Analysis review: O notation, Taylor series, and linear algebra

Patrick Breheny

August 31

Introduction

One final lecture of analysis review, in which we go over three indispensable tools that we will use constantly in the remainder of the course:

- O, o notation
- Taylor series expansions
- Linear algebra

o-notation: Motivation

- When investigating the asymptotic behavior of functions, it is often convenient to replace unwieldy expressions with compact notation
- For example, if we encountered the mathematical expression

$$x^2 + a - a,$$

  we would obviously want to replace it with x^2 since $a - a = 0$
- However, what if we encounter something like

$$x^2 + \frac{5\theta}{\sqrt{n}} - \frac{3\theta}{n+5}?$$

- We can no longer just replace this with x^2

o-notation: Motivation (cont'd)

- However, as n gets larger, the expression gets closer and closer to x^2
- It would be convenient to have a shorthand notation for this, something like $x^2 + o_n$, where o_n represents some quantity that becomes negligible as n becomes large
- This is the basic idea behind o-notation, and its simplifying powers become more apparent as the mathematical expression we are dealing with becomes more complicated:

$$\frac{x^2 + \frac{5\theta}{\sqrt{n}} - \frac{3\theta}{n+5}}{(n^2+5n-2)/(n^2-3n+1)} + \frac{\exp\{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2\}}{2\sqrt{n}\theta\int_0^\infty g(s)ds}$$

o-notation

- There is where something called o-notation comes in: a formal way of handling terms that effectively "cancel out" as we take limits
- **Definition:** A sequence of numbers X_n is said to be $o(1)$ if it converges to zero. Likewise, X_n is said to be $o(r_n)$ if

$$\frac{X_n}{r_n} \to 0$$

  as $n \to \infty$.
- When the rate is constant, o notation is pretty straightforward:

$$x^2 + \frac{5\theta}{\sqrt{n}} - \frac{3\theta}{n+5} = x^2 + o(1)$$

o-notation remarks

- When the rate is not constant, expressions are a bit harder to think about – it helps to go over some cases:
- For example:
  - $X_n \to \infty$, but $r_n \to \infty$ even faster:

$$n = o(n^2)$$

  - $r_n \to 0$, but $X_n \to 0$ even faster:

$$\frac{1}{n^2} = o(1/n)$$

- Note that this extension to non-constant rates are what makes o different from and more flexible than simpler statements about limits (convergence to a constant)

O-notation

- A very useful companion of o-notation is O-notation, which denotes whether or not a term remains bounded as $n \to \infty$
- **Definition:** A sequence of numbers X_n is said to be $O(1)$ if there exist M and N such that

$$|X_n| < M$$

for all $n > N$. Likewise, X_n is said to be $O(r_n)$ if there exist M and N such that for all $n > N$,

$$\left| \frac{X_n}{r_n} \right| < M.$$

O-notation remarks

- For example,

$$\frac{\exp\{-\frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2\}}{2\sqrt{n}\theta\int_0^\infty g(s)ds} = O(n^{-1/2})$$

- Note that $X_n = O(1)$ does not necessarily mean that X_n is bounded, just that it is eventually bounded
- Note also that just because a term is $O(1)$, this does not necessarily mean that it has a limit; for example,

$$\sin\left(\frac{n\pi}{2}\right) = O(1),$$

even though the sequence does not converge

O-notation remarks (cont'd)

- You may encounter the phrase "x_n is of order r_n"; this is a bit ambiguous
- The author may mean that $x_n = O(r_n)$
- However, it might also mean something stronger: that there exist constants m and M such that

$$m \leq \left| \frac{x_n}{r_n} \right| \leq M$$

  for large enough n; i.e., the ratio is bounded above but also bounded below
- In other words, $x_n = O(r_n)$ but in addition $x_n \neq o(r_n)$; some authors use the notation $x_n \asymp r_n$ to denote this situation

Informative-ness of o and O notation

- There are typically many ways of writing an expression using O notation, although not all of them will be equally informative
- For example, if $x_n = 1 + \frac{1}{n}$, then all of the following are true:

$$x_n = 1 + o(1)$$
$$x_n = 1 + O(1) \qquad \text{(least informative)}$$
$$x_n = 1 + O(\tfrac{1}{n}) \qquad \text{(most informative)}$$

Algebra of O, o notation

O, o-notation are useful in combination because simple rules govern how they interact with each other

