Brief review of linear algebra, plus some additional facts

1. $\mathbf{R}^n \equiv \left\{ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \;\middle|\; x_i \text{ real} \right\}.$

2. A linear combination of vectors $\mathbf{x}_1 \cdots, \mathbf{x}_k$ in $\mathbf{R}^n$ is a vector of the form $\mathbf{x} = c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$ where $c_1, ..., c_k$ are scalars.

3. A set of vectors $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ is linearly dependent if some nonzero linear combination of them yields the zero vector; otherwise it is linearly independent.

4. A subset $\mathbf{V} \subset \mathbf{R}^n$ is a subspace of $\mathbf{R}^n$ if $\mathbf{x}, \mathbf{y} \in \mathbf{V} \implies c\mathbf{x} + d\mathbf{y} \in \mathbf{V}$, i.e., if $\mathbf{V}$ is closed under linear combination.

5. A set of vectors $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ in a subspace $\mathbf{V}$ is a basis for $\mathbf{V}$ if each $\mathbf{x} \in \mathbf{V}$ can be expressed uniquely as a linear combination of $\mathbf{x}_1, ...\mathbf{x}_k$. The dimension of $\mathbf{V}$, denoted $\dim(\mathbf{V})$, is the number of vectors in a basis for $\mathbf{V}$.

Result: A set of vectors in $\mathbf{R}^n$ is a basis for $\mathbf{R}^n$ iff there are n of them in all and they are linearly independent.

6. span($\mathbf{x}_1, ..., \mathbf{x}_k$) is the set of all linear combinations of $\mathbf{x}_1, ..., \mathbf{x}_k$.

7. $\mathbf{R}^{m,n} \equiv \left\{ A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \;\middle|\; a_{i,j} \text{ real} \right\}.$

8. For $A \in \mathbf{R}^{m,n}$...

column space(A) $\equiv$ span(columns of A).

row space(A) $\equiv$ span(rows of A).

range(A) $\equiv \{\mathbf{y} \in \mathbf{R}^m \mid \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in R^n\}$.

null(A) $\equiv \{\mathbf{x} \in \mathbf{R}^n \mid A\mathbf{x} = 0\}$.

rank(A) $\equiv$ dim(column space(A)) (= dim(row space (A))

9. The transpose of $A \in \mathbf{R}^{m,n}$ is a matrix $A^T \in \mathbf{R}^{n,m}$ whose i, j^{th} entry is $a_{j,i}$.

10. $A \in \mathbf{R}^{n,n}$ is nonsingular if there exists a matrix $A^{-1} \in \mathbf{R}^{n,n}$ such that $AA^{-1} = I$, where I is the $n \times n$ identity matrix. If A^{-1} exists, it is also the case that $A^{-1}A = I$.

11. *Result*: $(AB)^T = B^T A^T$; $(AB)^{-1} = B^{-1}A^{-1}$ (assuming A and B are nonsingular).

12. The inner product (or dot product, or scalar product) of two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ is:

$$(\mathbf{x}, \mathbf{y}) \equiv \sum_{i=1}^{n} x_i y_i \ \ (= \mathbf{x}^T \mathbf{y}).$$

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

13. Two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ are orthogonal if $(\mathbf{x}, \mathbf{y}) = 0$.

14. $\{\mathbf{x}_1, \ldots, x_k\}$ is an orthogonal set of vectors if $(\mathbf{x}_i, \mathbf{x}_j) = 0$ for $i \neq j$.

15. $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is an orthonormal set if

$$(\mathbf{x}_i, \mathbf{x}_j) = \delta_{i,j} \equiv \begin{cases} 1, \ i = j \\ 0, \ i \neq j. \end{cases}$$

Result: An orthonormal set of vectors is linearly independent. In fact, if $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ is an orthornormal set and $\mathbf{x} = c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$, then $c_i = (\mathbf{x}, \mathbf{x}_i), \ i = 1, ..., k$.

