Some linear algebra

Recall the convention that, for us, all vectors are column vectors.

1. Symmetric matrices

Let $A$ be a real $n \times n$ matrix. Recall that a complex number $\lambda$ is an eigenvalue of $A$ if there exists a real and nonzero vector $x$—called an eigenvector for $\lambda$—such that $Ax = \lambda x$. Whenever $x$ is an eigenvector for $\lambda$, so is $ax$ for every real number $a$.

The characteristic polynomial $\chi_A$ of matrix $A$ is the function

$$\chi_A(\lambda) := \det(\lambda I - A),$$

defined for all complex numbers $\lambda$, where $I$ denotes the $n \times n$ identity matrix. It is not hard to see that a complex number $\lambda$ is an eigenvalue of $A$ if and only if $\chi_A(\lambda) = 0$. We see by direct computation that $\chi_A$ is an $n$th-order polynomial. Therefore, $A$ has precisely $n$ eigenvalues, thanks to the fundamental theorem of algebra. We can write them as $\lambda_1, \ldots, \lambda_n$, or sometimes more precisely as $\lambda_1(A), \ldots, \lambda_n(A)$.

1. The spectral theorem. The following important theorem is the starting point of our discussion. It might help to recall that vectors $x_1, \ldots, x_k \in \mathbb{R}^n$ are orthonormal if $x_i'x_j = 0$ when $i \neq j$ and $x_i'x_i = \|x_i\|^2 = 1$.

**Theorem 1.** If $A$ is a real and symmetric $n \times n$ matrix, then $\lambda_1, \ldots, \lambda_n$ are real numbers. Moreover, there exist $n$ orthonormal eigenvectors $v_1, \ldots, v_n$ that correspond respectively to $\lambda_1, \ldots, \lambda_n$. 

I will not prove this result, as it requires developing a good deal of elementary linear algebra that we will not need. Instead, let me state and prove a result that is central for us.

**Theorem 2** (The spectral theorem). Let $A$ denote a symmetric $n \times n$ matrix with real eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Define $D := \text{diag}(\lambda_1, \ldots, \lambda_n)$ to be the diagonal matrix of the $\lambda_i$’s and $P$ to be the matrix whose columns are $v_1$ through $v_n$ respectively; that is,

$$D := \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \quad P := (v_1, \ldots, v_n).$$

Then $P$ is orthogonal [$P' = P^{-1}$] and $A = PDP^{-1} = PDP'$.

**Proof.** $P$ is orthogonal because the orthonormality of the $v_i$’s implies that

$$P'P = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = I.$$

Furthermore, because $Av_j = \lambda_j v_j$, it follows that $AP = PD$, which is another way to say that $A = PDP^{-1}$.

Recall that the *trace* of an $n \times n$ matrix $A$ is the sum $A_{1,1} + \cdots + A_{n,n}$ of its diagonal entries.

**Corollary 3.** If $A$ is a real and symmetric $n \times n$ matrix with real eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$\text{tr}(A) = \lambda_1 + \cdots + \lambda_n \quad \text{and} \quad \text{det}(A) = \lambda_1 \times \cdots \times \lambda_n.$$

**Proof.** Write $A$, in spectral form, as $PDP^{-1}$. Since the determinant of $P^{-1}$ is the reciprocal of that of $A$, it follows that $\text{det}(A) = \text{det}(D)$, which is clearly $\lambda_1 \times \cdots \times \lambda_n$. In order to compute the trace of $A$ we compute directly also:

$$\text{tr}(A) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{i,j} \left(DP^{-1}\right)_{i,j} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} P_{i,j}D_{i,k}P_{j,k}^{-1} = \sum_{i=1}^{n} \sum_{k=1}^{n} \left(P^{-1}P\right)_{i,k}D_{i,k} = \sum_{i=1}^{n} D_{i,i} = \text{tr}(D),$$
2. Positive-semidefinite matrices

which is \( \lambda_1 + \cdots + \lambda_n. \)

2. The square-root matrix. Let \( A \) continue to denote a real and symmetric \( n \times n \) matrix.

Proposition 4. There exists a complex and symmetric \( n \times n \) matrix \( B \)—called the square root of \( A \) and written as \( A^{1/2} \) or even sometimes as \( \sqrt{A} \)—such that \( A = B^2 := BB \).

