Economics 204 Summer/Fall 2018
Lecture 10–Friday August 3, 2018

Diagonalization of Symmetric Real Matrices (from Handout)

Definition 1 Let
$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$
A basis $V = \{v_1, \ldots, v_n\}$ of $\mathbf{R}^n$ is *orthonormal* if $v_i \cdot v_j = \delta_{ij}$.

In other words, a basis is orthonormal if each basis element has unit length ($\|v_i\|^2 = v_i \cdot v_i = 1$ for each i), and distinct basis elements are perpendicular ($v_i \cdot v_j = 0$ for $i \neq j$).

Remark: Suppose that $x = \sum_{j=1}^n \alpha_j v_j$ where $\{v_1, \ldots, v_n\}$ is an orthonormal basis of $\mathbf{R}^n$. Then
$$\begin{aligned} x \cdot v_k &= \left(\sum_{j=1}^n \alpha_j v_j \right) \cdot v_k \\ &= \sum_{j=1}^n \alpha_j (v_j \cdot v_k) \\ &= \sum_{j=1}^n \alpha_j \delta_{jk} \\ &= \alpha_k \end{aligned}$$
so
$$x = \sum_{j=1}^n (x \cdot v_j) v_j$$

Example: The standard basis of $\mathbf{R}^n$ is orthonormal.

Recall that for a real $n \times m$ matrix A, $A^\top$ denotes the transpose of A: the $(i,j)^{th}$ entry of $A^\top$ is the $(j,i)^{th}$ entry of A. So the i^{th} row of $A^\top$ is the i^{th} column of A.

Definition 2 A real $n \times n$ matrix A is *unitary* if $A^\top = A^{-1}$.

Theorem 3 *A real $n \times n$ matrix A is unitary if and only if the columns of A are orthonormal.*

Proof: Let v_j denote the j^{th} column of A.
$$\begin{aligned} A^\top = A^{-1} &\iff A^\top A = I \\ &\iff v_i \cdot v_j = \delta_{ij} \quad \forall i, j \\ &\iff \{v_1, \ldots, v_n\} \text{ is orthonormal} \end{aligned}$$

$\blacksquare$

If A is unitary, let V be the set of columns of A and W be the standard basis of $\mathbf{R}^n$. Since A is unitary, it is invertible, so V is a basis of $\mathbf{R}^n$.

$$A^\top = A^{-1} = Mtx_{V,W}(id)$$

Since V is orthonormal, the transformation between bases W and V preserves all geometry, including lengths and angles.

Theorem 4 *Let $T \in L(\mathbf{R}^n, \mathbf{R}^n)$ and W be the standard basis of $\mathbf{R}^n$. Suppose that $Mtx_W(T)$ is symmetric. Then the eigenvectors of T are all real, and there is an orthonormal basis $V = \{v_1, \ldots, v_n\}$ of $\mathbf{R}^n$ consisting of eigenvectors of T, so that $Mtx_W(T)$ is diagonalizable:*

$$Mtx_W(T) = Mtx_{W,V}(id) \cdot Mtx_V(T) \cdot Mtx_{V,W}(id)$$

where $Mtx_V T$ is diagonal and the change of basis matrices $Mtx_{V,W}(id)$ and $Mtx_{W,V}(id)$ are unitary.

Proof: (*Sketch*) The proof of the theorem requires a lengthy digression into the linear algebra of *complex* vector spaces. Here is a very brief outline.

1. Let $M = Mtx_W(T)$.

2. The inner product in $\mathbf{C}^n$ is defined as follows:

$$x \cdot y = \sum_{j=1}^{n} x_j \cdot \overline{y_j}$$

   where $\bar{c}$ denotes the complex conjugate of any $c \in \mathbf{C}$; note that this implies that $x \cdot y = \overline{y \cdot x}$. The usual inner product in $\mathbf{R}^n$ is the restriction of this inner product on $\mathbf{C}^n$ to $\mathbf{R}^n$.

