Introduction
Prerequisites:
Things you must know:
Linear algebra: matrices, linear systems, gaussian elimination.
Vector spaces: Subspaces, linearly independent vectors, generators (spanning sets), bases, dimension.
Diagonal form of a matrix: eigenvalues, eigenvectors, eigenspaces.
Scalar products. Orthogonal vectors. Orthonormal bases. Orthogonal matrices. Spectral theorem (symmetric matrices are diagonalizable). Quadratic forms.
Analytic geometry of plane: conic sections.
Analytic geometry of space: lines, planes, distances, circles, spheres.

Outline of this part of course:
Complements of linear algebra: non-finitely generated vector spaces, orthonormal bases, projections, condition number of a matrix.
Complements of analytic geometry: curves and surfaces. Changes of coordinates. Quadrics.

Let us begin by revisiting some notions of linear algebra concerning vector spaces.
A basis in a vector space is a sequence of vectors $v_1, ..., v_n$ which are at the same time
- linearly independent (i.e. none of them can be written as a linear combination of the others)
- generators, (or a spanning set for the space) (i.e. any vector in the space can be written as a linear combination of them).

Problem: Does every vector space own a basis? The answer is: no. There are big vector spaces in which there is not a "finite" sequence of vectors who can span the whole space.

The simplest example is the following: Let us consider the set $\mathbb{R}[x] = \{$polynomials in one indeterminate x with real coefficients$\}$. This is a $\mathbb{R}$-vector space which doesn't have a "finite" basis.
Proof: Suppose $P_1(x), ..., P_n(x)$ were a basis for the space, and let m be maximum degree of the n polynomials, then any polynomial of degree bigger than m cannot be written as a linear combination of $P_1, ..., P_n$ □

There are vector spaces bigger than $\mathbb{R}[x]$:
$C[a, b]$ is the vector space of functions defined in an interval $[a, b] \subset \mathbb{R}$
$C^0[a, b]$ is the vector space of continuous functions defined in an interval $[a, b] \subset \mathbb{R}$
$C^1[a, b]$ is the vector space of continuous functions $f(x)$ defined in an interval $[a, b] \subset \mathbb{R}$, which admit first derivative $f'(x)$ in the interval and the derivative $f'(x)$ is too continuous.
$C^n[a, b]$ is the vector space of continuous functions $f(x)$ defined in an interval $[a, b] \subset \mathbb{R}$, which admit derivatives $f'(x), f''(x), ..., f^{(n)}(x)$ in the interval and all derivatives are continuous.
$C^\infty[a, b]$ is the vector space of continuous functions $f(x)$ defined in an interval $[a, b] \subset \mathbb{R}$, which admit derivatives $f^{(n)}(x)$ of any order in the interval and all derivatives are continuous.

If a vector space V has a "finite" basis $v_1, ..., v_n$, we say that V is finitely generated. Otherwise we have to slightly modify the definition of basis.

Definition 1: Let V be a $\mathbb{K}$-vector space. We say that a family $\{v_i\}_{i \in I}$ of vectors of V is free (or that they are linearly independent) if every finite subset of the family is formed by linearly independent vectors.
Definition 2: Let V be a $\mathbb{K}$-vector space. We say that a family $\{v_i\}_{i \in I}$ of vectors of V spans V (or that they are generators of V) if any vector of V is a linear combination of a finite number of vectors of the family.
Definition 3: A Hamel basis of V is a family $\{v_i\}_{i \in I}$ of vectors of V which are free and span V.

Example: $\mathbb{R}[x]$. The simplest Hamel basis of $\mathbb{R}[x]$ is $\{x^n\}_{n \in \mathbb{N}}$, the set of all monomials with coefficient 1. Every polynomial is a linear combination of a finite number of monomials.
There are other bases of $\mathbb{R}[x]$, which are useful in many questions, as we shall see.

It can be proved that every vector space has at least one Hamel basis, but in many cases it is virtually impossible to describe it. By instance the space $C^0(\mathbb{R})$ of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ has many Hamel bases, but they are too big to be described.

Just a brief outline: Some $\mathbb{R}$-vector spaces, the Hilbert spaces, have another kind of basis, the so-called Hilbert bases. Every vector is a non-finite summation of vectors of the Hilbert basis. Since non-finite summations require new definitions of limits and convergence of series, this theory is rather complicated and a whole chapter of functional analysis is devoted to Hilbert spaces.

