ECEU692 Intro to SSI Course Notes

Some selected topics in linear algebra for SSI (3)

Profs. DiMarzio and Brooks
Northeastern University

Spring 2004

Outline of those Selected Topics

- Some subsurface problems that lead to linear systems of equations
- Some different ways of looking at n equations in n unknowns: in particular, some geometric interpretations
- The geometry of when square linear systems have:
  - exact solutions,
  - no solutions, and/or
  - non-unique solutions
- **Problems where you have more equations than unknowns**
- **Problems where you have more unknowns than equations**
- **A quick discussion of solving linear systems in Matlab**

Outline of these Notes: Linear Systems that are not Square

- What about errors in the model?
- Overdetermined systems: motivation and geometry
- "Solving" overdetermined systems via least-squares
- What about when you can't make enough measurements ?
- Underdetermined systems: motivation and geometry
- A few words on underdetermined systems
- A few words on solving linear systems in Matlab

Reflectance Spectroscopy Again: Some Errors in our Setup

- In the reflectance spectroscopy problem, we have several sources of error:
  - We may not know, or be able to account for, all the chromophores present.
  - We may not have exact values for the spectra (true of your lab experiment).
  - There will be error in what we assumed to be known and constant (also true of your lab experiment).
  - There will be noise in our measurements: in optics, both electronic and shot noise (as again true of your lab experiment).
  - ...
- Note that the first 3 are examples of "model error",
- While the last item is an example of "measurement error"
- The result is that the linear system does not really describe reality.

More on thse Errors

- Model error means our model:
  1. does not describe the physics perfectly accurately ,
  2. so, does not perfectly predict (noise-free) measurements given true concentrations
- Consequence: even with perfect measurements, not really true that

$$\begin{bmatrix} b_{\lambda_1} \\ b_{\lambda_2} \\ \vdots \\ b_{\lambda_n} \end{bmatrix} = \begin{bmatrix} \kappa_1(\lambda_1) & \kappa_2(\lambda_1) & \dots & \kappa_m(\lambda_1) \\ \kappa_1(\lambda_2) & \kappa_2(\lambda_2) & \dots & \kappa_m(\lambda_2) \\ \vdots & \vdots & \dots & \vdots \\ \kappa_1(\lambda_n) & \kappa_2(\lambda_n) & \dots & \kappa_m(\lambda_n) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

- While "measurement error", in this case, has
  1. a signal-independent additive component (electronic noise), and
  2. a signal-dependent, additive component (shot noise).
- Measurement noise can also be multiplicative (say, a gain factor that times the signal) or even more non-linear (say, clipping from saturation).

DOT Again: Some Errors in our Setup

- Error sources:
  - Diffusion equation is an approximation.
  - Linearization is an approximation.
  - Estimation of background absorption and scattering are approximate.
  - Position of sources/detectors and coupling of light to tissue are approximate.
  - Discretization introduces approximations.
  - There will be noise in our measurements: in optics, both electronic and shot noise.
  - ...
- Again, we have both model error and measurement error.

One Way to Deal with Error: Use More Equations than Unknowns

Let's talk about reflectance spectroscopy problem:

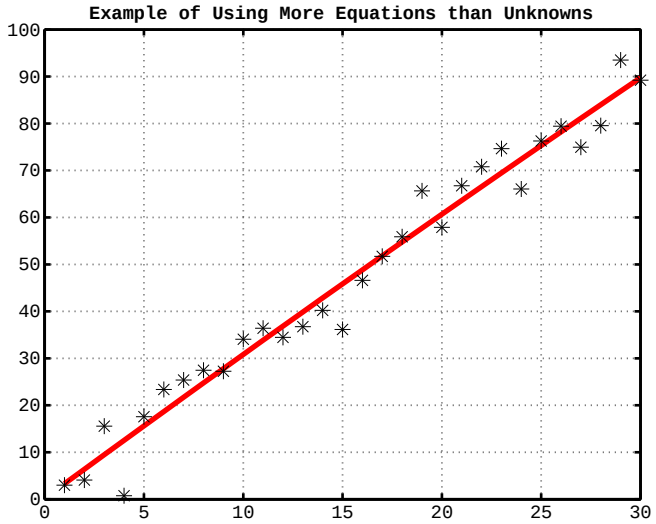
- If everything were perfect, there would be no reason to take more measurements than there are unknown chromophores in model.
- In the case of noise and model error, though
  - Solving with the exact number of equations fits
    1. model error, and
    2. noise

