Least squares estimation under linear constraints - A projective proof

Shaikh Ammar • 6 Dec 2025

Introduction

Consider the regression problem

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \mathrm{var}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}.$$

The least squares estimation of $\boldsymbol{\beta}$ under the constraint $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ generally involves solving the Lagrangian. Like many optimisation problems in linear models, a purely linear algebraic proof involving projections exists. Such projective proofs are typically more enlightening and less tedious. Seber and Lee (2003) contains such a projective proof, but the proof is far from enlightening and if anything, even more tedious than the Lagrangian method. I present a non-standard proof that is not only less tedious but the use of projections gives way easily to the sum of squared identities used in F-test.

Inner products induced by a matrix

In the euclidiean inner product space, the orthogonal projector into the column space of a full-column-rank matrix $\mathbf{X}$ is given by $\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. This form obviously does not carry over to different inner products.

Lemma. *Let* $\mathbf{B}$ *be a positive-definite matrix. Consider the inner product:*

$$\langle\cdot,\cdot\rangle_{\mathbf{B}} : (\mathbf{x},\mathbf{y}) \mapsto \mathbf{x}^{\mathsf{T}}\mathbf{B}\mathbf{x}.$$

The orthogonal projector into the column space of a full-column-rank matrix $\mathbf{X}$ *under this inner product is given by*

$$\mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{B}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{B}.$$

Proof. Let $\mathbf{P}$ be the required projector. Then, since column space of $\mathbf{P}$ must equal that of $\mathbf{X}$, $\mathbf{P} = \mathbf{X}\mathbf{T}$ for some $\mathbf{T}$. We have that for every $\mathbf{y}, \mathbf{z}$,

$$\begin{aligned} &\langle \mathbf{X}\mathbf{z}, (\mathbf{I}-\mathbf{P})\mathbf{y}\rangle = 0 \\ \iff\ & \mathbf{z}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{B}(\mathbf{I}-\mathbf{P})\mathbf{y} = 0. \end{aligned}$$

This must be true for every vector $\mathbf{z}$ and $\mathbf{y}$, and in particular for the canonical basis vectors, which implies

$$\begin{aligned} \mathbf{X}^{\mathsf{T}}\mathbf{B}(\mathbf{I}-\mathbf{P}) &= 0 \\ \iff \quad \mathbf{X}^{\mathsf{T}}\mathbf{B} - \mathbf{X}^{\mathsf{T}}\mathbf{B}\mathbf{X}\mathbf{T} &= 0 \\ \iff \quad (\mathbf{X}^{\mathsf{T}}\mathbf{B}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{B} - \mathbf{T} &= 0 \end{aligned}$$

as desired. □

In the euclidean inner product space, the column space of a matrix $\mathbf{X}$ is orthogonal to its row space. A similar result exists for this inner product.

Lemma. *Under the preceding inner product, the orthogonal complement of* $\mathcal{C}(\mathbf{X})$ *is* $\mathcal{N}(\mathbf{B}^{-1}\mathbf{X}^{\mathsf{T}})$.

Proof. Let $\mathbf{z}$ be orthogonal to $\mathcal{C}(\mathbf{X})$. Then for all vectors $\mathbf{y}$,

$$\begin{aligned} \mathbf{z}^{\mathsf{T}}\mathbf{B}\mathbf{X}\mathbf{y} &= 0 \\ \iff \quad (\mathbf{X}^{\mathsf{T}}\mathbf{B}\mathbf{z})^{\mathsf{T}}\mathbf{y} &= 0 \\ \iff \quad \mathbf{X}^{\mathsf{T}}\mathbf{B}\mathbf{z} &= 0, \end{aligned}$$

where the last statement follows from the fact that the previous holds true for all $\mathbf{y}$. □

Projections into affine subspaces

Standard treatments of linear algebra generally only talk about projections into *linear* subspaces. Whenever a projection into an *affine* subspace is required, the affine subspace is converted into a linear subspace by an appropriate shift in the origin. However, treating affine projections as bona fide projections in their own right can be more instructive.