The proof of Proposition 4 is more important than its statement. So let us prove this result.

Proof. Apply the spectral theorem and write \( A = PDP^{-1} \). Since \( D \) is a diagonal matrix, its square root can be defined unambiguously as the following complex-valued \( n \times n \) diagonal matrix:

\[
D^{1/2} := \begin{pmatrix}
\lambda_1^{1/2} & 0 & 0 & \cdots & 0 \\
0 & \lambda_2^{1/2} & 0 & \cdots & 0 \\
0 & 0 & \lambda_3^{1/2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_n^{1/2}
\end{pmatrix}
\]

Define \( B := PD^{1/2}P^{-1} \), and note that

\[
B^2 = PD^{1/2}P^{-1}PD^{1/2}P^{-1} = PDP^{-1} = A,
\]
since \( P^{-1}P = I \) and \( (D^{1/2})^2 = D \). \( \square \)

2. Positive-semidefinite matrices

Recall that an \( n \times n \) matrix \( A \) is **positive semidefinite** if it is symmetric and

\[
x'Ax \geq 0 \quad \text{for all } x \in \mathbb{R}^n.
\]

Recall that \( A \) is **positive definite** if it is symmetric and

\[
x'Ax > 0 \quad \text{for all nonzero } x \in \mathbb{R}^n.
\]

**Theorem 5.** A symmetric matrix \( A \) is positive semidefinite if and only if all of its eigenvalues are \( \geq 0 \). \( A \) is positive definite if and only if all of its eigenvalues are \( > 0 \). In the latter case, \( A \) is also nonsingular.

The following is a ready consequence.

**Corollary 6.** All of the eigenvalues of a variance-covariance matrix are always \( \geq 0 \).

Now let us establish the theorem.
Proof of Theorem 5. Suppose $A$ is positive semidefinite, and let $\lambda$ denote one of its eigenvalues, together with corresponding eigenvector $x$. Since $0 \leq x^tAx = \lambda \|x\|^2$ and $\|x\| > 0$, it follows that $\lambda \geq 0$. This proves that all of the eigenvalues of $A$ are nonnegative. If $A$ is positive definite, then the same argument shows that all of its eigenvalues are $> 0$. Because $\det(A)$ is the product of all $n$ eigenvalues of $A$ (Corollary 3), it follows that $\det(A) > 0$, whence $A$ is nonsingular.

This proves slightly more than half of the proposition. Now let us suppose that all eigenvalues of $A$ are $\geq 0$. We write $A$ in spectral form $A = PD\text{P}^t$, and observe that $D$ is a diagonal matrix of nonnegative numbers. By virtue of its construction, $x^tAx = \left(D_{1/2}^tPx\right)^t\left(D_{1/2}^tPx\right) = \left(D_{1/2}^tPx\right)^2$, which is $\geq 0$. Therefore, $A$ is positive semidefinite.

If all of the eigenvalues of $A$ are $> 0$, then (1) tells us that

$$x^tAx = \left(D_{1/2}^tPx\right)^2 = \sum_{j=1}^n \left(D_{1/2}^tPx\right)_j^2 = \sum_{j=1}^n \lambda_j (\left[Px\right]_j)^2,$$

which is $\geq 0$. Therefore, $A$ is positive semidefinite.

