \boldsymbol{A}}, \tag{2.128}$$

where $\boldsymbol{A}^{\dagger}$ is the cofactor matrix of $\boldsymbol{A}$ (see eq. (2.20)). Let $\boldsymbol{A} = (\boldsymbol{a}_1, ..., \boldsymbol{a}_n)$. From the cofactor expansion formula (2.14), the following *Cramer formula* is obtained:

$$x_i = \frac{|\boldsymbol{a}_1, ..., \overset{(i)}{\boldsymbol{b}}, ..., \boldsymbol{a}_n|}{\det \boldsymbol{A}}. \tag{2.129}$$

The numerator on the right-hand side is the determinant of the matrix obtained by replacing the ith column of $\boldsymbol{A}$ by $\boldsymbol{b}$.

If $\det \boldsymbol{A}$ is very close to 0, a small perturbation of $\boldsymbol{b}$ can causes a large perturbation to the solution $\boldsymbol{x}$. If this occurs, the linear equation (2.126) is said to be *ill-conditioned*; otherwise, it is *well-conditioned*. If $\boldsymbol{b}$ is perturbed into $\boldsymbol{b} + \Delta\boldsymbol{b}$, the solution $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$ is perturbed by $\Delta\boldsymbol{x} = \boldsymbol{A}^{-1}\Delta\boldsymbol{b}$. Applying eq. (2.119), we obtain $\|\Delta\boldsymbol{x}\| \leq \|\boldsymbol{A}^{-1}\|_s\|\Delta\boldsymbol{b}\|$. From eq. (2.126), we have $\|\boldsymbol{b}\| \leq \|\boldsymbol{A}\|_s\|\boldsymbol{x}\|$. Combining these, we obtain

$$\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \mathrm{cond}(\boldsymbol{A})\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}, \tag{2.130}$$

where

$$\mathrm{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|_s\|\boldsymbol{A}^{-1}\|_s = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{2.131}$$

Here, $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and the smallest singular values of $\boldsymbol{A}$, respectively (see eq. (2.118)). The number $\mathrm{cond}(\boldsymbol{A})$ is called the *condition number*[14] and measures the ill-posedness of the linear equation (2.126)—the equation becomes more ill-conditioned as $\mathrm{cond}(\boldsymbol{A})$ becomes larger.

Suppose eq. (2.126) is consistent but only r ($\leq m$) of the m component equations are independent, i.e., the matrix $\boldsymbol{A}$ has rank r. Theoretically, the

[14] The condition number can also be defined for a singular matrix $\mathbf{A}$ in the form $\mathrm{cond}(\mathbf{A}) = \|\mathbf{A}\|_s\|\mathbf{A}^{-}\|_s = \lambda_{\max}/\lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the largest and the smallest of the nonnegative singular values of $\mathbf{A}$.

solution is given in the form of eq. (2.127). However, if the elements of the matrix $\boldsymbol{A}$ and the components of the vector $\boldsymbol{b}$ are supplied by a physical measurement, all the m equations may be independent because of noise. As a result, eq. (2.126) may become ill-conditioned or inconsistent. In such a case, a well-conditioned equation that gives a good approximation to $\boldsymbol{x}$ is obtained by "projecting" both sides of eq. (2.126) onto the eigenspace of $\boldsymbol{A}$ defined by the largest r singular values. The solution of the projected equation is given in terms of the rank-constrained generalized inverse in the form

$$\hat{\boldsymbol{x}} = (\boldsymbol{A})_r^- \boldsymbol{b} + \mathcal{N}_{(\boldsymbol{A})_r^-}. \tag{2.132}$$

The rank r is estimated either by an a priori theoretical analysis or by appropriately thresholding the singular values of $\boldsymbol{A}$ a posteriori.

2.3.3 Quadratic optimization

A. Least-squares optimization

Let $\boldsymbol{A}$ be an mn-matrix, and $\boldsymbol{b}$ an m-vector. Consider the *least-squares optimization* for n-vector $\boldsymbol{x}$ in the form

$$J[\boldsymbol{x}] = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \to \min. \tag{2.133}$$

Application of the singular value decomposition to $\boldsymbol{A}$ yields the general solution in the following form:

$$\hat{\boldsymbol{x}} = \boldsymbol{A}^- \boldsymbol{b} + \mathcal{N}_{\boldsymbol{A}}. \tag{2.134}$$

If $\boldsymbol{x}$ is constrained to be in $\mathcal{N}_{\boldsymbol{A}}^{\perp}$, the solution is uniquely given by $\hat{\boldsymbol{x}} = \boldsymbol{A}^- \boldsymbol{b}$. The *residual* $J[\hat{\boldsymbol{x}}]$ is given by

$$J[\hat{\boldsymbol{x}}] = \|\boldsymbol{P}_{\mathcal{R}_A} \boldsymbol{b}\|^2. \tag{2.135}$$

Evidently, the residual is 0 if and only if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable.