Now we recall the definition of scalar product in real vector spaces.
If V is a $\mathbb{R}$-vector space, a scalar product is a pairing $\langle v, w\rangle$ which has the following properties:
1) $\langle v, w\rangle = \langle w, v\rangle$ (symmetry)
2) $\langle av + bw, u\rangle = a\langle v, u\rangle + b\langle w, u\rangle$ (bilinearity)
3) $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if $v = 0$

Examples: $\mathbb{R}^n$ has the so-called euclidean scalar product:
$\langle (x_1, ..., x_n), (y_1, ..., y_n)\rangle = x_1y_1 + \cdots + x_ny_n$

$\mathbb{R}^n$ has many other scalar product, as we shall see

In V_3, the space of geometric vectors there is the well-known scalar product $v \cdot w =\mid v \mid \cdot \mid w \mid \cos\theta$. When V_3 is identified with $\mathbb{R}^3$ via an orthonormal basis, this product becomes the euclidean product in $\mathbb{R}^3$.

Let us consider the space $C^0[a, b]$ of continuous functions defined in an interval $[a, b] \subset \mathbb{R}$. We can define a scalar product this way:
$\langle f(x), g(x)\rangle = \int_a^b f(x)g(x)dx$
It is easy to verify properties 1) and 2). As for part 3), it is clear that for any function $f(x)$ we have $\int_a^b f^2(x)dx \geq 0$. Let us now prove that
$\int_a^b f^2(x)dx = 0$ if and only if $f \equiv 0$ (i.e. f is the constant null function).
Suppose not, than there is a $x_0 \in [a, b]$ such that $f(x_0) \neq 0$. There must be an interval surrounding x_0 in which $f(x) \neq 0$ and in this interval $\int f^2(x)dx$ is not 0. So $\int_a^b f^2(x)dx$ is not 0, since $f^2(x)$ is not negative.

There are other scalar products in function spaces.

Definition: A norm in a $\mathbb{R}$-vector space V is a function $\parallel . \parallel: V \to \mathbb{R}$ such that
1) $\parallel av \parallel=\mid a \mid\parallel v \parallel$ for any $a \in \mathbb{R}$
2) $\parallel v + w \parallel\leq\parallel v \parallel + \parallel w \parallel$ (triangular inequality)
3) $\parallel v \parallel\geq 0$ and $\parallel v \parallel= 0$ if and only if $v = 0$

Every scalar product induces a norm this way: $\parallel v \parallel= \sqrt{\langle v, v\rangle}$

The Cauchy-Schwarz inequality
Let v, w be vectors in V (V vector space with a scalar product), then
$\mid \langle v, w\rangle \mid\leq\parallel v \parallel\parallel w \parallel$
Proof: If $w = 0$ the inequality is trivial. So let us suppose $w \neq 0$.
For every $x \in \mathbb{R}$ let us consider the vector $v + xw$ and the nonnegative number $\parallel v + xw \parallel^2$
We have:
$0\leq\parallel v + xw \parallel^2= \langle v + xw, v + xw\rangle = \langle v, v\rangle + 2\langle v, w\rangle x + \langle w, w\rangle x^2$
This is an inequality regarding a second degree polynomial in x.
So the discriminant of the polynomial cannot be positive, that is:
$\Delta/4 = \langle v, w\rangle^2 - \parallel v \parallel^2\parallel w \parallel^2\leq 0$
Since we are dealing with positive numbers, this is the goal of our proof. □

As a consequence, if $v, w \neq 0$, we have $-1 \leq \frac{\langle v, w\rangle}{\parallel v \parallel\parallel w \parallel} \leq 1$ and we can define in any space the angle between two vectors.
$\cos(v^\wedge w) = \frac{\langle v, w\rangle}{\parallel v \parallel\parallel w \parallel}$

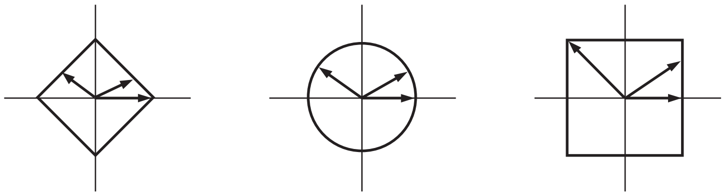
Norms

Every scalar product induces a norm, but there are norms which are not inducted by any scalar product.

Example 1: In $C^0[a,b]$ we can define the following norm $\| f(x) \| = \sup\{| f(x) |\}$ (sup is the supremum). This norm is not induced by any scalar product.