Definition (Affine subspace)**.** An *affine subspace* is a set of the form $S' = \mathbf{u} + S$, where S is a linear subspace. The orthogonal projection of $\mathbf{x}$ into S' is a vector $\hat{\mathbf{x}}$ such that

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{y},$$

where $\hat{\mathbf{x}} \in S'$, and $\mathbf{y} \in S^{\perp}$.

It can be easily verified that projections into affine subspaces are unique as well. This leads to an *affine* projector operator $\mathbf{P'}$. The following is also easy to verify.

Theorem. *Let* $\mathbf{P}'$ *be the orthogonal projector into* S'*, and* $\mathbf{P}$ *the orthogonal projector into* S*. Then,*

$$\mathbf{P}'\mathbf{x} = \mathbf{u} + \mathbf{P}(\mathbf{x} - \mathbf{u})$$

Affine projections—like linear projections—minimise norm of the difference.

Theorem. *The minimisation problem*

$$\begin{aligned} \min \quad & \|\mathbf{x} - \mathbf{z}\| \\ \text{s.t.} \quad & \mathbf{z} \in S' \end{aligned}$$

is attained at the orthogonal projection of $\mathbf{x}$ *into* S'*.*

Proof. Minimising $\|\mathbf{x} - \mathbf{z}\|$ over $\mathbf{z} \in S'$ is equivalent to minimising $\|(\mathbf{x} - \mathbf{u}) - (\mathbf{z} - \mathbf{u})\|$ over $\mathbf{z} - \mathbf{u} \in S$. This is attained $\mathbf{z} - \mathbf{u} = \mathbf{P}(\mathbf{x} - \mathbf{u})$, or equivalently $\mathbf{z} = \mathbf{P}'\mathbf{x}$. □

If a linear subspace is included in a larger one, iteratively projecting first, into the larger one, and then into the smaller one is equivalent to directly projecting into the smaller one. We have a similar result for affine subspaces.

Theorem. *Let* $S' = \mathbf{u} + S$ *be an affine subspace included in a linear subspace* T*, where* $\mathbf{u} \in T$*. Let* $\mathbf{P}'$*,* $\mathbf{P}$ *and* $\mathbf{Q}$ *be the orthogonal projectors of* S'*,* S *and* T *respectively. Then,*

$$\mathbf{P}' = \mathbf{P}'\mathbf{Q}$$

Proof. For any vector $\mathbf{x}$,

$$\begin{aligned} \mathbf{P}'(\mathbf{Q}\mathbf{x}) &= \mathbf{P}(\mathbf{Q}\mathbf{x} - \mathbf{u}) + \mathbf{u} \\ &= \mathbf{P}\mathbf{Q}\mathbf{x} - \mathbf{P}\mathbf{u} + \mathbf{u} \\ &= \mathbf{P}\mathbf{x} - \mathbf{P}\mathbf{u} + \mathbf{u} \\ &= \mathbf{P}(\mathbf{x} - \mathbf{u}) + \mathbf{u} \\ &= \mathbf{P}'\mathbf{x}, \end{aligned}$$

where we have used the fact that $\mathbf{P} = \mathbf{P}\mathbf{Q}$ since $S = S' - \mathbf{u} \subseteq T$. □

Proof of the estimator

Now, let's go back to regression problem. Let $\mathbf{X}^{\mathsf{L}}$ be a left-inverse of $\mathbf{X}$, and $\boldsymbol{\beta}_0$ be a particular solution to the constraint. Then, since $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$ if and only if $\mathbf{A}\mathbf{X}^{\mathsf{L}}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{c}$,

$$\begin{aligned} & \mathbf{A}\boldsymbol{\beta} = \mathbf{c} \\ \iff \quad & \boldsymbol{\beta} \in \boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}) \\ \iff \quad & \mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(A\mathbf{X}^{\mathsf{L}})) \end{aligned}$$

We thus want the solution to the problem

$$\begin{aligned} \min \quad & \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\| \\ \text{s.t.} \quad & \mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}\mathbf{X}^{\mathsf{L}})) \end{aligned}$$

