Lecture slides for

Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares

Developed by Stephen Boyd   Lieven Vandenberghe
Modified by John Duchi

8. Linear equations

Outline

Linear functions

Linear function models

Linear equations

Balancing chemical equations

Superposition

- $f : \mathbf{R}^n \to \mathbf{R}^m$ means f is a function that maps n-vectors to m-vectors
- we write $f(x) = (f_1(x), \ldots, f_m(x))$ to emphasize components of $f(x)$
- we write $f(x) = f(x_1, \ldots, x_n)$ to emphasize components of x
- f satisfies *superposition* if for all x, y, α, β

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

(this innocent looking equation says a lot ...)

- such an f is called *linear*

Matrix-vector product function

- with A an $m \times n$ matrix, define f as $f(x) = Ax$
- f is linear:

$$\begin{aligned} f(\alpha x + \beta y) &= A(\alpha x + \beta y) \\ &= A(\alpha x) + A(\beta y) \\ &= \alpha(Ax) + \beta(Ay) \\ &= \alpha f(x) + \beta f(y) \end{aligned}$$

- converse is true: if $f : \mathbf{R}^n \to \mathbf{R}^m$ is linear, then

$$\begin{aligned} f(x) &= f(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) \\ &= x_1 f(e_1) + x_2 f(e_2) + \cdots + x_n f(e_n) \\ &= Ax \end{aligned}$$

with $A = \begin{bmatrix} f(e_1) & f(e_2) & \cdots & f(e_n) \end{bmatrix}$

Examples

- reversal: $f(x) = (x_n, x_{n-1}, \ldots, x_1)$

$$A = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & ⋰ & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}$$

- running sum: $f(x) = (x_1,\ x_1 + x_2,\ x_1 + x_2 + x_3,\ \ldots,\ x_1 + x_2 + \cdots + x_n)$

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

Affine functions

- function $f : \mathbf{R}^n \to \mathbf{R}^m$ is *affine* if it is a linear function plus a constant, *i.e.*,

$$f(x) = Ax + b$$

- same as:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

  holds for all x, y, and α, β with $\alpha + \beta = 1$

- can recover A and b from f using

$$\begin{aligned} A &= \begin{bmatrix} f(e_1) - f(0) & f(e_2) - f(0) & \cdots & f(e_n) - f(0) \end{bmatrix} \\ b &= f(0) \end{aligned}$$

- affine functions sometimes (incorrectly) called linear

Outline

Linear functions

Linear function models

Linear equations

Balancing chemical equations

Linear and affine functions models

- in many applications, relations between n-vectors and m-vectors are *approximated* as linear or affine
- sometimes the approximation is excellent, and holds over large ranges of the variables (*e.g.*, electromagnetics)
- sometimes the approximation is reasonably good over smaller ranges (*e.g.*, aircraft dynamics)
- in other cases it is quite approximate, but still useful (*e.g.*, econometric models)

Price elasticity of demand

- n goods or services
- prices given by n-vector p, demand given as n-vector d
- $\delta_i^{\text{price}} = (p_i^{\text{new}} - p_i)/p_i$ is fractional changes in prices
- $\delta_i^{\text{dem}} = (d_i^{\text{new}} - d_i)/d_i$ is fractional change in demands
- *price-demand elasticity model*: $\delta^{\text{dem}} = E\delta^{\text{price}}$

- what do the following mean?

$$E_{11} = -0.3, \qquad E_{12} = +0.1, \qquad E_{23} = -0.05$$

Taylor series approximation

- suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable
- first order Taylor approximation $\hat{f}$ of f near z:

$$\begin{aligned}\hat{f}_i(x) &= f_i(z) + \frac{\partial f_i}{\partial x_1}(z)(x_1 - z_1) + \cdots + \frac{\partial f_i}{\partial x_n}(z)(x_n - z_n) \\ &= f_i(z) + \nabla f_i(z)^T(x - z)\end{aligned}$$

- in compact notation: $\hat{f}(x) = f(z) + Df(z)(x - z)$
- $Df(z)$ is the $m \times n$ *derivative* or *Jacobian* matrix of f at z

$$Df(z)_{ij} = \frac{\partial f_i}{\partial x_j}(z), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

- $\hat{f}(x)$ is a very good approximation of $f(x)$ for x near z
- $\hat{f}(x)$ is an affine function of x

Regression model

- regression model: $\hat{y} = x^T\beta + v$
  - x is n-vector of features or regressors
  - β is n-vector of model parameters; v is offset parameter
  - (scalar) $\hat{y}$ is our prediction of y
- now suppose we have N *examples* or *samples* $x^{(1)}, \ldots, x^{(N)}$, and associated responses $y^{(1)}, \ldots, y^{(N)}$
- associated predictions are $\hat{y}^{(i)} = (x^{(i)})^T\beta + v$
- write as $\hat{y}^{\mathrm{d}} = X^T\beta + v\mathbf{1}$
  - X is feature matrix with columns $x^{(1)}, \ldots, x^{(N)}$
  - y^{d} is N-vector of responses $(y^{(1)}, \ldots, y^{(N)})$
  - $\hat{y}^{\mathrm{d}}$ is N-vector of predictions $(\hat{y}^{(1)}, \ldots, \hat{y}^{(N)})$
- *prediction error* (vector) is $y^{\mathrm{d}} - \hat{y}^{\mathrm{d}} = y^{\mathrm{d}} - X^T\beta - v\mathbf{1}$

Outline

Linear functions

Linear function models

Linear equations

Balancing chemical equations

Systems of linear equations

- set (or *system*) of m linear equations in n variables $x_1, \ldots, x_n$:

$$\begin{array}{rcl} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n & = & b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n & = & b_2 \\ & \vdots & \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n & = & b_m \end{array}$$

- n-vector x is called the variable or unknowns
- A_{ij} are the *coefficients*; A is the coefficient matrix
- b is called the *right-hand side*
- can express very compactly as $Ax = b$

Systems of linear equations

- systems of linear equations classified as
  - under-determined if $m < n$ (A wide)
  - square if $m = n$ (A square)
  - over-determined if $m > n$ (A tall)
- x is called a *solution* if $Ax = b$
- depending on A and b, there can be
  - no solution
  - one solution
  - many solutions
- we'll see how to solve linear equations later

Outline

Linear functions

Linear function models

Linear equations

Balancing chemical equations

Chemical equations

- a chemical reaction involves p reactants, q products (molecules)
- expressed as

$$a_1 R_1 + \cdots + a_p R_p \longrightarrow b_1 P_1 + \cdots + b_q P_q$$

- $R_1, \ldots, R_p$ are reactants
- $P_1, \ldots, P_q$ are products
- $a_1, \ldots, a_p, b_1, \ldots, b_q$ are positive coefficients
- coefficients usually integers, but can be scaled
  - *e.g.*, multiplying all coefficients by $1/2$ doesn't change the reaction

Example: electrolysis of water

$$2H_2O \longrightarrow 2H_2 + O_2$$

- one reactant: water (H_2O)
- two products: hydrogen (H_2) and oxygen (O_2)
- reaction consumes 2 water molecules and produces 2 hydrogen molecules and 1 oxygen molecule

Balancing equations

- each molecule (reactant/product) contains specific numbers of (types of) atoms, given in its formula
  - *e.g.*, $\mathrm{H_2O}$ contains two H and one O
- *conservation of mass*: total number of each type of atom in a chemical equation must *balance*
- for each atom, total number on LHS must equal total on RHS
- *e.g.*, electrolysis reaction is balanced:
  - 4 units of H on LHS and RHS
  - 2 units of O on LHS and RHS
- finding (nonzero) coefficients to achieve balance is called *balancing* equations

Reactant and product matrices

- consider reaction with m types of atoms, p reactants, q products
- $m \times p$ *reactant matrix* R is defined by

$$R_{ij} = \text{number of atoms of type } i \text{ in reactant } R_j,$$

  for $i = 1, \ldots, m$ and $j = 1, \ldots, p$
- with $a = (a_1, \ldots, a_p)$ (vector of reactant coefficients)

$$Ra = \text{(vector of) total numbers of atoms of each type in reactants}$$

- define product $m \times q$ matrix P in similar way
- m-vector Pb is total numbers of atoms of each type in products
- conservation of mass is $Ra = Pb$

Balancing equations via linear equations

- conservation of mass is

$$\begin{bmatrix} R & -P \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

- simple solution is $a = 0$, $b = 0$
- to find a nonzero solution, set any coefficient (say, a_1) to be 1
- balancing chemical equations can be expressed as solving a set of $m + 1$ linear equations in $p + q$ variables

$$\begin{bmatrix} R & -P \\ e_1^T & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = e_{m+1}$$

(we ignore here that a_i and b_i should be nonnegative integers)

Conservation of charge

- can extend to include charge, *e.g.*, $\mathrm{Cr}_2\mathrm{O}_7^{2-}$ has charge -2
- *conservation of charge*: total charge on each side of reaction must balance
- we can simply treat charge as another type of atom to balance

Example

$$a_1 Cr_2O_7^{2-} + a_2 Fe^{2+} + a_3 H^+ \longrightarrow b_1 Cr^{3+} + b_2 Fe^{3+} + b_3 H_2O$$

- 5 atoms/charge: Cr, O, Fe, H, charge
- reactant and product matrix:

$$R = \begin{bmatrix} 2 & 0 & 0 \\ 7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 2 & 1 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 3 & 3 & 0 \end{bmatrix}$$

- balancing equations (including $a_1 = 1$ constraint)

$$\begin{bmatrix} 2 & 0 & 0 & -1 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -2 \\ -2 & 2 & 1 & -3 & -3 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Balancing equations example

- solving the system yields

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 14 \\ 2 \\ 6 \\ 7 \end{bmatrix}$$

- the balanced equation is

$$Cr_2O_7^{2-} + 6Fe^{2+} + 14H^+ \longrightarrow 2Cr^{3+} + 6Fe^{3+} + 7H_2O$$