MTH 420/520 Spring 2019
Class notes

Malgorzata Peszyńska

Department of Mathematics, Oregon State University

Spring 2019

In this class ...

- We describe certain common discrete and continuous models and discuss the corresponding methods
  1. Equilibrium models: SPD systems, minimization of convex functions, basics of the calculus of variations
  2. Regression, and classification: Least Squares, and dimension reduction (SVD, PCA)
  3. Saddle point systems: Lagrange multipliers, constrained optimization
  4. Oscillatory systems and analyses: Fourier analysis and applications
- For each model, students learn the challenges, and derive the methods of analysis and solution
- Students use theory and MATLAB to solve the problem

This class is ...

- NOT a class on (abstract) linear algebra

  but you will use the concepts all the time.

  If needed, please review material from MTH 341. You can also take MTH 342 or MTH 443 to prove some of the facts we use in this class.

- NOT a class on differential equations (ODEs or PDEs)

  but many models will be formulated as ODEs or PDEs

  If needed, please review material from MTH 256. Other classes of interest include MTH 4/582, MTH 621-2-3.

- NOT a physics or chemistry or CS or ... class

  but many models will require intuition from college level classes

- NOT a numerical analysis or programming class ...

  but we will use MATLAB in examples and HW.

  If needed, please consult resources such as

- NOT a machine learning (data science) class

  but you will see some examples and their mathematical models

- NOT a WIC

  but you will be expected to write at the MTH 4XX level. Typing is encouraged!

References and math you will need

- [GS] G. Strang, Introduction to Applied Math, Wellesley-Cambridge, 1986
- [LO] J.D. Logan, "Applied Mathematics", Wiley 1987
- [HA] P.S. Hansen, "Discrete Inverse Problems. Insight and Algorithms", SIAM 2010
- [DL] D.P. O'Leary, "Scientific Computing with Case Studies", SIAM 2009
- [CM] Cleve Moler's books and materials; http://www.mathworks.com/moler/
- [GE] A. Geron, Hands-on Machne Learning with Scikit-Learn & TensorFlow, O'Reilly, 2017

If needing a refresher on MTH 341, please consult https://tinyurl.com/yc8lxz2u. For those interested in the theory behind the concepts, please consult https://www.math.brown.edu/~treil/papers/LADW/LADW_2017-09-04.pdf or https://www.math.ucdavis.edu/~anne/linear_algebra/mat67_course_notes.pdf (some of the material is covered in MTH 342).

You will be expected to work through the examples in these course notes. Please contact the instructor immediately if you are struggling: identify what "math" or "computational skills" you are missing.

Linear algebra notation and MATLAB

We will work primarily in $\mathbb{R}^n$ (as opposed to C^n).

- Column vectors can be written as $u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ or $u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$. In these class notes we will blend the latter notation with the shorthand MATLAB notation for column vectors `u=[1;2;3]=[1,2,3]'` so we will write $u = [1;2;3]$ or $u = [1,2,3]^T$.
  In turn, $v = [0,-1,5]$ will denote a row vector so that $v^T = [0;-1;5]$.
- Dot product (inner product) of $a, b \in \mathbb{R}^3$ is $a^T b = \sum_j a_j b_j$; in MATLAB `a'*b`. We will also write $(u, v) = v^T u$. Dot product is symmetric
- Norm of a vector $\| x \|_2$ is the usual Euclidean norm. We will also consider alternative norms.
- For matrices, we write $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ as $A = [1,2,3;4,5,6]$. In MATLAB you can also write the same matrix as `A=[1,2 3;4,5 6]`.

Pbm. I.1 MATLAB expertise. Work through

http://math.oregonstate.edu/~mpesz/teaching/420-520_S19/matlab_intro.html

LA concepts to be fluent in

1. **Vectors:** linearly independent, orthogonal ($u^T v = 0$), orthonormal (if $\|u\| = 1$, basis. Angle between vectors.

`Pbm.` I.2 Consider $u = [1,2,0]^T$, $v = [-1,0,5]^T$. Check if (u,v) are linearly independent, orthogonal, orthonormal; check their angle. Find a third vector w if possible so that (u,v,w) is a basis for $\mathbb{R}^3$. Write $a = [2,0,-5]^T$ in this basis.
Find a vector $\tilde{v}$ so that $(u,\tilde{v})$ is an orthogonal (orthonormal) set (an orthogonal set spanning $Col(u,\tilde{v})$) ? Repeat for $(u,\tilde{v},\tilde{w})$ for some $\tilde{w}$ that you would determine. Is it possible to do this for $(u,\tilde{v},a,\tilde{w})$? Check how to use MATLAB to check your answers.

