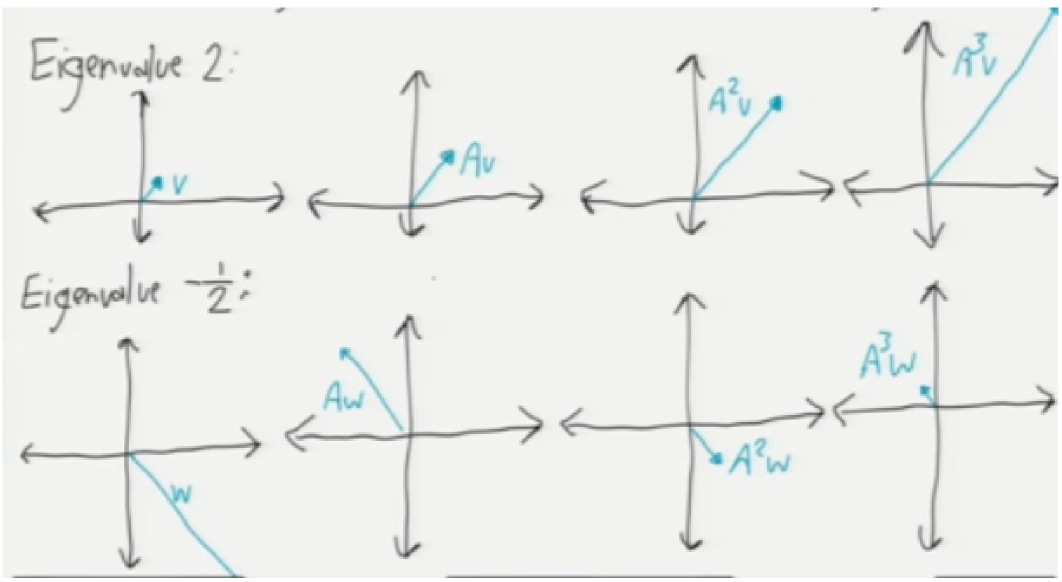
8 Eigenvectors and the Anisotropic Multivariate Gaussian Distribution

EIGENVECTORS

[I don't know if you were properly taught about eigenvectors here at Berkeley, but I sure don't like the way they're taught in most linear algebra books. So I'll start with a review. You all know the definition of an eigenvector:]

Given square matrix A, if $Av = \lambda v$ for some vector $v \neq 0$, scalar λ, then
v is an eigenvector of A and λ is the eigenvalue of A associated w/v.

[But what does that mean? It means that v is a magical vector that, after being multiplied by A, still points in the *same direction*, or in exactly the *opposite direction*.]



Draw this figure by hand (eigenvectors.png)

[For most matrices, most vectors don't have this property. So the ones that do are special, and we call them eigenvectors.]
[Clearly, when you scale an eigenvector, it's still an eigenvector. Only the direction matters, not the length. Let's look at a few consequences.]

Theorem: if v is eigenvector of A w/eigenvalue λ,
then v is eigenvector of A^k w/eigenvalue λ^k [we will use this later]

Proof: $A^2 v = A(\lambda v) = \lambda^2 v$, etc.

Theorem: moreover, if A is invertible,
then v is eigenvector of A^{-1} w/eigenvalue $1/\lambda$

Proof: $A^{-1}v = \frac{1}{\lambda}A^{-1}Av = \frac{1}{\lambda}v$ [look at the figures above, but go from right to left.]

[Stated simply: When you invert a matrix, the eigenvectors don't change, but the eigenvalues get inverted. When you square a matrix, the eigenvectors don't change, but the eigenvalues get squared.]

[Those theorems are pretty obvious. The next theorem is not obvious at all.]

Spectral Theorem: every real, symmetric $n \times n$ matrix has real eigenvalues and n eigenvectors that are mutually orthogonal, i.e., $v_i^\top v_j = 0 \quad$ for all $i \neq j$

[This takes about a page of math to prove.
One minor detail is that a matrix can have more than n eigenvector directions. If two eigenvectors happen to have the same eigenvalue, then every linear combination of those eigenvectors is also an eigenvector. Then you have infinitely many eigenvector directions, but they all span the same plane. So you just arbitrarily pick two vectors in that plane that are orthogonal to each other. By contrast, the set of eigenvalues is always uniquely determined by a matrix, including the multiplicity of the eigenvalues.]

