Week 4

Everything "is" matrix multiplication:

matrix multiplication	covector / vector / scalar version		
Example	Example	operation	
$\underbrace{\begin{bmatrix}0&1\\2&3\\4&5\end{bmatrix}}_{3\times2}\underbrace{\begin{bmatrix}1\\1\end{bmatrix}}_{2\times1}=\underbrace{\begin{bmatrix}1\\5\\9\end{bmatrix}}_{3\times1}$	$\underbrace{\begin{bmatrix}0&1\\2&3\\4&5\end{bmatrix}}_{3\times2}\underbrace{\begin{pmatrix}1\\1\end{pmatrix}}_{\in\mathbb{R}^2}=\underbrace{\begin{pmatrix}1\\5\\9\end{pmatrix}}_{\in\mathbb{R}^3}$	$\underbrace{A\mathbf{x}}_{\in\mathbb{R}^m}$	matrix-vector multiplication $A\in\mathbb{R}^{m\times n},\mathbf{x}\in\mathbb{R}^n$ (Definition 2.4)
$\underbrace{\begin{bmatrix}1&2\end{bmatrix}}_{1\times2}\underbrace{\begin{bmatrix}3\\4\end{bmatrix}}_{2\times1}=\underbrace{\begin{bmatrix}11\end{bmatrix}}_{1\times1}$	$\underbrace{\begin{pmatrix}1&2\end{pmatrix}}_{\in(\mathbb{R}^2)^*}\underbrace{\begin{pmatrix}3\\4\end{pmatrix}}_{\in\mathbb{R}^2}=\underbrace{11}_{\in\mathbb{R}}$	$\underbrace{\mathbf{v}^\top\mathbf{w}}_{\in\mathbb{R}}$	scalar product $\mathbf{v},\mathbf{w}\in\mathbb{R}^m$ (Definition 1.9, Definition 1.19)
$\underbrace{\begin{bmatrix}1&1&1\end{bmatrix}}_{1\times3}\underbrace{\begin{bmatrix}0&1\\2&3\\4&5\end{bmatrix}}_{3\times2}=\underbrace{\begin{bmatrix}6&9\end{bmatrix}}_{1\times2}$	$\underbrace{\begin{pmatrix}1&1&1\end{pmatrix}}_{\in(\mathbb{R}^3)^*}\underbrace{\begin{bmatrix}0&1\\2&3\\4&5\end{bmatrix}}_{3\times2}=\underbrace{\begin{pmatrix}6&9\end{pmatrix}}_{\in(\mathbb{R}^2)^*}$	$\underbrace{\mathbf{y}^\top A}_{\in(\mathbb{R}^n)^*}$	covector-matrix multiplication $\mathbf{y}^\top\in(\mathbb{R}^m)^*,A\in\mathbb{R}^{m\times n}$ (Definition 2.43)
$\underbrace{\begin{bmatrix}3\\4\\5\end{bmatrix}}_{3\times1}\underbrace{\begin{bmatrix}1&2\end{bmatrix}}_{1\times2}=\underbrace{\begin{bmatrix}3&6\\4&8\\5&10\end{bmatrix}}_{3\times2}$	$\underbrace{\begin{pmatrix}3\\4\\5\end{pmatrix}}_{\in\mathbb{R}^3}\underbrace{\begin{pmatrix}1&2\end{pmatrix}}_{\in(\mathbb{R}^2)^*}=\underbrace{\begin{bmatrix}3&6\\4&8\\5&10\end{bmatrix}}_{3\times2}$	$\underbrace{\mathbf{v}\mathbf{w}^\top}_{m\times n}$	outer product $\mathbf{v}\in\mathbb{R}^m,\mathbf{w}\in\mathbb{R}^n$ (Definition 2.44)
$\underbrace{\begin{bmatrix}3\\4\end{bmatrix}}_{2\times1}\underbrace{\begin{bmatrix}2\end{bmatrix}}_{1\times1}=\underbrace{\begin{bmatrix}6\\8\end{bmatrix}}_{2\times1}$	$\underbrace{\begin{pmatrix}3\\4\end{pmatrix}}_{\in\mathbb{R}^2}\underbrace{2}_{\in\mathbb{R}}=\underbrace{\begin{pmatrix}6\\8\end{pmatrix}}_{\in\mathbb{R}^2}$	$\underbrace{\mathbf{v}\lambda}_{\in\mathbb{R}^m}$	scalar multiplication (the other way around) $\mathbf{v}\in\mathbb{R}^m,\lambda\in\mathbb{R}$ (Definition 1.3).
$\underbrace{\begin{bmatrix}2\end{bmatrix}}_{1\times1}\underbrace{\begin{bmatrix}3&4\end{bmatrix}}_{1\times2}=\underbrace{\begin{bmatrix}6&8\end{bmatrix}}_{1\times2}$	$\underbrace{2}_{\in\mathbb{R}}\underbrace{\begin{pmatrix}3&4\end{pmatrix}}_{\in(\mathbb{R}^2)^*}=\underbrace{\begin{pmatrix}6&8\end{pmatrix}}_{\in(\mathbb{R}^2)^*}$	$\underbrace{\lambda\mathbf{v}^\top}_{\in(\mathbb{R}^m)^*}$	scalar multiplication (covector version) $\mathbf{v}^\top\in\mathbb{R}^m,\lambda\in\mathbb{R}$
$\underbrace{\begin{bmatrix}2\end{bmatrix}}_{1\times1}\underbrace{\begin{bmatrix}3\end{bmatrix}}_{1\times1}=\underbrace{\begin{bmatrix}6\end{bmatrix}}_{1\times1}$	$\underbrace{2}_{\in\mathbb{R}}\underbrace{3}_{\in\mathbb{R}}=\underbrace{6}_{\in\mathbb{R}}$	$\underbrace{\lambda\mu}_{\in\mathbb{R}}$	standard multiplication $\lambda,\mu\in\mathbb{R}$

