Math Review

Jesse Perla
University of British Columbia

January 7, 2019

1 Linear Algebra

1.1 Vectors

- Vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$: column vector
- Transposing the vector: $x^T = x' = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$: row vector
- Notation for a vector of reals: $x \in \mathbb{R}^n$

1.2 Matrices

- $A = \begin{bmatrix} A_{11}, & \dots & A_{1m} \\ A_{21}, & \dots & A_{2m} \\ A_{n1}, & \dots & A_{nm} \end{bmatrix}$: $n \times m$ matrix
  - Sometimes write $[A_{ij}], i = 1, \dots, n, j = 1, \dots, m$.
- Matrix Transpose:
  - Definition: The transpose of an n-by-m matrix A is the m-by-n matrix, denoted by A^T or A'
  - A^T is formed by turning rows into columns and vice versa: $\left(A^T\right)_{i,j} = A_{j,i}$
  - For example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -6 & 7 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 2 & -6 \\ 3 & 7 \end{bmatrix} \tag{1}$$

- Note: $(A^T)^T = A$

Matrix addition/subtraction

- Definition: If A and B are two $n \times m$ matrices, then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots a_{2m} + b_{2m} \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots a_{nm} + b_{nm} \end{bmatrix} \tag{2}$$

$$A - B = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \cdots a_{1m} - b_{1m} \\ a_{21} - b_{21} & a_{22} - b_{22} & \cdots a_{2m} - b_{2m} \\ a_{n1} - b_{n1} & a_{n2} - b_{n2} & \cdots a_{nm} - b_{nm} \end{bmatrix} \tag{3}$$

- Example 1:

$$\begin{bmatrix} 1 & 3 & 2 \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 5 \\ 7 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 5 \\ 7 & 5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 3 & 2 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+0 & 3+0 & 2+5 \\ 1+7 & 0+5 & 1+1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 7 \\ 8 & 5 & 2 \end{bmatrix} \tag{4}$$

  Example 2:

$$\begin{bmatrix} 1 & 6 \\ 9 & 3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 6 \\ 9 & 2 \end{bmatrix} \tag{5}$$

- Properties:
  - Commutativity of Addition : $A + B = B + A$
  - Associativity : $(A + B) + C = A + (B + C) = A + B + C$
  - $(A + B)^T = A^T + B^T$

Matrix Multiplication (not commutative!)

- The key: do algebra being consistent with order on the right vs. left side of each equation. Only reorder where commutatively holds (e.g. addition, scalar multiplication...)
- For product $\underbrace{C}_{n\times p} = \underbrace{A}_{n\times m} \cdot \underbrace{B}_{m\times p}$: the inner sizes "$m$" need to match, where each element $C_{ik} = \sum_{j=1}^{m} A_{ij} B_{jk}$

- Example 1:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} c_{11} = a_{11}b_{11} + a_{12}b_{21} & c_{12} = a_{11}b_{12} + a_{12}b_{22} \\ c_{21} = a_{21}b_{11} + a_{22}b_{21} & c_{22} = a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} \tag{6}$$

  Example 2:

$$\begin{bmatrix} \underline{\mathbf{2}} & \underline{\mathbf{3}} & \underline{\mathbf{4}} \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & \underline{\mathbf{1000}} \\ 1 & \underline{\mathbf{100}} \\ 0 & \underline{\mathbf{10}} \end{bmatrix} = \begin{bmatrix} 3 & \underline{\mathbf{2340}} \\ 0 & 1000 \end{bmatrix} \tag{7}$$

- A is $n \times m$ matrix, B is $m \times p$ matrix and C is $p \times n$ matrix, α, β are scalar, then matrices have properties:

  - $(AB)^T = B^T A^T$
  - Associativity: $(AB)\,C = A\,(BC)$; $(\alpha\beta)\,A = \alpha\,(\beta A)\,, \alpha\,(AB) = (\alpha A)\,B$
  - Distributivity: $(A+B)\,C = AC + BC$ and $C\,(A+B) = CA + CB$; $(\alpha+\beta)\,A = \alpha A + \beta A$
  - Commutativity <u>holds</u> for scalar multiplication: $\alpha \cdot A = A \cdot \alpha$

$$2 \cdot \begin{bmatrix} 1 & 8 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 8 \\ 4 & -2 \end{bmatrix} \cdot 2 = \begin{bmatrix} 1 \cdot 2 & 8 \cdot 2 \\ 4 \cdot 2 & -2 \cdot 2 \end{bmatrix} = \begin{bmatrix} 2 & 16 \\ 8 & -4 \end{bmatrix} \tag{8}$$