2. **Square matrices:** singular, invertible; determinants; inverse matrix. Eigenvalues and eigenvectors. Diagonalizable matrices. *Algebraic multiplicity and geometric multiplicity of an eigenvalue.*

3. **Rectangular matrix** A. Consider $Col(A), Range(A), Ker(A), Null(A)$, A^T. Linear solvable systems, underdetermined and overdetermined systems. (For systems, use some unit vector on the right hand side).

`Pbm.` I.3 Work with matrices below to practice concepts in (2-3). Use hand calculations and MATLAB.
$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$,
$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. $A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Equilibrium model: spring-mass system

Example I.1 (Spring-mass system)

Variables and coefficients (data and unknowns)

n masses: $m_i, i = 1 \ldots n$, external forces f_i, displacements x_i,
$n+1$ springs: elasticity constant $c_j; j = 1, \ldots n+1$; elongations e_j,
internal forces y_j.

Draw a picture.

Model connects data and unknowns with physics laws:

Define $e_j = x_j - x_{j-1}$, use convention $x_0 = 0, x_{n+1} = 0$. $(e = Ax)$
Hooke's law $y_j = c_j e_j$. $(y = Ce)$
Newton's law $f_i + y_{i+1} - y_i = 0$ $(f = A^T y)$

Check.

Count the unknowns and equations. Check dimensions.

Solve the EQ problem

Summarize and solve.

Decide on the unknowns. Choose $x = (x_i)_i$; define $K = A^T CA$.

$$Kx = f. \tag{1}$$

Example I.2 (Equilibrium as minimization)

Equation (1) is related to finding the minimizer of the potential energy function

$$x = \text{argmin } \phi(r), \;\; \phi(r) = \tfrac{1}{2} r^T K r - r^T f \tag{2}$$

(Elastic energy $\sum_j \frac{1}{2} c_j e_j^2$: each $c_j e_j^2$ arises from the work of internal force ce done on the stretch de; integrate $\int cede$; balanced by the external work on the masses $\sum_i x_i f_i$).

Analyze: are (1) and (2) solvable? Are they equivalent?

Yes and yes.
We show this later via the spd property of the matrix K.

Other equilibrium (stationary) models with structure as in (1)

Example I.3 (Models in 1d)

Heat conduction. Diffusion equation. (Linear) elasticity. Electric circuits. Flow in porous media.

`Pbm. I.4 Connect the laws below to the applications. What are the unknowns? Are they constitutive or balance laws?`

`Energy conservation. Darcy's law. Fourier's law. Fick's law.`

`Kirchoff's law. Hooke's law. Ohm's law. Mass conservation.`

Example I.4 (Discrete (network) models)

Primary variables: scalar, defined at the nodes of a network, connected by an incidence matrix A (representing a directed graph or a network)
Secondary variables: fluxes. defined on the edges.

Connect them by constitutive and balance laws.
Summarize. Analyze. Solve.

Does the order ("direction" in network) matter?

Example I.5 (Spring-mass system, revisited)

Rewrite constitutive equations for spring-mass system changing the signs in some of the definitions of e_j. (or the interpretation of x_i or of f_i). For example, write $e_2 = x_1 - x_2$ instead of $e_2 = x_2 - x_1$.
Check properties of matrix K. Better yet: check the entries!
Solve the problem, apply the new interpretation of x_i, y_j, f_i.

Convince yourself that the physical solution does not depend on the signs you prescribed in your definitions.

`Pbm. I.5 What if there are no walls? One wall only?`

Example I.6 (Network)

Draw a triangle with vertices $1, 2, 3$; assign to these some "potential" values x_1, x_2, x_3, and some sources f_1, f_2, f_3. Assign some direction to the edges e_1, e_2, e_3. Write the model, with appropriate definitions of e_j given by the direction of the edges (e.g., if edge e_1 is from node 1 to 2, write $e_1 = -x_1 + x_2$). Also, write $y_j = c_j e_j$ for every edge. Close with balance laws $f_i = \dots$ (total flow in minus total flow out).
Check properties of the matrix K. Is the problem $Kx = f$ solvable?