For all $x \in \mathbb{R}^n$,

$$x^tAx = \left(D_{1/2}^tPx\right)^2 = \sum_{j=1}^n \left(D_{1/2}^tPx\right)_j^2 = \sum_{j=1}^n \lambda_j (\left[Px\right]_j)^2,$$

where $\lambda_j > 0$ for all $j$. Therefore,

$$x^tAx \geq \min_{1 \leq j \leq n} \lambda_j \cdot \sum_{j=1}^n (\left[Px\right]_j)^2 = \min_{1 \leq j \leq n} \lambda_j \cdot x^tP^tPx = \min_{1 \leq j \leq n} \lambda_j \cdot \|x\|^2.$$

Since $\min_{1 \leq j \leq n} \lambda_j > 0$, it follows that $x^tAx > 0$ for all nonzero $x$. This completes the proof.

Let us pause and point out a consequence of the proof of this last result.

Corollary 7. If $A$ is positive semidefinite, then its extremal eigenvalues satisfy

$$\min_{1 \leq j \leq n} \lambda_j = \min_{\|x\| > 0} \frac{x^tAx}{\|x\|^2}, \quad \max_{1 \leq j \leq n} \lambda_j = \max_{\|x\| > 0} \frac{x^tAx}{\|x\|^2}.$$

Proof. We saw, during the course of the previous proof, that

$$\min_{1 \leq j \leq n} \lambda_j \cdot \|x\|^2 \leq x^tAx \quad \text{for all} \quad x \in \mathbb{R}^n.$$

Optimize over all $x$ to see that

$$\min_{1 \leq j \leq n} \lambda_j \leq \min_{\|x\| > 0} \frac{x^tAx}{\|x\|^2}.$$
3. The rank of a matrix

But \( \min_{1 \leq j \leq n} \lambda_j \) is an eigenvalue for \( A \); let \( z \) denote a corresponding eigenvector in order to see that
\[
\min_{1 \leq j \leq n} \lambda_j \leq \min_{\|x\| > 0} \frac{x'Ax}{\|x\|^2} \leq \frac{z'Az}{\|z\|^2} = \min_{1 \leq j \leq n} \lambda_j.
\]
So both inequalities are in fact equalities, and hence follows the formula for the minimum eigenvalue. The one for the maximum eigenvalue is proved similarly. \( \square \)

Finally, a word about the square root of positive semidefinite matrices:

**Proposition 8.** If \( A \) is positive semidefinite, then so is \( A^{1/2} \). If \( A \) is positive definite, then so is \( A^{1/2} \).

**Proof.** We write, in spectral form, \( A = PD P' \) and observe [by squaring it] that \( A^{1/2} = PD^{1/2} P' \). Note that \( D^{1/2} \) is a real diagonal matrix since the eigenvalues of \( A \) are \( \geq 0 \). Therefore, we may apply (1) to \( A^{1/2} \) [in place of \( A \)] to see that \( x'A^{1/2}x = \|D^{1/2}Px\|^2 \geq 0 \) where \( D^{1/2} \) denotes the [real] square root of \( D^{1/2} \). This proves that if \( A \) is positive semidefinite, then so is \( A^{1/2} \). Now suppose there exists a positive definite \( A \) whose square root is not positive definite. It would follow that there necessarily exists a nonzero \( x \in \mathbb{R}^n \) such that \( x'A^{1/2}x = \|D^{1/2}Px\|^2 = 0 \). Since \( D^{1/2}Px = 0 \),
\[
D^{1/2}Px = D^{1/2}D^{1/2}Px = 0 \quad \Rightarrow \quad x'Ax = \|D^{1/2}Px\|^2 = 0.
\]
And this contradicts the assumption that \( A \) is positive definite. \( \square \)

3. The rank of a matrix

Recall that vectors \( v_1, \ldots, v_k \) are **linearly independent** if
\[
c_1v_1 + \cdots + c_kv_k = 0 \quad \Rightarrow \quad c_1 = \cdots = c_k = 0.
\]
For instance, \( v_1 := (1, 0)' \) and \( v_2 := (0, 1)' \) are linearly independent 2-vectors.