3. Given any complex matrix A, define A^* to be the matrix whose $(i,j)^{th}$ entry is $\overline{a_{ji}}$; in other words, A^* is formed by taking the complex conjugate of each element of the transpose of A. It is easy to verify that given $x, y \in \mathbf{C}^n$ and a complex $n \times n$ matrix A, $Ax \cdot y = x \cdot A^* y$. Since M is real and symmetric, $M^* = M$.

4. If M is real and symmetric, and $\lambda \in \mathbf{C}$ is an eigenvalue of M, with eigenvector $x \in \mathbf{C}^n$, then

$$\begin{aligned} \lambda |x|^2 &= \lambda (x \cdot x) \\ &= (\lambda x) \cdot x \\ &= (Mx) \cdot x \\ &= x \cdot (M^* x) \end{aligned}$$

$$
\begin{aligned}
&= x \cdot (Mx) \\
&= x \cdot (\lambda x) \\
&= \overline{(\lambda x) \cdot x} \\
&= \overline{\lambda(x \cdot x)} \\
&= \overline{\lambda |x|^2} \\
&= \bar{\lambda} |x|^2
\end{aligned}
$$

which proves that $\lambda = \bar{\lambda}$, hence $\lambda \in \mathbf{R}$.

5. If M is real (not necessarily symmetric) and $\lambda \in \mathbf{R}$ is an eigenvalue, then $\det(M - \lambda I) = 0 \Rightarrow \exists v \in \mathbf{R}^n$ s.t. $(M - \lambda I)v = 0$, so there is at least one real eigenvector. Symmetry implies that, if λ has multiplicity m, there are m independent real eigenvectors corresponding to λ (but unfortunately we don't have time to show this). Thus, there is a basis of eigenvectors, hence M is diagonalizable over $\mathbf{R}$.

6. If M is real and symmetric, eigenvectors corresponding to distinct eigenvalues are orthogonal: Suppose that $Mx = \lambda x$ and $My = \rho y$ with $\rho \neq \lambda$. Then

$$
\begin{aligned}
\lambda(x \cdot y) &= (\lambda x) \cdot y \\
&= (Mx) \cdot y \\
&= (Mx)^\top y \\
&= \left(x^\top M^\top\right) y \\
&= \left(x^\top M\right) y \\
&= x^\top (My) \\
&= x^\top (\rho y) \\
&= x \cdot (\rho y) \\
&= \rho(x \cdot y)
\end{aligned}
$$

so $(\lambda - \rho)(x \cdot y) = 0$; since $\lambda - \rho \neq 0$, we must have $x \cdot y = 0$.

7. Using the Gram-Schmidt method, we can get an orthonormal basis of eigenvectors:

   - Let $X_\lambda = \{x \in \mathbf{R}^n : Mx = \lambda x\}$, the set of all eigenvectors corresponding to λ. Notice that if $Mx = \lambda x$ and $My = \lambda y$, then

     $$M(\alpha x + \beta y) = \alpha Mx + \beta My = \alpha \lambda x + \beta \lambda y = \lambda(\alpha x + \beta y)$$

     so X_λ is a vector subspace. Thus, given any basis of X_λ, we wish to find an orthonormal basis of X_λ; all elements of this orthonormal basis will be eigenvectors corresponding to λ.

   - Suppose X_λ is m-dimensional and we are given independent vectors $x_1, \ldots, x_m \in X_\lambda$. The Gram-Schmidt method finds an orthonormal basis $\{v_1, \ldots, v_m\}$ for X_λ.

   - Let $v_1 = \frac{x_1}{|x_1|}$. Note that $|v_1| = 1$.

- Suppose we have found an orthonormal set $\{v_1, \ldots, v_k\}$ such that span $\{v_1, \ldots, v_k\} =$ span $\{x_1, \ldots, x_k\}$, with $k < m$. Let

$$y_{k+1} = x_{k+1} - \sum_{j=1}^{k} (x_{k+1} \cdot v_j) v_j, \;\; v_{k+1} = \frac{y_{k+1}}{|y_{k+1}|}$$

