Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 7

Linear Equations

One pint of good wine costs 50 gold pieces, while one pint of poor wine costs 10. Two pints of wine are bought for 30 gold pieces. How much of each kind of wine was bought?

Jiuzhang Suanshu, 200 B.C.

Outline

1. Introduction
   - History
   - Examples

2. Set of solutions of linear equations
   - Solution set
   - Use case in optimization
   - Overdetermined, underdetermined, and square systems

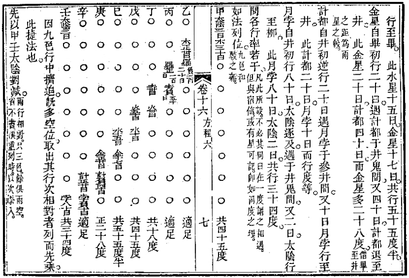
3. Solving linear systems
   - Solving via SVD
   - Solving via QR

Introduction

- Linear equations describe the most basic form of relationship among variables in an engineering problem.
- Systems of linear equations are ubiquitous in all branches of science: they appear for instance in elastic mechanical systems, relating forces to displacements, in resistive electrical networks, relating voltages to currents, in curve fitting, in many geometrical problems such as triangulation, trilateration, and localization from relative position measurements, in discrete-time dynamical systems relating input and output signals, etc.
- Linear equations form the core of linear algebra, and often arise as constraints in optimization problems.
- They are also an important building block of optimization methods, since many optimization algorithms rely on solution of a set of linear equations as a key step in the algorithm's iterations.

History

An early example



Linear equations have been around for thousands of years. The picture on the left shows a 17th century Chinese text that explains the ancient art of fangcheng (rectangular arrays).

Examples

An elementary 3×2 system

- The following is an example of a system of 3 equations in 2 unknowns:

$$\begin{aligned} x_1 + 4.5x_2 &= 1, \\ 2x_1 + 1.2x_2 &= -3.2, \\ -0.1x_1 + 8.2x_2 &= 1.5. \end{aligned}$$

- This system can be written in vector format as $Ax = y$, where A is a 3×2 matrix, and y is a 3-vector:

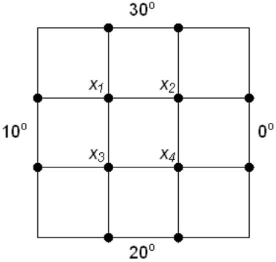
$$A = \begin{bmatrix} 1 & 4.5 \\ 2 & 1.2 \\ -0.1 & 8.2 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ -3.2 \\ 1.5 \end{bmatrix}.$$

- A solution to the linear equations is a vector $x \in \mathbb{R}^2$ that satisfies the equations.
- In the present example, it can be readily verified by hand calculations that the equations have no solution, i.e., the system is *infeasible*.

Examples

Temperature distribution

In some applications we wish to estimate the temperature inside an object (say, an engine) by sensing the temperature on its boundary.



In this 2D example we wish to estimate the temperature at the nodes of the grid, based on four measurements at the boundary.

Examples

Temperature distribution: mean-value property

We can use the (discretized) mean-value property, which states that the temperature at any inside node point is the average of the temperature of its neighbors. This leads to the linear system

$$
\begin{aligned}
x_1 &= \frac{1}{4}(10 + 30 + x_2 + x_3), \\
x_2 &= \frac{1}{4}(x_1 + 30 + 0 + 4), \\
x_3 &= \frac{1}{4}(10 + x_1 + x_4 + 20), \\
x_4 &= \frac{1}{4}(x_3 + x_2 + 0 + 20).
\end{aligned}
$$

Examples

Polynomial interpolation

- Consider the problem of interpolating a given set of points (x_i, y_i), $i = 1, \ldots, m$, with a polynomial of degree $n-1$

$$p(x) = a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

- The polynomial interpolates the i-th point if and only if $p(x_i) = y_i$, and each of such conditions is a linear equation on the polynomial coefficients a_j, $j = 0, \ldots, n-1$.
- An interpolating polynomial is hence found if the following system of linear equations in the a_j variables has a solution:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \cdots & \cdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},$$

where the matrix of coefficients on the left has a so-called Vandermonde structure.