B. Unconstrained quadratic optimization

Let $\boldsymbol{C}$ be a positive semi-definite (nn)-matrix, and $\boldsymbol{d}$ an n-vector. Consider the quadratic optimization for n-vector $\boldsymbol{x}$ in the form

$$J[\boldsymbol{x}] = \frac{1}{2}(\boldsymbol{x}, \boldsymbol{C}\boldsymbol{x}) + (\boldsymbol{d}, \boldsymbol{x}) \to \min. \tag{2.136}$$

If $\boldsymbol{x}$ is constrained to be in $\mathcal{N}_{\boldsymbol{C}}^{\perp}$, the solution is uniquely given in the following form:

$$\hat{\boldsymbol{x}} = -\boldsymbol{C}^- \boldsymbol{d}. \tag{2.137}$$

The residual is

$$J[\hat{\boldsymbol{x}}] = -\frac{1}{2}(\boldsymbol{d}, \boldsymbol{C}^- \boldsymbol{d}). \tag{2.138}$$

C. Constrained quadratic optimization

Let $\boldsymbol{S}$ be a positive semi-definite (nn)-matrix. Consider the quadratic optimization for n-vector $\boldsymbol{x}$ in the form

$$J[\boldsymbol{x}] = \frac{1}{2}(\boldsymbol{x}, \boldsymbol{S}\boldsymbol{x}) \rightarrow \min. \tag{2.139}$$

Evidently, $\boldsymbol{x} = \boldsymbol{0}$ is a solution (but not necessarily unique) if no constraint is imposed on $\boldsymbol{x}$. The following three types of constraint are important:

- If $\boldsymbol{x}$ is constrained to be a unit vector ($\|\boldsymbol{x}\| = 1$), the solution is given by any unit eigenvector $\hat{\boldsymbol{x}}$ of $\boldsymbol{S}$ for the smallest eigenvalue $\lambda_{\min}$ (see eqs. (2.86)); the residual is $J[\hat{\boldsymbol{x}}] = \lambda_{\min}$ (see eq. (2.95)).
- If $\boldsymbol{x}$ is constrained by $(\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}) = 1$ for a positive definite (nn)-matrix $\boldsymbol{G}$, the solution is given by any unit generalized eigenvector $\hat{\boldsymbol{x}} \in \mathcal{N}_{\boldsymbol{S}}^{\perp}$ of $\boldsymbol{S}$ with respect to $\boldsymbol{G}$ for the smallest generalized eigenvalue $\lambda_{\min}$; the residual is $J[\hat{\boldsymbol{x}}] = \lambda_{\min}$. If $\boldsymbol{S}$ is of full rank, the same conclusion is obtained even though $\boldsymbol{G}$ is not of full rank (see eq. (2.104)).
- Suppose $\boldsymbol{x}$ is constrained by a linear equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A}$ is an mn-matrix and $\boldsymbol{b}$ is an m-vector. If

  1. $\boldsymbol{x}$ is constrained to be in $\mathcal{N}_{\boldsymbol{S}}^{\perp}$, and
  2. the constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is *satisfiable* for $\boldsymbol{x} \in \mathcal{N}_{\boldsymbol{S}}^{\perp}$, i.e., at least one $\boldsymbol{x}_0 \in \mathcal{N}_{\boldsymbol{S}}^{\perp}$ exists such that $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$,

  then the solution is uniquely given in the following form:

  $$\hat{\boldsymbol{x}} = \boldsymbol{S}^{-}\boldsymbol{A}^{\top}(\boldsymbol{A}\boldsymbol{S}^{-}\boldsymbol{A}^{\top})^{-}\boldsymbol{b}. \tag{2.140}$$

  The residual is

  $$J[\hat{\boldsymbol{x}}] = \frac{1}{2}(\boldsymbol{b}, (\boldsymbol{A}\boldsymbol{S}^{-}\boldsymbol{A}^{\top})^{-}\boldsymbol{b}). \tag{2.141}$$

2.3.4 Matrix inner product and matrix norm

The *matrix inner product* of mn-matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ is defined by

$$(\boldsymbol{A}; \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}^{\top}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\top}) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}. \tag{2.142}$$

Evidently, $(\boldsymbol{A}; \boldsymbol{B}) = (\boldsymbol{B}; \boldsymbol{A})$. If $(\boldsymbol{A}; \boldsymbol{B}) = 0$, matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be *orthogonal*. An (nn)-matrix is orthogonal to any $[nn]$-matrix; an $[nn]$-matrix is orthogonal to any (nn)-matrix. The following identities are easy to prove:

$$(\boldsymbol{A}; \boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{B}^{\top}\boldsymbol{A}; \boldsymbol{C}) = (\boldsymbol{A}\boldsymbol{C}^{\top}; \boldsymbol{B}),$$

$$(\boldsymbol{a}, \boldsymbol{A}\boldsymbol{b}) = (\boldsymbol{a}\boldsymbol{b}^\top; \boldsymbol{A}), \qquad (\boldsymbol{a}\boldsymbol{b}^\top; \boldsymbol{c}\boldsymbol{d}^\top) = (\boldsymbol{a}, \boldsymbol{c})(\boldsymbol{b}, \boldsymbol{d}). \tag{2.143}$$

The *(Euclidean) matrix norm*[15] of an mn-matrix is defined by

$$\|\boldsymbol{A}\| = \sqrt{(\boldsymbol{A}; \boldsymbol{A})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}{}^2}. \tag{2.144}$$

We define the *normalization* $N[\,\cdot\,]$ of an nn-matrix $\boldsymbol{A}$ as follows (see the second of eqs. (2.7)):

$$N[\boldsymbol{A}] = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|}. \tag{2.145}$$

The *Schwarz inequality* and the *triangle inequality* hold in the same way as in the case of vectors:

$$-\|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\| \leq (\boldsymbol{A}; \boldsymbol{B}) \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|, \tag{2.146}$$

$$\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|. \tag{2.147}$$

In both inequalities, equality holds if and only if there exists a real number t such that $\boldsymbol{A} = t\boldsymbol{B}$ or $\boldsymbol{B} = \boldsymbol{O}$.