  - Commutativity <u>does not hold</u> for matrix multiplication: $AB \neq BA$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix} \tag{9}$$

    whereas

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 0 \end{bmatrix} \tag{10}$$

    Note: number of columns of A must equal number of rows of B

- Matrix-vector multiplication in terms of dot-products:

  - Note that matrix multiplication can be written in terms of dot products

– Example stacking dot-products:

$$\begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix}\cdot\begin{bmatrix}3\\-1\end{bmatrix}=\begin{bmatrix}\begin{bmatrix}1 & 2\end{bmatrix}\cdot\begin{bmatrix}3 & -1\end{bmatrix}\\\begin{bmatrix}3 & 4\end{bmatrix}\cdot\begin{bmatrix}3 & -1\end{bmatrix}\\\begin{bmatrix}5 & 6\end{bmatrix}\cdot\begin{bmatrix}3 & -1\end{bmatrix}\end{bmatrix}=\begin{bmatrix}1\\5\\9\end{bmatrix}$$

– We will sometimes write the product of a matrix and a vector as $A \cdot x$, using the dot product

  * $x \cdot B = \left(B^T \cdot x^T\right)^T$. Being a little sloppy with row vs. column major for vectors, this means that $x \cdot B = B^T \cdot x$

Matrix Inverse

- Define Identity Matrix:

$$I=\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix} \tag{11}$$

  – Only "1" s down the diagonal

  – Square of size $n \times n$

  – $I \cdot A = A \cdot I = A$ for all $n \times n,\ A$ and $I.$ i.e. caqn commute

- If A is square, and a square F satisfies $FA = I$, then:

  – F is called the inverse of A and is denoted A^{-1}

  – The matrix A is called invertible or nonsingular

  – $A \cdot A^{-1} = I$, where all are $n \times n.$ Prove it!

  – I won't ask you to calculate non-trivial inverses.

  – Trivial ones such as:

$$I=\begin{bmatrix}a & 0\\0 & b\end{bmatrix}\begin{bmatrix}a^{-1} & 0\\0 & b^{-1}\end{bmatrix}=\begin{bmatrix}a^{-1} & 0\\0 & b^{-1}\end{bmatrix}\begin{bmatrix}a & 0\\0 & b\end{bmatrix} \tag{12}$$

- Unlike for scalars, A/B or $\frac{A}{B}$ is meaningless. i.e. is it $A^{-1}B$ or BA^{-1}?

Systems of Equations and Matrix Equations

- $\underbrace{A}_{n\times n}\cdot\underbrace{x}_{n\times 1}=\underbrace{b}_{n\times 1}$, system of n equations

- Left multiply both sides by A^{-1} (making sure to maintain order!): $A^{-1} \cdot A \cdot x = A^{-1} \cdot b \Rightarrow$ $\boxed{x = A^{-1} \cdot b}$

- Example of solving $Ax = b$ by using the inverse of A:

$$\begin{cases} 3x_1 + 4x_2 = 3 \\ 5x_1 + 6x_2 = 7 \end{cases} \Rightarrow \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 3 \\ 7 \end{bmatrix} \tag{13}$$

- Example of selecting from a vector: Given $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ how would I extract the second element (useful for *observations* of a vector of data)? Use a vector with a single 1 and 0 otherwise,

$$x_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{14}$$

2 Functional equation

2.1 Differences between functional equation and equations

- Equations usually define the relationship between one or more variables. We can solve the equation and find out the values of the variables which fulfill the equation (it may not be unique).
  - Abstractly, take some $f : \mathbb{R} \to \mathbb{R}$ and let the equation $f(x) = 1$. Then the solution to the equation is the set of $x^* \in \mathbb{R}$ such that $f(x^*) = 1$. For example, if $f(x) = x^2$ then $x^* = 1$.
  - Single variable equation: $x^2 - 5x = 0$. i.e. solution is a single x, unique in that case.
  - Multi-variable equation: $2x + 7y = 3$ i.e. solution may be an x, y if system of equations, or set of x and y is under-determined.