Redo the problem by changing the direction of the edges. The matrix K should not change.

The problem is not solvable unless we fix (ground) one of x_i. Apply and redo.

Symmetric positive definite matrices

Definition I.7

$A \in \mathbb{R}^{n\times n}$ is symmetric if

$$A = A^T \tag{3}$$

Let $A \in \mathbb{R}^{n\times n}$ be symmetric. It is positive definite (spd) if

$$x^T A x > 0, \forall x \in \mathbb{R}^n, x \neq 0. \tag{4}$$

One can also define semidefinite (nonnegative definite) matrices. These satisfy $x^T A x \geq 0$.

`Pbm. I.6 Work out conditions that guarantee that` $A \in \mathbb{R}^{2\times 2}$ `is spd, using the definition` (4).

Hint: start by assuming $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *with some* $a, b, c \in \mathbb{R}$.

Write out what $x^T Ax$ *is. Now, how do you guarantee the inequality* (4) *holds for any* x*? Find the conditions on* a, b, c.

`Answer: We must have` $a > 0$, `and` $ac - b^2 > 0$.

Characterization of spd matrices

Assume $A = A^T$.

Theorem I.8 (Characterization of spd matrices (a-b))

(a) *A is spd iff its leading principal minors are all positive (this connects to the pivots found during LU decomposition)*

The kth leading principal minor of a matrix is the determinant of its upper-left $k \times k$ submatrix.

(b) *A is spd iff its eigenvalues are all positive (this connects to diagonalization of A and an equivalent inner product on $\mathbb{R}^n$).*

Recall the connection between the eigenvalues, the trace, and the determinant of the matrix, i.e., $\lambda_1 + \lambda_2 = \ldots$ and $\lambda_1\lambda_2 = \ldots$.

`Pbm. I.7 Determine the conditions that guarantee that` $A \in \mathbb{R}^{2\times 2}$ `is spd using the characterization (a) and (b). Compare to the conditions you found in Pbm I.6.`

Spd matrix and convexity

Recall the concept of functions on $\mathbb{R}$ that are "convex up" from calculus. Generalize this concept to $\mathbb{R}^n$.
(It can be also generalized to functionals = functions of functions).

Definition I.9

$\phi : V \to \mathbb{R}$ is strictly convex if its graph lies always strictly below its secants, or

$$\phi(tx + (1-t)y) < t\phi(x) + (1-t)\phi(y), \;\; \forall t \in (0,1), \; \forall x, y \in \mathbb{R}^n. \tag{5}$$

The vector $tx + (1-t)y$ is called a convex combination of x, y.

Pbm. I.8 Start with $n = 1$. Convince yourself that $\phi(x) = ax^2$ is convex using (5), as long as $a > 0$. What happens if $a = 0$? What about $ax^2 + dx + f$?

Pbm. I.9 Consider $A = [a, 0; 0, c]$ and $\phi(x) = ax_1^2 + cx_2^2$. What conditions on a, c guarantee that $\phi(\cdot)$ is convex?

Convexity and spd condition, cd

Consider again a general $A \in \mathbb{R}^{n\times n}$ with $A = A^T$.

Theorem I.10 (Another characterization of spd $A = A^T$ is spd.)

(c) *A is spd iff the function $\phi(x) = x^T Ax$ is strictly convex.*

Pbm. I.10 Determine the conditions that guarantee that $A \in \mathbb{R}^{2\times 2}$ is spd using the characterization (c).

Hint: there is a lot of algebra involved. Write out $\phi(x)$ and check the convexity condition (5).

Pbm. I.11 Recall multivariable calculus problem: checking for the extreme values of $\phi(x) = ax_1^2 + 2bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f$.
Write the equations to find the critical point.
The second derivative of ϕ is the Hessian $D^2\phi = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. What conditions guarantee that you have a minimum of ϕ at the critical point?
Compare with Pbm.I 6.

We will not be proving the general case $\mathbb{R}^{n\times n}$.

Back to the equilibrium model (1) and (2)

Now we want to know if the problem (1) is well-posed, i.e. that it has a solution, and that the solution is unique. We also want to known the connection between (1) and (2).