Let $\boldsymbol{U}$ be an n-dimensional orthogonal matrix. From eqs. (2.52) and the first of eqs. (2.143), it is immediately seen that for arbitrary nn-matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

$$(\boldsymbol{U}\boldsymbol{A}; \boldsymbol{U}\boldsymbol{B}) = (\boldsymbol{A}\boldsymbol{U}; \boldsymbol{B}\boldsymbol{U}) = (\boldsymbol{A}; \boldsymbol{B}). \tag{2.148}$$

Letting $\boldsymbol{A} = \boldsymbol{B}$, we obtain

$$\|\boldsymbol{U}\boldsymbol{A}\| = \|\boldsymbol{A}\boldsymbol{U}\| = \|\boldsymbol{A}\|. \tag{2.149}$$

Further letting $\boldsymbol{A} = \boldsymbol{I}$, we see that

$$\|\boldsymbol{U}\| = \sqrt{n}. \tag{2.150}$$

A nonsingular nn-matrix $\boldsymbol{T}$ defines a mapping from an nn-matrix $\boldsymbol{A}$ to an nn-matrix in the form

$$\boldsymbol{A}' = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}. \tag{2.151}$$

---

[15] Some authors use different terminologies such as the *Frobenius norm*, the *Schur norm*, and the *Schmidt norm*. In general, the norm $\|\mathbf{A}\|$ can be defined arbitrarily as long as (i) $\|\mathbf{A}\| \geq 0$, equality holding if and only if $\mathbf{A} = \mathbf{O}$, (ii) $\|c\mathbf{A}\| = |c| \cdot \|\mathbf{A}\|$ for any scalar c, and (iii) the triangle inequality (2.147) holds. There exist other definitions that satisfy these—the *1-norm* $\|\mathbf{A}\|_1 = \Sigma_{i=1}^n \max_j |A_{ij}|$, the ∞-*norm* $\|\mathbf{A}\|_\infty = \Sigma_{j=1}^n \max_i |A_{ij}|$, and the *spectral norm* $\|\mathbf{A}\|_s$ defined by eq. (2.118), for instance. If $\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{x}\|$ holds, the matrix norm $\|\mathbf{A}\|$ is said to be *consistent* with the vector norm $\|\mathbf{x}\|$. The spectral norm $\|\mathbf{A}\|_s$ is consistent with the Euclidean norm $\|\mathbf{x}\|$, and the 1-norm $\|\mathbf{A}\|_1$ and the ∞-norm $\|\mathbf{A}\|_\infty$ are consistent with the 1-norm $\|\mathbf{x}\|_1$ and the ∞-norm $\|\mathbf{x}\|_\infty$, respectively (see Footnote 4 in Section 2.1).

This is a one-to-one and onto mapping and is called the *similarity transformation*[16].

A function $f(\cdot)$ of a matrix is called an *invariant* with respect to similarity transformations if $f(\boldsymbol{A}') = f(\boldsymbol{A})$ for an arbitrary nonsingular matrix $\boldsymbol{T}$. The trace and the determinant are typical invariants:

$$\mathrm{tr}(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}) = \mathrm{tr}\boldsymbol{A}, \qquad \det(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}) = \det \boldsymbol{A}. \tag{2.152}$$

Eq. (2.67) implies that any symmetric matrix is mapped to a diagonal matrix by an appropriate similarity transformation; the transformation is defined by an orthogonal matrix. Hence, if $\boldsymbol{A}$ is a symmetric matrix with eigenvalues $\{\lambda_i\}$, any invariant with respect to similarity transformations is a function of $\{\lambda_i\}$. Eqs. (2.67) and (2.149) imply that

$$\|\boldsymbol{A}\| = \sqrt{\sum_{i=1}^{n} \lambda_i^2}. \tag{2.153}$$

Hence, $\|\boldsymbol{A}\|$ is also an invariant with respect to similarity transformation.

In three dimensions, $\mathrm{tr}\boldsymbol{A}$, $\det \boldsymbol{A}$, and $\|\boldsymbol{A}\|$ can uniquely determine the three eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ of a (33)-matrix $\boldsymbol{A}$ (see eqs. (2.68)). Hence, the three invariants $\{\mathrm{tr}\boldsymbol{A}, \det \boldsymbol{A}, \|\boldsymbol{A}\|\}$ are an ***invariant basis*** in the sense that any invariant can be expressed in terms of them.

A nonsingular nn-matrix $\boldsymbol{T}$ defines a mapping from an (nn)-matrix $\boldsymbol{A}$ to an (nn)-matrix in the form

$$\boldsymbol{A}' = \boldsymbol{T}^\top \boldsymbol{A}\boldsymbol{T}. \tag{2.154}$$

This is a one-to-one and onto mapping and called the *congruence transformation*[17]. The pair (p, q) consisting of the number p of positive eigenvalues and the number q of negative eigenvalues of an (nn)-matrix $\boldsymbol{A}$ is called the *signature* of $\boldsymbol{A}$. Under a congruence transformation, the signature does not change (*Sylvester's law of inertia*). Hence, the rank is also preserved. It follows that a positive definite symmetric matrix is always transformed to a positive definite symmetric matrix; a positive semi-definite symmetric matrix is always transformed to a positive semi-definite matrix of the same rank.

The congruence transformation defined by an orthogonal matrix $\boldsymbol{U}$ coincides with the similarity transformation defined by $\boldsymbol{U}$, and the matrix inner product and the matrix norm are also preserved:

$$(\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}; \boldsymbol{U}^\top \boldsymbol{B}\boldsymbol{U}) = (\boldsymbol{A}; \boldsymbol{B}), \qquad \|\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}\| = \|\boldsymbol{A}\|. \tag{2.155}$$

[16] Similarity transformations define a group of transformations isomorphic to $GL(n)$, the group of nonsingular matrices under multiplication.