-

$$\begin{aligned} \text{span}\,\{v_1, \ldots, v_{k+1}\} &= \text{span}\,\{v_1, \ldots, v_k, v_{k+1}\} \\ &= \text{span}\,\{v_1, \ldots, v_k, y_{k+1}\} \\ &= \text{span}\,\{v_1, \ldots, v_k, x_{k+1}\} \\ &= \text{span}\,\{x_1, \ldots, x_k, x_{k+1}\} \end{aligned}$$

- For $i = 1, \ldots, k$,

$$\begin{aligned} y_{k+1} \cdot v_i &= \left(x_{k+1} - \sum_{j=1}^{k} (x_{k+1} \cdot v_j) v_j \right) \cdot v_i \\ &= x_{k+1} \cdot v_i - \sum_{j=1}^{K} (x_{k+1} \cdot v_j)(v_j \cdot v_i) \\ &= x_{k+1} \cdot v_i - \sum_{j=1}^{K} (x_{k+1} \cdot v_j) \delta_{ij} \\ &= x_{k+1} \cdot v_i - x_{k+1} \cdot v_i \\ &= 0 \\ v_{k+1} \cdot v_i &= \frac{y_{k+1} \cdot v_i}{|y_{k+1}|} \\ &= \frac{0}{|y_{k+1}|} \\ &= 0 \\ |v_{k+1}| &= \frac{|y_{k+1}|}{|y_{k+1}|} \\ &= 1 \end{aligned}$$

■

Application to Quadratic Forms

Consider a quadratic form

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \alpha_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j \tag{1}$$

Let

$$\alpha_{ij} = \begin{cases} \frac{\beta_{ij}}{2} & \text{if } i < j \\ \frac{\beta_{ji}}{2} & \text{if } i > j \end{cases}$$

Let

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix}$$

so

$$f(x) = x^\top A x$$

Example: Let

$$f(x) = \alpha x_1^2 + \beta x_1 x_2 + \gamma x_2^2$$

Let

$$A = \begin{pmatrix} \alpha & \frac{\beta}{2} \\ \frac{\beta}{2} & \gamma \end{pmatrix}$$

so A is symmetric and

$$\begin{aligned} & (x_1, x_2) \begin{pmatrix} \alpha & \frac{\beta}{2} \\ \frac{\beta}{2} & \gamma \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ = \; & (x_1, x_2) \begin{pmatrix} \alpha x_1 + \frac{\beta}{2} x_2 \\ \frac{\beta}{2} x_1 + \gamma x_2 \end{pmatrix} \\ = \; & \alpha x_1^2 + \beta x_1 x_2 + \gamma x_2^2 \\ = \; & f(x) \end{aligned}$$

Returning to the general quadratic form in Equation (1), A is symmetric, so let $V = \{v_1, \ldots, v_n\}$ be an orthonormal basis of eigenvectors of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$\begin{aligned} A &= U^\top D U \\ \text{where } D &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\ \text{and } U &= Mtx_{V,W}(id) \text{ is unitary} \end{aligned}$$

The columns of $U^\top$ (the rows of U) are the coordinates of $v_1, \ldots, v_n$, expressed in terms of the standard basis W.

Given $x \in \mathbf{R}^n$, recall

$$x = \sum_{i=1}^{n} \gamma_i v_i \text{ where } \gamma_i = x \cdot v_i$$

Then

$$f(x) = f\left(\sum \gamma_i v_i\right)$$

$$
\begin{aligned}
&= \left(\sum \gamma_i v_i\right)^\top A \left(\sum \gamma_i v_i\right) \\
&= \left(\sum \gamma_i v_i\right)^\top U^\top D U \left(\sum \gamma_i v_i\right) \\
&= \left(U \sum \gamma_i v_i\right)^\top D \left(U \sum \gamma_i v_i\right) \\
&= \left(\sum \gamma_i U v_i\right)^\top D \left(\sum \gamma_i U v_i\right) \\
&= (\gamma_1, \ldots, \gamma_n) D \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} \\
&= \sum \lambda_i \gamma_i^2
\end{aligned}
$$

