IS5152 Data-driven decision making
Preliminary Concepts and Notation

1. **Vectors**

   (a) n-vector
   An **n-vector or n-dimensional vector $\mathbf{x}$**, for any positive integer n, is an n-tuple,

   $$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

   of real numbers. The real number x_i is the ith component or element of the vector $\mathbf{x}$.

   (b) $\mathbb{R}^n$ is the n-dimensional Euclidean space. For all integer n, it is the set of all n-vectors.

2. **Vector addition and multiplication by a real number**
   Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. The sum $\mathbf{x} + \mathbf{y}$ is defined by

   $$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

   and the multiplication by a real number $\alpha\mathbf{x}$ is defined by

   $$\alpha\mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

3. **Linear dependence and independence**
   The vector $\mathbf{x}^1, \mathbf{x}^2, \ldots \mathbf{x}^m \in \mathbb{R}^n$ are said to be linearly independent if

   $$\left.\begin{array}{rcl} \lambda^1\mathbf{x}^{\mathbf{1}} + \lambda^2\mathbf{x}^2 + \ldots + \lambda^m\mathbf{x}^m & = & 0 \\ \lambda^1, \lambda^2, \ldots, \lambda^m & \in & \mathbb{R} \end{array}\right\} \Rightarrow \lambda^1 = \lambda^2 = \ldots \lambda^m = 0$$

   otherwise they are linearly dependent.

4. **Linear combination**
   The vector $\mathbf{x} \in \mathbb{R}^n$ is a **linear combination** of $\mathbf{x}^1, \mathbf{x}^2, \ldots \mathbf{x}^m \in \mathbb{R}^n$ if $\mathbf{x} = \lambda^1\mathbf{x}^1 + \lambda^2\mathbf{x}^2 + \ldots + \lambda^m\mathbf{x}^m$ for some $\lambda^1, \lambda^2, \ldots \lambda^m \in \mathbb{R}$.

5. **Scalar product**
   The **scalar (or dot) product xy** of two vectors $\mathbf{x}, \mathbf{y} \in \mathrm{I\!R^n}$ is defined by

$$\mathbf{xy} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

   It is normally denoted as $\mathbf{x}^t\mathbf{y}$ for clarity.

6. **Norm of a vector**
   The **p-norm** of a vector $\mathbf{x} \in \mathrm{I\!R^n}$ is defined by

$$\|\mathbf{x}\|_p = (|x_1|^p + \ldots + |x_n|^p)^{1/p}$$

   of which

   - $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$
   - $\|\mathbf{x}\|_2 = (|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2)^{1/2}$ and
   - $\|\mathbf{x}\|_\infty = \max_i |x_i|$
     are the most important. When the subscript is dropped from the norm, we mean the **2** -norm.

7. **Distance between two points**
   Let $\mathbf{x}, \mathbf{y} \in \mathrm{I\!R^n}$. The nonnegative number $\delta(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$ is called the Euclidean distance between the two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathrm{I\!R^n}$.

8. **Matrices**
   $\mathrm{I\!R^{m\times n}}$ denotes the space of all m-by-n real matrices.

$$\mathbf{A} \in \mathrm{I\!R^{m\times n}} \Longleftrightarrow \mathbf{A} = [\mathrm{A_{ij}}] = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \ldots & A_{mn} \end{pmatrix}$$

   The linear system of the following type will be frequently encountered:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n &= b_2 \\ \ldots & \quad . \\ A_{m1}x_1 + A_{m2}x_2 + \ldots + A_{mn}x_n &= b_m \end{aligned}$$

   where A_{ij} and b_i, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, are given real numbers. If we let $\mathbf{A}_i$ denote an n-vector whose n components are A_{ij}, $j = 1, 2, ..., n$, and if we let $\mathbf{x} \in \mathrm{I\!R^n}$, then the above system is equivalent to