    along with the model and the data
  - If we use some "extra" measurements, we can hope to "average" out the error
- The measurement noise often
  - has zero average value, so the mean of the noisy measurements is the true (noise-free) measurement.
  - is uncorrelated with the true measurement.

  ... At least that's often a reasonable assumption ...

So Let's Hedge Our Bets

- Want an approximate solution that takes more measurements into account,
- Hope to average out the + and − errors to decrease their effect.
- How about the model error?
  - May have components whose error is zero mean
  - But much less likely to be uncorrelated with measurment



- Ideally, the error goes down as $1/\sqrt{n}$ for n measurements.

Rectangular (non-Square) Systems: Overdetermined Case

Consider the case where we have more equations than unknowns:

- $\mathbf{Ax} = \mathbf{b}$, with $\mathbf{A}$
  1. having m rows and n columns (we say it is m by n ($m \times n$)) (say, $\mathbf{3 \times 2}$)
  2. Here $m > n$
- This is know as an *overdetermined* (or tall, thin) set of equations
- Since we have only n $(\mathbf{2})$ columns: largest possible dimension of the column space is n $(\mathbf{2})$.
- But the columns are vectors of length n $(\mathbf{3})$.
- So we don't have enough columns to fill out all of $\mathcal{R}^m$ $(\mathcal{R}^3)$.
- This means that the dimension of the *range* (or column space) must be $\leq n < m$, since the column space is not "big enough" to fill out $\mathcal{R}^m$.

What are the Implications?

- Remember from square systems, two bad things can happen:
  1. Inconsistent measuremnts (not in the column space)
  2. Invisible solutions (in the null space)
- In our case:
  - There are possible measurements $\mathbf{b}$ that don't fit our model— $\mathbf{b}$ could be *any* vector in $\mathcal{R}^m$, but our model only can explain those lying in a n-dimensional subspace (a plane in 3-space).
  - In fact, if measurements are noisy, the probability is very high that $\mathbf{b}$ is not in the column space.
- So in reflectance spectroscopy using "extra" wavelengths, we don't expect to be able to solve the linear equation exactly.

What about the nullspace ?

- Remember that nullspace vectors are non-zero solution vectors ($\mathbf{x}$'s) that produce zero $\mathbf{b}$ (measurememnts).
- Thus they have the same number of elements as the number of columns, $\boldsymbol{n}$ $(\mathbf{2})$.
- Now, if the columns of $\mathbf{A}$ are linearly independent, we can't find a set of $\{\boldsymbol{x_i}\}$'s that make them $\mathbf{0}$
- So there are no non-zero vectors in the nullspace
- This means we don't have to worry about "invisible" components in the solution (non-uniqueness)
- In this case:
  - if $\mathbf{b}$ happens to be in the column space it is a unique solution.
  - if $\mathbf{b}$ is not in the column space there is no (exact) solution.
  - Interpretation: in this last case the problem is "overconstrained". You don't have enough degrees of freedom to solve it.

And if the Columns are not Independent . . . ?

- Now, if the columns of $\mathbf{A}$x are linearly dependent as well as the rows:
  - Then the dimension of the column space is $< n$, which itself is $< m$.
  - So we have some possible concentrations that we can't see in the measurements.
  - Thus in this case:
    - if $\mathbf{b}$ is in the column space there are an infinite number of solutions.
    - if $\mathbf{b}$ is not in the column space there is no (exact) solution.
- Summary: in the $\mathbf{3 \times 2}$ case:
  1. if the columns are independent, and if $\mathbf{b}$ is in the plane they define, there is a unique solution
  2. if the columns are not independent, and if $\mathbf{b}$ is in the line they define, there are infinitely many solutions
  3. if $\mathbf{b}$ is not in the column space, whether plane (independent) or line (dependent), there is no exact solution

What Do We Do About Systems with No Exact Solution?