First, check that $K = A^T C A$ is spd, as long as C and A are full rank.

Pbm. I.12 Show that for any nonsingular diagonal matrix C, $A^T C A$ is symmetric.

Next show it is nonnegative definite.

Now assume that C and A in the equilibrium problem are full rank.

Conclude K is of full rank. Show directly that K is spd.

Calculations on previous slides suggest now that $\phi(x)$ defined by (2) is convex.

Thus we must find its critical point, the same as its minimizer.

Pbm. I.13 Follow the calculations to convince yourself that the critical point of $\phi(\cdot)$ is $x : Kx = f$ ((1) holds).

In the end, since K is spd, it is nonsingular, and (1) has a unique solution. We also find that (1) is equivalent to (2).

Discrete vs continuum models for our example

Let $n \to \infty$ (springs $\to$ rod)

Displacements x_i, $i = 1, \dots n$.
Fix $x_0 = 0, x_{n+1} = 0$
Elongation e_j
Difference $e_i = x_i - x_{i-1}$
Hooke's law $y_j = c_j e_j$
...
Model ((1))
$Kx = f$

Displacements $u(x)$, $x \in (0, l)$.
Boundary conditions $u(0) = 0, u(l) = 0$.
Strain e
Derivative $e = \frac{d}{dx} u$
Stress-strain $\sigma = ce$
...
Model (limit of (1)) we get

$$-\frac{d}{dx}\left(c(x)\frac{d}{dx}u(x)\right) = f \tag{6}$$

Example I.11 (Beyond 1 spatial dimension)

In more than one spatial dimension the model becomes

$$-\nabla \cdot (c\nabla u) = f, \; x \in \Omega \tag{7}$$

and requires boundary conditions on $\partial\Omega$.

Continuous problem cast as a minimization

In the discrete spring-mass model (or other such stationary model) recall that (1) defines the minimizer for (2).
What about the continuous model (6)?

Define $J(u) = \int_0^l \frac{1}{2}c(u')^2 - fu$. It is known that (6) defines the minimizer of $J(\cdot)$ over the set of sufficiently smooth functions satisfying $u(0) = u(l) = 0$. The equation (6) is the so-called Euler-Lagrange equation for the functional $J(\cdot)$.

To derive the Euler-Lagrange equations for some $J(\cdot)$, we proceed as follows. Calculate the first variation of $J(\cdot)$ $\frac{1}{t}(J(u+tv) - J(u))$, and take a limit (if possible) as $t \to 0$. Interpret this quantity $J'(u)(v)$ as the directional (Gateau) derivative of J at u in the direction of v.

Set $J'(u)(v) = 0$ for any v. Work (by integration by parts) to identify the equation that the minimizer u must satisfy.

Details in variational calculus: spaces and boundary conditions

Define space of functions $V = C_0^2 = C^2[0,1] \cap \{v : v(0) = 0 = v(1)\}$.

Example I.12 (Minimize $J(u) = \int_0^1 (u')^2 - fu$ over V)

Following the strategy outlined above, (recall $v \in V$)

$$J'(u)(v) = \int_0^1 u'v' - fv = 0.$$

Integrating by parts and using boundary conditions we obtain

$$\int_0 (-u'' - f)v = 0$$

Use FLCV (below)to conclude $-u'' = f$. Apply BC to solve for u.

Lemma I.13 (FLCV (Fundamental Lemma of Calculus of Variations))

If $\phi(\cdot) \in C^0[a,b]$ and $\int_a^b \phi(x)\psi(x)dx = 0$ for every $\psi \in C_0^2[a,b]$, then $\phi(x) = 0$, $\forall x \in [a,b]$.

Different BC

Example I.14 (Same functional, different V)

Consider $\int_0^1 (u')^2 - fu$ over $V = C^2[0,1] \cap \{v : v(0) = 0\}$.
Following calculations as above we derive ...

$$\int_0^1 (-u'' - f)v + u'(1)v(1) = 0 \tag{8}$$

Now test this equation with v which happens to satisfy $v(1) = 0$. We conclude $-u'' = f$. From this it follows from (8) that $u'(1)v(1) = 0$ regardless what the value of $v(1)$ is, i.e., $u'(1) = 0$ (this is the natural or Neumann condition).