The **column rank** of a matrix \( A \) is the maximum number of linearly independent column vectors of \( A \). The **row rank** of a matrix \( A \) is the maximum number of linearly independent row vectors of \( A \). We can interpret these definitions geometrically as follows: First, suppose \( A \) is \( m \times n \) and define \( \mathcal{L}(A) \) denote the linear space of all vectors of the form \( c_1v_1 + \cdots + c_nv_n \), where \( v_1, \ldots, v_n \) are the column vectors of \( A \) and \( c_1, \ldots, c_n \) are real numbers. We call \( \mathcal{L}(A) \) the column space of \( A \).

We can define the row space \( \mathcal{R}(A) \), of \( A \) similarly, or simply define \( \mathcal{R}(A) := \mathcal{L}(A') \).
Lemma 9. For every $m \times n$ matrix $A$,

$$G(A) = \{ Ax : x \in \mathbb{R}^n \}, \quad R(A) := \{ x'A : x \in \mathbb{R}^m \}. $$

We can think of an $m \times n$ matrix $A$ as a mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$; namely, we can think of matrix $A$ also as the function $f_A(x) := x \mapsto Ax$. In this way we see that $G(A)$ is also the “range” of the function $f_A$.

**Proof.** Let us write the columns of $A$ as $a_1, a_2, \ldots, a_n$. Note that $y \in G(A)$ if and only if there exist $c_1, \ldots, c_n$ such that $y = c_1a_1 + \cdots + c_na_n = Ac$, where $c := (c_1, \ldots, c_n)'$. This shows that $G(A)$ is the collection of all vectors of the form $Ax$, for $x \in \mathbb{R}^n$. The second assertion [about $R(A)$] follows from the definition of $R(A)$ equalling $G(A')$ and the already-proven first assertion. \hfill\Box

It then follows, from the definition of dimension, that

$$\text{column rank of } A = \dim G(A), \quad \text{row rank of } A = \dim R(A).$$

**Proposition 10.** Given any matrix $A$, its row rank and column rank are the same. We write their common value as $\text{rank}(A)$.

**Proof.** Suppose $A$ is $m \times n$ and its column rank is $r$. Let $b_1, \ldots, b_r$ denote a basis for $G(A)$ and consider the matrix $m \times r$ matrix $B := (b_1, \ldots, b_r)$. Write $A$, columnwise, as $A := (a_1, \ldots, a_n)$. For every $1 \leq j \leq n$, there exists $c_{1,j}, \ldots, c_{r,j}$ such that $a_j = c_{1,j}b_1 + \cdots + c_{r,j}b_r$. Let $C := (c_{i,j})$ be the resulting $r \times n$ matrix, and note that $A = BC$. Because $A_{i,j} = \sum_{k=1}^r B_{i,k}C_{k,j}$, every row of $A$ is a linear combination of the rows of $C$. In other words, $R(A) \subseteq R(C)$ and hence the row rank of $A$ is $\leq \dim R(C) = r = \text{the column rank of } A$. Apply this fact to $A'$ to see that also the row rank of $A'$ is $\leq$ the column rank of $A'$; equivalently that the column rank of $A$ is $\leq$ the row rank of $A$. \hfill\Box

**Proposition 11.** If $A$ is $n \times m$ and $B$ is $m \times k$, then

$$\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)).$$

**Proof.** The proof uses an idea that we exploited already in the proof of Proposition 10: Since $(AB)_{i,l} = \sum_{v=1}^m A_{i,v}B_{v,l}$, the rows of $AB$ are linear combinations of the rows of $B$; that is $R(AB) \subseteq R(B)$, whence $\text{rank}(AB) \leq \text{rank}(B)$. Also, $G(AB) \subseteq G(A)$, whence $\text{rank}(AB) \leq \text{rank}(A)$. These observations complete the proof. \hfill\Box
Proposition 12. If \( A \) and \( C \) are nonsingular, then
\[
\text{rank}(ABC) = \text{rank}(B),
\]
provided that the dimensions match up so that \( ABC \) makes sense.

Proof. Let \( D := ABC \); our goal is to show that \( \text{rank}(D) = \text{rank}(B) \).