We can use them as a basis for $\mathbb{R}^n$.

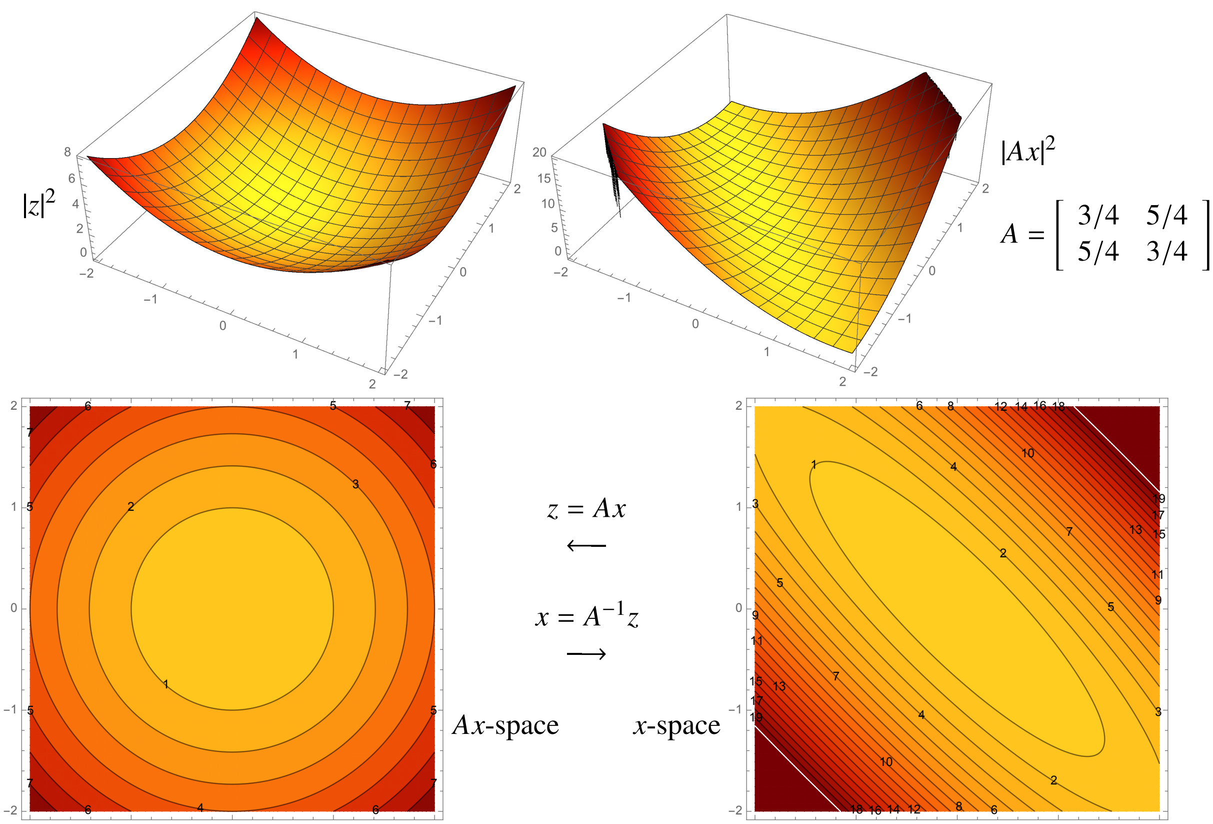
Quadratic Forms

[My favorite way to visualize a symmetric matrix is to graph something called *the quadratic form*, which shows how applying the matrix affects the length of a vector. The following example uses the same two eigenvectors and eigenvalues as above.]