16. An orthogonal matrix is a real matrix whose transpose is its inverse. Example:

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \qquad \text{(rotation through angle } \theta \text{ in } R^2\text{)}$$

Results:

(i) The product of orthogonal matrices is orthogonal.
Proof: Let Q_1, Q_2 be orthogonal. Then

$$(Q_1Q_2)^T(Q_1Q_2) = Q_2^TQ_1^TQ_1Q_2 = Q_2^TIQ_2 = Q_2^TQ_2 = I.$$

(ii) A real matrix is orthogonal iff its columns (rows) comprise an orthonormal set of vectors.

Proof: Note that matrix multiplication can be defined in terms of inner products as follows: $(AB)_{i,j} =$ (row i of A, column j of B). Thus Q is orthogonal $\iff Q^TQ = I \iff (Q^TQ)_{i,j} = \delta_{i,j} \iff$ (row i of Q^T, column j of Q) $= \delta_{i,j} \iff (q^{(i)}, q^{(j)}) = \delta_{i,j}$, where $q^{(j)}$ is the j^{th} column of Q.

17. Determinant. Recursive definition:

$$\det A_{n\times n} = \begin{cases} a_{1,1} \quad \text{if } n = 1 \\ \displaystyle\sum_{j=1}^{n} (-1)^{(1+j)} a_{1,j} \cdot \det A'(1,j) \quad \text{if } n > 1, \end{cases}$$

where $A'(i,j)$ is the $n-1$ by $n-1$ matrix gotten by deleting row i, column j from A.

Results:

(i) Any row or column can be used for cofactor expansion (not just row 1, as in above definition).

(ii) Addition of a multiple of a row (column) of A to another row (column) of A has no effect on $\det A$.

(iii) Interchanging two rows (columns) of A changes the sign of $\det A$.

(iv) $\det(AB) = \det A \cdot \det B$.

(v) The determinant of a lower (upper) triangular matrix is the product of the entries on its main diagonal.

Example: Pencil and paper computation of $\det A$ using cofactor expansions:

$$\det \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 2 & 1 & 2 \\ 0 & 3 & -2 & 0 \\ 0 & -1 & 2 & 1 \end{bmatrix} = -\det \begin{bmatrix} 1 & -1 & 0 \\ 3 & -2 & 0 \\ -1 & 2 & 1 \end{bmatrix} = -\det \begin{bmatrix} 1 & -1 \\ 3 & -2 \end{bmatrix} = -1.$$

Computer algorithm for computing $\det A$ for a general matrix: Apply Gaussian elimination to A, reducing it to an upper triangular matrix U. Then set

$$\det A = (-1)^m \cdot u_{1,1} \cdots u_{n,n},$$

where m is the number of row interchanges performed.

18. $A_{n\times n}\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} \in \mathbf{R}^n$...

$\iff$ A is nonsingular, i.e., A^{-1} exists

$\iff$ $\det(A) \neq 0$

$\iff$ $A\mathbf{x} = 0$ has only the trivial solution $\mathbf{x} = 0$

$\iff$ the columns (rows) of A are linearly independent

$\iff$ $\text{rank}(A) = n$.

19. An eigenvector of $A_{n\,\times\,n}$ is a vector $\mathbf{v} \neq 0$ which is transformed by A into a multiple of itself, i.e., $A\mathbf{v} = \lambda\mathbf{v}$. The scalar λ is an eigenvalue of A.

20. Characteristic polynomial of $A_{n\,\times\,n}$:

$$f(\lambda) \equiv \det(\lambda I - A). \tag{*}$$

Results:

(i) $f(\lambda)$ is an n^{th} degree polynomial in λ whose roots are the eigenvalues of A:

$$f(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_n). \tag{**}$$

(ii) $\det(A) = \Pi_{i=1}^n \lambda_i$; trace(A) $\equiv \sum_{i=1}^n a_{i,i} = \sum_{i=1}^n \lambda_i$.
(These can be shown by equating the λ^0 and λ^1 coefficients in representations (*) and (**) for $f(\lambda)$.)

(iii) If $f(\lambda)$ has no multiple roots, A has n linearly independent eigenvectors, one per eigenvalue.

(iv) If $A_{n\times n}$ is symmetric, its eigenvalues are <u>real</u> and it has <u>n orthonormal</u> eigenvectors.