[17] Congruence transformations define a group of transformations isomorphic to $GL(n)$, the group of nonsingular matrices under multiplication.

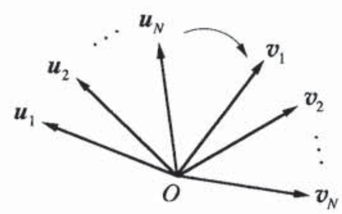

Fig. 2.5. Finding a rotation that maps one set of vectors to another.

2.3.5 Optimal rotation fitting

rotational part of Procrustes method

Let $\{\boldsymbol{u}_\alpha\}$ and $\{\boldsymbol{v}_\alpha\}$, $\alpha = 1, ..., N$, be two sets of n-vectors. Consider the problem of finding a rotation $\boldsymbol{R}$ such that

$$\sum_{\alpha=1}^{N} W_\alpha \|\boldsymbol{u}_\alpha - \boldsymbol{R}\boldsymbol{v}_\alpha\|^2 \to \min, \tag{2.156}$$

where W_α are nonnegative weights (Fig. 2.5). Since $\|\boldsymbol{R}\boldsymbol{v}_\alpha\| = \|\boldsymbol{v}_\alpha\|$, the right-hand side can be rewritten as $\sum_{\alpha=1}^{N} W_\alpha \|\boldsymbol{u}_\alpha\|^2 - 2\sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{u}_\alpha, \boldsymbol{R}\boldsymbol{v}_\alpha)$ $+ \sum_{\alpha=1}^{N} W_\alpha \|\boldsymbol{v}_\alpha\|^2$. Hence, if we define the *correlation matrix*

$$\boldsymbol{A} = \sum_{\alpha=1}^{N} W_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^\top, \tag{2.157}$$

the problem can be rewritten as follows (see the second of eqs. (2.143)):

$$(\boldsymbol{A}; \boldsymbol{R}) \to \max. \tag{2.158}$$

This problem can also be viewed as finding a rotation matrix $\boldsymbol{R}$ that is the closest to a given matrix $\boldsymbol{A}$ in the matrix norm:

$$\|\boldsymbol{R} - \boldsymbol{A}\| \to \min. \tag{2.159}$$

In fact, eqs. (2.144) and (2.150) imply that $\|\boldsymbol{R} - \boldsymbol{A}\|^2 = \|\boldsymbol{R}\|^2 - 2(\boldsymbol{R}; \boldsymbol{A}) + \|\boldsymbol{A}\|^2$ $= n - 2(\boldsymbol{A}; \boldsymbol{R}) + \|\boldsymbol{A}\|^2$, so minimizing $\|\boldsymbol{R} - \boldsymbol{A}\|$ is equivalent to maximizing $(\boldsymbol{A}; \boldsymbol{R})$

Let $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the singular value decomposition of $\boldsymbol{A}$. The solution of the optimization (2.159) is given by

$$\boldsymbol{R} = \boldsymbol{V}\mathrm{diag}(1, ..., 1, \det(\boldsymbol{V}\boldsymbol{U}^\top))\boldsymbol{U}^\top. \tag{2.160}$$

If the optimization is conducted over orthogonal matrices (i.e., if $\det \boldsymbol{R} = 1$ is not required), the solution is given by

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{U}^\top, \tag{2.161}$$

2.4 Matrix and Tensor Algebra

2.4.1 Direct sum and tensor product

For an m-vector $\boldsymbol{a} = (a_i)$ and an n-vector $\boldsymbol{b} = (b_i)$, the $(m+n)$-vector $(a_1, ..., a_m, b_1, ..., b_n)^\top$ is called the *direct sum* of $\boldsymbol{a}$ and $\boldsymbol{b}$ and denoted by $\boldsymbol{a} \oplus \boldsymbol{b}$. For an mm-matrix $\boldsymbol{A}$ and an nn-matrix $\boldsymbol{B}$, the $(m+n)(m+n)$-matrix that has $\boldsymbol{A}$ and $\boldsymbol{B}$ as diagonal blocks in that order and zero elements elsewhere is called the *direct sum* of $\boldsymbol{A}$ and $\boldsymbol{B}$ and denoted by $\boldsymbol{A} \oplus \boldsymbol{B}$. Direct sums of more than two vectors or more than two matrices are defined similarly:

$$\boldsymbol{a} \oplus \cdots \oplus \boldsymbol{b} = \begin{pmatrix} \boldsymbol{a} \\ \vdots \\ \boldsymbol{b} \end{pmatrix}, \qquad \boldsymbol{A} \oplus \cdots \oplus \boldsymbol{B} = \begin{pmatrix} \boldsymbol{A} & & \\ & \ddots & \\ & & \boldsymbol{B} \end{pmatrix}.$$

direct sum = stacking

Let $\boldsymbol{A}$ be an mm-matrix, and $\boldsymbol{B}$ an nn-matrix. Let $\boldsymbol{u}$ and $\boldsymbol{a}$ be m-vectors, and $\boldsymbol{v}$ and $\boldsymbol{b}$ n-vectors. The following relations are obvious:

$$(\boldsymbol{A} \oplus \boldsymbol{B})(\boldsymbol{u} \oplus \boldsymbol{v}) = (\boldsymbol{A}\boldsymbol{u}) \oplus (\boldsymbol{B}\boldsymbol{v}),$$
$$(\boldsymbol{a} \oplus \boldsymbol{b}, \boldsymbol{u} \oplus \boldsymbol{v}) = (\boldsymbol{a}, \boldsymbol{u}) + (\boldsymbol{b}, \boldsymbol{v}). \tag{2.163}$$