Examples

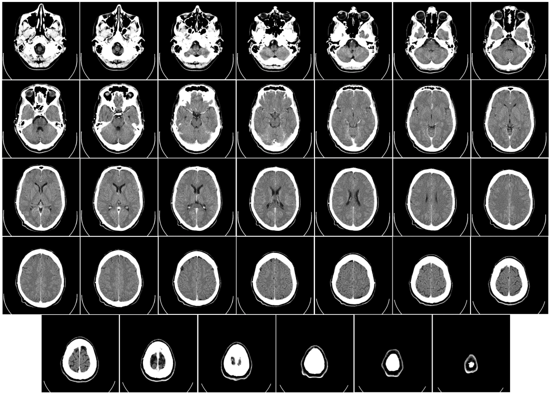
CAT scan imaging

- Tomography means reconstruction of an image from its sections. The word comes from the greek "tomos" (slice) and "graph" (description). The problem arises in many fields, ranging from astronomy to medical imaging.
- Computerized Axial Tomography (CAT) is a medical imaging method that processes large amounts of two-dimensional X-ray images in order to produce a three-dimensional image. The goal is to picture, for example, the tissue density of the different parts of the brain, in order to detect anomalies, such as brain tumors.
- Typically, the X-ray images represent "slices" of the part of the body that is examined. Those slices are indirectly obtained via axial measurements of X-ray attenuation
- In CAT for medical imaging, one uses axial (line) measurements to get two-dimensional images (slices), and from those slices one may proceed to digitally reconstruct a three-dimensional view. Here, we focus on the process that produces a single two-dimensional image from axial measurements.

Examples

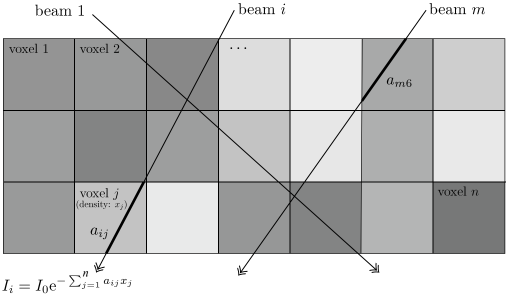
CAT scan imaging

CAT scan slices of a human brain (Source: Wikipedia).



Example: CAT scan imaging

- In CAT-based medical imaging, a number of X rays are sent through the tissues to be examined along different directions, and their intensity after they have traversed the tissues is captured by a receiver sensor.
- For each direction, we record the attenuation of the X ray, by comparing the intensity of the X ray at the source to the intensity after the X ray has traversed the tissues, at the receiver's end.
- To a reasonable degree of approximation, the log-ratio of the intensities at the source and at the receiver is linear in the densities of the tissues traversed.



Examples

CAT scan imaging

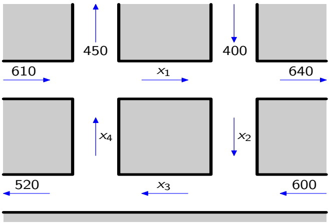
- A (typically large) number m of beams of intensity I_0 at the source travel across the tissue: the i-th beam, $i = 1, \ldots, m$, has a path of length a_{ij} through voxel j, $j = 1, \ldots, n$.
- The log-attenuation of the i-th beam intensity due to the j-th voxel is proportional to the density of the voxel x_j times the length of the path, that is $a_{ij} x_j$.
- The total log-attenuation for beam i is therefore given by the sum of the log-attenuations:

$$y_i = \log \frac{I_0}{I_i} = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m,$$

  where I_i is the intensity of the i-th beam at the receiver end.
- Recovering the densities x_j from the y_i measurements thus amounts to solving a system of linear equations of the form $y = Ax$, where $A \in \mathbb{R}^{m,n}$. A is typically a “fat” matrix, i.e., the system is underdetermined.