Pbm. I.14 Follow the Euler-Lagrange strategy for ϕ defined in (2). You should get (1).

Pbm. I.15 Follow the Euler-Lagrange strategy for $J(u) = \int_0^1 \frac{1}{2}u^2 - u$ for $u \in C^2[0,1]$. What is the extremal?
Try also $J(u) = \int_0^1 \frac{1}{2}e^x (u')^2 - sin(x)u$.

Generalizations of Euler-Lagrange equations

The general form of Euler-Lagrange equations for the functional $J(u) = \int_a^b L(x, u(x), u'(x))dx$ over $u \in C^2[a, b], u(a) = A, u(b) = B$ is

$$L_u - \frac{d}{dx} L_{u'} = 0. \tag{9}$$

Its solution, the critical point of $J(\cdot)$ is called the stationary solution, since the equation provides the necessary but not always the sufficient condition for the minimum.
The difficulty most of the time is not deriving the EL equations, but rather solving them.

These equations can be generalized to $\mathbb{R}^d$, $d > 1$, to higher order derivatives, and multiple functions.

`Pbm. I.16 Classical examples include the brachistochrone problem, minimal surface problem, Fermat problem in geometrical optics, the Plateau problem, and systems of reaction-diffusion equations. Look these up and discuss (for extra credit).`

Convexity of functionals

Above you have gained some working experience with $J(u)$, for $u \in V = C^2[a,b]$.
We call $J(\cdot)$ a functional (i.e., a function of functions with values in $\mathbb{R}$).
Recall the definition of convexity (5) which we now restate for functionals and say that $J : V \to \mathbb{R}$ is convex if

$$J(tu + (1-t)v) < tJ(u) + (1-t)J(v), \;\; \forall t \in (0,1), \; \forall u, v \in V. \quad (10)$$

Pbm. I.17 Show that $J(u) = \int_0^1 u^2$ is convex. What about the Dirichlet functional $J(u) = \int_0^1 \frac{1}{2}(u')^2? - uf$?

One can prove that if a functional is convex (and if some additional conditions hold), then its extremal found by Euler-Lagrange equations is the minimizer of the functional.

Norms and inner products

Recall $x \cdot y = y^T x = \sum_i x_i y_i$ is the "dot product" or scalar product of two vectors $x, y \in \mathbb{R}^n$. This notion is extended to any vector space V, and denoted by (x, y), for $x, y \in V$.

An inner product on any space V (with scalars in $\mathbb{R}$) should be symmetric, bilinear, and positive definite. (A more general notion is defined on $\mathbb{C}^n$ and with scalars in $\mathbb{C}$.)

Euclidean norm of a vector $\| x \|_2 = \sqrt{x^T x} = \sqrt{(x, x)} = \sqrt{\sum x_i^2}$. (Sometimes we skip the subscript "2").

Pbm. II.1 Check if $(x, y)_{my} = \sum 2x_i y_i$ is an inner product. Also, for what $A \in \mathbb{R}^{2\times 2}$, is $(x, y)_{yours} = y^T A x$ an inner product?

(Answer to the second: A must be spd).

Example II.1 (Inner product space of functions)

For $f, g \in V = C[0, 1]$, define $(f, g) = \int_0^1 f(x) g(x) dx$.

Pbm. II.2 Check that the inner product for functions satisfies the desired properties.

Least squares (LSQ) and normal equations

Consider some data $(x_i, y_i)_{i=1}^n$, with $n \geq 2$. Assume that there is a linear trend $y_i \approx \theta x_i + b$. We try to identify this trend i.e., to find θ, b.

This is known as a "linear regression" problem, or "linear least squares". *In machine learning, b is called a "bias" and θ is called a feature weight, and (x_i, y_i) are known as training data.*

To find θ, b, we see that the system of equations $y_i = \theta x_i + b$ is overdetermined, thus there is no solution. We can, however, try to minimize $\| y - (\theta x + b) \|$.
If the norm is Euclidean, this is the LSQ problem, where we minimize $\| y - (\theta x + b) \|_2^2$, or try to find

$$\text{argmin} \quad \phi([\theta, b]) = \sum_i (y_i - \theta x_i - b)^2. \tag{11}$$

Pbm. II.3 Use calculus to find the minimizer of $\phi([\theta, b])$.