A set of real numbers $\mathcal{T} = (T_{i_1 i_2 \cdots i_r})$, $i_1, i_2, ..., i_r = 1, ..., n$, with r indices running over n-dimensional coordinates is called a *tensor* of *dimension* n and *degree* r. If each index corresponds to coordinates of a different dimensionality, $\mathcal{T}$ is called a tensor of *mixed dimensions* or a *mixed tensor*. If index i_k runs over 1, ..., n_k for $k = 1$, ..., r, the tensor is said to be of *type* $n_1 n_2 \cdots n_r$. A tensor of type $n_1 n_2 \cdots n_r$ is also referred to as an $n_1 n_2 \cdots n_r$-tensor. If $T_{i_1 i_2 \cdots i_r}$ is symmetric with respect to indices i_k and i_{k+1}, the type is written as $i_1 \cdots (i_k i_{k+1}) \cdots i_r$; If $T_{i_1 i_2 \cdots i_r}$ is antisymmetric with respect to indices i_k and i_{k+1}, the type is written as $i_1 \cdots [i_k i_{k+1}] \cdots i_r$; Scalars, vectors, and matrices are tensors of degrees 0, 1, and 2, respectively.

The *tensor product* of tensor $\mathcal{A} = (A_{i_1 \cdots i_r})$ of degree r and tensor $\mathcal{B} = (B_{i_1 \cdots i_s})$ of degree s is a tensor $\mathcal{C} = (C_{i_1 \cdots i_{r+s}})$ of degree $r+s$ defined by

$$C_{i_1 \cdots i_{r+s}} = A_{i_1 \cdots i_r} B_{i_1 \cdots i_s}. \tag{2.164}$$

This is symbolically written as

$$\mathcal{C} = \mathcal{A} \otimes \mathcal{B}. \tag{2.165}$$

The following identities hold for scalar c and vectors $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$c \otimes \boldsymbol{u} = c\boldsymbol{u}, \qquad \boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^\top. \tag{2.166}$$

derive these from 2.164 (see (2.1)...)

2.4.2 Cast in three dimensions

A. 33-matrices

The elements of a 33-matrix $\boldsymbol{A} = (A_{ij})$ are rearranged into a 9-vector

$$\boldsymbol{a} = \begin{pmatrix} A_{11} \\ A_{12} \\ \vdots \\ A_{33} \end{pmatrix}, \tag{2.167}$$

which can be written as $\boldsymbol{a} = (a_\kappa)$ with

$$a_\kappa = A_{(\kappa-1)\mathrm{div}3+1,(\kappa-1)\mathrm{mod}3+1}. \tag{2.168}$$

The symbols 'div' and 'mod' denote integer division and integer remainder, respectively. Conversely, a 9-vector $\boldsymbol{a} = (a_\kappa)$ is rearranged into a 33-matrix

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix}, \tag{2.169}$$

which can be written as $\boldsymbol{A} = (A_{ij})$ with

$$A_{ij} = a_{3(i-1)+j}. \tag{2.170}$$

The above *type transformation* or *cast* is denoted by

$$\boldsymbol{a} = \mathrm{type}_9[\boldsymbol{A}], \qquad \boldsymbol{A} = \mathrm{type}_{33}[\boldsymbol{a}]. \tag{2.171}$$

The norm is preserved by cast:

$$\|\boldsymbol{a}\| = \|\boldsymbol{A}\|. \tag{2.172}$$

The left-hand side designates the vector norm, whereas the right-hand side designates the matrix norm. The cast can be extended to tensors:

- A 3333-tensor $\mathcal{T} = (T_{ijkl})$ is cast, by rearranging the elements with respect to the indices i and j, into a mixed tensor ${}^*\mathcal{T} = ({}^*T_{\kappa kl})$ of type 933, which is denoted by $\mathrm{type}_{933}[\mathcal{T}]$; the inverse cast is $\mathcal{T} = \mathrm{type}_{3333}[{}^*\mathcal{T}]$.
- A 3333-tensor $\mathcal{T} = (T_{ijkl})$ is cast into a tensor $\mathcal{T}^* = (T^*_{ij\kappa})$ of type 339, which is denoted by $\mathrm{type}_{339}[\mathcal{T}]$; the inverse cast is $\mathcal{T} = \mathrm{type}_{3333}[\mathcal{T}^*]$.
- If both operations are applied, $\mathcal{T} = (T_{ijkl})$ is cast into a 99-matrix $\boldsymbol{T} = (T_{\kappa\lambda})$, which is denoted by $\mathrm{type}_{99}[\mathcal{T}]$; the inverse cast is $\mathcal{T} = \mathrm{type}_{3333}[\boldsymbol{T}]$.

what will be the geometrical significance of the matrix norms?