Example 2: In $\mathbb{R}^n$ there are many norms and some are useful in many contexts. The main are
2-norm, the one induced by euclidean product. It is denoted this way $\| (x_1,...,x_n) \|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$.
1-norm, defined this way: $\| (x_1,...,x_n) \|_1 = | x_1 | + \cdots + | x_n |$
p-norm ($p \in \mathbb{N}, n \geq 1$): $\| (x_1,...,x_n) \|_p = (| x_1 |^p + \cdots + | x_n |^p)^{1/p}$
∞-norm : $\| (x_1,...,x_n) \|_\infty = \max\{| x_1 |,...,| x_n |\}$. It can be proved that $\| v \|_\infty = \lim_{p\to\infty} \| v \|_p$

It can be useful and amusing to sketch all the vectors in $\mathbb{R}^2$ which have norm equal to 1, using 1-norm, 2-norm and ∞-norm.



Let us go back to spaces with a scalar product. From now on, unless otherwise stated, all vector spaces will be spaces with a scalar product.

We say that two vectors $v, w \in V$ are orthogonal if $\langle v, w\rangle = 0$
Example: In $C^0[0, 2\pi]$ with the product defined by the integral, the functions $\sin(x)$ and $\cos(x)$ are orthogonal. More, all the functions $\sin(nx)$, $\cos(mx)$ ($n, m \in \mathbb{N}$) are mutually orthogonal.

Proposition: Let $v_1,...,v_n$ be vectors which are:
1) pairwise orthogonal 2) different from 0 (the null vector).
Then $v_1,...,v_n$ are linearly independent.

Proof: Suppose $a_1v_1 + \cdots + a_nv_n = 0$. Multiply each side of the relation by v_1 and use bilinear property
$a_1\langle v_1, v_1\rangle + \cdots + a_n\langle v_n, v_1\rangle = \langle 0, v_1\rangle$. By hypothesis, all the terms of first side vanish except the first $a_1 \| v_1 \|^2 = 0$. But $v_1 \neq 0$, and so $\| v_1 \| \neq 0$. We conclude that $a_1 = 0$.
By repeating the same calculation and multiplying each side by $v_2,...,v_n$ we get subsequently $a_1 = a_2 = \cdots = a_n = 0$. □

Definition: A basis $\{v_i\}_{i\in I}$ of a vector space is called orthonormal if
1) the vectors are pairwise orthogonal
2) the norm of each vector is 1: $\| v_i \| = 1 \quad \forall i$

Does every vector space have an orthonormal basis?

The Gram-Schmidt algorithm:
Given $v_1,..,v_n$ linearly independent vectors, build $v_1'',...,v_n'$' such that
1) $L\{v_1,..,v_i\} = L\{v_1'',..,v_i''\}$ for $1 \leq i \leq n$
2) $v_1'',...,v_n''$ are orthonormal

G-S process is well-known
$v_i' = v_i - \langle v_1'', v_i\rangle v_1'' - ... - \langle v_{i-1}'', v_i\rangle v_{i-1}''$
$v_i'' = \dfrac{v_i'}{\| v_i' \|}$

So, any finitely generated vector space has at least one orthonormal basis.

Example: Let V be the subspace of $\mathbb{R}[x]$ consisting of all polynomials whose degree is less or equal to n ($n \geq 1$)
$V = \{P(x) \mid \deg(P) \leq n\}$. We can think of V as a subspace of $C^0[-1,1]$ with the scalar product defined by the integral.

We want to find an orthonormal basis of V by applying G-S process to the basis $1, x, x^2, x^3, ...$
We have choosen the interval $[-1,1]$ for sake of simplicity, but the following computation can be carried out for every interval.
Let us notice that any monomial of odd degree is orthogonal to any monomial of even degree, since their product is a odd function in $[-1,1]$.
Then the G-S algorithm runs this way:
$p_0'' = 1/\sqrt{2}$
$p_1'' = x/\sqrt{2/3}$ (1 and x are already orthogonal)
$p_2' = x^2 - \langle 1, x^2\rangle/2$ (x^2 and x are already orth) $= x^2 - 1/3$
$p_2'' = (x^2 - 1/3)/\sqrt{8/45} = (3x^2 - 1)\ \sqrt{5/8}$
$p"_3 = (5x^3 - 3x)\sqrt{7/2}/2$
ans so on.
These polynomial form an orthonormal infinite basis for $\mathbb{R}$[x]