<u>Examples</u>

$$1.\ A=\begin{bmatrix}1&1&0\\0&2&1\\0&0&3\end{bmatrix}\qquad\rightarrow\qquad \det(\lambda I-A)=(\lambda-1)(\lambda-2)(\lambda-3)$$

$$\lambda_1=1,\ \mathbf{v}_1=\begin{pmatrix}1\\0\\0\end{pmatrix};\qquad \lambda_2=2,\ \mathbf{v}_2=\begin{pmatrix}1\\1\\0\end{pmatrix};\qquad \lambda_3=3,\ \mathbf{v}_3=\begin{pmatrix}1\\2\\2\end{pmatrix}.$$

$$2.\ A=\begin{bmatrix}1&0&1\\2&2&1\\-1&0&0\end{bmatrix}\qquad\rightarrow\qquad \det(\lambda I-A)=(\lambda-2)(\lambda^2-\lambda+1)$$

$$\lambda_1=2,\ \mathbf{v}_1=\begin{pmatrix}0\\1\\0\end{pmatrix};\qquad \lambda_2=\frac{1}{2}+\frac{\sqrt{3}}{2}i,\ \mathbf{v}_2=\begin{pmatrix}-\frac{1}{2}-\frac{\sqrt{3}}{2}i\\-\frac{1}{2}+\frac{\sqrt{3}}{2}i\\1\end{pmatrix};$$

$$\lambda_3=\frac{1}{2}-\frac{\sqrt{3}}{2}i,\ \mathbf{v}_3=\begin{pmatrix}-\frac{1}{2}+\frac{\sqrt{3}}{2}i\\-\frac{1}{2}+\frac{\sqrt{3}}{2}i\\1\end{pmatrix}.$$

$$3.\ A=\begin{bmatrix}0&1\\0&0\end{bmatrix}\qquad\rightarrow\qquad \det(\lambda I-A)=\lambda^2$$

$$\lambda_1=\lambda_2=0,\ \mathbf{v}_1=\begin{pmatrix}1\\0\end{pmatrix};\qquad \text{only one eigenvector.}$$

$$4.\ A=\begin{bmatrix}3&-1&0\\-1&2&-1\\0&-1&3\end{bmatrix}\qquad\rightarrow\qquad \det(\lambda I-A)=(\lambda-4)(\lambda-3)(\lambda-1)$$

$$\lambda_1=4,\ \mathbf{v}_1=\frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\end{pmatrix};\quad \lambda_2=3,\ \mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix};\quad \lambda_3=1,\ \mathbf{v}_3=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\2\\1\end{pmatrix}.$$

(v) Suppose $A_{n\times n}$ has n linearly independent eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$ with corresponding eigenvalues $\lambda_1,\ldots,\lambda_n$. Define matrices S,Λ as follows:

$$S\equiv[\mathbf{v}_1,\ldots,\mathbf{v}_n]\,;\quad \Lambda\equiv\begin{bmatrix}\lambda_1&&\\&\ddots&\\&&\lambda_n\end{bmatrix}.$$

Then $\Lambda = S^{-1}AS,\ A = S\Lambda S^{-1}$.

Proof:

$$\begin{aligned} AS = A\left[\mathbf{v}_1, \ldots, \mathbf{v}_n\right] = \left[A\mathbf{v}_1, \ldots, A\mathbf{v}_n\right] &= \left[\lambda_1\mathbf{v}_1, \ldots, \lambda_n\mathbf{v}_n\right] \\ &= \left[\mathbf{v}_1, \ldots, \mathbf{v}_n\right]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = S\Lambda. \end{aligned}$$

Note: If A is symmetric and $\{\mathbf{v}_i\}$ are an orthonormal set, then S is an orthogonal matrix - call it Q - and the above becomes $\Lambda = Q^TAQ,\ A = Q\Lambda Q^T$.

21. Gerschgorin's circle theorem. Let $A_{n\times n}$ have complex (or real) entries. To each row of A, associate a disk

$$D_i = \{\lambda \in C \big|\ |\lambda - a_{ii}| \le \sum_{j\ne i} |a_{ij}|\}$$

in the complex plane. Then:
(i) λ an eigenvalue of $A \implies \lambda \in D_i$ for some i.
(ii) If $\cup_i D_i$ separates into disjoint sets, the number of eigenvalues in each set equals the number of disks in that set.