B. (33)-matrices

The elements of a (33)-matrix $\boldsymbol{S} = (S_{ij})$ are rearranged into a 6-vector

$$\boldsymbol{s} = \begin{pmatrix} S_{11} \\ S_{22} \\ S_{33} \\ \sqrt{2}S_{23} \\ \sqrt{2}S_{31} \\ \sqrt{2}S_{12} \end{pmatrix}. \tag{2.173}$$

Conversely, a 6-vector $\boldsymbol{s} = (s_\kappa)$ is rearranged into a (33)-matrix

$$\boldsymbol{S} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}s_1 & s_6 & s_5 \\ s_6 & \sqrt{2}s_2 & s_4 \\ s_5 & s_4 & \sqrt{2}s_3 \end{pmatrix}. \tag{2.174}$$

This cast is denoted by

$$\boldsymbol{s} = \text{type}_6[\boldsymbol{S}], \qquad \boldsymbol{S} = \text{type}_{(33)}[\boldsymbol{s}]. \tag{2.175}$$

The norm is preserved by cast:

$$\|\boldsymbol{s}\| = \|\boldsymbol{S}\|. \tag{2.176}$$

The cast can be extended to tensors:

- A (33)33-tensor $\mathcal{L} = (L_{ijkl})$ is cast, by rearranging the elements with respect to the indices i and j, into a mixed tensor ${}^*\mathcal{L} = ({}^*L_{\kappa kl})$ of type 633, which is denoted by $\text{type}_{633}[\mathcal{L}]$; the inverse cast is $\mathcal{L} = \text{type}_{(33)33}[{}^*\mathcal{L}]$.
- A 33(33)-tensor $\mathcal{N} = (S_{ijkl})$ is cast to a mixed tensor $\mathcal{N}^* = (S^*_{ij\kappa})$ of type 336, which is denoted by $\text{type}_{336}[\mathcal{N}]$; the inverse cast is $\mathcal{N} = \text{type}_{33(33)}[\mathcal{N}^*]$.
- If both operations are applied, a (33)(33)-tensor $\mathcal{M} = (M_{ijkl})$ is cast to a 66-matrix $\boldsymbol{M} = (M_{\kappa\lambda})$, which is denoted by $\text{type}_{66}[\mathcal{M}]$. In elements,

$$\boldsymbol{M} = \begin{pmatrix} M_{1111} & M_{1122} & M_{1133} & \sqrt{2}M_{1123} & \sqrt{2}M_{1131} & \sqrt{2}M_{1112} \\ M_{2211} & M_{2222} & M_{2233} & \sqrt{2}M_{2223} & \sqrt{2}M_{2231} & \sqrt{2}M_{2212} \\ M_{3311} & M_{3322} & M_{3333} & \sqrt{2}M_{3323} & \sqrt{2}M_{3331} & \sqrt{2}M_{3312} \\ \sqrt{2}M_{2311} & \sqrt{2}M_{2322} & \sqrt{2}M_{2333} & 2M_{2323} & 2M_{2331} & 2M_{2312} \\ \sqrt{2}M_{3111} & \sqrt{2}M_{3122} & \sqrt{2}M_{3133} & 2M_{3123} & 2M_{3131} & 2M_{3112} \\ \sqrt{2}M_{1211} & \sqrt{2}M_{1222} & \sqrt{2}M_{1233} & 2M_{1223} & 2M_{1231} & 2M_{1212} \end{pmatrix} \tag{2.177}$$

The inverse cast is $\mathcal{M} = \text{type}_{(33)(33)}[\boldsymbol{M}]$.

C. [33]-matrices

The elements of a [33]-matrix $\boldsymbol{W} = (W_{ij})$ are rearranged into a 3-vector

$$\boldsymbol{w} = \begin{pmatrix} W_{32} \\ W_{13} \\ W_{21} \end{pmatrix}, \tag{2.178}$$

which can be written as $\boldsymbol{w} = (w_\kappa)$ with

$$w_\kappa = -\frac{1}{2}\sum_{i,j=1}^{3} \epsilon_{\kappa ij} W_{ij}. \tag{2.179}$$

Conversely, a 3-vector $\boldsymbol{w} = (w_\kappa)$ is rearranged into a [33]-matrix

$$\boldsymbol{W} = \begin{pmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{pmatrix} = \boldsymbol{w} \times \boldsymbol{I}, \tag{2.180}$$

which can be written as $\boldsymbol{W} = (W_{ij})$ with

$$W_{ij} = -\sum_{k=1}^{3} \epsilon_{ij\kappa} w_\kappa. \tag{2.181}$$

This cast is denoted by

$$\boldsymbol{w} = \text{type}_3[\boldsymbol{W}], \qquad \boldsymbol{W} = \text{type}_{[33]}[\boldsymbol{w}]. \tag{2.182}$$

The following identities hold, where $\boldsymbol{r}$ is an arbitrary 3-vector:

$$\|\boldsymbol{W}\| = \sqrt{2}\|\boldsymbol{w}\|, \qquad \boldsymbol{W}\boldsymbol{r} = \boldsymbol{w} \times \boldsymbol{r}. \tag{2.183}$$

The cast can be extended to tensors:

- A [33]33-tensor $\mathcal{P} = (P_{ijkl})$ is cast, by rearranging the elements with respect to the indices i and j, into a mixed tensor ${}^*\mathcal{P} = ({}^*P_{\kappa kl})$ of type 333, which is denoted by $\text{type}_{333}[\mathcal{P}]$; the inverse cast is $\mathcal{P} = \text{type}_{[33]33}[{}^*\mathcal{P}]$.
- A 33[33]-tensor $\mathcal{Q} = (Q_{ijkl})$ is cast to a mixed tensor $\mathcal{Q}^* = (Q^*_{ij\kappa})$ of type 333, which is denoted by $\text{type}_{333}[\mathcal{Q}]$; the inverse cast is $\mathcal{Q} = \text{type}_{33[33]}[\mathcal{Q}^*]$.
- If both operations are applied, a [33][33]-tensor $\mathcal{R} = (R_{ijkl})$ is cast to a 33-matrix $\boldsymbol{R} = (R_{\kappa\lambda})$, which is denoted by $\text{type}_{33}[\mathcal{R}]$. In elements,

$$\boldsymbol{R} = \begin{pmatrix} R_{3232} & R_{3213} & R_{3221} \\ R_{1332} & R_{1313} & R_{1321} \\ R_{2132} & R_{2113} & R_{2121} \end{pmatrix}. \tag{2.184}$$

The inverse cast is $\mathcal{R} = \text{type}_{[33][33]}[\boldsymbol{R}]$.