- Consider an overdetermined system of equations:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b},$$

  with $\mathbf{A}$ being $m \times n$, $m > n$.

- We know from previous discussion that, in general, $\mathbf{b}$ will not be in the range space of $\mathbf{A}$, since we don't have enough length m vectors in the $n < m$ columns of $\mathbf{A}$ to fill out $\mathcal{R}^m$.

- We could choose some set of n "best" equations and solve it, but
  - Need to choose "best" from $\binom{m}{n}$ possibilities.
  - Throwing away potentially useful measurements.

- Instead: Let's find a solution $\mathbf{x}^*$ such that we minimize some measure of the difference between $\mathbf{A} \cdot \mathbf{x}^*$ and $\mathbf{b}$.

- Then $\mathbf{x}^*$ will be provide an "optimal fit" to our measurements, in whatever way we define "optimal".

Least Squares Solutions

- Want a solution that comes "close" to fitting model to measurements.
- How do we measure "close"? How about making the simple differences (errors) between actual measurement and $\mathbf{A} \cdot \mathbf{x}$,

$$\mathbf{e} = \mathbf{b} - \mathbf{A} \cdot \mathbf{x}$$

  as small as possible?
- So we need to make the vector $\mathbf{e}$ "short": but how do we measure "length" of a vector?'
  - In vectorspeak, length is called a 'norm'.
  - There are many vector norms, but we'll use the 2-norm:
    - For a vector $\mathbf{x} = [x_1,\ x_2,\ \ldots,\ x_n]^T$,

$$\|\mathbf{x}\|_2 = \left[\sum_{k=1}^{n} x_k^2\right]^{1/2}.$$

    - Corresponds to usual Euclidean sense of length, by far most commonly used
    - Others include maximum error and using first power (absolute values) instead of second power (squares and square roots)

Why least-squares?

- We minimize the sums of the squared errors, so this is called "least-squares"
- reasonable: intuitive definition of "size" of error
- convenient: taking derivatives leads to linear equations
- optimal: in certain common statistical settings, especially where Gaussian distributions are used.
- related to 'fitting a line to data', 'regression lines', etc.
- aside: a computational trick: for a vector $\mathbf{x}$

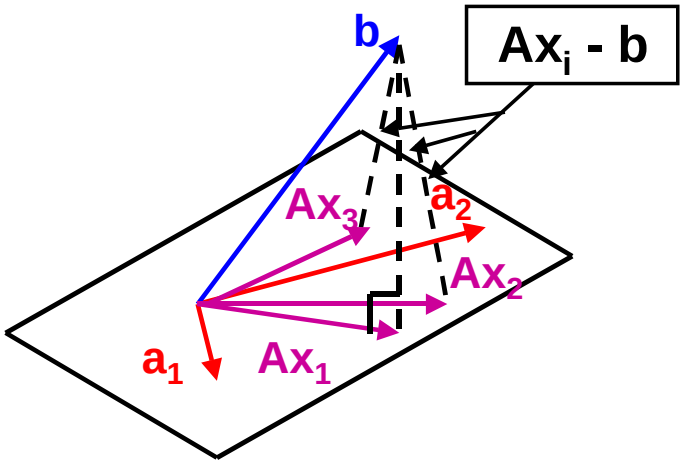
$$\|\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{x}$$

.

- We will look at solution first geometrically, then algebraically.