Examples

Traffice flows



$$x_1 = ?, \quad x_2 = ?, \quad x_3 = ?, \quad x_4 = ?$$

A basic traffic flow estimation problem involves inferring the amount of cars going through links based on information on the amount of cars passing through neighboring links.

Examples

Traffic example: flow equations

At each intersection, the incoming traffic has to match the outgoing traffic:

$$
\begin{array}{lrcl}
\text{Intersection A:} & x_4 + 610 & = & x_1 + 450, \\
\text{Intersection B:} & x_1 + 400 & = & x_2 + 640, \\
\text{Intersection C:} & x_2 + 600 & = & x_3, \\
\text{Intersection D:} & x_3 & = & x_4 + 520.
\end{array}
$$

We can write this in matrix format: $Ax = y$, with

$$
A = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad y = \begin{pmatrix} -160 \\ 240 \\ -600 \\ 520 \end{pmatrix}.
$$

The matrix A is nothing else than the incidence matrix associated with the graph that has the intersections as nodes and links as edges.

Examples

Balance equations in chemistry [1]

Chemical reaction with m types of atoms / charges, p reactants, q products:

$$a_1 R_1 + \ldots + a_p R_p \to b_1 P_1 + \ldots + b_q P_q$$

Reactant matrix: for $i = 1, \ldots, m$, $j = 1, \ldots, p$

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

Define $m \times q$ **product matrix** P similarly.

With $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$ (vectors of reactant and product coefficients)

- Ra = (vector of) total numbers of atoms of each type in reactants;
- Pb is total numbers of atoms of each type in products;
- conservation of mass is Ra = Pb.

[1] From Boyd & Vandenberghe, `http://web.stanford.edu/~boyd/vmls/vmls-slides.pdf`

Examples

Balance equations in chemistry

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

5 equations corresponding to each atom Cr, O, Fe, H, or charge.

$$R = \begin{pmatrix} 2 & 0 & 0 \\ 7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 2 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 3 & 3 & 0 \end{pmatrix}.$$

Solving $Ra = Pb$, with $a_1 = 1$ (WLOG, since we can always scale a, b), yields

$$a = (1, 6, 14), \quad b = (2, 6, 7).$$

Example

Linear dynamical systems

$$x_{t+1} = A_t x_t, \quad t = 0, 1, 2, 3, \dots$$

where $A_t \in \mathbb{R}^{n \times n}$ matrices

- system is called time-invariant if $A_t = A$ doesnt depend on time;
- from a know initial conditions x_0, we can simulate evolution;
- model can be extended to include inputs and offset

$$x_{t+1} = A_t x_t + B_t u_t + c_t, \quad t = 0, 1, 2, 3, \dots$$

- model can be extended to auto-regressive model

$$x_{t+1} = A_t x_t + A_{t-1} x_{t-1} + \dots + A_{t-m} x_{t-m}, \quad t = m, m+1, \dots$$

Example

Steady-state solution in linear dynamical systems

Consider a linear system with constant input:

$$x_{t+1} = Ax_t + b,$$

with $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$. (Trajectory $(x_t)_{t \geq 0}$ is sometimes called "step response".)

Steady-state solution: limit of x_t as $t \to +\infty$, satisfies

$$(I - A)x_\infty = b$$

Population dynamics

Model population distribution:

- $x_t \in \mathbb{R}^{100}$ gives population distribution in year t
- $x_t(i)$ is the number of people with age $i-1$ in year t (say, on January 1)
- total population in year t is $\mathbf{1}^\top x_t$
- number of people age 70 or older in year t is $(0_{70}; \mathbf{1}_{30})^\top x_t$

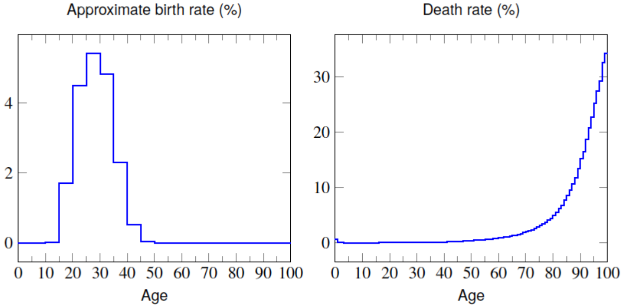
Dynamics parameters:

- birth rate $b \in \mathbb{R}^{100}$, death (or mortality) rate $d \in \mathbb{R}^{100}$;
- b_i (resp. d_i) is the number of births (resp. deaths) per person with age $i-1$;
- b and d can vary with time, but well assume they are constant.

Dynamical equations

$x_t = Ax_t$, where

$$A = \begin{pmatrix} b_1 & b_2 & \dots & b_{99} & b_{100} \\ 1-d_1 & 0 & \dots & 0 & 0 \\ 0 & 1-d_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-d_{99} & 0 \end{pmatrix}$$



Set of solutions of linear equations

- Generic linear equations can be expressed in vector format as

$$Ax = y,$$

where $x \in \mathbb{R}^n$ is the vector of unknowns, $y \in \mathbb{R}^m$ is a given vector, and $A \in \mathbb{R}^{m,n}$ is a matrix containing the coefficients of the linear equations.

- Key issues are: existence, uniqueness of solutions; characterization of the *solution set*:

$$S \doteq \{x \in \mathbb{R}^n : \ Ax = y\}.$$

- Let $a_1, \ldots, a_n \in \mathbb{R}^m$ denote the columns of A, i.e. $A = [a_1 \ \cdots \ a_n]$, and notice that the product Ax is nothing but a linear combination of the columns of A, with coefficients given by x:

$$Ax = x_1 a_1 + \cdots + x_n a_n.$$

- Ax always lies in $\mathcal{R}(A)$.
- Thus, $S \neq \emptyset \Leftrightarrow y \in \mathcal{R}(A)$.

Set of solutions of linear equations

- The linear equation
$$Ax = y, \quad A \in \mathbb{R}^{m,n}$$
admits a solution if and only if $\mathrm{rank}([A\ y]) = \mathrm{rank}(A)$.
- When this existence condition is satisfied, the set of all solutions is the affine set
$$S = \{x = \bar{x} + Nz\},$$
where $\bar{x}$ is any vector such that $A\bar{x} = y$, and $N \in \mathbb{R}^{n,n-r}$ is a matrix whose columns span the nullspace of A (hence $AN = 0$).
- In particular, the system has a unique solution if $\mathrm{rank}([A\ y]) = \mathrm{rank}(A)$ and $\mathcal{N}(A) = \{0\}$.

Use case in optimization

Finding the solution set (or determine the set is empty) is ometimes useful in optimization. Consider an optimization problem with linear equality constraints:

$$\min_x \; f_0(x) \;:\; Ax = b,$$

with $f_0 : \mathbb{R}^n \to \mathbb{R}$, and $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$ the variable. We assume that the problem is feasible, that is, the solution set of $Ax = y$ is not empty.

Since the solution set is affine, any solution is of the form $x_0 + Nz$, with x_0 a particular solution and N a matrix whose columns span the nullspace of A.

We can formulate the above problem as an *unconstrained* one:

$$\min_z \; f_0(x_0 + Nz).$$

Overdetermined, underdetermined, and square systems

Overdetermined systems

- The system $Ax = y$ is said to be *overdetermined* when it has more equations than unknowns, i.e., when matrix A has more rows then columns ("skinny" matrix): $m > n$.
- Assume that A is full column rank, that is $\mathrm{rank}(A) = n$. Then, $\dim \mathcal{N}(A) = 0$, hence the system has either one or no solution at all.
- Indeed, the most common case for overdetermined systems is that $y \notin \mathcal{R}(A)$, so that no solution exists.
- In this case, it is often useful to introduce a notion of approximate solution, that is a solution that renders minimal some suitable measure of the mismatch between Ax and y (more on this later!)