2.4.3 *Linear mapping of matrices in three dimensions*

A. 33-matrices

moral: tensors can be recast (for computations, not meaning)

A 3333-tensor $\mathcal{T} = (T_{ijkl})$ defines a line[illegible] 33-matrix: matrix $\boldsymbol{A} = (A_{ij})$ is mapped to matrix $\boldsymbol{A}' = (A'_{ij})$ in the form

$$A'_{ij} = \sum_{k,l=1}^{3} T_{ijkl} A_{kl}. \tag{2.185}$$

This mapping is denoted by

$$\boldsymbol{A}' = \mathcal{T}\boldsymbol{A}. \tag{2.186}$$

The identity mapping $\mathcal{I} = (I_{ijkl})$ is given by

$$I_{ijkl} = \delta_{ik}\delta_{jl}. \tag{2.187}$$

The similarity transformation $\boldsymbol{A}' = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ defined by a nonsingular matrix $\boldsymbol{T} = (T_{ij})$ maps a 33-matrix $\boldsymbol{A}$ to a 33-matrix (see eq. (2.151)). This mapping can be written as $\boldsymbol{A}' = \mathcal{T}\boldsymbol{A}$, where the tensor $\mathcal{T} = (T_{ijkl})$ is defined by

$$T_{ijkl} = T^{-1}_{ik} T_{lj}. \tag{2.188}$$

Here, T^{-1}_{ik} denotes the (ik) element of $\boldsymbol{T}^{-1}$.

If a 3333-tensor $\mathcal{T}$ is cast into a 99-matrix $\boldsymbol{T}$ and if 33-matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ are cast into 9-vectors $\boldsymbol{a}$ and $\boldsymbol{a}'$, respectively, the mapping $\boldsymbol{A}' = \mathcal{T}\boldsymbol{A}$ is identified with

$$\boldsymbol{a}' = \boldsymbol{T}\boldsymbol{a}, \tag{2.189}$$

which is a linear mapping from a 9-vector $\boldsymbol{a}$ to a 9-vector $\boldsymbol{a}'$. Hence, the mapping $\mathcal{T}$ is nonsingular if and only if the 99-matrix $\boldsymbol{T}$ obtained by cast is nonsingular. The inverse $\mathcal{T}^{-1}$ of a nonsingular mapping $\mathcal{T}$ is given through the cast:

$$\mathcal{T}^{-1} = \text{type}_{3333}[\text{type}_{99}[\mathcal{T}]^{-1}]. \tag{2.190}$$

If mapping $\mathcal{T}$ is singular, its generalized inverse is also defined through the same cast:

$$\mathcal{T}^{-} = \text{type}_{3333}[\text{type}_{99}[\mathcal{T}]^{-}]. \tag{2.191}$$

A 33-matrix $\boldsymbol{A}$ is an *eigenmatrix* of a 3333-tensor $\mathcal{T}$ for eigenvalue λ if

$$\mathcal{T}\boldsymbol{A} = \lambda\boldsymbol{A}. \tag{2.192}$$

Eigenvalues and eigenmatrices are computed by solving the eigenvalue problem of the (99)-matrix obtained by cast: if $\boldsymbol{T} = \text{type}_{99}[\mathcal{T}]$ and $\boldsymbol{a} = \text{type}_9[\boldsymbol{A}]$, eq. (2.192) reads

$$\boldsymbol{T}\boldsymbol{a} = \lambda\boldsymbol{a}. \tag{2.193}$$

B. (33)-matrices

A (33)(33)-tensor $\mathcal{M} = (M_{ijkl})$ defines a linear mapping from a (33)-matrix to a (33)-matrix: matrix $\boldsymbol{S}$ is mapped to matrix $\boldsymbol{S}' = \mathcal{M}\boldsymbol{S}$ in the form eq. (2.185). The identity mapping $\mathcal{I} = (I_{ijkl})$ is given by

$$I_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il}). \tag{2.194}$$

The congruence transformation $\boldsymbol{S}' = \boldsymbol{T}^{-1}\boldsymbol{S}\boldsymbol{T}$ defined by a nonsingular 33-matrix $\boldsymbol{T} = (T_{ij})$ maps a (33)-matrix $\boldsymbol{S}$ to a (33)-matrix (see eq. (2.154)). This mapping can be written as $\boldsymbol{S}' = \mathcal{M}\boldsymbol{S}$, where the tensor $\mathcal{M} = (M_{ijkl})$ is defined by

$$M_{ijkl} = \frac{1}{2}(T_{ki}T_{lj} + T_{kj}T_{li}). \tag{2.195}$$

If a (33)(33)-tensor $\mathcal{M}$ is cast into a 66-matrix $\boldsymbol{M}$ and if (33)-matrices $\boldsymbol{S}$ and $\boldsymbol{S}'$ are cast into 6-vectors $\boldsymbol{s}$ and $\boldsymbol{s}'$, respectively, the mapping $\boldsymbol{S}' = \mathcal{M}\boldsymbol{S}$ is identified with

$$\boldsymbol{s}' = \boldsymbol{M}\boldsymbol{s}, \tag{2.196}$$

which is a linear mapping from 6-vector $\boldsymbol{s}$ to 6-vector $\boldsymbol{s}'$. Hence, the mapping $\mathcal{M}$ is nonsingular if and only if the 66-matrix $\boldsymbol{M}$ obtained by cast is nonsingular. The inverse $\mathcal{M}^{-1}$ and the generalized inverse $\mathcal{M}^{-}$ are defined through the cast:

$$\mathcal{M}^{-1} = \text{type}_{(33)(33)}[\text{type}_{66}[\mathcal{M}]^{-1}], \tag{2.197}$$

$$\mathcal{M}^{-} = \text{type}_{(33)(33)}[\text{type}_{66}[\mathcal{M}]^{-}]. \tag{2.198}$$

Eigenvalues and eigenmatrices are also defined and computed through the cast.