Owing to the preceding theorem, we can first project $\mathbf{y}$ into $\mathcal{C}(\mathbf{X})$, and then into $\mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}\mathbf{X}^{\mathsf{L}}))$. The first projection is just the least squares projection $\mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the least squares coefficient. The second projection is the minimisation problem

$$\begin{aligned} \min \quad & \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}\| \\ \text{s.t.} \quad & \mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}\mathbf{X}^{\mathsf{L}})), \end{aligned}$$

but since $\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}\| = \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{\mathbf{X}^{\mathsf{T}}\mathbf{X}}$, this is equivalent to

$$\begin{aligned} \min \quad & \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{\mathbf{X}^{\mathsf{T}}\mathbf{X}} \\ \text{s.t.} \quad & \boldsymbol{\beta} \in \boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}), \end{aligned}$$

This is attained at the projection of $\hat{\boldsymbol{\beta}}$ into $\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A})$ under the inner product induced by $\mathbf{X}^{\mathsf{T}}\mathbf{X}$, or equivalently $\boldsymbol{\beta}_0$ plus the projection of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ into $\mathcal{N}(\mathbf{A})$. By the lemmas in the discussion of inner product, the projector into $\mathcal{N}(\mathbf{A})$ is the projector into the orthogonal complement of $\mathcal{C}((\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}})$ which is

$$\mathbf{I} - (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{A}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{A}$$

We thus have the final solution

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\beta}_0 + (\mathbf{I} - (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{A}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{A})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &= \hat{\boldsymbol{\beta}} - (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{A}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{A}^{\mathsf{T}})^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}), \end{aligned}$$

where it has been used that $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{c}$.

Decomposition of sum of squares

Let $\hat{\mathbf{y}}$ be the least squares projection of $\mathbf{y}$, and $\mathbf{y}_H$ the projection of $\mathbf{y}$ into the constrained affine subspace

$$\mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}\mathbf{X}^{\mathsf{L}})).$$

We have easily from the Pythagorean theorem that

$$\|\mathbf{y} - \mathbf{y}_H\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{y}_H\|^2$$

since $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathcal{C}(\mathbf{X})$ and $\hat{\mathbf{y}} - \mathbf{y}_H$ is in $\mathcal{C}(\mathbf{X})$. The projection approach shows clearly that $\mathbf{y}_H$ minimises $\|\hat{\mathbf{y}} - \mathbf{y}_H\|$ over the constrained space. This identity is usually written as

$$\mathrm{RSS}_H = \mathrm{RSS} + (\mathrm{RSS}_H - \mathrm{RSS}),$$

and the F-statistic is given by

$$\frac{(\mathrm{RSS}_H - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p)},$$

where p is the number of columns in $\mathbf{X}$ (including the column of 1s), and q the number of rows in $\mathbf{A}$ (i.e., the number of linear constraints). The numerator is just

$$\|\hat{\mathbf{y}} - \mathbf{y}_H\|^2 = \|(\mathbf{P} - \mathbf{P}_H)\mathbf{y}\|^2,$$

where $\mathbf{P}$ and $\mathbf{P}_H$ are the projectors into the $\mathcal{C}(\mathbf{X})$ and $\mathcal{C}(\mathbf{X}) \cap (\mathbf{X}\boldsymbol{\beta}_0 + \mathcal{N}(\mathbf{A}\mathbf{X}^{\mathsf{L}}))$. We thus have the F-statistic as

$$\frac{n-p}{q} \cdot \frac{\mathbf{y}^{\mathsf{T}}(\mathbf{P} - \mathbf{P}_H)\mathbf{y}}{\mathbf{y}^{\mathsf{T}}(\mathbf{I} - \mathbf{P})\mathbf{y}}$$

References

Rao, A. R., and P. Bhimasankaram. 2000. *Linear Algebra*. Texts and Readings in Mathematics. Hindustan Book Agency Gurgaon. https://doi.org/10.1007/978-93-86279-01-9.

Seber, G. A. F, and A. J. Lee. 2003. *Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley.