C106A Math Notes

Contents

1 Overview

This document is meant to be a brief overview of some of the math concepts you will find useful as you enter EECS/BioE/MechE C106A/206A Introduction to Robotics. This review is not a substitute for a real linear algebra class. Treat this document as a way of reviewing important facts and formulas and identifying topics that may be worth revisiting if they feel rusty.

2 Linear Algebra

2.1 Notation

- $\mathbb{R}^{m\times n}$ is the set of matrices with m rows and n columns with entries drawn from the real numbers.

- $\mathbb{R}^n$ is the set of n dimensional vectors. By convention, we treat $x \in \mathbb{R}^n$ as a column vector (an $n \times 1$ matrix). When we wish to denote a row vector, we do so by referring to it as x^T, the transpose of a column vector x.

- For $A \in \mathbb{R}^{m\times n}$ we denote by A_{ij} or a_{ij} the entry in the i^{th} row and j^{th} column of A.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- For $A \in \mathbb{R}^{m\times n}$ we denote by A_{ij} or a_{ij} the entry in the i^{th} row and j^{th} column of A.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- For $x \in \mathbb{R}^n$ we denote by x_i the i^{th} entry of vector x. We denote the j^{th} column of A by a_j.

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

and the i^{th} row of A by a_i^T

$$A = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$$

2.2 Matrix Multiplication

Here we survey some of the important ways of conceptualizing vector-vector, matrix-vector, and matrix-matrix products.

The *product* of two matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$ is a matrix $C = AB \in \mathbb{R}^{m\times p}$ with entries

$$C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

For this product to be defined, we need A to have as many columns as B does rows. Note that from the above formula it is evident that matrix multiplication is not commutative. Indeed, depending on the dimensions of the two matrices, the product BA may not even be defined. Even when it is defined, in general $AB \neq BA$.

The following properties are important to remember:

1. Matrix multiplication is associative: $(AB)C = A(BC)$.
2. Matrix multiplication distributes over addition: $A(B + C) = AB + AC$
3. Matrix multiplication is in general *not* commutative. It is possible that $AB \neq BA$.

2.2.1 Vector-Vector Products

1. Inner Product

For $x, y \in \mathbb{R}^n$, we denote the product $x^T y$ the *inner product* or *dot product* of x and y.

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

Note the result is a scalar and that this is simply a special case of the matrix multiplication formula when $m = p = 1$.

2. Outer Product

As a dual to the inner product, for $x, y \in \mathbb{R}^n$ we define the *outer product* of x and y as the matrix xy^T. The result is a matrix with entry $x_i y_j$ in row i, column j.

2.2.2 Matrix-Vector Products

Given $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$ their product $y = Ax$ is a vector in $\mathbb{R}^m$. There are two ways to conceptualize this product:

1. Column-centric Perspective

Let the columns of A be a_i. Then the vector $y = Ax$ can be written as

$$y = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

In other words, y is a *linear combination of the columns of A.* The coefficients of this linear combination are the entries of the vector x.

2. Row-centric Perspective

Let the rows of A be a_i^T. Then the vector $y = Ax$ can be written as

$$y = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

In other words, each entry of y is the dot product between x and a row of A.

2.2.3 Matrix-Matrix Products

We return to the multiplication of two general matrices, $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$. There are once again a few different ways of looking at the product $C = AB \in \mathbb{R}^{m \times p}$.

1. Write A in terms of its rows a_i^T and B in terms of its columns b_i. Then we can write each entry of C as an inner (dot) product

$$C_{ij} = a_i^T b_j$$

2. Write A in terms of its columns a_i and B in terms of its rows b_j^T. Then the product $C = AB$ can be written as the *sum of outer products*

$$C = \sum_{i=1}^{n} a_i b_i^T$$

   Note that each element of the above summation is a matrix of shape $m \times p$, and so the sum makes sense.

3. Write B in terms of its columns b_i. Then the i^{th} *column* of $C = AB$ is the vector Ab_i.

$$C = A \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

This is a very useful interpretation of matrix multiplication that you should be well aware of. In particular, this means that if we think of A as implementing some meaningful linear operator on vectors, then taking the product AB amounts to operating on *each column* of B according to A.

4. Analogously to the above, if we write A in terms of its *rows* a_i^T, then the i^{th} *row* c_i^T of C is given by the row vector $a_i^T B$.

2.3 The Identity Matrix

The square matrix with ones along the diagonal and zeros elsewhere is called the *Identity Matrix*. The $n \times n$ identity matrix is notated I_n. When the dimension is obvious from context, we drop the subscript and simply write I. Formally, I_n is the matrix with entries $I_{i,j} = 1$ if $i = j$ and 0 elsewhere.

The identity matrix has the property that for any matrix A for which the following products make sense, we have $IA = A$ and $AI = A$. As a transformation, the identity matrix implements the identity function.

2.4 Transpose

The *transpose* of a matrix $A \in \mathbb{R}^{m\times n}$ is denoted A^T and is an element of $\mathbb{R}^{n\times m}$, and is constructed by swapping the rows and columns of A. In particular, if $B = A^T$, then by definition $B_{ij} = A_{ji}$.

The transpose admits the following properties:

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(AB)^T = B^T A^T$. For products of more than 2 matrices, the arguments simply flip order.
4. $(cA)^T = cA^T$ for scalar c.

2.5 Determinant

The *determinant* is a function $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined on square matrices. Geometrically, the determinant of a square matrix $A \in \mathbb{R}^{n\times n}$ is the volume of the image of the unit square under the transformation $T(x) = Ax$. It is a polynomial in the entries of the matrix A, and is hence a differentiable function in each entry.

The following recursive algorithm for computing the determinant is important to know. It lets us write the determinant of an $n\times n$ matrix in terms of the determinant of $(n-1)\times(n-1)$ matrices, with the base case being that the determinant of a scalar (a "1 x 1" matrix) is the scalar itself.

Let $A(i|j)$ denote the $(n-1) \times (n-1)$ matrix you get if you remove the i'th row and j'th column from A. Then pick any row i. The following is a a formula for $\det A$.

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det A(i|j)$$

This formula is called the *cofactor expansion of the determinant along row i*. It is good to know how to use this formula for the $n = 2, 3$ cases. For $n = 2$,

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

and for $n = 3$,

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{array}{c} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}$$

The determinant admits the following properties:

1. $\det(I) = 1$.
2. If we multiply a single row by a scalar t, the determinant also gets scaled by t.
3. Swapping any pair or rows or any pair of columns flips the sign of the determinant.
4. $\det(tA) = t^n \det(A)$.
5. $\det(AB) = \det(A) \det(B)$.
6. $\det(A^T) = \det(A)$.
7. $\det(A^{-1}) = 1/\det(A)$

Note that from property (6) it follows that we can just as easily perform the cofactor expansion along a *column* rather than a row without changing the formula.

When $n = 3$, we have the following additional equivalent formulas that compute the determinant in terms of the columns a_i of $A \in \mathbb{R}^{3\times3}$:

$$\begin{aligned} \det(A) &= (a_1 \times a_2) \cdot a_3 \\ &= (a_3 \times a_1) \cdot a_2 \\ &= (a_2 \times a_3) \cdot a_1 \end{aligned}$$

where $\times$ is the *cross product* for 3D vectors and $\cdot$ is the dot product or inner product. This expression is called the *scalar triple product* of the vectors a_1, a_2, a_3.

2.6 Vector Norms

The *norm* of a vector $x \in \mathbb{R}^n$ is denoted $||x||$ and is a measure of the "length" of the vector x. Formally, a function $|| \cdot || : \mathbb{R}^n \to \mathbb{R}$ is called a *norm* if it satisfies the following defining properties

1. $||x|| \geq 0$ for all x.
2. $||x|| = 0$ if and only if $x = 0$.
3. $||\lambda x|| = |\lambda| \cdot ||x||$ for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.
4. [Triangle inequality] $||x + y|| \leq ||x|| + ||y||$ for all $x, y \in \mathbb{R}^n$

The norm of a vector x can also be thought of as the distance of the point x in Euclidean space from the origin. The most commonly used norm is the Euclidean ℓ_2 norm, which is essentially the familiar "Pythagorean" notion of distance

$$||x||_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

Where we also have the *very* useful property that

$$||x||_2^2 = x^T x$$

We can generalize this notion of distance to the family of p-norms. The ℓ_p norm for $p \geq 1$ is defined as

$$||x||_p = (x_1^p + \cdots x_n^p)^{1/p}$$

Special cases that are worth remembering on their own are the ℓ_2 norm from above, the ℓ_1 norm

$$||x||_1 = |x_1| + \cdots + |x_n|$$

and the ℓ_∞ norm

$$||x||_\infty = \max_i |x_i|$$

2.7 Linear Transformations

A function $T : \mathbb{R}^m \to \mathbb{R}^n$ is called a *linear* transformation if it *preserves vector addition and scalar multiplication.* In other words, T is a linear transformation if and only if

1. For all $x, y \in \mathbb{R}^m$, $T(x + y) = T(x) + T(y)$
2. For all $x \in \mathbb{R}^m$ and $c \in \mathbb{R}$, $T(cx) = cT(x)$

Note that it also follows that $T(0) = 0$.