C. [33]-matrices

derive the 2!

If a [33][33]-tensor $\mathcal{R}$ is cast into a 33-matrix $\boldsymbol{R}$ and if [33]-matrices $\boldsymbol{W}$ and $\boldsymbol{W}'$ are cast into 3-vectors $\boldsymbol{w}$ and $\boldsymbol{w}'$, respectively, the mapping $\boldsymbol{W}' = \mathcal{R}\boldsymbol{W}$ is identified with

$$\boldsymbol{w}' = 2\boldsymbol{R}\boldsymbol{w}, \tag{2.199}$$

which is a linear mapping from 3-vector $\boldsymbol{w}$ to 3-vector $\boldsymbol{w}'$. Hence, the mapping $\mathcal{R}$ is nonsingular if and only if the 33-matrix $\boldsymbol{R}$ obtained by cast is nonsingular. The inverse $\mathcal{R}^{-1}$ and the generalized inverse $\mathcal{R}^{-}$ are defined through the cast:

$$\mathcal{R}^{-1} = \frac{1}{4}\text{type}_{[33][33]}[\text{type}_{33}[\mathcal{R}]^{-1}], \tag{2.200}$$

$$\mathcal{R}^{-} = \frac{1}{4}\text{type}_{[33][33]}[\text{type}_{33}[\mathcal{R}]^{-}]. \tag{2.201}$$

Eigenvalues and eigenmatrices are also defined and computed through the cast.

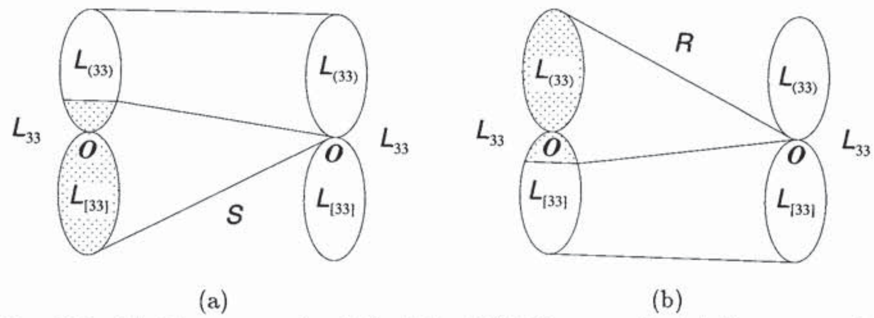

Fig. 2.6. (a) Linear mapping defined by (33)(33)-tensor $\mathcal{S}$. (a) Linear mapping defined by [33][33]-tensor $\mathcal{R}$.

D. Matrix spaces

The nine-dimensional linear space $\mathcal{L}_{33}$ of all 33-matrices is the direct sum of the six-dimensional subspace $\mathcal{L}_{(33)}$ of all (33)-matrices and the three-dimensional subspace $\mathcal{L}_{[33]}$ of all [33]-matrices (Fig. 2.6). The two subspaces are orthogonal complements of each other (see Section 2.3.4):

$$\mathcal{L}_{33} = \mathcal{L}_{(33)} \oplus \mathcal{L}_{[33]}, \qquad \mathcal{L}_{(33)} \perp \mathcal{L}_{[33]}. \tag{2.202}$$

This is because any 33-matrix $\boldsymbol{A}$ is uniquely decomposed into a (33)-matrix $\boldsymbol{A}_s$ and a [33]-matrix $\boldsymbol{A}_a$:

$$\boldsymbol{A} = \boldsymbol{A}_s + \boldsymbol{A}_a, \qquad (\boldsymbol{A}_s; \boldsymbol{A}_a) = 0, \tag{2.203}$$

$$\boldsymbol{A}_s = S[\boldsymbol{A}], \qquad \boldsymbol{A}_s = A[\boldsymbol{A}]. \tag{2.204}$$

Here, the *symmetrization operator* $S[\cdot]$ and the *antisymmetrization operator* $A[\cdot]$ are defined as follows:

$$S[\boldsymbol{A}] = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^\top), \qquad A[\boldsymbol{A}] = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^\top). \tag{2.205}$$

We observe the following:

- If a (33)(33)-tensor $\mathcal{S}$ is viewed as a 3333-tensor, the linear mapping it defines is singular: its null space includes $\mathcal{L}_{[33]}$, and its range is a subspace of $\mathcal{L}_{(33)}$ (Fig. 2.6a). Hence, it always has eigenvalue 0, whose multiplicity is at least 3.

- If a [33][33]-tensor $\mathcal{R}$ is viewed as a 3333-tensor, the linear mapping it defines is also singular: its null space includes $\mathcal{L}_{(33)}$, and its range is a subspace of $\mathcal{L}_{[33]}$ (Fig. 2.6b). Hence, it always has eigenvalue 0, whose multiplicity is at least 6.