- Functional equations provide an expression where we are solving for an entire function, not just a single value. As before, there may a unique function which solves the functional equation, or a set of functions.
  - For example, take the functional equation $[f(x)]^2 - x^2 = 0$. The goal is not to find an x^* but rather a function $f(x)$ which holds for all x.
  - In this simple example, note that the function defined by $f(x) = x$ fulfills this equation for all $x \in \mathbb{R}$. As does $f(x) = -x$.

2.2 Undetermined coefficients in functional equations

- Example 1: Given $\partial f(z) = z$. Guess that $f(z) = C_1 z^2 + C_2$ solves this equation. Then:

  $$\partial f(z) = 2C_1 z = z \Rightarrow C_1 = \frac{1}{2} \tag{15}$$

  So $C_1 = \frac{1}{2}$ and C_2 is indeterminate.

- Example 2: Now do it with a difference equation. Let the difference equation be:

  $$z_{t+1} = z_t + 1 \tag{16}$$

  Guess the solution is of the form $z_t = C_1 t + C_2$. So we have:

  $$z_{t+1} = z_t + 1 \Rightarrow C_1(t+1) + C_2 = C_1 t + C_2 \Rightarrow C_1 = 1 \tag{17}$$

  So we have $C_1 = 1$ and C_2 is indeterminate. What if we add that $z_0 = 1$? Show that this pins down C_2

3 Review of Optimization

3.1 Unconstrained Optimization

$$\max_x \; f(x) \tag{18}$$

First order necessary condition

$$\boldsymbol{\partial} f(x) = \underbrace{\boxed{f'(x) = 0}}_{\text{if convex, etc.}} \tag{19}$$

where $\boldsymbol{\partial} f(x) = \frac{d}{dx} f(x)$: derivative of $f(x)$ with respect to x.

3.2 Constrained Optimization

Can always convert to this canonical form (i.e., order of inequalities, max vs. min, etc.) then apply formulas,

$$\max_x \; f(x) \tag{20}$$

$$\text{s.t.} \;\; g(x) \geq 0 \leftarrow \text{ may or may not bind} \tag{21}$$

$$h(x) = 0 \leftarrow \text{ always binds} \tag{22}$$

Solution Method: Form a Lagrangian:

$$\mathcal{L} = f(x) + \mu g(x) + \lambda h(x) \tag{23}$$

where μ and λ are called Lagrange multipliers.

First-order Necessary Conditions:

$$\boldsymbol{\partial} \mathcal{L}(x) = 0 \tag{24}$$

$$\boldsymbol{\partial} f(x) + \mu \boldsymbol{\partial} g(x) + \lambda \boldsymbol{\partial} h(x) = 0 \tag{25}$$

$$g(x) \geq 0, \;\; h(x) = 0 \tag{26}$$

$$\mu \geq 0 \tag{27}$$

$$\mu \cdot g(x) = 0 \text{ i.e., } \quad \underbrace{\mu = 0}_{\substack{\text{constraint} \\ \text{doesn't bind}}} \quad \text{or } g(x) = 0 \tag{28}$$

Any $\{x, \mu, \lambda\}$ that fulfils these conditions solves this problem.

Example:

$$\max \;\; -\frac{1}{2}(x+1)^2 \tag{29}$$

$$\text{s.t.}\, x \geq 0 \tag{30}$$

The Lagrange setup is:

$$\mathcal{L} = -\frac{1}{2}(x+1)^2 + \mu x \tag{31}$$

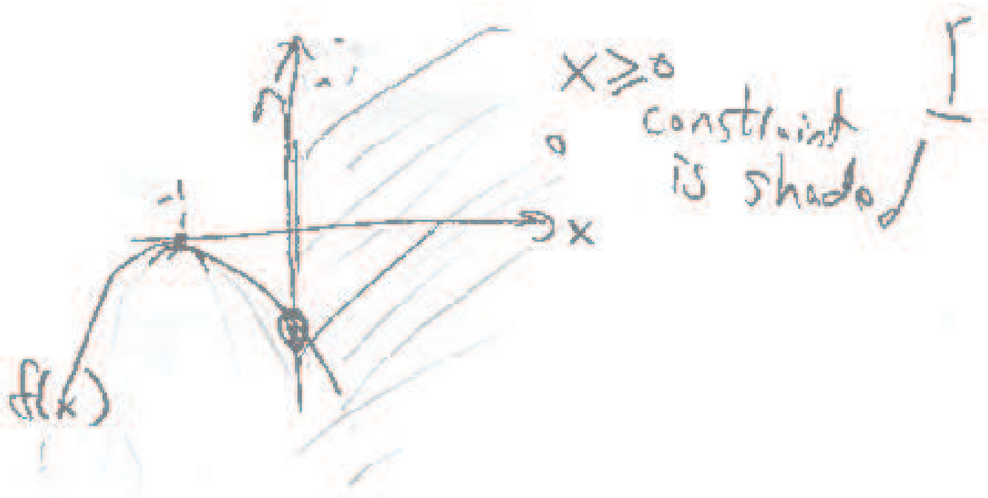

Figure 1: Graphical representation of the example problem

FONC:

$$[x] : -(x+1) + \mu = 0,\ \ \mu \geq 0 \tag{32}$$

or $(x+1) = \mu$ and $\mu x = 0$, (either $x = 0$ or $\mu = 0$)

$$\Rightarrow x+1 > 0 \text{ if } \mu > 0 \text{ and } x = 0 \tag{33}$$
$$x+1 = 0 \text{ if } \mu = 0 \tag{34}$$
$$x \geq 0 \tag{35}$$

Solution:

$$\text{If } \mu = 0 \Rightarrow x = -1,\ \text{contradicting the constraint } x \geq 0 \tag{36}$$
$$\Rightarrow\ \ \mu > 0 \text{ and } x = 0 \tag{37}$$
$$\Rightarrow \mu = (1+0) = 1 \tag{38}$$

Shorthand for this problem (typical of linear constraints):

$$-(x+1) \leq 0, = 0 \text{ if } x > 0 \tag{39}$$

Another Example:

$$\max\ \ f(x) \tag{40}$$
$$\text{s.t.}\, x \leq m \tag{41}$$

Lagrangian:
First, reorder constraint as:

$$x - m \leq 0 \Rightarrow m - x \geq 0 \text{ (multiplying inequality by -1 switches directions)}$$

Then:

$$\mathcal{L} = f(x) + \lambda(m - x) \tag{42}$$

FONC:

$$[x] : f'(x) - \lambda = 0 \tag{43}$$
$$\lambda(m - x) = 0, \ \ \lambda \geq 0 \tag{44}$$

The *Kuhn-Tucker* conditions (`https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker`
are:

$$\text{i.e., if } \lambda > 0 \Rightarrow m - x = 0 \Rightarrow x = m \qquad \text{(binding)} \tag{45}$$
$$\text{i.e., if } \lambda = 0 \Rightarrow f'(x) = 0 \qquad \text{(nonbinding)} \tag{46}$$

Linear = Corners (if at all): Consider for some $a \in \mathbb{R}$,

$$\min \ ax \tag{47}$$
$$\text{s.t.}\ x \geq 1 \tag{48}$$

Draw and see the cases:

- $a > 0$: max is at $x = 1$
- $a = 0$: max is indeterminate
- $a < 0$: max is at $x = \infty$. i.e., doesn't exist

More formally, first convert to a max problem (i.e. multiply objective by -1), then form the Lagrangian,

$$\mathcal{L} = -ax + \mu(x - 1) \tag{49}$$

FONC gives a system of equalities and inequalities,

$$-a + \mu = 0 \tag{50}$$

$$(x-1) \geq 0 \tag{51}$$

$$\mu \geq 0 \tag{52}$$

$$\mu(x-1) = 0 \tag{53}$$

Rearrange (50),

$$\mu = a \tag{54}$$

Cases:

- If $a < 0$, then (52) is contradicted for any x. i.e., no solution to the system of equations.
- If $a = 0$, then $\mu = 0$. Checking the system: (53) holds for any x (notation, $\forall\, x$). (52) is fulfilled. Finally, (51) is fulfilled for any $x \geq 1$. To summarize, the solution holds: $\mu = 0$ and $\forall\, x \geq 1$
- If $a > 0$, then $\mu = a > 0$. From (53), since $\mu > 0$, it is necessary for $x = 1$. Checking the solution: (51) to (53) all fulfilled.