$$\mathbf{A}_i\mathbf{x} = b_i, \quad i = 1, 2, \ldots, m$$

where $\mathbf{A}_i\mathbf{x}$ is interpreted as the scalar product of $\mathbf{A}_i$ and $\mathbf{x}$. If we further let $\mathbf{Ax}$ denote an m-vector whose m components are $\mathbf{A}_i\mathbf{x}$, $i = 1, 2, \ldots, m$, and $\mathbf{b}$ an m-vector whose m components are b_i, then the linear system can simply be written as

$$\mathbf{Ax} = \mathbf{b}$$

The i-th row of the matrix $\mathbf{A}$ will be denoted by $\mathbf{A}_i$ and will be an n-vector, hence

$$\mathbf{A}_i = \begin{pmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ A_{in} \end{pmatrix}$$

for $i = 1, 2, \ldots, m$. The j-th column of the matrix $\mathbf{A}$ will be denoted by $\mathbf{A}_{.j}$ and will be an m-vector. Hence

$$\mathbf{A}_{.j} = \begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix}$$

The above linear system can be written still in another form as follows

$$\sum_{j=1}^{n} \mathbf{A}_{.j} x_j = \mathbf{b}$$

The transpose of the matrix $\mathbf{A}$ is denoted by $\mathbf{A}^t$ and is the $n \times m$ matrix defined by

$$\mathbf{A}^t = \begin{pmatrix} A_{11} & \ldots & A_{m1} \\ \vdots & \ldots & \vdots \\ A_{1n} & \ldots & A_{mn} \end{pmatrix}$$

There are several important types of square matrices. We say that $\mathbf{A} \in \mathrm{I\!R}^{\mathrm{n} \times \mathrm{n}}$ is

- *the identity matrix* iff $A_{ij} = 1$, if $i = j$; $A_{ij} = 0$ otherwise, $i, j = 1, 2, \ldots, n$. The identity matrix is denoted by $\mathbf{I}$.
- *the inverse* of the matrix B iff $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. The matrix $\mathbf{A}$ is denoted by $\mathbf{B}^{-1}$, and $\mathbf{B}$ is said to be **invertible**.
- *symmetric* if $\mathbf{A}^t = \mathbf{A}$.
- *positive definite* if $\mathbf{x}^t\mathbf{Ax} > 0$, $\mathbf{0} \neq \mathbf{x} \in \mathrm{I\!R}^{\mathrm{n}}$.
- *positive semi-definite* if $\mathbf{x}^t\mathbf{Ax} \geq 0$, $\mathbf{x} \in \mathrm{I\!R}^{\mathrm{n}}$.

If $\mathbf{A}$ is an $n \times n$ matrix, the following statements are equivalent

- $\mathbf{A}$ is invertible.
- $\mathbf{Ax} = 0$ has only the trivial solution.
- $\mathbf{Ax} = \mathbf{b}$ is consistent for every n-vector b.
- $det(\mathbf{A}) \neq 0$.
- $\mathbf{A}$ has rank n.
- The rows of $\mathbf{A}$ are linearly independent.
- The columns of $\mathbf{A}$ are linearly independent.

9. **Function**

- A **function** f is a single-valued mapping from a set X into a set Y. That is for each $\mathbf{x} \in X$, the image set $f(\mathbf{x})$ consists of a single element of Y. The domain of F is X, and we say that f is defined on X. The range of f is $f(X) = \bigcup_{\mathbf{x} \in X} f(\mathbf{x})$.
- **Numerical function**
  A numerical function θ is a function from a set X into $\mathbb{R}$.
- **Vector function**
  An m-dimensional vector function f is a function from a set X into $\mathbb{R}^m$. The m components of the vector $f(\mathbf{x})$ are denoted by $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})$. Each f_i is a numerical function on X.
- **Linear vector function on $\mathbb{R}^n$**
  An m-dimensional vector function f defined on $\mathbb{R}^n$ is said to be linear if