The equation for the level sets of f is

$$\sum_{i=1}^{n} \lambda_i \gamma_i^2 = C$$

- If $\lambda_i \geq 0$ for all i, the level set is an ellipsoid, with principal axes in the directions $v_1, \ldots, v_n$. The length of the principal axis along v_i is $\sqrt{C/\lambda_i}$ if $C \geq 0$ (if $\lambda_i = 0$, the level set is a degenerate ellipsoid with principal axis of infinite length in that direction). The level set is empty if $C < 0$. See Figure 1.
- If $\lambda_i \leq 0$ for all i, the level set is an ellipsoid, with principal axes in the directions $v_1, \ldots, v_n$. The length of the principal axis along v_i is $\sqrt{C/\lambda_i}$ if $C \leq 0$ (if $\lambda_i = 0$, the level set is a degenerate ellipsoid with principal axis of infinite length in that direction). The level set is empty if $C > 0$.
- If $\lambda_i > 0$ for some i and $\lambda_j < 0$ for some j, the level set is a hyperboloid. For example, suppose $n = 2$, $\lambda_1 > 0$, $\lambda_2 < 0$. The equation is

$$
\begin{aligned}
C &= \lambda_1 \gamma_1^2 + \lambda_2 \gamma_2^2 \\
&= \left(\sqrt{\lambda_1}\gamma_1 + \sqrt{|\lambda_2|}\gamma_2\right)\left(\sqrt{\lambda_1}\gamma_1 - \sqrt{|\lambda_2|}\gamma_2\right)
\end{aligned}
$$

  This is a hyperbola with asymptotes

$$
\begin{aligned}
0 &= \sqrt{\lambda_1}\gamma_1 + \sqrt{|\lambda_2|}\gamma_2 \\
\Rightarrow \gamma_1 &= -\sqrt{\frac{|\lambda_2|}{\lambda_1}}\gamma_2 \\
0 &= \left(\sqrt{\lambda_1}\gamma_1 - \sqrt{|\lambda_2|}\gamma_2\right) \\
\Rightarrow \gamma_1 &= \sqrt{\frac{|\lambda_2|}{\lambda_1}}\gamma_2
\end{aligned}
$$

See Figure 2. This proves the following corollary of Theorem 4.

Corollary 5 *Consider the quadratic form (1).*

1. *f has a global minimum at 0 if and only if $\lambda_i \geq 0$ for all i; the level sets of f are ellipsoids with principal axes aligned with the orthonormal eigenvectors $v_1, \ldots, v_n$.*
2. *f has a global maximum at 0 if and only if $\lambda_i \leq 0$ for all i; the level sets of f are ellipsoids with principal axes aligned with the orthonormal eigenvectors $v_1, \ldots, v_n$.*
3. *If $\lambda_i < 0$ for some i and $\lambda_j > 0$ for some j, then f has a saddle point at 0; the level sets of f are hyperboloids with principal axes aligned with the orthonormal eigenvectors $v_1, \ldots, v_n$.*

Section 3.4. Linear Maps between Normed Spaces

Definition 6 Suppose X, Y are normed vector spaces and $T \in L(X, Y)$. We say T is *bounded* if

$$\exists \beta \in \mathbf{R} \text{ s.t. } \|T(x)\|_Y \leq \beta \|x\|_X \quad \forall x \in X$$

Note this implies that T is Lipschitz with constant β.

Theorem 7 (Thms. 4.1, 4.3) *Let X, Y be normed vector spaces and $T \in L(X, Y)$. Then*

$$\begin{array}{ll} & T \textit{ is continuous at some point } x_0 \in X \\ \iff & T \textit{ is continuous at every } x \in X \\ \iff & T \textit{ is uniformly continuous on } X \\ \iff & T \textit{ is Lipschitz} \\ \iff & T \textit{ is bounded} \end{array}$$

Proof: Suppose T is continuous at x_0. Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that

$$\|z - x_0\| < \delta \Rightarrow \|T(z) - T(x_0)\| < \varepsilon$$

Now suppose x is any element of X. If $\|y - x\| < \delta$, let $z = y - x + x_0$, so $\|z - x_0\| = \|y - x\| < \delta$.

$$\begin{array}{ll} & \|T(y) - T(x)\| \\ = & \|T(y - x)\| \\ = & \|T(y - x + x_0 - x_0))\| \\ = & \|T(z) - T(x_0)\| \\ < & \varepsilon \end{array}$$

which proves that T is continuous at every x, and uniformly continuous.