$|z|^2 \quad = z^\top z \qquad \Leftarrow$ quadratic; isotropic; isosurfaces are spheres

$|Ax|^2 \quad = x^\top A^2 x \qquad \Leftarrow$ quadratic form of the matrix $A^2 \quad$ (A symmetric)

anisotropic; isosurfaces are ellipsoids



circlebowl.pdf, ellipsebowl.pdf, circles.pdf, ellipses.pdf

[Both figures at left graph $|z|^2$, and both figures at right graph $|Ax|^2$.
(Draw the stretch direction $(1, 1)$ with eigenvalue 2 and the shrink direction $(1, -1)$ with eigenvalue $-\frac{1}{2}$ on the ellipses at bottom right.)]

[The matrix A maps the ellipses on the left to the circles on the right. They're stretching along the direction with eigenvalue 2, and shrinking along the direction with eigenvalue $-1/2$. You can also think of this process in reverse, if you remember that A^{-1} has the same eigenvectors but reciprocal eigenvalues. The matrix A^{-1} maps the circles to the ellipses, shrinking along the direction with eigenvalue 2, and stretching along the direction with eigenvalue $-1/2$. I can prove that formally.]

$|Ax|^2 = 1$ is an ellipsoid with axes $v_1, v_2, \ldots, v_n$ and
radii $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$
because if v_i has length $1/\lambda_i$, $|Av_i|^2 = |\lambda_i v_i|^2 = 1 \quad \Rightarrow \quad v_i$ lies on the ellipsoid

[The reason the ellipsoid radii are the reciprocals of the eigenvalues is that it is the matrix A^{-1} that maps the spheres to ellipsoids. So each axis of the spheres gets scaled by 1/eigenvalue.]

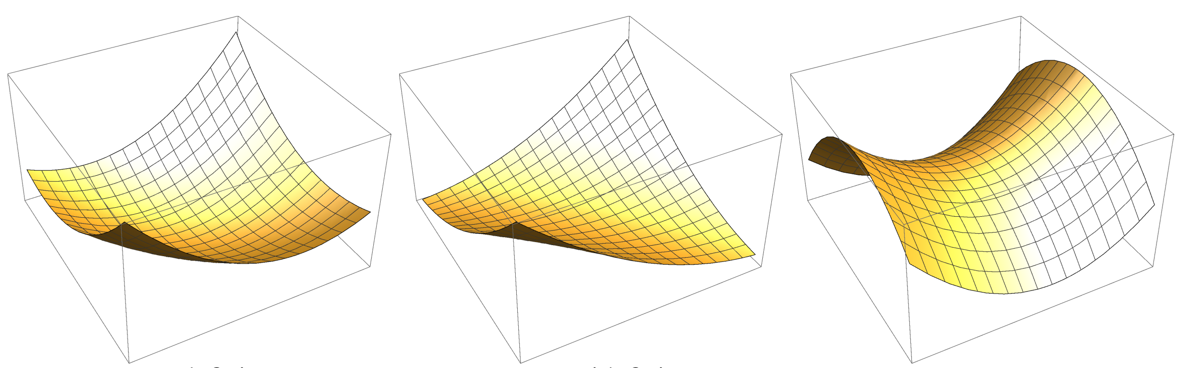
bigger eigenvalue $\Leftrightarrow$ steeper hill $\Leftrightarrow$ shorter ellipsoid radius
[$\uparrow$ bigger curvature, to be precise]

Alternate interpretation: ellipsoids are "spheres" $\{x : d(x, \text{center}) = \text{isovalue}\}$ in the <u>distance metric</u> A^2.
Call $M = A^2$ a <u>metric tensor</u> because the "distance" between points x & x' in this metric is

$$d(x, x') = |Ax - Ax'| = \sqrt{(x - x')^\top M (x - x')}$$

[This is the Euclidean distance in the Ax-space, but we think of it as an alternative metric for measuring distances in the original space.]
[I'm calling M a "tensor" because that's standard usage in Riemannian geometry, but don't worry about what "tensor" means. For our purposes, it's a matrix.]