$$f(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$$

  where $\mathbf{A}$ is some fixed $m \times n$ matrix and $\mathbf{b}$ is some fixed vector in $\mathbb{R}^m$.
- **Gradient of a numerical function**
  Let θ be a numerical function defined on $\mathbb{R}^n$, the n-dimensional vector of partial derivatives of θ with respect to $x_1, x_2, \ldots, x_n$ at $\bar{\mathbf{x}}$ is called the **gradient** of θ at $\bar{\mathbf{x}}$ and is denoted by $\nabla\theta(\bar{\mathbf{x}})$, that is,

$$\nabla\theta(\bar{\mathbf{x}}) = (\partial\theta(\bar{\mathbf{x}})/\partial x_1, \partial\theta(\bar{\mathbf{x}})/\partial x_2, \ldots, \partial\theta(\bar{\mathbf{x}})/\partial x_n)^t$$

- **Hessian of a numerical function**
  Let θ be a numerical function defined in $\mathbb{R}^n$, if θ is twice differentiable, the $n \times n$ matrix $\nabla^2\theta(x)$ is called the **Hessian** matrix of θ at $\mathbf{x}$ and its ijth element is written as

$$[\nabla^2\theta(\mathbf{x})]_{ij} = \partial^2\theta(\mathbf{x})/\partial x_i \partial x_j, \quad i, j = 1, 2, \ldots n$$

- **Jacobian of a vector function**
  Let f be an m-dimensional vector function in $\mathbb{R}^n$. The $m \times n$ matrix Jacobian of f at $\overline{\mathbf{x}}$ is

$$\nabla f(\overline{\mathbf{x}}) = \begin{pmatrix} \partial f_1(\overline{\mathbf{x}})/\partial x_1 & \dots & \partial f_1(\overline{x})/\partial x_n \\ \vdots & & \vdots \\ \partial f_m(\overline{\mathbf{x}})/\partial x_1 & \dots & \partial f_m(\overline{\mathbf{x}})/\partial x_n \end{pmatrix}$$

- **Convex function**
  A numerical function θ defined on the convex set $\Gamma \subset \mathbb{R}^n$ is said to be **convex** (with respect to Γ) if and only if

$$\left.\begin{array}{rcl} \mathbf{x}, \overline{\mathbf{x}} & & \in \Gamma \\ 0 \leq & \lambda & \leq 1 \end{array}\right\} \Rightarrow (1-\lambda)\theta(\overline{\mathbf{x}}) + \lambda\theta(\mathbf{x}) \geq \theta[(1-\lambda)\overline{\mathbf{x}} + \lambda\mathbf{x}]$$

- **Strictly convex function**
  A numerical function θ defined on the convex set $\Gamma \subset \mathbb{R}^n$ is said to be **strictly convex** (with respect to Γ) if and only if

$$\left.\begin{array}{rcl} \mathbf{x}, \overline{\mathbf{x}}, \mathbf{x} \neq \overline{\mathbf{x}} & & \in \Gamma \\ 0 < & \lambda & < 1 \end{array}\right\} \Rightarrow (1-\lambda)\theta(\overline{\mathbf{x}}) + \lambda\theta(\mathbf{x}) > \theta[(1-\lambda)\overline{\mathbf{x}} + \lambda\mathbf{x}]$$

- **Twice differentiable convex and concave function**
  Let θ be a numerical twice-differentiable function on $\mathbb{R}^n$, θ is convex if and only if $\nabla^2\theta(\mathbf{x})$ is positive semidefinite, that is, for each $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{y}\nabla^2\theta(\mathbf{x})\mathbf{y} \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^n$$

  θ is concave if and only if $\nabla^2\theta(x)$ is negative semidefinite, that is, for each $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{y}\nabla^2\theta(\mathbf{x})\mathbf{y} \leq 0 \quad \forall \mathbf{y} \in \mathbb{R}^n$$

REFERENCES:

- Howard, Anton. *Elementary Linear Algebra*, Wiley, 1981.
- Mangasarian, O.L. *Nonlinear Programming*, SIAM, 1994.