We claim that T is bounded if and only if T is continuous at 0. Suppose T is not bounded. Then

$$\exists \{x_n\} \text{ s.t. } \quad \|T(x_n)\| > n\|x_n\| \quad \forall n$$

Note that $x_n \neq 0$. Let $\varepsilon = 1$. Fix $\delta > 0$ and choose n such that $\frac{1}{n} < \delta$. Let

$$\begin{aligned}
x_n' &= \frac{x_n}{n\|x_n\|} \\
\|x_n'\| &= \frac{\|x_n\|}{n\|x_n\|} \\
&= \frac{1}{n} \\
&< \delta \\
\|T(x_n') - T(0)\| &= \|T(x_n')\| \\
&= \frac{1}{n\|x_n\|}\|T(x_n)\| \\
&> \frac{n\|x_n\|}{n\|x_n\|} \\
&= 1 \\
&= \varepsilon
\end{aligned}$$

Since this is true for every δ, T is not continuous at 0. Therefore, T continuous at 0 implies T is bounded. Now, suppose T is bounded, so find M such that $\|T(x)\| \leq M\|x\|$ for every $x \in X$. Given $\varepsilon > 0$, let $\delta = \varepsilon/M$. Then

$$\begin{aligned}
\|x - 0\| < \delta &\Rightarrow \|x\| < \delta \\
&\Rightarrow \|T(x) - T(0)\| = \|T(x)\| < M\delta \\
&\Rightarrow \|T(x) - T(0)\| < \varepsilon
\end{aligned}$$

so T is continuous at 0.

Thus, we have shown that continuity at some point x_0 implies uniform continuity, which implies continuity at every point, which implies T is continuous at 0, which implies that T is bounded, which implies that T is continuous at 0, which implies that T is continuous at some x_0, so all of the statements except possibly the Lipschitz statement are equivalent.

Suppose T is bounded, with constant M. Then

$$\begin{aligned}
\|T(x) - T(y)\| &= \|T(x - y)\| \\
&\leq M\|x - y\|
\end{aligned}$$

so T is Lipschitz with constant M; conversely, if T is Lipschitz with constant M, then T is bounded with constant M. So all the statements are equivalent. ■

Every linear map on a finite-dimensional normed vector space is bounded (and thus continuous, uniformly continuous, and Lipschitz continuous).

Theorem 8 (Thm. 4.5) *Let X, Y be normed vector spaces with $\dim X = n$. Every $T \in L(X, Y)$ is bounded.*

Proof: See de la Fuente. ■

Definition 9 A *topological isomorphism* between normed vector spaces X and Y is a linear transformation $T \in L(X, Y)$ that is invertible (one-to-one, onto), continuous, and has a continuous inverse.

Two normed vector spaces X and Y are *topologically isomorphic* if there is a topological isomorphism $T : X \to Y$.

Suppose X and Y are normed vector spaces. We define

$$\begin{aligned} B(X,Y) &= \{T \in L(X,Y) : T \text{ is bounded}\} \\ \|T\|_{B(X,Y)} &= \sup\left\{\frac{\|T(x)\|_Y}{\|x\|_X}, x \in X, x \neq 0\right\} \\ &= \sup\{\|T(x)\|_Y : \|x\|_X = 1\} \end{aligned}$$

Theorem 10 (Thm. 4.8) *Let X, Y be normed vector spaces. Then*

$$\Big(B(X,Y), \|\cdot\|_{B(X,Y)}\Big)$$

is a normed vector space.