The Geometry of Least Squares

Consider visualize-able dimensions: 3×2 system:

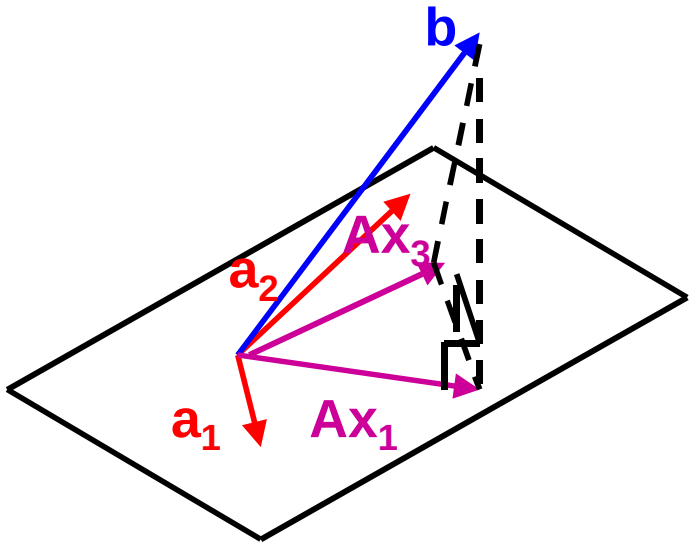


- Red vectors $\mathbf{a_i}$ are columns, and plane is column space.
- Blue vector $\mathbf{b}$ is right-hand side.
- Pink vectors $\mathbf{Ax_i}$ are approximations resulting from using various $\mathbf{x}$'s as possible solutions (note they have to be in the column space (the plane)
- Which $\mathbf{Ax_i}$ vector in the plane gives the shortest error vector

$$\|\mathbf{A} \cdot \mathbf{x}^* - \mathbf{b}\|_2?$$

Why Be Normla?

Consider any other solution besides the one for which the error vector is orthogonal (normal) to the column space:



- We can decompose that error into a component in the column space **plus** one orthogonal to the column space.
- By the triangle inequality, the orthogonal error is shorter.

So What Geometry Does This Tell Us?

- For the least-squares optimal solution, the error is orthogonal to its approximation to $\mathbf{b}$:

$$(\mathbf{A} \cdot \mathbf{x}^*)^T(\mathbf{A} \cdot \mathbf{x}^* - \mathbf{b}) = \mathbf{0}.$$

- Using the fact that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, we get

$$\mathbf{x}^{*T}\mathbf{A}^T(\mathbf{A} \cdot \mathbf{x}^* - \mathbf{b}) = \mathbf{0}.$$

- If columns of $\mathbf{A}$ are independent, and we want $\mathbf{x}^* \neq \mathbf{0}$, finding $\mathbf{x}^*$,

$$\mathbf{A}^T\mathbf{A}\mathbf{x}^* - \mathbf{A}^T\mathbf{b} = \mathbf{0},$$

or

$$\mathbf{A}^T\mathbf{A}\mathbf{x}^* = \mathbf{A}^T\mathbf{b}.$$

- Note that the fact of the orthogonality of the error (also called the "residual") and the projected (onto the column space) solution is *enough* to specify the solution.
- One example of very powerful and widely used Projection Theorem.

What About Solving This System?

$$\mathbf{A}^T\mathbf{A} \cdot \mathbf{x}^* = \mathbf{A}^T\mathbf{b}.$$

- Called the "normal equations".
- Extremely common and important in many fields: social science, experimental analysis in physical science, statistics, signal processing, communications, ...
- When do they have a (unique) solution? When the nullspace of $\mathbf{A}^T\mathbf{A}$ is empty!
- FACT: the nullspace of $\mathbf{A}^T\mathbf{A}$ is the same as the nullspace of $\mathbf{A}$ itself.
- Proof next, but first what does this mean?
  - Recall that $\mathbf{A}$ is $m \times n$, $m > n$.
  - From earlier, this means that its nullspace is empty if the n columns are linearly independent (full column rank).