Two applications of the previous proposition together yield \( \text{rank}(D) \leq \text{rank}(AB) \leq \text{rank}(B) \). And since \( B = A^{-1}DC^{-1} \), we have also \( \text{rank}(B) \leq \text{rank}(A^{-1}D) \leq \text{rank}(D) \). \( \square \)

Corollary 13. If \( A \) is an \( n \times n \) real and symmetric matrix, then \( \text{rank}(A) = \) the total number of nonzero eigenvalues of \( A \). In particular, \( A \) has full rank if and only if \( A \) is nonsingular. Finally, \( \mathcal{G}(A) \) is the linear space spanned by the eigenvectors of \( A \) that correspond to nonzero eigenvalues.

Proof. We write \( A \), in spectral form, as \( A = PDP^{-1} \), and apply the preceding proposition to see that \( \text{rank}(A) = \text{rank}(D) \), which is clearly the total number of nonzero eigenvalue of \( A \). Since \( A \) is nonsingular if and only if all of its eigenvalues are nonzero, \( A \) has full rank if and only if \( A \) is nonsingular.

Finally, suppose \( A \) has rank \( k \leq n \); this is the number of its nonzero eigenvalues \( \lambda_1, \ldots, \lambda_k \). Let \( v_1, \ldots, v_n \) denote orthonormal eigenvectors such that \( v_1, \ldots, v_k \) are eigenvectors that correspond to \( \lambda_1, \ldots, \lambda_k \) and \( v_{k+1}, \ldots, v_n \) are eigenvectors that correspond to eigenvalues \( 0 \) [Gram–Schmidt]. And define \( \mathcal{E} \) to be the span of \( v_1, \ldots, v_k \); i.e.,
\[
\mathcal{E} := \{c_1v_1 + \cdots + c_kv_k : c_1, \ldots, c_k \in \mathbb{R}\}.
\]

Our final goal is to prove that \( \mathcal{E} = \mathcal{G}(A) \), which we know is equal to the linear space of all vectors of the form \( Ax \).

Clearly, \( c_1v_1 + \cdots + c_kv_k = Ax \), where \( x = \sum_{j=1}^{k} (c_j/\lambda_j)v_j \). Therefore, \( \mathcal{E} \subseteq \mathcal{G}(A) \). If \( k = n \), then this suffices because in that case \( v_1, \ldots, v_k \) is a basis for \( \mathbb{R}^n \), hence \( \mathcal{E} = \mathcal{G}(A) = \mathbb{R}^n \). If \( k < n \), then we can write every \( x \in \mathbb{R}^n \) as \( a_1v_1 + \cdots + a_nv_n \), so that \( Ax = \sum_{j=1}^{k} a_j\lambda_jv_j \in \mathcal{E} \). Thus, \( \mathcal{G}(A) \subseteq \mathcal{E} \) and we are done. \( \square \)

Let \( A \) be \( m \times n \) and define the null space [or "kernel"] of \( A \) as
\[
\mathcal{N}(A) := \{x \in \mathbb{R}^n : Ax = 0\}.
\]
Note that \( \mathcal{N}(A) \) is the linear span of the eigenvectors of \( A \) that correspond to eigenvalue \( 0 \). The other eigenvectors can be chosen to be orthogonal to these, and hence the preceding proof contains the facts
that: (i) Nonzero elements of $N(A)$ are orthogonal to nonzero elements of $G(A)$; and (ii)
\[
\dim N(A) + \text{rank}(A) = n \quad (\text{the number of columns of } A).
\]  
(5)

**Proposition 14.** $\text{rank}(A) = \text{rank}(A'A) = \text{rank}(AA')$ for every $m \times n$ matrix $A$.

**Proof.** If $Ax = 0$ then $A'Ax = 0$, and if $A'Ax = 0$, then $\|Ax\|^2 = x'A'Ax = 0$. In other words, $N(A) = N(A'A)$. Because $A'A$ and $A$ both have $n$ columns, it follows from (5) that $\text{rank}(A'A) = \text{rank}(A)$. Apply this observation to $A'$ to see that $\text{rank}(A') = \text{rank}(AA')$ as well. The result follows from this and the fact that $A$ and $A'$ have the same rank (Proposition 10). \qed

### 4. Projection matrices

A matrix $A$ is said to be a projection matrix if: (i) $A$ is symmetric; and (ii) $A$ is "idempotent"; that is, $A^2 = A$.