The famous Legendre polynomials are proportional to these (so they are pairwise orthogonal, but don't have unitary norm)

$$q_n = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n = \frac{(2n!)}{2^n (n!)^2} p_n$$

Projections
Let V be a vector space and let W be a subspace of V. If $v \in V$, we say that the vector p is the (orthogonal) projection of v in W if
1) $p \in W$
2) $\langle p - v, w\rangle = 0$ for every $w \in W$
It can be proved that p exists and that it is unique.
To calculate p, we must have an orthonormal basis $w_1, ..., w_n$ of W. Then $p = \langle w_1, v\rangle w_1 + \cdots + \langle w_n, v\rangle w_n$
A straightforward calculation proves that this p is orthogonal to $w_1, ..., w_n$ and consequently to any vector $w \in W$.
It can be proved that p is, among the vectors of W, the one which is closest to v.
To be clear: if we define the distance between two vectors this way: $\mathrm{dist}\,(u, v) = \| v - u \|$,
then it can be proved that: for any $w \in W$ we have $\mathrm{dist}\,(w, v) \geq \mathrm{dist}\,(w, p)$.

Orthogonal matrices
A square $n \times n$ matrix P is called orthogonal if its columns form an orthonormal basis for $\mathbb{R}^n$ (with the euclidean product).
Caution! the name is misleading; the correct name should be orthonormal matrix, but we cannot overcome a old habit.
Properties of orthogonal matrices
1) $P^{-1} = P^T$ the transpose e the inverse are the same matrix (some books give this one as definition of orthogonal matrix)
2) The rows too form an orthonormal basis for $\mathbb{R}^n$.
3) $\det(P) = \pm 1$. So there are to kinds of orth. matrices. The ones with det=1 which are called rotation matrices, the others are called symmetry matrices. The latter ones are mainly used in algorithms which are used to calculate the eigenvalues of a matrix, the former ones are used in changes of coordinates.

The spectral theorem
We recall this famous theorem:
Let A be a symmetric $n \times n$ real matrix. Then
1) A has exactly n eigenvalues (each one must be considered as many times as its multiplicy as root of the characteristic polynomial of A).
2) For each eigenvalue λ we have $\dim(V_\lambda) =$ multiplicity of λ (V_λ the eigenspace).
3) If λ_1 and λ_2 are different eigenvalues of A and v_1 and v_2 are corresponding eigenvectors, then $\langle v_1, v_2\rangle = 0$.

As a consequence of this theorem, if A be a symmetric $n \times n$ real matrix then there exists an orthogonal matrix P and a diagonal matrix D such that
$P^T A P = D$
The columns of P form an orthonormal basis of $\mathbb{R}^n$, in which every vector is an eigenvector, while the numbers in the diagonal of D are the eigenvalues.

Definite positive matrices
A symmetric $n \times n$ real matrix A is said to be definite positive if for every column vector $v \in \mathbb{R}^n$ $(v \neq 0)$ we have
$v^T A v > 0$
This means that the quadratic form $Q(v) = v^T A v$ associated to A has only positive values.
It can be proved that A is definite positive if and only if all the (real) eigenvalues of A are positive.

Exercise: Let A be a definite positive $n \times n$ matrix, then we can define a new scalar product in $\mathbb{R}^n$ this way:
$\langle v, w\rangle_* = v^T A w$ (vectors in $\mathbb{R}^n$ are column vectors).
We must prove the 3 properties of scalar products.
1) $\langle v, w\rangle_* = v^T A w = (v^T A w)^T = w^T A^T v = w^T A v = \langle w, v\rangle_*$
2) is a consequence of elementary properties of operations between matrices
3) obvious by the definition of definite positive matrix

The Sylvester theorem
In order to check if a symmetric matrix A is definite positive or not, the best thing is to use Sylvester's law of inertia which can be described in this elementary way.
Apply gaussian elimination to A both to its rows and columns the same way.
To be clear: if you make the elementary operation $R_i \to R_i + kR_j$, you should make the corresponding operation $C_i \to C_i + kC_j$ in order to preserve symmetry, and so on.
This way, at the end of the algorithm you get a diagonal matrix.
Sylvester's law of inertia states that the eigenvalues of this diagonal matrix have the same signs as the eigenvalues of A.
A is positive definite if and only if all the numbers in the diagonal of the diagonal matrix obtained this way are positive.