4 Probability

4.1 Discrete Random Variable

- A **random variable** is a number whose value depends upon the outcome of a random experiment. Mathematically, a random variable X is a real-valued function on S, the space of outcomes (which can be a very abstract set):

$$X : S \to \mathbb{R} \tag{55}$$

- A **discrete random variable** X has finite or countably many values x_s for $s = 1, 2, \cdots$
- The probabilities $\mathbb{P}\left(X = x_i\right)$ with $s = 1, 2, \cdots$ are called the **probability mass function** of X, which has the following properties:
  - For all s, $\mathbb{P}\left(X = x_s\right) \geq 0$
  - For any $B \subseteq S$, $\mathbb{P}\left(X \in B\right) = \sum\limits_{x_s \in B} \mathbb{P}\left(X = x_s\right)$
  - $\sum\limits_{s} \mathbb{P}\left(X = x_s\right) = 1$

- Assume that X is a discrete random variable with possible values x_s. Then, the **expectation** of X is defined as:

$$\mathbb{E}[X] = \sum_s x_s \mathbb{P}(X = x_s) \tag{56}$$

4.2 Expectations and Vectors

Assume that there are n states, i.e. $x_1, \ldots x_n$. List of values for states of the world:

$$x \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Now, list out all of the probabilities in a vector, $\phi \in \mathbb{R}^n$

$$\phi \equiv \begin{bmatrix} \mathbb{P}(X = x_1) \\ \mathbb{P}(X = x_2) \\ \vdots \\ \mathbb{P}(X = x_n) \end{bmatrix}$$

Then the expectation can be written as a vector-vector dot product:

$$\mathbb{E}[X] = \sum_{i=1}^{I} \phi_i x_i = \phi \cdot x$$

As an example, assume the probability of unemployment is $\phi_1 = 0.1$, income from unemployment insurance is $x_1 = \$15,000$; probability of employment is $\phi_2 = 0.9$, income from employment is $x_2 = \$40,000$. Then expected income (or average across states of world):

$$\mathbb{E}[X] = (0.1 \times 15,000) + (0.9 \times 40,000)$$

More generally, The value $\mathbb{E}[X]$ is the mean of the random variable X, over whatever probability distribution you are interested in. This generalizes to other functions of a random variable. e.g. $\mathbb{E}[X^2] = \phi \cdot x^2$.

4.3 Joint Distributions

For discrete random variable, consider if there are multiple events yielding random variables X and Y (e.g. $i \in \{L, H\}$ and $j \in A, B$). The **joint probability distribution** is the probability that both events occur,

$$\mathbb{P}\left(X = x_i \text{ and } Y = y_j\right)$$

such that $\sum_i \sum_j \mathbb{P}\left(X = x_i \text{ and } Y = y_j\right) = 1$.

- The **marginal probability** is the distribution of one random variable if we ignore the other one. For example, the probability that x_i occurs (regardless of the y_j outcome) just sums over the probabilities in the joint distribution with y_j,

$$\mathbb{P}\left(X = x_i\right) = \sum_j \mathbb{P}\left(X = x_i \text{ and } Y = y_j\right)$$

- The **conditional probability** is the distribution of one random variable if we know the other has occurred. For example, if we know $Y = y_j$ then the probability that x_i occurs is written as (and reorganized as),

$$\mathbb{P}\left(X = x_i \,\middle|\, Y = y_j\right) = \frac{\mathbb{P}\left(X = x_i \text{ and } Y = y_j\right)}{\mathbb{P}\left(Y = y_j\right)} \quad = \frac{\mathbb{P}\left(X = x_i \text{ and } Y = y_j\right)}{\sum_k \mathbb{P}\left(X = x_k \text{ and } Y = y_j\right)} \tag{57}$$

- A **conditional expectation** is when one of the multiple events is known (e.g. which Y occurred), and finds the expectation over the other event. It is denoted, $\mathbb{E}\left[X \mid Y\right]$.
  - For example in the above if we know that $Y = y_j$,

$$\mathbb{E}\left[X \mid Y = y_j\right] = \sum_i x_i \mathbb{P}\left(X = x_i \,\middle|\, Y = y_j\right) \tag{58}$$

  - This will be especially useful for agents making forecasts of the future given knowledge of events today

- Events X and Y has **statistical independence** if

$$\mathbb{P}\left(X = x_i \text{ and } Y = y_j\right) = \mathbb{P}\left(X = x_i\right)\mathbb{P}\left(Y = y_j\right)$$

From (57), we see that if $P(Y = y_i) > 0$ then independence implies $\mathbb{P}\left(X = x_i \,\middle|\, Y = y_j\right) = \mathbb{P}\left(X = x_i\right)$.