A square matrix B is	<u>positive definite</u>	if $w^\top Bw > 0$ for all $w \neq 0$.	$\Leftrightarrow$ all eigenvalues positive
	<u>positive semidefinite</u>	if $w^\top Bw \geq 0$ for all w.	$\Leftrightarrow$ all eigenvalues nonnegative
	<u>indefinite</u>	if +ve eigenvalue & $-$ve eigenvalue	
	invertible	if no zero eigenvalue	



pos definite pos semidefinite indefinite

$\boxed{\text{posdef.pdf, possemi.pdf, indef.pdf}}$
[Examples of quadratic forms for positive definite, positive semidefinite, and indefinite matrices. Positive eigenvalues correspond to axes where the curvature goes up; negative eigenvalues correspond to axes where the curvature goes down. (Draw the eigenvector directions, and draw the flat trough in the positive semidefinite bowl.)]

[Our metric is A^2, so its eigenvalues are the squares of the eigenvalues of A, so A^2 cannot have a negative eigenvalue. Therefore, A^2 is positive semidefinite. A^2 might have an eigenvalue of zero, but if it does, it is technically not a "metric," and the isosurfaces are cylinders instead of ellipsoids. We're going to use these ideas to define Gaussian distributions, and for that, we'll need a strictly positive definite metric.]

Special case: A & M are diagonal $\Leftrightarrow$ eigenvectors are coordinate axes
$\Leftrightarrow$ ellipsoids are axis-aligned

[Draw axis-aligned isocontours for a diagonal metric.]

Building a Quadratic

[There are a lot of applications where you're given a matrix, and you want to extract the eigenvectors and eigenvalues. But I, personally, think it's more intuitive to go in the opposite direction. Suppose you have an ellipsoid in mind. Suppose you pick the ellipsoid axes and the radius along each axis, and you want to create the matrix that will fulfill the ellipsoid of your dreams.]

Choose n mutually orthogonal **unit** n-vectors $v_1, \ldots, v_n$ [so they specify an orthonormal coordinate system]
Let $V = [v_1 \quad v_2 \quad \ldots \quad v_n]$ $\Leftarrow n \times n$ matrix
Observe: $V^\top V = I$ [off-diagonal 0's because the vectors are orthogonal]
[diagonal 1's because they're unit vectors]

$$\Rightarrow \quad V^\top = V^{-1} \quad \Rightarrow \quad VV^\top = I$$

V is orthonormal matrix: acts like rotation (or reflection)

Choose some inverse radii λ_i:

$$\text{Let } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$ [diagonal matrix of eigenvalues]

Defn. of "eigenvector": $AV = V\Lambda$
[This is the same definition of eigenvector I gave you at the start of the lecture, but this is how we express it in matrix form, so we can cover all the eigenvectors in one statement.]

$\Rightarrow AVV^\top = V\Lambda V^\top$ [which leads us to . . .]

Theorem: $A = V\Lambda V^\top = \sum_{i=1}^{n} \lambda_i \underbrace{v_i v_i^\top}_{\text{outer product: } n \times n \text{ matrix, rank 1}}$ has chosen eigenvectors/values

This is a matrix factorization called the eigendecomposition.
Λ is the diagonalized version of A.
$V^\top$ rotates the ellipsoid to be axis-aligned.
[So now, if you want an ellipsoid or paraboloid with specified axes and radii, you know how to write it as an equation.]

Observe: $M = A^2 = V\Lambda V^\top V\Lambda V^\top = V\Lambda^2 V^\top$
Given a SPD metric tensor M, we can find a symmetric square root $A = M^{1/2}$:

compute eigenvectors/values of M
take square roots of M's eigenvalues
reassemble matrix A

[This will be useful to know when we try to understand the multivariate Gaussian distribution in its full generality. The first step of this algorithm—computing the eigenvectors and eigenvalues of a matrix—is much harder than the remaining two steps.]

<u>ANISOTROPIC GAUSSIANS</u>

[Let's revisit the multivariate Gaussian distribution, with different variances along different directions.]

$X \sim N(\mu, \Sigma)$ [X and μ are d-vectors. X is random with mean μ.]