Proof: See de la Fuente. ■

Theorem 11 (Thm. 4.9) *Let $T \in L(\mathbf{R}^n, \mathbf{R}^m)$ $(= B(\mathbf{R}^n, \mathbf{R}^m))$ with matrix $A = (a_{ij})$ with respect to the standard bases. Let*

$$M = \max\{|a_{ij}| : 1 \leq i \leq m, 1 \leq j \leq n\}$$

Then

$$M \leq \|T\| \leq M\sqrt{mn}$$

Proof: See de la Fuente. ■

Theorem 12 (Thm. 4.10) *Let $R \in L(\mathbf{R}^m, \mathbf{R}^n)$ and $S \in L(\mathbf{R}^n, \mathbf{R}^p)$. Then*

$$\|S \circ R\| \leq \|S\|\|R\|$$

Proof: See de la Fuente. ■

Define

$$\Omega(\mathbf{R}^n) = \{T \in L(\mathbf{R}^n, \mathbf{R}^n) : T \text{ is invertible}\}$$

Theorem 13 (Thm. 4.11') *Suppose $T \in L(\mathbf{R}^n, \mathbf{R}^n)$ and E is the standard basis of $\mathbf{R}^n$. Then*

$$\begin{aligned} & T \textit{ is invertible} \\ \Leftrightarrow \quad & \ker T = \{0\} \\ \Leftrightarrow \quad & \det\left(Mtx_E(T)\right) \neq 0 \\ \Leftrightarrow \quad & \det\left(Mtx_{V,V}(T)\right) \neq 0 \textit{ for every basis } V \\ \Leftrightarrow \quad & \det\left(Mtx_{V,W}(T)\right) \neq 0 \textit{ for every pair of bases } V, W \end{aligned}$$

Theorem 14 (Thm. 4.12) *If $S, T \in \Omega(\mathbf{R}^n)$, then $S \circ T \in \Omega(\mathbf{R}^n)$ and*

$$(S \circ T)^{-1} = T^{-1} \circ S^{-1}$$

Theorem 15 (Thm. 4.14) *Let $S, T \in L(\mathbf{R}^n, \mathbf{R}^n)$. If T is invertible and*

$$\|T - S\| < \frac{1}{\|T^{-1}\|}$$

then S is invertible. In particular, $\Omega(\mathbf{R}^n)$ is open in $L(\mathbf{R}^n, \mathbf{R}^n) = B(\mathbf{R}^n, \mathbf{R}^n)$.

Proof: See de la Fuente. ■

Theorem 16 (4.15) *The function $(\cdot)^{-1} : \Omega(\mathbf{R}^n) \to \Omega(\mathbf{R}^n)$ that assigns T^{-1} to each $T \in \Omega(\mathbf{R}^n)$ is continuous.*