2.7.1 Matrices as Linear Transformations

Given $A \in \mathbb{R}^{m \times n}$, it is easy to check that the function $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(x) = Ax$ is a linear transformation. In fact, the following converse is also true. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear map. Then there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that $T(x) = Ax$ for all x. In other words, the set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be identified with the set of $m \times n$ matrices.

Therefore, we can think of matrices as more than just a static array of numbers. Matrices are *dynamic*. They *acts* on vectors. In fact, every matrix is uniquely determined by the transformation that it implements.

In particular, if $b_1, ..., b_n$ is the standard basis for $\mathbb{R}^n$, then the matrix A corresponding to the linear map T is that matrix which has the vector $T(b_i)$ for its i'th column. In this way we can compute the matrix representation of any given linear map T.

2.7.2 Range and Rank

The ***range*** of a linear transformation is the set of vectors in its codomain that get mapped onto by some vector in its domain. It is the image of the domain under the linear transformation.

$$\text{range}(T) = \{v \in \mathbb{R}^m \text{ such that } \exists\ u \in \mathbb{R}^n \text{ with } v = T(u)\}$$

The range of a linear transformation is a vector subspace of its co-domain.

Since we can identify every linear transformation with a matrix and vice versa, we speak of the ***range space*** of a matrix as the range of the linear transformation it represents. Since we saw earlier that the matrix-vector product $T(x) = Ax$ amounts to taking a linear combination of the columns of A with coefficients from the vector x, it follows that the range space of A is *the span of the columns of A.*

The range space of A is denoted $\mathcal{R}(A)$.

The dimension of the range space of a matrix is called the ***rank*** of the matrix. It is equal to the size of the largest subset of the columns of A that is linearly independent. With some abuse of terminology, we also call this the *number of linearly independent columns of A.*

We say that A has ***full column rank*** when it has linearly independent columns. Likewise, we say it has ***full row rank*** when it has linearly independent rows. It can be proven (though we do not do so here) that the **row rank and column rank of any matrix are the same.**

The rank of a matrix admits the following properties:

1. $A \in \mathbb{R}^{m \times n}$ has $\text{rank}(A) \leq \min(m, n)$. When equality is achieved, A is **full rank**.
2. $\text{rank}(A) = \text{rank}(A^T)$
3. $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$
4. $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$

2.7.3 Nullspace and Nullity

The ***nullspace*** of a matrix A is the set of vectors that get mapped to zero by $T(x) = Ax$. It is a vector space, and is denoted $\mathcal{N}(A)$. The dimension of this space for a matrix A is called the ***nullity*** of A.

2.7.4 Rank-Nullity Theorem

As it turns out, there is a trade-off between the rank and the nullity of a matrix. In particular, for $A \in \mathbb{R}^{m \times n}$, the following holds:

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n$$

This fact illustrates that, as vector spaces, the "larger" the range of a matrix, the "smaller" its nullspace will be, and vice-versa. This further illustrates that if A is full column rank, its nullspace is 0 dimensional, and so its only element is the zero vector. When this is the case, we say that A has a *trivial nullspace.*

2.7.5 Inverse

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ can be represented by a square matrix $A \in \mathbb{R}^n$.

T is a bijection if and only if A is full rank.

To see this, assume that A is not full rank. Then by the rank-nullity theorem, there are nonzero vectors in its nullspace. If we are given a y and have some x such that $T(x) = y$, we can construct $x' = x + v$ for any v in the nullspace of A and we will get a new vector x' such that $T(x') = y$, and so T cannot be bijective. On the other hand if A is full rank, then columns are linearly independent and form a basis for its range. Since the columns are n linearly independent vectors in $\mathbb{R}^n$, they must form a basis for all of $\mathbb{R}^n$. Moreover, we know that every element in the range of A can be written as a linear combination of its columns. Since the columns are linearly independent, this linear combination is unique. So, T is both surjective and injective, and hence bijective.

When T is bijective, its inverse map T^{-1} exists and is also a linear map. The matrix associated with T^{-1} is denoted called A^{-1} and is referred to as the *inverse matrix* or simply *inverse* of A. It is the unique matrix that satisfies

$$AA^{-1} = A^{-1}A = I$$

The inverse admits the following properties

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^{-1})^T = (A^T)^{-1}$

Invertible Matrix Theorem: The following statements are equivalent.

