

ECE 586: Vector Space Methods

Lecture 18: Frequently Asked Questions

Henry D. Pfister
Duke University

November 8, 2022

1 Best Approximation Formulas

Why do we have the normal equations in Slide 1? The normal equations provide the matrix representation $G\underline{s} = \underline{t}$ for the best approximation problem. With this matrix representation, the solution of best approximation is derived by the pseudoinverse $(A^H A)^{-1} A^H$ of A as well as the projection matrix $P = A(A^H A)^{-1} A^H$ for the range of A .

On Slide 6, how can we represent the minimum mean-squared error by the inverse of G ? I don't quite understand the second term of it. Let $A = [X_1, X_2, \dots, X_n]$, $\hat{Y} = A\underline{s}$. Then,

$$\begin{aligned}\widehat{Y\hat{Y}} &= (A\underline{s})^H (A\underline{s}) \\ &= (AG^{-1}\underline{t})^H (AG^{-1}\underline{t}) \\ &= \underline{t}^H G^{-1} A^H A G^{-1} \underline{t} \\ &= \underline{t}^H G^{-1} G G^{-1} \underline{t} \\ &= \underline{t}^H G^{-1} \underline{t}.\end{aligned}$$

How does the Gramian inverse method compare against the Moore-Penrose inverse? Are they equivalent, does one only work under certain conditions? The Moore-Penrose inverse A^+ of a matrix A is the generalization of the inverse matrix. In particular, if A has linearly independent columns (which corresponds to the best approximation problem), then matrix $A^H A$ is invertible and A^+ is computed as $A^+ = (A^H A)^{-1} A^H$. If A has linearly independent rows (which corresponds to the minimum-norm solution problem), then matrix AA^H is invertible and A^+ is computed as $A^+ = A^H (A^H A)^{-1}$.

What is the use of projection operators? As in, what is the motivation behind needing them? Conceptually, projections are used to approximate things (e.g., general functions by polynomials). Projection operators allow us to see that the approximation mapping is linear and to understand its properties. The projection matrix $P = A(A^H A)^{-1} A^H$ projects a vector $\underline{v} \notin \mathcal{R}(A)$ onto $\mathcal{R}(A)$. Notable applications includes linear regression, linear minimum mean-squared estimation, etc. Please refer to section 4.4 in lecture notes for more details.

On Slide 8 and 9, what implies that the minimum-norm solution lies in the column space of $B = A^H$? Conceptually, this is best understood using the four fundamental subspaces described in the Video 20 slides. The minimum-norm solution should be orthogonal to the null space of A and this implies that it must lie in the range of A^H .

In this presentation, the result comes from the theorem on Slide 7, which states that the minimum-norm vector $\underline{w} \in V$ satisfying $\langle \underline{w}, \underline{w}_j \rangle = c_i$ for $i = 1, \dots, n$ also satisfies $\underline{w} = \sum_{i=1}^n s_i \underline{w}_i \in W$. Let $A = [\underline{w}_1, \underline{w}_2, \dots, \underline{w}_n]$, the matrix representation of $\langle \underline{w}, \underline{w}_j \rangle = c_i$ is $A^H \underline{w} = \underline{c}$. The theorem states that

the minimum-norm solution \underline{w} for the underdetermined linear system $A^H \underline{w} = \underline{c}$ is in the column space of A : $W = \mathcal{R}(A)$.

On Slide 3, why is \underline{t} called the cross-correlation vector? In signal processing, cross-correlation is a measure of similarity of two signals as a function of the displacement of one relative to the other (From Wikipedia <https://en.wikipedia.org/wiki/Cross-correlation>). For a vector of samples, this is also known as a sliding dot product or sliding inner-product. For a random vector, the definition is different but closely related.

For the optimal filtering problems described in EF 4.4.3-4.4.4, the RHS \underline{t} vector equals elements of the cross-correlation between the input and the desired output. Together, these two examples show the connection between deterministic least-squares problems and optimal linear filtering of random signals.

Historically, I think this term originated from these important problems. Now, it is still used for the general case even though it may not be well motivated.

For the weighted inner product, the normal equations look the same but solve a different problem. What is the difference more specifically? For a weighted inner product, the weighting is included in the definition of the inner product. Thus, the normal equations (using that inner product) look the same. But, one can also write a weighted inner product using the standard inner product and weighting matrix. Using this approach, the solution looks different but it mathematically equivalent. This is described in EF 4.3.2 - 4.3.3.

I saw positive-definite matrices in my undergrad linear algebra course and was told they are important. But, the only other place I've seen them is in this class. Can you give some common examples of their applications? This property plays a big role in two related areas. For the optimization of functions $f: \mathbb{R}^n \rightarrow \mathbb{R}$, the matrix of 2nd partial derivatives H (i.e., the Hessian matrix) determines if a stationary point is a local minimum. In particular, it is local minimum if H is positive definite and it is not a local minimum if H has a negative eigenvalue. If it has a zero eigenvalue, then higher-order derivatives are required to be sure.

Positive-definite matrices also appear regularly as Gramian matrices $G = A^H A$ for an arbitrary A . Optimization and vector spaces come together when one is minimizing the second-order expansion of a function at a point where the Hessian H is positive definite. In this case, the answer can be viewed either as a optimization problem or as a minimum-norm projection problem using the weighted inner product defined by H .