The nullspace of $\mathbf{A}^T\mathbf{A}$

- We want to show that it is simply the nullspace of $\mathbf{A}$. To do this we need to show that
  1. if a vector $\mathbf{y}$ is in the nullspace of $\mathbf{A}$ it's in the nullspace of $\mathbf{A}^T\mathbf{A}$, and
  2. if a vector $\mathbf{y}$ is in the nullspace of $\mathbf{A}^T\mathbf{A}$ it's in the nullspace of $\mathbf{A}$.
- First, if $\mathbf{y}$ is in the nullspace of $\mathbf{A}$, then

$$(\mathbf{A}^T\mathbf{A})\mathbf{y} = \mathbf{A}^T(\mathbf{A}\mathbf{y}) = \mathbf{A}^T\mathbf{0} = \mathbf{0},$$

  so the first point is satisfied.
- To show the second point, assume $(\mathbf{A}^T\mathbf{A})\mathbf{y} = \mathbf{0}$.
- Then if we premultiply this by $\mathbf{y}^T$, we get

$$\mathbf{y}^T\mathbf{A}^T\mathbf{A}\mathbf{y} = \mathbf{0}.$$

- But this expression is the 2-norm of the vector $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{y}$, so if the vector $\tilde{\mathbf{y}}$ has length $\mathbf{0}$ then it must be the zero vector itself.

So What's the Algorithm?

Assuming we know that $\mathbf{A}$ is full column rank:

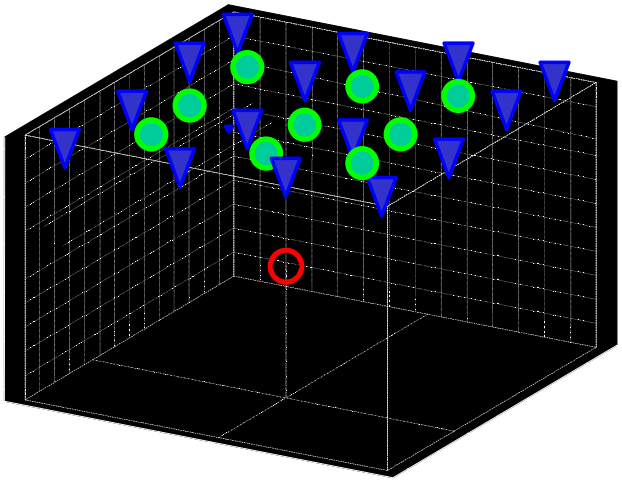
- Compute the matrix $\mathbf{A}^T\mathbf{A}$
- Compute the vector $\mathbf{A}^T\mathbf{b}$
- Solve the system $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$
- Note that $\mathbf{A}^T\mathbf{A}$ is a square (and symmetric) matrix so we can solve this system by ordinary methods.
- In good numerical practice, there are ways to do this that avoid explicit computation of $\mathbf{A}^T\mathbf{A}$.
- Note: in the case that $\mathbf{A}$ is square and invertible, this reduces to the original problem $\mathbf{b} = \mathbf{A}\mathbf{x}$, since
  1. $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ for any two invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,
  2. Then $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{A}^{-1}\boldsymbol{A}^{-T}\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Let's Take Stock of Where We Are

- We have a geometric picture of when matrix-vector problems have a unique solution
- We have a method to solve problems with more equations than unknowns
- Some things still to address
  1. Suppose we can't get enough equations, that is, we have more unknowns (columns) than measurements (rows)?
  2. Suppose we have enough (or more than enough) equations—at least as many measurements (rows) as unknowns (columns)—but the columns are not linearly indepedent
  3. Suppose the columns or rows are linearly independent, but just barely ...
- We'll deal with the first one of these here

Why Use Less Equations than Unknowns?

This is never a desired situation, however if we are limited by a physical situation as to where we can make measurements:



DOT Reflectance Simulation Geometry: 9 sources, 16 detectors, = 144 measurements, about 10 times as many voxels

More on Less (Equations . . .)

- When we can't get enough measurements because of:
  1. the shape of the part of the body of interest
  2. need to stay above the surface of the water
  3. not being able to sample the ground very often
  4. the cost/time to acquire data is important
  5. and we want to maintain a certain resolution in our model of the subsurface phenomenon
- then we face this problem:
  Our system of equations has fewer equations than unknowns ($\boldsymbol{m}$ ¡ $\boldsymbol{n}$ if matrix is $\boldsymbol{m} \times \boldsymbol{n}$).
- Problem is underconstrained: called *underdetermined*.
- Matrix is "short and fat" not "tall and thin"