Example

$$A = \begin{bmatrix} -10 & -3 & 2 \\ 1 & 2 & -1 \\ -1 & 3 & 5 \end{bmatrix}$$

Proof of (i): Suppose λ is an eigenvalue of $A_{n\times n}$. Then

$$A\mathbf{x} = \lambda\mathbf{x}, \quad \text{for some } \mathbf{x} \ne 0.$$

Equivalently,

$$\sum_{j=1}^{n} a_{i,j}x_j = \lambda x_i, \ \ i = 1, \ldots, n.$$

Choose k such that $|x_k| = \max_i |x_i|$. We will show that $\lambda \in D_k$. The k^{th} of the above equations can be written

$$(\lambda - a_{k,k})x_k = \sum_{j\ne k} a_{k,j}x_j.$$

Thus

$$|\lambda - a_{k,k}| \le \sum_{j\ne k} |a_{k,j}\frac{x_j}{x_k}| = \sum_{j\ne k} |a_{k,j}|.$$

22. spectral radius of A:

$$\rho(A) \equiv \max_i |\lambda_i(A)|.$$

Result:

$$\rho(A) \le \max_i \left\{ \sum_{j=1}^{n} |a_{i,j}| \right\}.$$

This follows from part (i) of Gerschgorin.

23. *Results*:
(i) The eigenvalues of A^{-1} are the reciprocals of those of A. ($A\mathbf{v} = \lambda\mathbf{v} \implies A^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}$.)
(ii) If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k. ($A\mathbf{v} = \lambda\mathbf{v} \implies A^k\mathbf{v} = \lambda^k\mathbf{v}$.)
(iii) Let $p(x) = c_0 + c_1 x + \cdots + c_k x^k$ be a polynomial in x. For a matrix argument A, define $p(A) = c_0 I + c_1 A + \cdots + c_k A^k$ (also a matrix). Assuming λ is an eigenvalue of A, then $p(\lambda)$ is an eigenvalue of $p(A)$. ($A\mathbf{v} = \lambda\mathbf{v} \implies p(A)\mathbf{v} = p(\lambda)\mathbf{v}$.)
(iv) Q orthogonal $\implies Q^T A Q$ has the same eigenvalues as A.

24. A matrix $A \in \mathbf{R}^{n,n}$ is positive definite if $\mathbf{x}^T A \mathbf{x} > 0$ for all nonzero vectors $\mathbf{x} \in \mathbf{R}^n$.

Result: A symmetric matrix is positive definite if and only if its eigenvalues are positive.
Proof: Write $A = Q\Lambda Q^T$, Q orthogonal. Then $\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y}$, $\mathbf{y} = Q\mathbf{x}$. Thus $\mathbf{x}^T A \mathbf{x} > 0$ for all $\mathbf{x} \ne 0 \iff \mathbf{y}^T\Lambda\mathbf{y} > 0$ for all $\mathbf{y} \ne 0 \iff \lambda_i(A) > 0, i = 1, ..., n$.

25. vector norm: a function $\| \cdot \|$ from $\mathbf{R}^n$ to nonnegative scalars satisfying the following axioms:
(i) $\|\mathbf{x}\| > 0$ if $\mathbf{x} \ne 0$, $\|0\| = 0$.
(ii) $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$, α a scalar.
(iii) $\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$.

Examples: $\|\mathbf{x}\|_1 \equiv \sum_{i=1}^n |x_i|$
$\|\mathbf{x}\|_2 \equiv \sqrt{\sum_{i=1}^n x_i^2}$
$\|\mathbf{x}\|_\infty \equiv \max_i |x_i|$

26. *Results*:
(i) If $Q_{n\times n}$ is orthogonal, then $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.
(ii) $|(\mathbf{x}, \mathbf{y})| \le \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$ (Schwarz inequality)

27. matrix norm: a function, $\| \cdot \|$, from $\mathbf{R}^{m,n}$ to nonnegative scalars satisfying the following axioms:
(i) $\|A\| > 0$ if $A \ne 0$, $\|0\| = 0$.
(ii) $\|\alpha A\| = |\alpha| \cdot \|A\|$, α a scalar.
(iii) $\|A + B\| \le \|A\| + \|B\|$.