Theorem: For $a \leq b$:

$$O(1) + O(1) = O(1)$$
$$o(1) + o(1) = o(1)$$
$$o(1) + O(1) = O(1)$$
$$O(1)O(1) = O(1)$$
$$O(1)o(1) = o(1)$$
$$\{1 + o(1)\}^{-1} = O(1)$$
$$O\{O(1)\} = O(1)$$
$$o\{O(1)\} = o(1)$$
$$o(r_n) = r_n o(1)$$
$$O(r_n) = r_n O(1)$$
$$O(n^a) + O(n^b) = O(n^b)$$
$$o(n^a) + o(n^b) = o(n^b)$$

Remarks

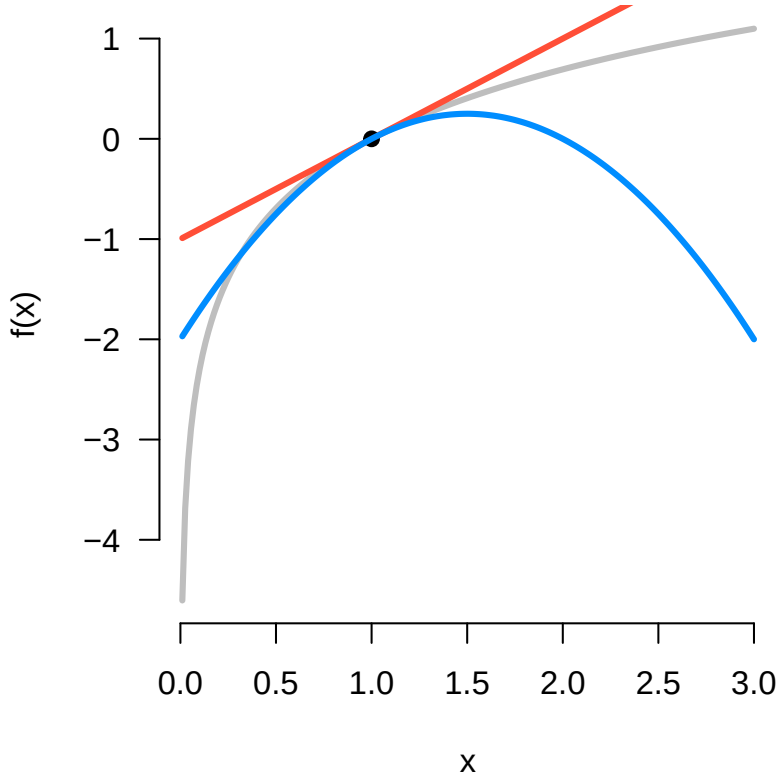
- O, o "equations" are meant to be read left-to-right; for example, $O(\sqrt{n}) = O(n)$ is a valid statement, but $O(n) = O(\sqrt{n})$ is not
- **Exercise:** Determine the order of

$$n^{-2}\left\{(-1)^n \sqrt[n]{2} + (1 + \tfrac{1}{n})^n\right\}.$$

- As we will see in a week or two, there are stochastic equivalents of these concepts, involving convergence in probability and being bounded in probability
- As such, we won't do a great deal with O, o-notation right now, but will use the stochastic equivalents extensively

Taylor series: Introduction

- As we will see (many times!), it is useful to be able to approximate a complicated function with a simple polynomial:



- This is the idea behind Taylor series expansions

Taylor series: Introduction (cont'd)

- It is difficult to overstate the importance of Taylor series expansions to statistical theory, and for that reason we are now going to cover them fairly extensively
- In particular, Taylor's theorem comes in a number of versions, and it is worth knowing several of them, since they come up in statistics quite often
- Furthermore, students often have not seen the multivariate versions of these expansions

Taylor's theorem

- **Theorem (Taylor):** Suppose n is a positive integer and $f : \mathbb{R} \to \mathbb{R}$ is n times differentiable at a point x_0. Then

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n(x, x_0),$$

  where the remainder R_n satisfies

$$R_n(x, x_0) = o(|x - x_0|^n) \text{ as } x \to x_0$$

- If $f^{(n+1)}(x_0)$ exists, you could also say that R_n is $O(|x - x_0|^{n+1})$
- This form of the remainder is sometimes called the *Peano* form

Taylor's theorem: Lagrange form

- **Theorem (Taylor):** Suppose $f : \mathbb{R} \to \mathbb{R}$ is $n+1$ times differentiable on an open interval containing x_0. Then for any point x in that interval, there exists $\bar{x} \in (x, x_0)$:

$$R_n(x, x_0) = \frac{f^{(n+1)}(\bar{x})}{(n+1)!}(x - x_0)^{n+1}.$$

- This is also known as the *mean-value form*, as the mean value theorem is the central idea in proving the result