Summary of Underdetermined Case

- Rows are length n vetors,and we have m of them
- But $m < n$, so we have "not enough" row vectors to go in all directions of the higher-dimensional row space
- This means there must be vectors in $\mathcal{R}^n$ which are orthogonal to all the rows
- And this means that a non-empty nullspace always exists.
- Which in turn (see previous slide) means that solutions are not unique, since we can add to any solution a second solution in the nullspace and not be able to see it.
- And in fact there are an infinite number of solutions since we can always scale the null-space component of a solution
- In DOT terms, this would mean that there are different ways the absorption coefficient could be distributed that would look the exactly the same to our measurements, even with no measurement noise.
- Therefore we need some way to choose a *unique* solution.

The Most Common Solution for Underdetermined Systems

- Most common choice is to pick the *smallest* solution from the infinite number of possible solutions: called **min-norm**.
- Intuitively, this means the null-space component is zero (since otherwise the solution would be longer . . .)
- Leads to the following:

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

- Note: in the case that $\mathbf{A}$ is square and invertible, this reduces to the original problem.
- The inverse in this equation exists when $\mathbf{A}$ has rank m (that is, it is full row rank).
- No particular reason to choose min-norm, and often does not give a good solution !!

So What's This Algorithm?

Assuming we know that $\mathbf{A}$ is full row rank: we want $\mathbf{x}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$

- Compute the matrix $\mathbf{A}\mathbf{A}^T$
- Solution 1:
  1. Compute $(\mathbf{A}\mathbf{A}^T)^{-1}$
  2. Multiply it by $\boldsymbol{A}^T$
  3. Multiply this by the measurement vector
- Computing matrix inverses is not the best numerical practice
  1. It's more (computational) work than necessary
  2. It can amplify numerical errors
- So here's Solution 2:
  1. Solve the system $\mathbf{A}\mathbf{A}^T\mathbf{u} = \mathbf{b}$ for $\mathbf{u}$.
  2. Multiply $\mathbf{A}^T\mathbf{u}$ to get $\mathbf{x}^*$

Comparison of Least-Squares and Min-Norm Solutions

- Least-squares for overdetermined system—more measurements than unknowns
- Min-norm for underdetermined system—more unknowns than measurements
- Least squares equation is:

$$\mathbf{A}^T\mathbf{A} \cdot \mathbf{x}^* = \mathbf{A}^T\mathbf{b}.$$

- Min-norm is:

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

- Both reduce to the same equation for square and full-rank $\mathbf{A}$
- Key remaining questions: what about if system is
  1. Square but not full-rank
  2. Not square and neither full column rank nor full row rank
  3. Most important: Technically full rank but . . . oh so close to not being so

  To be continued . . .

Some Comments about Solving Systems in MATLAB

- You can explictly compute the inverse with the `inv(A)` command.
- However this is generally not the best thing to do.
  - It's better to stop the row operations at upper triangular form and use back-substitution.
  - Otherwise you do more work than necessary to get there using Gaussian elimination, and
  - Results are more sensitive to numerical error (accumulation of round-off error, for instance).
- Better to use Matlab command \.
- This allows Mathworks to use the best current algorithms they can find and you to know nothing about it ...

Details on the Matlab \ command

- $x = A \backslash b$
- Solves square systems.
- For overdetermined systems, calculates least-squares solution.
- For underdetermined system, does *not* find min-norm solution: rather it finds one which minimizes the number of non-zero elements (maximally sparse)
- What about the $(\mathbf{A}\mathbf{A}^T)^{-1}$ we need for underdetermined min-norm solutions? Remember Solution 2 from a couple slides ago:
  1. Set $(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$ equal to some vector $\mathbf{u}$.
  2. Then $\mathbf{u}$ is the solution to $(\mathbf{A}\mathbf{A}^T)\mathbf{u} = \mathbf{b}$.
  3. So solve this system for $\mathbf{u}$, *e.g.* by using $\mathbf{u} = \mathbf{A}\mathbf{A}^T \backslash \mathbf{b}$.
  4. Then premultiply $\mathbf{u}$ by $\mathbf{A}^T$