Proof: See de la Fuente. ■

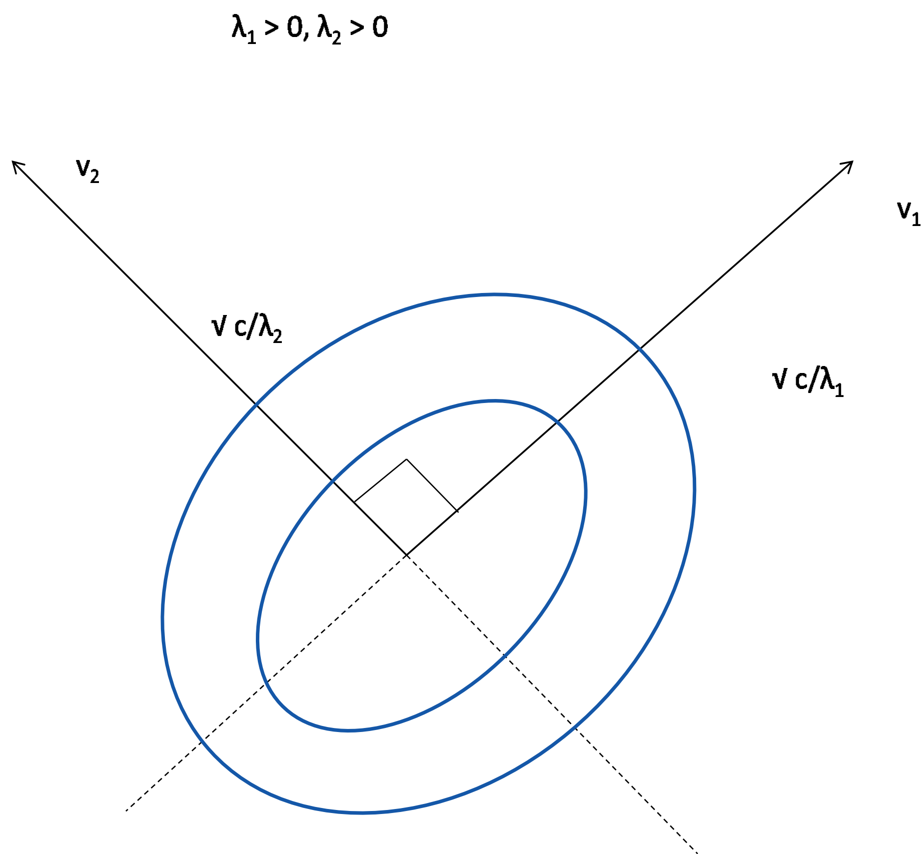

Figure 1: If $\lambda_1, \lambda_2 > 0$ and $C > 0$, the level set is an ellipsoid, with principal axes in the directions v_1, v_2. The length of the principal axis along v_i is $\sqrt{C/\lambda_i}$.

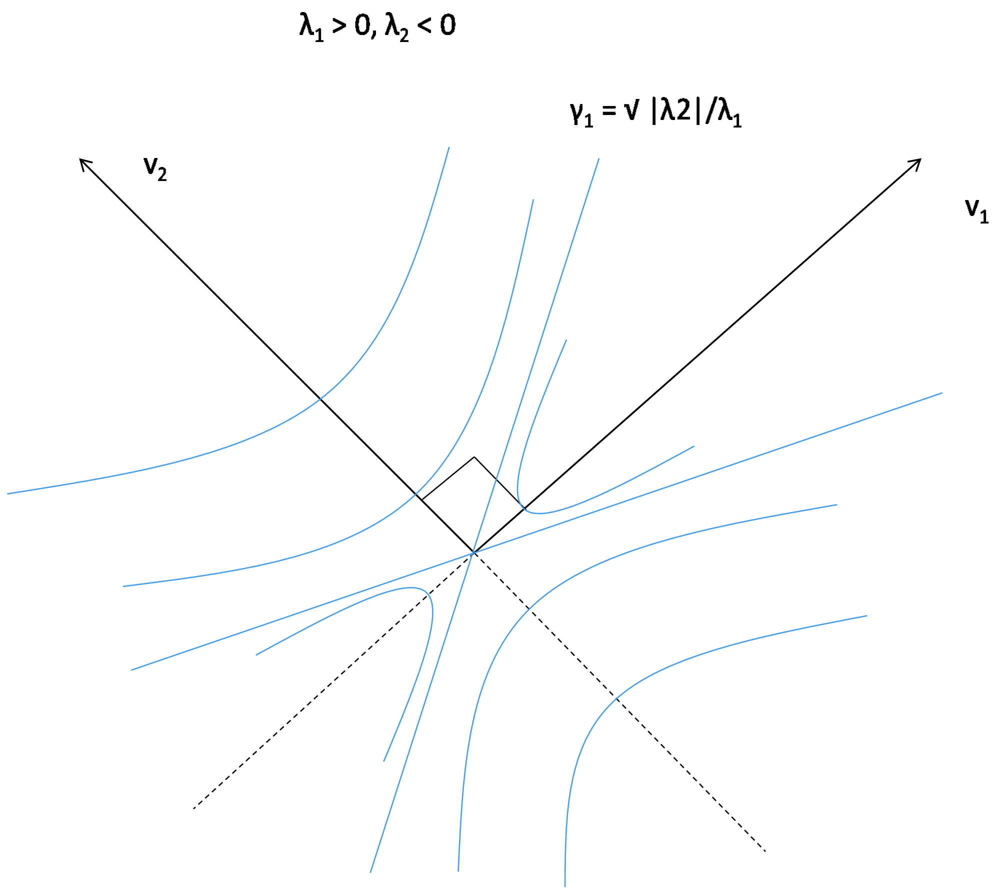

Figure 2: If $\lambda_1 > 0$ and $\lambda_2 < 0$, the level set is a hyperbola with asymptotes $\gamma_1 = \sqrt{\frac{|\lambda_2|}{\lambda_1}}\gamma_2$.