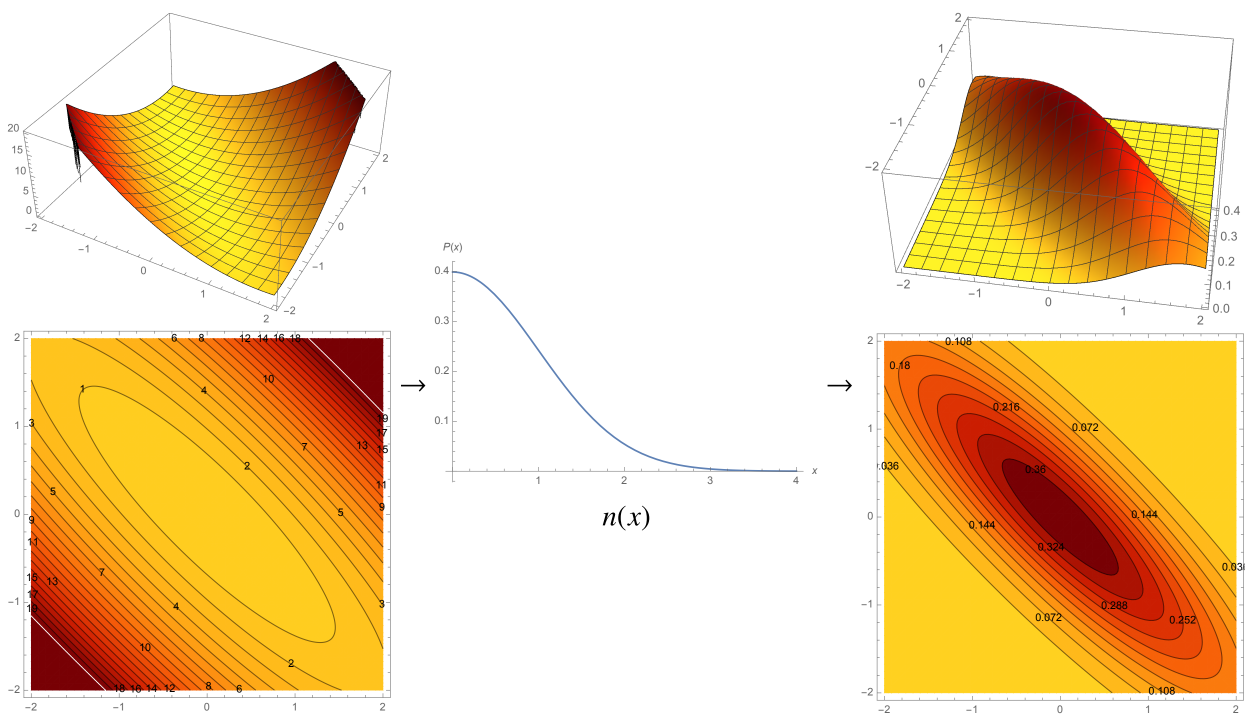
$$P(x) = \frac{1}{(\sqrt{2\pi})^d \underbrace{\sqrt{|\Sigma|}}_{\uparrow \text{ determinant of } \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)\right)$$

Σ is the $d \times d$ SPD <u>covariance matrix</u>.
Σ^{-1} is the $d \times d$ SPD <u>precision matrix</u>; serves as metric tensor.

Write $P(x) = \underset{\substack{\uparrow \\ \mathbb{R} \to \mathbb{R},\ \text{exponential}}}{n}(\underset{\substack{\uparrow \\ \mathbb{R}^d \to \mathbb{R},\ \text{quadratic}}}{q}(x))$, where $q(x) = (x-\mu)^\top \Sigma^{-1} (x-\mu)$

[Now $q(x)$ is a function we understand—it's just a quadratic bowl centered at μ whose curvature is represented by the metric tensor Σ^{-1}. $q(x)$ is the squared distance from μ to x under this metric. The other function $n(\cdot)$ is like a 1D Gaussian with a different normalization constant. This mapping $n(\cdot)$ does not change the isosurfaces.]

Principle: given $n : \mathbb{R} \to \mathbb{R}$, isosurfaces of $n(q(x))$ are same as $q(x)$ (different isovalues), except that some might be combined [if n maps them to the same value]



$q(x)$ $P(x) = n(q(x))$

ellipsebowl.pdf, ellipses.pdf, gauss.pdf, gauss3d.pdf, gausscontour.pdf

[(Show this figure on a separate "whiteboard" for easy reuse next lecture.) A paraboloid (left) becomes a bivariate Gaussian (right) after you compose it with the univariate Gaussian (center).]