(Mixed) products (e.g. $\mathbf{x}^\top A^\top A\mathbf{x}$) can be evaluated (using matrix multiplication) as if

- vectors in $\mathbb{R}^m$ were $m\times1$ matrices,
- covectors in $(\mathbb{R}^n)^*$ were $1\times n$ matrices,
- scalars in $\mathbb{R}$ were 1×1 matrices.

Only $[\cdots]$ must suitably be adapted to $(\cdots)$ or $\cdots$.
Associativity still holds, bracketing does not matter (Lemma 2.42).
Transpose a product: transpose individual factors and multiply them in reverse order (Lemma 2.40); $\lambda^\top = \lambda$.
Examples:

$$\underbrace{\underbrace{\mathbf{x}^\top A^\top}_{\text{covector vector}}\underbrace{A\mathbf{x}}_{\text{covector vector}}}_{\text{scalar}} = (A\mathbf{x})^\top(A\mathbf{x}) = \|A\mathbf{x}\|^2$$

Alternative bracketings:

$$\underbrace{\underbrace{\mathbf{x}^\top}_{\text{covector}}\underbrace{A^\top A}_{\text{covector matrix}}\underbrace{\mathbf{x}}_{\text{vector}}}_{\text{covector}}\underbrace{}_{\text{scalar}} \quad \text{or} \quad \underbrace{\underbrace{\mathbf{x}^\top}_{\text{covector}}\underbrace{A^\top A}_{\text{matrix vector}}\underbrace{\mathbf{x}}_{\text{vector}}}_{\text{vector}}\underbrace{}_{\text{scalar}}$$

$$\underbrace{\underbrace{\mathbf{v}^\top\mathbf{w}}_{\text{scalar}}\underbrace{\mathbf{v}^\top\mathbf{w}}_{\text{scalar}}}_{\text{scalar}} = (\mathbf{v}^\top\mathbf{w})^2$$

$$\underbrace{\underbrace{\mathbf{v}^\top}_{\text{covector}}\underbrace{\mathbf{w}\mathbf{v}^\top}_{\text{matrix}}\underbrace{\mathbf{w}}_{\text{vector}}}_{\text{vector}}\underbrace{}_{\text{scalar}} \leftarrow \text{may be very inefficient!}$$

Most efficient bracketing can be computed by *dynamic programming*.