The problem of conditioning
Let A be a square non-singular matrix and let
$Au = b$
be a linear system whose coefficient matrix is A. Then let x be its solution.
Let us now consider this other system
$Au = (b + \delta b)$
where the coefficient matrix A is the same and $b + \delta b$ is a column matrix very close to b, that is such that δb is "very small" and let $x + \delta x$ be the solution of this system.
The problem is: if $b + \delta b$ is close to b, is $x + \delta x$ close to x?
In other words: if δb is "small", is δx "small"?
The answer will be: it depends on A.
There are matrices for which the answer to the question is: yes, if δb is small then δx is small; these matrices will be called well-conditioned matrices.
There are matrices for which the answer is: no, even if δb is very small, δx can be large; these matrices will be called ill-conditioned matrices.

How do we decide how small is a vector? Obviously by computing its norm.
But in $\mathbb{R}^n$ there are many norms and as we shall see, in some cases norms other than the euclidean norm can be more suitable for this kind of problem.
Let us choose some norm in $\mathbb{R}^n$.
We have
$Ax = b \qquad A(x + \delta x) = b + \delta b \qquad Ax + A\delta x = b + \delta b \quad$ and so $\quad A\delta x = \delta b$
Applying norms (any norm) we get $\| Ax \| = \| b \|$ and $\| A\delta x \| = \| \delta b \|$
Now we want to compare $\| x \|$ with $\| b \|$ and $\| \delta x \|$ with $\| \delta b \|$, so we must find some relation between $\| Ax \|$ and $\| x \|$ and between $\| A\delta x \|$ and $\| \delta x \|$.
We introduce now a new object:
The matrix norm
Let A is a square non-singular matrix and let $\| - \|$ be any norm in $\mathbb{R}^n$
Definition: The matrix norm (opposed to well-known vector-norm) of A is

$$\| A \| = \max_{v \in \mathbb{R}^n, v \neq 0} \frac{\| Av \|}{\| v \|}$$

This definition gives us the relation between $\| Av \|$ and $\| v \|$
$\| A \| \geq \| Av \| / \| v \|$ for every $v \in \mathbb{R}^n$ and so $\| A \| \| v \| \geq \| Av \|$
So this is what we were looking for, but from a practical point of view it is impossible to calculate $\| A \|$ using this definition. We'll have to calculate $\| A \|$ in some other way.
We shall investigate later this crucial problem, forfor now let us return to the problem of conditioning

We have $\| Ax \| = \| b \|$ and $\| A \| \| x \| \geq \| Ax \| = \| b \|$

Then we have $A\,\delta x = \delta b$, but for our purpose it is better to write it this way: $\delta x = A^{-1}b$ and so $\| \delta x \| = \| A^{-1}b \| \leq \| A^{-1} \| \| \delta b \|$

We will not compare $\| \delta x \|$ with $\| \delta b \|$, rather we shall compare $\| \delta x \| / \| x \|$ and $\| \delta b \| / \| b \|$, since relative variations are more important than absolute variations. We have
$\| A \| \| x \| \geq \| b \| \qquad \| \delta x \| \leq \| A^{-1} \| \| \delta b \|$
Now write these inequalities this way:

$$\frac{\| 1 \|}{\| x \|} \leq \frac{\| A \|}{\| b \|} \qquad \| \delta x \| \leq \| A^{-1} \| \| \delta b \|$$

Finally:

$$\frac{\| \delta x \|}{\| x \|} \leq \| A \| \| A^{-1} \| \frac{\| \delta b \|}{\| b \|}$$

Let's now state:
Definition: If A is a square non-singular real matrix and $\| \cdot \|$ is a matrix norm, the number

$$\text{cond}\,(A) = \| A \| \| A^{-1} \|$$

is called the condition number of A with respect to the chosen norm.
So the inequality is

$$\frac{\| \delta x \|}{\| x \|} \leq \text{cond}\,(A) \frac{\| \delta b \|}{\| b \|}$$

If $\text{cond}(A)$ is small (as we shall see, $\text{cond}(A) \geq 1$) then a small variation in b induces a small variation in x.