1. A is invertible.
2. A^{-1} exists.
3. A has a trivial nullspace. i.e. the equation $Ax = 0$ has the unique solution $x = 0$.
4. A is full rank.
5. The equation $Ax = b$ has exactly one solution for each b, which is given by $x = A^{-1}b$.
6. The columns of A form a basis for $\mathbb{R}^n$.
7. $\det A \neq 0$.

2.7.6 Change of Basis

A ***basis*** $\mathcal{B} = \{b_1, ..., b_n\}$ for $\mathbb{R}^n$ is a set of n linearly independent vectors that together *span* $\mathbb{R}^n$. i.e. any vector can be expressed as a unique linear combination of the vectors in $\mathcal{B}$. When we write $v = \sum a_i b_i$ for scalar coefficients a_i, we say that $(a_1, ..., a_n)^T$ are *coordinates of* v written in the basis $\mathcal{B}$.

The *standard basis for* $\mathbb{R}^n$ is the one where we set b_i to be the vector with a 1 in the i^{th} entry and 0 elsewhere.

Whenever we write the coordinates of a vector as an n-tuple of scalars $v = (x_1, ..., x_n)^T$, implicit there is a choice of basis with respect to which these coordinates are given. All vectors are written with respect to that basis. It is easy to see that the coordinates of the basis vectors written with respect to themselves are exactly those of the standard basis.

Change of basis for vectors

Now let's say we wish to express a vector v whose coordinates $v_a = (\alpha_1, ..., \alpha_n)^T$ are given in a basis $\mathcal{A} = \{u_1, ..., u_n\}$, in a new basis $\mathcal{B} = \{w_1, ..., w_2\}$. We seek a set of coordinates $v_b = (\beta_1, ..., \beta_n)^T$ such that

$$\alpha_1 u_1 + \cdots \alpha_n u_n = \beta_1 w_1 + \cdots \beta_n w_n$$

Stack the vectors $\{u_1, ..., u_n\}$ into the columns of a matrix A, and likewise do the same for $\mathcal{B}$ to get a matrix B. Then we can re-write the above equation in matrix form. We seek a set of coordinates v_b such that

$$Av_a = Bv_b$$

Since A, B both have independent columns, we can write

$$v_b = B^{-1}Av_a$$

If v was originally given in the standard basis, then we would simply have

$$v_b = B^{-1}v_a$$

B^{-1} is called the *change of basis matrix* for $\mathcal{B}$, since it implements the linear transformation of mapping the coordinates of a vector in the standard basis to its coordinates in the basis $\mathcal{B}$. Note that in the expression $v_b = B^{-1}Av_a$, we simply transform v_a into the standard basis first using A and then transform the resulting vector to $\mathcal{B}$ using B^{-1}.

Change of basis for matrices

Now suppose we have a square matrix A which acts on and outputs vectors written in the standard basis, and we seek a matrix A' that performs the same transformation as A, but which does so on vectors written in the basis $\mathcal{B}'$ instead. We can do this as follows. Our simple approach will be to consider the linear transformation $T(x) = Ax$.

T expects inputs to be given in the standard basis, and outputs vectors in the standard basis as well. We seek a transformation T' which instead acts on and outputs vectors written in $\mathcal{B}$. So to implement T' we will simply transform the input point into the standard basis, pass the resulting vector through T, and then transform the result back into $\mathcal{B}$. This gives us

$$T'(x) = B^{-1}T(Bx)$$

where we recall that x is the coordinate of a vector written in basis $\mathcal{B}$. It is then easily seen that the matrix corresponding to T' is

$$A' = B^{-1}AB$$

Note that the same procedure can be used to find the matrix when the input and output spaces are required to be written in different bases. We simply need to use different basis change matrices for the two steps of the construction.

2.8 Eigenvalues and Eigenvectors

Given a square matrix $A \in \mathbb{R}^{n\times n}$ we say a scalar $\lambda \in \mathbb{C}$ is an ***eigenvalue*** with corresponding ***eigenvector*** $v \in \mathbb{C}^n$ if

$$Av = \lambda v$$

Geometrically, an eigenvector is a vector on which the effect of A is a simple scaling. To find eigenvalues, we rearrange the above equation a bit and note that it is equivalent to

$$(A - \lambda I)v = 0.$$

The above equation can only hold for nonzero v if $(A - \lambda I)$ has a non-trivial nullspace. Therefore, eigenvalues of A are exactly those scalars λ that make the matrix $(A - \lambda I)$ (or equivalently, $\lambda I - A$), non-invertible. Recall that this is equivalent to saying that $\det(\lambda I - A) = 0$.