Examples: $\|A\|_1 \equiv \max_{\mathbf{x}\ne 0} \frac{\|Ax\|_1}{\|\mathbf{x}\|_1}$; analogous definitions for $\|A\|_2, \|A\|_\infty$. Such norms are called *natural matrix norms*; they satisfy three other important properties:

(iv) $\|AB\| \leq \|A\| \cdot \|B\|$.
(v) $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$.
(iv) $\|I_{n\times n}\| = 1$.

An example of a matrix norm that is not a natural matrix norm is the *Frobenius* norm: $\|A\|_F \equiv \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2}$ $(= \sqrt{\text{trace}(A^TA)}$ via 20, part(ii).) Note that $\|I_{n\times n}\| = n$; thus property (iv) does not hold.

28. *Result*:

$$\|A\|_2 = \sqrt{\rho(A^TA)}$$

Proof: By definition,

$$\|A\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}.$$

Observe that

$$\|A\mathbf{x}\|_2^2 = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = (\mathbf{x}, A^T A\mathbf{x}) \tag{1}$$

and that A^TA is symmetric $[(A^TA)^T = A^T(A^T)^T = A^TA]$. Thus there exist orthonormal $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and real $\{\mu_1, ..., \mu_n\}$ such that $A^TA\mathbf{v}_i = \mu_i\mathbf{v}_i$, $i = 1, ..., n$, where $(\mathbf{v}_i, \mathbf{v}_j) = \delta_{i,j}$. Taking $\mathbf{x} = \mathbf{v}_i$ in (1) gives

$$\|A\mathbf{v}_i\|_2^2 = (\mathbf{v}_i, \mu_i\mathbf{v}_i) = \mu_i.$$

The eigenvalues of A^TA are therefore non-negative. We assume they are ordered as follows:

$$\mu_1 \geq ... \geq \mu_n \geq 0.$$

Any vector $\mathbf{x}$ can be written

$$\mathbf{x} = \sum_{i=1}^n c_i\mathbf{v}_i$$

for some set of c_i's. Using (1), this representation gives

$$\begin{aligned}\|A\mathbf{x}\|_2^2 &= (\sum_{i=1}^n c_i\mathbf{v}_i, A^TA\sum_{j=1}^n c_j\mathbf{v}_j) = (\sum_{i=1}^n c_i\mathbf{v}_i, \sum_{j=1}^n c_j\mu_j\mathbf{v}_j)\\ &= \sum_{i=1}^n\sum_{j=1}^n c_ic_j\mu_j(\mathbf{v}_i, \mathbf{v}_j) = \sum_{i=1}^n c_i^2\mu_i.\end{aligned}$$

Similarly,

$$\|\mathbf{x}\|_2^2 = (\sum_{i=1}^n c_i\mathbf{v}_i, \sum_{j=1}^n c_j\mathbf{v}_j) = \sum_{i=1}^n\sum_{j=1}^n c_ic_j(\mathbf{v}_i, \mathbf{v}_j) = \sum_{i=1}^n c_i^2.$$

Thus

$$\text{(2)} \qquad \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sqrt{\frac{\sum_{i=1}^n c_i^2 \mu_i}{\sum_{i=1}^n c_i^2}}.$$

It is now evident that $\max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$ occurs for $c_1 \neq 0$, $c_2 = \cdots = c_n = 0$, which corresponds to $\mathbf{x} =$ a multiple of $\mathbf{v}_1$. Thus by (2):

$$\|A\|_2 = \sqrt{\mu_1} = \sqrt{\rho(A^T A)}.$$

Note that if A is symmetric, $\rho(A^T A) = \rho(A^2) = [\rho(A)]^2$ (the eigenvalues of A^2 are the squares of those of A). Thus

$$\|A\|_2 = \rho(A) \quad \text{for a symmetric } A.$$

29. *Result*:

$$\|A\|_\infty = \max_i \left(\sum_{j=1}^n |a_{i,j}| \right).$$

Proof: By definition,

$$\|A\|_\infty = \max_{x\neq 0} \frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty}.$$