How to calculate cond (A)
First we make a simple remark:
If v is an eigenvector of A, we have $Av = \lambda v$ and so $\dfrac{\| Av \|}{\| v \|} = | \lambda |$
This shows that eigenvalues of A are closely related to the norm of A. The relation is very good if the norm is the 2-norm and if A is a symmetric matrix.
Let A be a symmetric matrix $n \times n$. Let us write the n eigenvalues of A ordered according to their absolute value:

$$| \lambda_1 | \leq | \lambda_2 | \leq \cdots \leq | \lambda_n |$$

So λ_1 is an eigenvalue of A whose absolute value is minimal and λ_n an eigenvalue of A whose absolute value is maximal.
It can be proved that
Proposition: if A is symmetric we have $\| A \|_2 = | \lambda_n |$.
Remark: If $| \lambda_1 | \leq | \lambda_2 | \leq \cdots \leq | \lambda_n |$ are the eigenvalues of A ordered according to their absolute value, then $| 1/\lambda_n | \leq \cdots \leq | 1/\lambda_2 | \leq | 1/\lambda_1 |$ are the eigenvalues of A^{-1} ordered according to their absolute value.
So we get: $\quad \| A^{-1} \| = | 1/\lambda_1 |$ and

$$\text{cond}_2\,(A) = \frac{| \lambda_n |}{| \lambda_1 |}$$

Remark: So $\text{cond}_2\,(A) \geq 1$. This is true for any matrix and for any matrix norm.

Let now A be non-symmetric; things are a little more complicated. Let us consider the matrix A^TA. This a symmetric definite positive matrix.
It is definite positive since for every v, we have $v^TA^TAv = (Av)^TAv = \| Av \|^2$
Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the ordered eigenvalues of A (they are all positive) and $s_1 \leq s_2 \leq \cdots \leq s_n$ be

their square roots. They are called the singular values of A.
It can be proved that

$$\text{cond}_2(A) = \frac{s_n}{s_1}$$

Remark 1: For a symmetric matrix this way of calculating condition number coincides with the previous one.
Remark 2: The matrices $A^T A = AA^T$ are not in general the same matrix, but have the same eigenvalues.
Remark 3: A matrix is well-conditioned if its condition number is 1 or close to 1. It can be proved that orthogonal matrices are always well-conditioned since their condition number (with respect to 2-norm) is 1.

Calculation of condition number involves computation of eigenvalues which takes often a long time.
This is the reason why, in many cases, 1-norm and ∞-norm are preferred to 2-norm since their computation takes a much shorter time.
In fact it can be proved that
- $\| A \|_1 = \max_j \sum_i | a_{ij} |$ (maximum 1-norm of column vectors of A)
- $\| A \|_\infty = \max_i \sum_j | a_{ij} |$ (maximum 1-norm of row vectors of A)

Example The following matrix A is apparently an innocent matrix; it is symmetric and $\det(A) = -1$

$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 0 \end{pmatrix}$ Let $b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ Solve the system $Ax = b$. Its solution is $x = (-3, 2, 0)$ Then let $b_1 = \begin{pmatrix} 1.1 \\ 1.9 \\ 1 \end{pmatrix}$

The solution of $Ax = b_1$ is now $x_1 = (1, -0.4\ ,\ 0.3)$ e though b and b_1 are very close, x and x_1 are not. This means that A should be ill-conditioned.
In fact the eigenvalues of A are about $\lambda_1 \simeq 0.0293$ $\lambda_2 \simeq -3.8646$ $\lambda_3 \simeq 8.8354$
So $\text{cond}\,(A) \simeq \dfrac{8.8354}{0.0293} \simeq 301.6885$ and is very big.
Let's examine in our case the inequality $\| \delta x \| \,/\, \| x \| \leq \text{cond}\,(A) \| \delta b \| \,/\, \| b \|$ which can be written as $\dfrac{\| \delta x \| \,/\, \| x \|}{\| \delta b \| \,/\, \| b \|} \leq \text{cond}\,(A)$.
$\| b \| = \sqrt{6} \simeq 2.4495$ $\quad \| \delta b \| = \| (0.1, -0.1, 0) \| \simeq 0.1414$ $\quad \| \delta b \| \,/\, \| b \| \simeq 0.0577$
$\| x \| = 3.6056$ $\quad \| \delta x \| = \| (4, -2.4, 0.3) \| \simeq 4.6744$ $\quad \| \delta x \| \,/\, \| x \| \simeq 1.2964$
So $\dfrac{\| \delta x \| \,/\, \| x \|}{\| \delta b \| \,/\, \| b \|} \simeq \dfrac{4.6744}{0.1414} \simeq 22.4551$. Though we are very far from cond (A), the difference between x and x_1 is plain.