Definition II.2 (The LSQ solution to an overdetermined pbm $Au \approx f$)

is the solution $\hat{u}$ to the normal equations (well-posed if A is full rank)

$$A^T A \hat{u} = A^T f \tag{12}$$

Pbm. II.4 Set $u = [\theta, b]^T$ and find A for (11) to be framed as $Au \approx f$.

Pbm. II.5 Set $u = [\theta; b]$, write $\phi(u)$ as $u^T K u + \dots$ for some K, and work out by hand its minimizer. Is $\phi(u)$ convex ?

Examples of LSQ

Example II.3 (LSQ in MATLAB)

Given the following data `t=[0.5, 1.5, 2, 5]; b = [3.1 6.8 10.3 25];`, solve the least squares problem for $x = [a; c]$ fitting the data (t, b) to the linear model $b = at + c$.

(i) Setup the matrix A and right hand side f.

(ii) Solve for x: `x=A\ f;` Compare with the use of normal equations.

(iii) Plot the curve and mark the points.
`plot(t,b,'*',t,a*t+c,'-');legend('data','linear fit');`

MATLAB function **`polyfit`** can be used to achieve the same as what you did in A. Learn how, and compare to A. Try the quadratic model $b = at^2 + dt + c$ next.

Example II.4 (Nonlinear data fitting)

Use data from Ex. II.3 for the model $b = ce^{at}$ describing the concentration of microbes that are growing with rate a in time t, with initial condition c. Can you use linear LSQ ? What if you knew a and wanted c ? What if you knew c and wanted a ? How can you combine these two ?

Pbm. II.6 Provide details why this exponential model might be a good model. Can you think of an appropriate differential equation describing microbe growth ?

Pbm. II.7 Learn how to use lsqnonlin (MATLAB) and apply to this example.

Examples. Beyond normal equations

Instead of the linear (strictly speaking, affine) model $y = \theta x + b$, one can have a multi-linear model $y = \sum_{k=1}^{K} \theta_k X_k + b$, where X_k are different variables (features), and θ_k are feature weights. (Typically b is set as θ_0, with $X_0 = 1$.)

Example II.5 ("Quality of life" model)

One can predict "happiness" y depending on variable X_1 which is the GDP/capita in a country, and variable X_2 which is the mean annual temperature in that country.
See data in `http:/goo.gl/0Eht9W` and `http:/goo.gl/j1MSKe`

If the number of variables K is really large, solving normal equations directly is not advantageous. One uses instead the QR algorithm, or an iterative algorithm such as a gradient based method.

Once the linear model is found, ... we can use this model to predict the output y for some new inputs x (or X). This is called "interpolation" or "extrapolation".

`Ex. For example, we can predict the ''happiness'' for a country knowing its temperature and GDP/capita.`

Orthogonal projections and LSQ

Recall how you calculate a projection $\hat{v}$ of a vector v on the direction of another vector q of unit length. The length of the projection is $\|\hat{v}\| = (q, v) = v^T q$, and the projection vector $\hat{v} = q(v, q) = qq^T v$.
If q is not of unit length, we replace q by $\frac{q}{\|q\|}$ in the calculations. We obtain

$$\hat{v} = \frac{q}{\|q\|} \frac{q^T}{\|q\|} v = \ldots = q(q^T q)^{-1} q^T v. \tag{13}$$

Pbm. II.8 Work out the formula for the projection of vector v on the space spanned by several orthogonal vectors $q_1, q_2, \ldots$ forming matrix Q.
Answer: $\hat{v} = QQ^T v = \sum_j q_j (v, q_j)$.
Repeat for the case when q_j form Q of full rank, but which is not orthogonal. Answer:

$$\hat{v} = Q(Q^T Q)^{-1} Q^T v = \sum_j q_j \frac{1}{(q_j, q_j)} (v, q_j). \tag{14}$$

The coefficients (q_j, v) of expansion of $\hat{v}$ in the basis of columns of Q are sometimes called Fourier coefficients.

One other way to find the LSQ solution to $Au \approx f$ is to realize that this problem has no solution unless f is the column space of A. So the solution is to calculate a projection $\hat{f}$ of f onto $Col(A)$, and then solve $A\hat{u} = \hat{f}$. This gives the LSQ solution

$$\hat{u} = (A^T A)^{-1} A^T f \tag{15}$$