Note that projection matrices are always positive semidefinite. Indeed, $x'A'Ax = x'A'Ax = \|Ax\|^2 \geq 0$

**Proposition 15.** If $A$ is an $n \times n$ projection matrix, then so is $I - A$. Moreover, all eigenvalues of $A$ are zeros and ones, and $\text{rank}(A) = \text{the number of eigenvalues that are equal to one}$.

**Proof.** $(I - A)^2 = I - 2A + A^2 = I - A$. Since $I - A$ is symmetric also, it is a projection. If $\lambda$ is an eigenvalue of $A$ and $x$ is a corresponding eigenvector, then $\lambda x = Ax = A^2x = \lambda Ax = \lambda^2x$. Multiply both sides by $x'$ to see that $\lambda \|x\|^2 = \lambda^2\|x\|^2$. Since $\|x\| > 0$, it follows that $\lambda \in \{0, 1\}$. The total number of nonzero eigenvalues is then the total number of eigenvalues that are ones. Therefore, the rank of $A$ is the total number of eigenvalues that are one. \qed

**Corollary 16.** If $A$ is a projection matrix, then $\text{rank}(A) = \text{tr}(A)$.

**Proof.** Simply recall that $\text{tr}(A)$ is the sum of the eigenvalues, which for a projection matrix, is the total number of eigenvalues that are one. \qed

Why are they called "projection" matrices? Or, perhaps even more importantly, what is a "projection"?
Lemma 17. Let $\Omega$ denote a linear subspace of $\mathbb{R}^n$, and $x \in \mathbb{R}^n$ be fixed. Then there exists a unique element $y \in \Omega$ that is closest to $x$; that is, 
$$
\|y - x\| = \min_{z \in \Omega} \|z - x\|. 
$$
The point $y$ is called the projection of $x$ onto $\Omega$.

Proof. Let $k := \dim \Omega$, so that there exists an orthonormal basis $b_1, \ldots, b_k$ for $\Omega$. Extend this to a basis $b_1, \ldots, b_n$ for all of $\mathbb{R}^n$ by the Gram–Schmidt method.

Given a fixed vector $x \in \mathbb{R}^n$, we can write it as $x := c_1 b_1 + \cdots + c_n b_n$ for some $c_1, \ldots, c_n \in \mathbb{R}$. Define $y := c_1 b_1 + \cdots + c_k b_k$. Clearly, $y \in \Omega$ and 
$$
\|y - x\|^2 = \sum_{i=k+1}^n c_i^2. 
$$
Any other $z \in \Omega$ can be written as $z = \sum_{i=1}^k d_i b_i$, and hence 
$$
\|z - x\|^2 = \sum_{i=1}^k (d_i - c_i)^2 + \sum_{i=k+1}^n c_i^2, 
$$
which is strictly greater than 
$$
\|y - x\|^2 = \sum_{i=k+1}^n c_i^2 
$$
unless $d_i = c_i$ for all $i = 1, \ldots, k$; i.e., unless $z = y$.