Matrix multiplication...
...in column notation (Definition 2.36) ...in row notation (Observation 2.45).

$$A\underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_b \\ | & | & & | \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} | & | & & | \\ A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_b \\ | & | & & | \end{bmatrix}}_{AB}$$

$$\underbrace{\begin{bmatrix} — & \mathbf{u}_1^\top & — \\ — & \mathbf{u}_2^\top & — \\ & \vdots & \\ — & \mathbf{u}_a^\top & — \end{bmatrix}}_{A} B = \underbrace{\begin{bmatrix} — & \mathbf{u}_1^\top B & — \\ — & \mathbf{u}_2^\top B & — \\ & \vdots & \\ — & \mathbf{u}_a^\top B & — \end{bmatrix}}_{AB}$$

CR decomposition:
Prime number decomposition: $1001 = 7 \cdot 11 \cdot 13$
Matrix decomposition: write a matrix A as product of matrices to understand A better!

Theorem 2.46: Let A be an $m \times n$ matrix of rank r (Definition 2.10). Let C be the $m \times r$ submatrix of A containing the independent columns. Then there is a unique $r \times n$ matrix R' such that

$$A = CR'.$$

Example ($r = 1$):

$$\underbrace{\begin{bmatrix} 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 2 \\ 3 \end{bmatrix}}_{C,2\times 1} \underbrace{\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}}_{R',1\times 3}.$$

Columns of R': scalars to write each column of A as a scalar multiple of the first column.

Proof. A and C have the same column space (Lemma 2.11) $\Rightarrow$ every column of A is a linear combination of the columns of C:

$$\underset{\text{column } j \text{ of } A}{\mathbf{v}_j} = C \underset{\text{vector of } r \text{ scalars}}{\mathbf{x}_j} \qquad : \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}}_{A} = C \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix}}_{R' \in \mathbb{R}^{r \times n}}$$

The columns of C are linearly independent: all $\mathbf{x}_j$ are unique (Lemma 1.24). So R' is unique. □

Example:

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 1 & 4 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

	columns of A			
	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$
	$\parallel$	$\parallel$	$\parallel$	$\parallel$
	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$
	$\parallel$	$\parallel$	$\parallel$	$\parallel$
$\mathbf{v}_1$	$1\mathbf{v}_1$	$2\mathbf{v}_1$		$3\mathbf{v}_1$
$\mathbf{v}_2$				
$\mathbf{v}_3$			$1\mathbf{v}_3$	$-2\mathbf{v}_3$
$\mathbf{v}_4$				
independent?	yes	no	yes	no

$$\underbrace{\begin{bmatrix} 1 & 2 & 0 & 3 \\ 2 & 4 & 1 & 4 \\ 3 & 6 & 2 & 5 \end{bmatrix}}_{A,\ 3\times4} = \underbrace{\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}}_{C,\ 3\times2} \underbrace{\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}}_{R',\ 2\times4}$$

Here we proved existence of $A = CR'$, *not* efficient computation (see Chapter 3).

$A = CR'$ and matrix compression:

$m \times n$ matrix A: mn entries

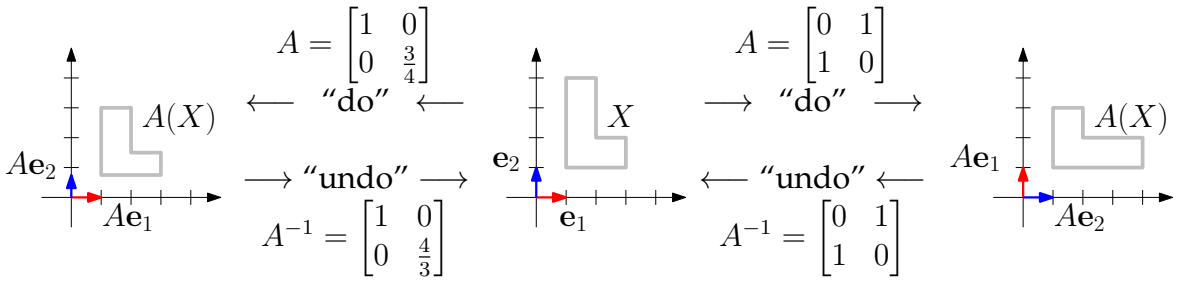
$m \times n$ matrix A of rank r: C has mr entries, R' has rn entries: $(m+n)r$ entries

If r is small, storing / transmitting C, R' instead of A is more efficient.

$$m = n = 1,000,\ r = 10: \qquad \begin{aligned} mn &: \ 1,000,000 \\ (m+n)r &: \ 20,000 \end{aligned}$$

Invertible and inverse matrices (Section 2.4)

T_A : input $\mapsto$ output ("do"). ? : output $\mapsto$ input ("undo")



$A = 0$: T_A : input $\mapsto \mathbf{0}$, cannot "undo" (no idea what the input was).

Definition 2.48: Let X, Y be sets and $f : X \to Y$ a function.

(i) f is *injective* if for every $y \in Y$, there is at most one $x \in X$ with $f(x) = y$.
("For every possible output, at most one input leads to it.")

(ii) f is *surjective* if for every $y \in Y$, there is at least one $x \in X$ with $f(x) = y$.
("For every possible output, at least one input leads to it.")

(iii) f is *bijective* (undoable) if f is both injective and surjective.
("For every possible output, exactly one input leads to it.")

(iv) The *inverse* of a bijective function f is the function

$$f^{-1} : Y \to X, \quad y \mapsto \text{the unique } x \in X \text{ such that } f(x) = y.$$

Fact 2.49: If $f : X \to Y$ is bijective, then $f^{-1} : Y \to X$ is also bijective, and $(f^{-1})^{-1} = f$. Moreover, $f^{-1} \circ f = \mathrm{id}$ (f^{-1} is undoing f) and $f \circ f^{-1} = \mathrm{id}$ (f is undoing f^{-1}).

$$\mathrm{id} : \text{input} \mapsto \text{input}$$

Matrix transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m, \mathbf{x} \mapsto A\mathbf{x}$:

- $m < n$ (A wide): not injective (Lemma 2.50): "More inputs than possible outputs."
- $m > n$ (A tall): not surjective (Lemma 2.51): "More possible outputs than inputs."
- $m = n$ (A square): potentially bijective, but not always (see $A = 0 \in \mathbb{R}^{m \times m}$)

Lemma 2.52: Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be a bijective linear transformation. Then its inverse $T^{-1} : \mathbb{R}^m \to \mathbb{R}^m$ is also a linear transformation (and bijective by Fact 2.49).

Proof. Prove linearity of T^{-1} (Definition 2.21):

$$T^{-1}(\lambda_1 \mathbf{y}_1 + \lambda_2 \mathbf{y}_2) = \lambda_1 T^{-1}(\mathbf{y}_1) + \lambda_2 T^{-1}(\mathbf{y}_2).$$

$\mathbf{x}_1 := T^{-1}(\mathbf{y}_1), \mathbf{x}_2 := T^{-1}(\mathbf{y}_2)$. ($T(\mathbf{x}_1) = \mathbf{y}_1$ and $T(\mathbf{x}_2) = \mathbf{y}_2$.)

Linearity of T:

$$T(\lambda_1 \underbrace{T^{-1}(\mathbf{y}_1)}_{\mathbf{x}_1} + \lambda_2 \underbrace{T^{-1}(\mathbf{y}_2)}_{\mathbf{x}_2}) = \lambda_1 \underbrace{\mathbf{y}_1}_{T(\mathbf{x}_1)} + \lambda_2 \underbrace{\mathbf{y}_2}_{T(\mathbf{x}_2)}.$$

Apply T^{-1} to both sides (and use $T^{-1}(T(\text{input})) = \text{input}$):

$$\lambda_1 T^{-1}(\mathbf{y}_1) + \lambda_2 T^{-1}(\mathbf{y}_2) = T^{-1}(\lambda_1 \mathbf{y}_1 + \lambda_2 \mathbf{y}_2).$$

□

How can we tell from A whether T_A is bijective?