Comparing the two forms

- Comparing the Basic and Lagrange forms for a second-order expansion,

$$f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + o(|x - x_0|^2)$$
$$f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(\bar{x})(x - x_0)^2$$

- We can see that in the second case, we have a simpler expression, but to obtain it, we require f'' to exist along the entire interval from x to x_0, not just at the point x_0

Example: Absolute value

- For example, consider approximating the function $f(x) = |x|$ at $x_0 = -0.1$
- Note that f' exists at x_0, but not at 0
- The basic form of Taylor's theorem says that if we get close enough to x_0, the approximation $f(-0.1) + f'(-0.1)(x + 0.1)$ becomes very accurate – indeed, the remainder is exactly zero for any x within 0.1 of x_0
- However, suppose $x = 0.2$; since f is not differentiable at zero, we are not guaranteed the existence of a point $\bar{x}$ such that

$$f(0.2) = f(-0.1) + 0.3f'(\bar{x});$$

and indeed in this case no such point exists

Lagrange bound

- One reason why the Lagrange form is more powerful is that it allows us to establish error bounds – to know exactly how close x must be to x_0 in order to ensure that the approximation error is less than ϵ
- In particular, if there exists an M such that $\left|f^{(n+1)}(x)\right| \leq M$ over the interval (x, x_0), then

$$|R_n(x)| \leq \frac{M}{(n+1)!} |x - x_0|^{n+1}$$

Multivariable forms of Taylor's theorem

- We now turn our attention to the multivariate case
- For the sake of clarity, I'll present the first- and second-order expansions for each of the previous forms, rather than abstract formulae involving $f^{(n)}$
- Lastly, I'll provide a form that goes out to third order, although higher orders are less convenient as they can't be represented compactly using vectors and matrices
- Note that these forms are only covering the case of scalar-valued functions $f : \mathbb{R}^d \to \mathbb{R}$; we will need results for the vector-valued case $f : \mathbb{R}^d \to \mathbb{R}^k$ as well, but we will go over that in a later lecture

Taylor's theorem

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable at a point $\mathbf{x}_0$. Then

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top (\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|)$$

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable at a point $\mathbf{x}_0$. Then

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2)$$

Taylor's theorem: Lagrange form

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\bar{\mathbf{x}}$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\bar{\mathbf{x}})^\top (\mathbf{x} - \mathbf{x}_0)$$

- **Theorem (Taylor):** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\bar{\mathbf{x}}$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \nabla^2 f(\bar{\mathbf{x}})(\mathbf{x} - \mathbf{x}_0)$$

- "$\bar{\mathbf{x}}$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$" means that there exists $w \in [0, 1]$ such that $\bar{\mathbf{x}} = w\mathbf{x} + (1 - w)(\mathbf{x}_0)$

Taylor's theorem: Third order

Theorem (Taylor): Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is three times differentiable on $N_r(\mathbf{x}_0)$. Then for any $\mathbf{x} \in N_r(\mathbf{x}_0)$, there exists $\bar{\mathbf{x}}$ on the line segment connecting $\mathbf{x}$ and $\mathbf{x}_0$ such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{j=1}^{d} \frac{\partial f(\mathbf{x}_0)}{\partial x_j}(x_j - x_{0j})$$
$$+ \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_j \partial x_k}(x_j - x_{0j})(x_k - x_{0k})$$
$$+ \frac{1}{6}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{\ell=1}^{d} \frac{\partial^3 f(\bar{\mathbf{x}})}{\partial x_j \partial x_k \partial x_\ell}(x_j - x_{0j})(x_k - x_{0k})(x_\ell - x_{0\ell}),$$

where $\partial f(\mathbf{x}_0)/\partial x_j$ is shorthand for $\partial f(\mathbf{x})/\partial x_j$ evaluated at $\mathbf{x}_0$

Linear algebra

- Our last mathematical topic to review/reference is linear algebra, which we will use right away in our next lecture on the multivariate normal distribution
- If this material is new to you, then I would encourage you to read this as well
- It is often useful to switch the rows and columns of a matrix around. The resulting matrix is called the *transpose* of the original matrix, and denoted with a superscript $^\top$ or an apostrophe $'$:

$$\mathbf{M} = \begin{bmatrix} 3 & 2 \\ 4 & -1 \\ -1 & 2 \end{bmatrix} \qquad \mathbf{M}^\top = \begin{bmatrix} 3 & 4 & -1 \\ 2 & -1 & 2 \end{bmatrix}$$