Usually, we have a $k$-dimensional linear subspace $\Omega$ of $\mathbb{R}^n$ that is the range of some $n \times k$ matrix $A$. That is, $\Omega = \{Ay : y \in \mathbb{R}^k\}$. Equivalently, 
$$
\Omega = \mathcal{G}(A). 
$$
In that case, 
$$
\min_{z \in \Omega} \|z - x\|^2 = \min_{y \in \mathbb{R}^k} \|Ay - x\|^2 = \min_{y \in \mathbb{R}^k} \left[ y'A'Ay - y'A'x - x'Ay + x'x \right]. 
$$
Because $y'A'x$ is a scalar, the preceding is simplified to 
$$
\min_{z \in \Omega} \|z - x\|^2 = \min_{y \in \mathbb{R}^k} \left[ y'A'Ay - 2y'A'x + x'x \right]. 
$$
Suppose that the $k \times k$ positive semidefinite matrix $A'A$ is nonsingular [so that $A'A$ and hence also $(A'A)^{-1}$ are both positive definite]. Then, we can relabel variables $\alpha := A'Ay$ to see that 
$$
\min_{z \in \Omega} \|z - x\|^2 = \min_{\alpha \in \mathbb{R}^k} \left[ (A'A)^{-1}\alpha - 2\alpha(A'A)^{-1}A'x + x'x \right]. 
$$
A little arithmetic shows that 
$$
(A' - A'x')(A'A)^{-1}(A' - A'x) 
= (A'A)^{-1}A'x - 2\alpha(A'A)^{-1}A'x + x'A(A'A)^{-1}A'x. 
$$
Consequently, 
$$
\min_{z \in \Omega} \|z - x\|^2 
= \min_{\alpha \in \mathbb{R}^k} \left[ (A' - A'x')(A'A)^{-1}(A' - A'x) - x'A(A'A)^{-1}A'x + x'x \right]. 
$$
The first term in the parentheses is $\geq 0$; in fact it is $> 0$ unless we select $\alpha = A'x$. This proves that the projection of $x$ onto $\Omega$ is obtained by setting $\alpha := A'x$, in which case the projection itself is $Ay =$
$A(A'A)^{-1}A'x$ and the distance between $y$ and $x$ is the square root of $\|x\|^2 - x'A(A'A)^{-1}A'x$.

Let $P_\Omega := A(A'A)^{-1}A'$. It is easy to see that $P_\Omega$ is a projection matrix. The preceding shows that $P_\Omega x$ is the projection of $x$ onto $\Omega$ for every $x \in \mathbb{R}^n$. That is, we can think of $P_\Omega$ as the matrix that projects onto $\Omega$. Moreover, the distance between $x$ and the linear subspace $\Omega$ [i.e., $\min_{z \in \mathbb{R}^n} \|z - x\|$] is exactly the square root of $x'x - x'P_\Omega x = x'(I - P_\Omega)x = \|(I - P_\Omega)x\|^2$, because $I - P_\Omega$ is a projection matrix. What space does it project into?

Let $\Omega^\perp$ denote the collection of all $n$-vectors that are perpendicular to every element of $\Omega$. If $z \in \Omega^\perp$, then we can write, for all $x \in \mathbb{R}^n$,

$$\|z - x\|^2 = \|z - (I - P_\Omega)x + P_\Omega x\|^2$$

$$= \|z - (I - P_\Omega)x\|^2 + \|P_\Omega x\|^2 - 2(z - (I - P_\Omega)x)'P_\Omega x$$

$$= \|z - (I - P_\Omega)x\|^2 + \|P_\Omega x\|^2,$$

since $z$ is orthogonal to every element of $\Omega$ including $P_\Omega x$, and $P_\Omega = P_\Omega^2$.

Take the minimum over all $z \in \Omega^\perp$ to find that $I - P_\Omega$ is the projection onto $\Omega^\perp$. Let us summarize our findings.

**Proposition 18.** If $A'A$ is nonsingular [equivalently, has full rank], then $P_{A(A'A)^{-1}A'} := A(A'A)^{-1}A'$ is the projection onto $\mathbb{R}(A)$, $I - P_{A(A'A)^{-1}A'}$ is the projection onto $\mathbb{R}^\perp(A)$, and we have

$$x = P_{A(A'A)^{-1}A'}x + P_{A(A'A)^{-1}A'}^\perp x, \quad \text{and} \quad \|x\|^2 = \|P_{A(A'A)^{-1}A'}x\|^2 + \|P_{A(A'A)^{-1}A'}^\perp x\|^2.$$

The last result is called the "Pythagorean property."