Lemma 2.53: Let A be an $m \times m$ matrix. The following three statements are equivalent.

(i) $T_A : \mathbb{R}^m \to \mathbb{R}^m$ is bijective.

(ii) There is an $m \times m$ matrix B such that $BA = I$.

(iii) The columns of A are linearly independent.

Proof structure: (i)⇒(ii)⇒(iii)⇒(i):

Proof.
(i)⇒(ii): If T_A is bijective, its inverse T_A^{-1} (Definition 2.48) is a matrix transformation T_B: T_A is a linear transformation (Observation 2.22), so T_A^{-1} is also linear transformation (Lemma 2.52) and a matrix transformation (Theorem 2.26). Fact 2.49:

$$\underbrace{T_B \circ T_A}_{\overset{\text{Cor. 2.37}}{=} T_{BA}} = \underbrace{\mathrm{id}}_{T_I}.$$

By Theorem 2.26 (unique matrix): $BA = I$.
(ii)⇒(iii): If $BA = I$, then $\mathbf{x} = \mathbf{0}$ is the only vector with $A\mathbf{x} = \mathbf{0}$ (so the columns of A are linearly independent by Obs. 2.5): Let $A\mathbf{x} = \mathbf{0}$. Then $\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = B\mathbf{0} = \mathbf{0}$.
(iii)⇒(i): If the columns of A are linearly independent, T_A is surjective: for each $\mathbf{y} \in \mathbb{R}^m$, there is at least one $\mathbf{x} \in \mathbb{R}^m$ with $T_A(\mathbf{x}) = A\mathbf{x} = \mathbf{y}$. True because the m columns of A span $\mathbb{R}^m$ (Lemma 1.28).
T_A is even bijective: there is exactly one such $\mathbf{x}$. True because $\mathbf{y}$ can be written as a linear combination of the columns of A in only one way (Lemma 1.24). □

In (ii), we could as well have said"$AB = I$":
Lemma 2.54: Let A, B be $m \times m$ matrices such that $BA = I$. Then also $AB = I$.

Definition and basic properties:

Definition 2.55: Let A be an $m \times m$ matrix. A is called *invertible* if there is an $m \times m$ matrix B such that $BA = I$. (By Lemma 2.53 this is the case if and only if A has linearly independent columns.) Otherwise, A is called *singular*.
Observation 2.56: In the definition, we could as well have said "$AB = I$" or "$AB = BA = I$" (Lemma 2.54). Also, B is unique.

Proof. Consider B, B' satisfying the conditions. Then

$$B = IB = (B'A)B = B'(AB) = B'I = B'.$$

□

Definition 2.57: Let A be an $m \times m$ matrix. A is invertible if and only if there exists an $m \times m$ matrix B such that $BA = I$ (or $AB = I$, or $AB = BA = I$). In this case, the matrix B is unique and called the *inverse* of A. We denote it by A^{-1}.

Observation 2.58: $(A^{-1})^{-1} = A$.

Case 1×1**.**

$$A = \begin{bmatrix} a \end{bmatrix} \quad \Rightarrow \quad A^{-1} = \begin{bmatrix} \frac{1}{a} \end{bmatrix} \quad \text{(if } a \neq 0\text{)}.$$

Case 2×2**.**

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \Rightarrow \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{(if } ad - bc \neq 0\text{)}.$$

Case $m \times m$**.** Inefficient formula (based on *Cramer's rule*); efficient computation in Chapter 3.

Lemma 2.59: Let A and B be invertible $m \times m$ matrices. Then AB is also invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. Check that $B^{-1}A^{-1}$ satisfies one of the conditions for the matrix AB:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

□

Extends to more matrices, for example $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

Lemma 2.60: Let A be an invertible $m \times m$ matrix. Then the transpose $A^\top$ is also invertible, and

$$\left(A^\top\right)^{-1} = \left(A^{-1}\right)^\top.$$