Linear and quadratic forms

- Matrix products involving linear and quadratic forms come up very often in statistics, and it is important to have an intuitive grasp on what they represent
- Here are some useful identities/relationships involving matrix products and their scalar representations:

$$\mathbf{a}^\top \mathbf{x} = \sum_i a_i x_i; \quad \mathbf{1}^\top \mathbf{x} = \sum_i x_i$$

$$\mathbf{A}^\top \mathbf{x} = (\sum_i a_{i1} x_i \quad \cdots \quad \sum_i a_{ik} x_i)^\top$$

$$\mathbf{a}^\top \mathbf{W} \mathbf{x} = \sum_i \sum_j a_i w_{ij} x_j; \quad \mathbf{a}^\top \mathbf{1} \mathbf{x} = \sum_i \sum_j a_i x_j$$

$$(\mathbf{AWB})_{ij} = \sum_k \sum_m a_{ik} w_{km} b_{mj}$$

Inverses

- **Definition:** The *inverse* of an $n \times n$ matrix $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, is the matrix satisfying $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.
- Note: We're sort of getting ahead of ourselves by saying that $\mathbf{A}^{-1}$ is "the" matrix satisfying $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$, but it is indeed the case that if a matrix has an inverse, the inverse is unique
- Some useful results:

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$
$$(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$$
$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
$$(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$$

Singular matrices

- However, not all matrices have inverses; for example

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

- There does not exist a matrix such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_2$
- Such matrices are said to be *singular*
- Remark: Only square matrices have inverses; an $n \times m$ matrix $\mathbf{A}$ might, however, have a *left inverse* (satisfying $\mathbf{B}\mathbf{A} = \mathbf{I}_m$) or *right inverse* (satisfying $\mathbf{A}\mathbf{B} = \mathbf{I}_n$)

Positive definite

- A related notion is that of a "positive definite" matrix, which applies to symmetric matrices
- **Definition:** A symmetric $n \times n$ matrix $\mathbf{A}$ is said to be *positive definite* if for all $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0 \qquad \text{if } \mathbf{x} \neq 0$$

- The two notions are related in the sense that if $\mathbf{A}$ is positive definite, then (a) $\mathbf{A}$ is not singular and (b) $\mathbf{A}^{-1}$ is also positive definite
- If $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$, then $\mathbf{A}$ is said to be *positive semidefinite*
- In statistics, these classifications are particularly important for variance-covariance matrices, which are always positive semidefinite (and positive definite, if they aren't singular)

Square root of a matrix

- These concepts are important with respect to knowing whether a matrix has a "square root"
- **Definition:** An $n \times n$ matrix $\mathbf{A}$ is said to have a *square root* if there exists a matrix $\mathbf{B}$ such that $\mathbf{BB} = \mathbf{A}$.
- **Theorem:** Let $\mathbf{A}$ be a positive definite matrix. Then there exists a unique matrix $\mathbf{A}^{1/2}$ such that $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
- Positive semidefinite matrices have square roots as well, although they aren't necessarily unique

Rank

- We also need to be familiar with the concept of matrix rank (there are many ways of defining rank; all are equivalent)
- **Definition:** The *rank* of a matrix is the dimension of its largest nonsingular submatrix.
- For example, the following 3×3 matrix is singular, but contains a nonsingular 2×2 submatrix, so its rank is 2:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & \cancel{3} \\ \cancel{2} & \cancel{4} & \cancel{6} \\ 1 & 0 & \cancel{1} \end{bmatrix}$$

- Note that a nonsingular $n \times n$ matrix has rank n, and is said to be *full rank*

Rank and multiplication

- There are many results and theorems involving rank; we're not going to cover them all, but it is important to know that rank cannot be increased through the process of multiplication
- **Theorem:** For any matrices $\mathbf{A}$ and $\mathbf{B}$ with appropriate dimensions, $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.
- In particular, $\text{rank}(\mathbf{A}^\top\mathbf{A}) = \text{rank}(\mathbf{AA}^\top) = \text{rank}(\mathbf{A})$.

Expectation and variance

- In addition, we need some results on expected values of vectors and functions of vectors
- First of all, we need to define expectation and variance as they pertain to random vectors
- **Definition:** Let $\mathbf{x} = (X_1 \; X_2 \; \cdots X_d)^\top$ denote a vector of random variables, then $\mathbb{E}(\mathbf{x}) = (\mathbb{E}X_1 \; \mathbb{E}X_2 \; \cdots \mathbb{E}X_d)^\top$. Meanwhile, $\mathbb{V}\mathbf{x}$ is a $d \times d$ matrix:

$$\mathbb{V}\mathbf{x} = \mathbb{E}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\} \text{ with elements}$$
$$(\mathbb{V}\mathbf{x})_{ij} = \mathbb{E}\left\{(X_i - \mu_i)(X_j - \mu_j)\right\},$$

where $\mu_i = \mathbb{E}X_i$. The matrix $\mathbb{V}\mathbf{x}$ is referred to as the *variance-covariance matrix* of $\mathbf{x}$.