Overdetermined, underdetermined, and square systems

Underdetermined systems

- The system $Ax = y$ is said to be *underdetermined* if it has more unknowns than equations, i.e., when matrix A has more columns than rows ("wide" matrix): $n > m$.
- Assume that A is full row rank, that is $\mathrm{rank}(A) = m$, and then $\mathcal{R}(A) = \mathbb{R}^m$, thus $\dim \mathcal{N}(A) = n - m > 0$.
- The system of linear equations is therefore solvable with infinite possible solutions, and the set of solutions has "dimension" $n - m$.
- Among all possible solutions, it is often of interest to single out one specific solution having minimum norm (more on this later!)

Overdetermined, underdetermined, and square systems

Square systems

- The system $Ax = y$ is said to be *square* when the number of equations is equal to the number of unknowns, i.e. when matrix A is square: $m = n$.
- If a square matrix is full rank, then it is invertible, and the inverse A^{-1} is unique and has the property that $A^{-1}A = I$.
- In the case of square full rank A the solution of the linear system is thus unique and it is formally written as

$$x = A^{-1}y.$$

Solving via SVD

"Rotate inputs and outputs to make system diagonal"

The linear equation $Ax = y$, where $A \in \mathbb{R}^{m,n}$ and $y \in \mathbb{R}^m$, can be fully analyzed via SVD. If $A = U\tilde{\Sigma}V^\top$ is the SVD of A (as in Thm. 1 of lecture 6), then $Ax = y$ is equivalent to

$$\tilde{\Sigma}\tilde{x} = \tilde{y}, \tag{1}$$

where $\tilde{x} \doteq V^\top x$, $\tilde{y} \doteq U^\top y$.

Since $\tilde{\Sigma}$ is a diagonal matrix:

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}, \quad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \succ 0.$$

the system (1) is very easy to solve/analyze.

Solution via SVD

The system $\tilde{\Sigma}\tilde{x} = \tilde{y}$ writes

$$\sigma_i \tilde{x}_i = \tilde{y}_i, \;\; i = 1, \ldots, r, \;\; 0 = \tilde{y}_i, \;\; i = r+1, \ldots, m.$$

Two cases can occur:

- If the last $m - r$ components of $\tilde{y}$ are not zero, then the above system is infeasible, and the solution set is empty. This occurs when y is not in the range of A.
- If $y \in \mathcal{R}(A)$, then the last set of conditions in the above system hold, and we can solve for $\tilde{x}$ with the first set of conditions: $\tilde{x}_i = \tilde{y}_i / \sigma_i$,$i = 1, \ldots, r$. The last $n - r$ components of $\tilde{x}$ are free, which corresponds to elements in the nullspace of A.
- If A is full column rank (its nullspace is reduced to $\{0\}$), then there is a unique solution.

Example

The system $Ax = y$, with

$$A = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

has no solution if $(y_3, y_4) \neq 0$. If $y_3 = y_4 = 0$, so that $y \in \mathcal{R}(A)$, we get the solution set $S = \{x_0 + Nz \ : \ z \in \mathbb{R}^2\}$, with

$$x_0 = \begin{pmatrix} \frac{y_1}{\sigma_1} \\ \frac{y_2}{\sigma_2} \\ 0 \\ 0 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Using QR

"Rotate outputs to make system triangular"

We may also use the QR decomposition to solve or analyze a system. This time, the idea is to reduce the system to a triangular one, and then solve by "backward" substitution (that is, solve one variable at a time, starting from the last one).

Assuming that A is square to simplify, and $A = QR$ with $Q^\top Q = I_n$, then $Ax = y$ is equivalent to $Rx = \tilde{y} \doteq Q^\top y$, which is a triangular system. Solving for x_n first, and substituting to solve for x_{n-1}, etc, leads to the solution.