The scalar function

$$p_A(\lambda) = \det(\lambda I - A)$$

is a degree n polynomial in λ and is called the ***characteristic polynomial*** of A. The eigenvalues of A, then, are exactly the *roots of this polynomial.*

Once an eigenvalue λ is found as a root of the characteristic polynomial, the corresponding eigenvector(s) can be found by solving for v in $(A - \lambda I)v = 0$.

2.8.1 Eigenspaces

Let $\lambda \in \mathbb{C}$ be a fixed eigenvector of A. There are, in general, infinitely many eigenvectors corresponding to λ, since if v is an eigenvector, it is easy to see that any scalar multiple cv is also an eigenvector. When this is the only ambiguity, we generally refer to "the eigenvector" for eigenvalue λ meaning an eigenvector of unit length.

However, this may not be the only kind of ambiguity in the eigenvector corresponding to λ. The set of all eigenvectors corresponding to an eigenvalue λ is referred to as E_λ and is, in general, a *vector space*. This is easy to see by recognizing that E_λ is the null space of $A - \lambda I$.

E_λ is called the ***eigenspace*** *of A corresponding to eigenvalue λ.* We can now see that in fact E_λ could be a space of any dimension at most n, depending on the rank of $A - \lambda I$. For instance, the identity matrix I has only one eigenvalue, 1. However, every vector in $\mathbb{R}^n$ is an eigenvector corresponding to this eigenvalue. In this case, E_1 is a space of dimension n and is in fact the whole parent space. In particular, $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ are all eigenvectors corresponding to the same eigenvalue, but are clearly not scalar multiples of each other.

2.8.2 Diagonalizability

Given the structure of eigenspaces, we in general have a lot of choice when picking eigenvectors for A. In particular, it *may* be the case that we can pick a set of n linearly independent eigenvectors. In this case, what we have is a basis for $\mathbb{R}^n$ made of eigenvectors of A. When this is possible, A is said to be *diagonalizable* because such a basis has the remarkable property that A becomes a diagonal matrix when written in this basis. In particular, there exists a *diagonal* matrix D such that

$$D = P^{-1}AP$$

or equivalently that we can write

$$A = PDP^{-1}$$

where $P = [v_1, ..., v_n]$ is constructed by setting the set of linearly independent eigenvectors of A as its columns, and D is the diagonal matrix where the i'th diagonal entry is λ_i, the eigenvalue corresponding to eigenvector v_i.

2.9 Applications

2.9.1 Systems of Linear Equations

A system of n linear equations in n unknown $x_1, ..., x_n$ can be written as a matrix equation for some constant matrix $A \in \mathbb{R}^{n \times n}$ and vector $b \in \mathbb{R}^n$. We seek to solve the following equation for $x \in \mathbb{R}^n$.

$$Ax = b$$

In general, this may have 0, 1, or infinitely many solutions. For a solution to exist, we require that b be in the range of A. If A is full rank, then it is invertible, and we have a unique solution $x = A^{-1}b$. If A is not invertible then there may be zero or infinitely many solutions depending on whether b is in the range of A.

2.9.2 Linear Least Squares

We often find ourselves concerned with situations where we have more equations than unknowns. In such a situation, the matrix A is rectangular, and the system is called *overconstrained.* As such, there may be no solution that satisfies $Ax = b$ exactly, so instead we seek the "best" solution, which we define to be the solution that minimizes the squared distance between Ax and b. To be precise, we seek to solve the following optimization problem.

$$\min_{x \in \mathbb{R}^n} ||Ax - b||_2^2$$

for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. When A is full column rank, this problem has a unique solution given by

$$x^* = (A^T A)^{-1} A^T b$$

This solution is found by taking the gradient of the objective $f(x) = ||Ax - b||_2^2$ and setting it to zero. Note that the condition that A have full column rank is necessary for $A^T A$ to be invertible.

2.9.3 Moore-Penrose Pseudo-inverse

Let's revisit the least squares solution to the overconstrained system $Ax = b$ from the previous section. Let us define the following operation on full column rank matrices A:

$$A^\dagger = (A^T A)^{-1} A^T$$

$A^\dagger$ is called the *left-pseudoinverse* of A. It is a generalization of the matrix inverse to all full-column rank matrices. It has the property that $A^\dagger A = I$ (i.e. it is a left inverse of A). This psuedo-inverse is unique when it exists, and when A is square, it is equal to the true inverse A^{-1}.

Comparing this matrix to the least squares solution from the previous section, we see that $A^\dagger b$ also gives us the least squares solution to the system $Ax = b$.

When A has full *row* rank instead, we can define a right-pseudoinverse given by

$$A^\dagger = A^T (A A^T)^{-1}$$

which satisfies $AA^\dagger = I$.