Linear and quadratic forms

- Letting $\mathbf{A}$ denote a matrix of constants and $\mathbf{x}$ a random vector with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$,

$$\begin{aligned}
\mathbb{E}(\mathbf{A}^\top \mathbf{x}) &= \mathbf{A}^\top \boldsymbol{\mu} \\
\mathbb{V}(\mathbf{A}^\top \mathbf{x}) &= \mathbf{A}^\top \boldsymbol{\Sigma} \mathbf{A} \\
\mathbb{E}(\mathbf{x}^\top \mathbf{A} \mathbf{x}) &= \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} + \operatorname{tr}(\mathbf{A}\boldsymbol{\Sigma}),
\end{aligned}$$

  where $\operatorname{tr}(\mathbf{A}) = \sum_i A_{ii}$ is the trace of $\mathbf{A}$
- Some useful facts about traces:

$$\begin{aligned}
\operatorname{tr}(\mathbf{AB}) &= \operatorname{tr}(\mathbf{BA}) \\
\operatorname{tr}(\mathbf{A} + \mathbf{B}) &= \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}) \\
\operatorname{tr}(c\mathbf{A}) &= c \operatorname{tr}(\mathbf{A}) \\
\operatorname{tr}(\mathbf{A}) &= \operatorname{rank}(\mathbf{A}) \quad \text{if } \mathbf{AA} = \mathbf{A}
\end{aligned}$$

Eigendecompositions

- Finally, we'll also take a moment to introduce some facts about eigenvalues
- The most important thing about eigenvalues is that they allow us to "diagonalize" a matrix: if $\mathbf{A}$ is a symmetric $d \times d$ matrix, then it can be factored into:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top,$$

  where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ of $\mathbf{A}$ and the columns of $\mathbf{Q}$ are its eigenvectors
- Furthermore, eigenvectors are orthonormal, so we have $\mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$

Eigenvalues and "size"

- This is very helpful from a conceptual standpoint, as it allows us to separate the "size" of a matrix ($\mathbf{\Lambda}$) from its "direction(s)" ($\mathbf{Q}$)
- For example, we have already seen that one measure of the size of a matrix is based on $\lambda_{\max}$ (for a symmetric matrix, its spectral norm is its largest eigenvalue)
- In addition, the trace and determinant, two other ways of quantifying the "size" of a matrix, are simple functions of the eigenvalues:
  - $\text{tr}(\mathbf{A}) = \sum_i \lambda_i$
  - $|\mathbf{A}| = \prod_i \lambda_i$

Eigenvalues and inverses

- Once one has obtained the eigendecomposition of $\mathbf{A}$, calculating its inverse is straightforward
- If $\mathbf{A}$ is not singular, then $\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^\top$; note that since $\mathbf{\Lambda}$ is diagonal, its inverse is trivial to calculate
- Even if $\mathbf{A}$ is singular, we can obtain something called a "generalized inverse": $\mathbf{A}^- = \mathbf{Q}\mathbf{\Lambda}^-\mathbf{Q}^\top$, where $(\mathbf{\Lambda}^-)_{ii} = \lambda_i^{-1}$ if $\lambda_i \neq 0$ and $(\mathbf{\Lambda}^-)_{ii} = 0$ otherwise
- Many other important properties of matrices can be deduced entirely from their eigenvalues:
  - $\mathbf{A}$ is positive definite if and only if $\lambda_i > 0$ for all i
  - $\mathbf{A}$ is positive semidefinite if and only if $\lambda_i \geq 0$ for all i
  - If $\mathbf{A}$ has rank r, then $\mathbf{A}$ has r nonzero eigenvalues and the remaining $d - r$ eigenvalues are zero

Extreme values

- Lastly, there is a connection between a matrix's eigenvalues and the extreme values of its quadratic form
- Let the eigenvalues $\lambda_1, \ldots, \lambda_d$ of $\mathbf{A}$ be ordered from largest to smallest. Over the set of all vectors $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$,

$$\max \mathbf{x}^\top \mathbf{A}\mathbf{x} = \lambda_1$$

and

$$\min \mathbf{x}^\top \mathbf{A}\mathbf{x} = \lambda_d$$