Now

$$|(A\mathbf{x})_i| = |\sum_{j=1}^n a_{i,j} x_j| \leq \sum_{j=1}^n |a_{i,j}|\,|x_j| \leq \left(\sum_{j=1}^n |a_{i,j}| \right) \|x\|_\infty \leq \max_i \left(\sum_{j=1}^n |a_{i,j}| \right) \|\mathbf{x}\|_\infty.$$

Thus

$$(*) \qquad \frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} \leq \max_i \left(\sum_{j=1}^n |a_{i,j}| \right).$$

Now choose k so that $\max_i \sum_{j=1}^n |a_{i,j}| = \sum_{j=1}^n |a_{k,j}|$ and define a vector x as follows:

$$x_j = \begin{cases} -1, & a_{k,j} < 0 \\ +1, & a_{k,j} \geq 0. \end{cases}$$

For this $\mathbf{x}$, it can easily be verified that the left and right hand sides of (*) are equal. This establishes the desired result.

30. *Result*: If $P_{m\times m}$ and $Q_{n\times n}$ are orthogonal and $A \in \mathbf{R}^{m,n}$, then $\|PAQ\|_2 = \|A\|_2$, $\|PAQ\|_F = \|A\|_F$.
Proof: $\|PAQ\|_2^2 = \rho((PAQ)^T PAQ) = \rho(Q^T A^T P^T PAQ) = \rho(Q^T A^T AQ) = \rho(A^T A) = \|A\|_2^2$. (see 23, part (iv))
Similarly, $\|PAQ\|_F^2 = \text{trace}((PAQ)^T(PAQ)) = \text{trace}(Q^T A^T AQ) = \text{trace}(A^T A) = \|A\|_F^2$ using 23, (iv) again and 20, (ii).

31. *Result*: Condition number bound
Suppose $A\mathbf{x} = \mathbf{b}$ and $(A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$. Then

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}\right) \qquad \text{provided } \kappa(A)\frac{\|\delta A\|}{\|A\|} < 1$$

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \lesssim \kappa(A)\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}\right) \qquad \text{provided } \kappa(A)\frac{\|\delta A\|}{\|A\|} \ll 1$$

Proof: We have

$$\begin{aligned}
&(A + \delta A)\,\mathbf{x} + (A + \delta A)\,\delta\mathbf{x} = \mathbf{b} + \delta\mathbf{b} \\
&(A + \delta A)\,\delta\mathbf{x} = -\delta A\ \mathbf{x} + \delta\mathbf{b} \\
&(I + A^{-1}\delta A)\,\delta\mathbf{x} = -A^{-1}\delta A\ \mathbf{x} + A^{-1}\delta\mathbf{b}.
\end{aligned}$$

Thus

$$\|\delta\mathbf{x}\|\ \left(1 - \|A^{-1}\|\,\|\delta A\|\right) \le \|(I + A^{-1}\delta A)\,\delta\mathbf{x}\| \le \|A^{-1}\|\,\|\delta A\|\,\|\mathbf{x}\| + \|A^{-1}\|\,\|\delta\mathbf{b}\|.$$

Therefore, provided $\|A^{-1}\|\,\|\delta A\| < 1$:

$$\begin{aligned}
\|\delta\mathbf{x}\| &\le \frac{\|A^{-1}\|\,\|\delta A\|\,\|\mathbf{x}\| + \|A^{-1}\|\,\|\delta\mathbf{b}\|}{1 - \|A^{-1}\|\,\|\delta A\|} \\
\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} &\le \frac{\kappa(A)}{1 - \kappa(A)\,\frac{\|\delta A\|}{\|A\|}}\left(\frac{\delta A\|}{\|A\|} + \frac{\|\delta\mathbf{b}\|}{\|A\|\,\|\mathbf{x}\|}\right).
\end{aligned}$$

The desired result follows from $\|\mathbf{b}\| \le \|A\|\,\|\mathbf{x}\|$.

Special case: if $\delta A = 0$, then

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \kappa(A)\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$