# Aritmetica del Computador

Hermes Pantoja Carhuavilca

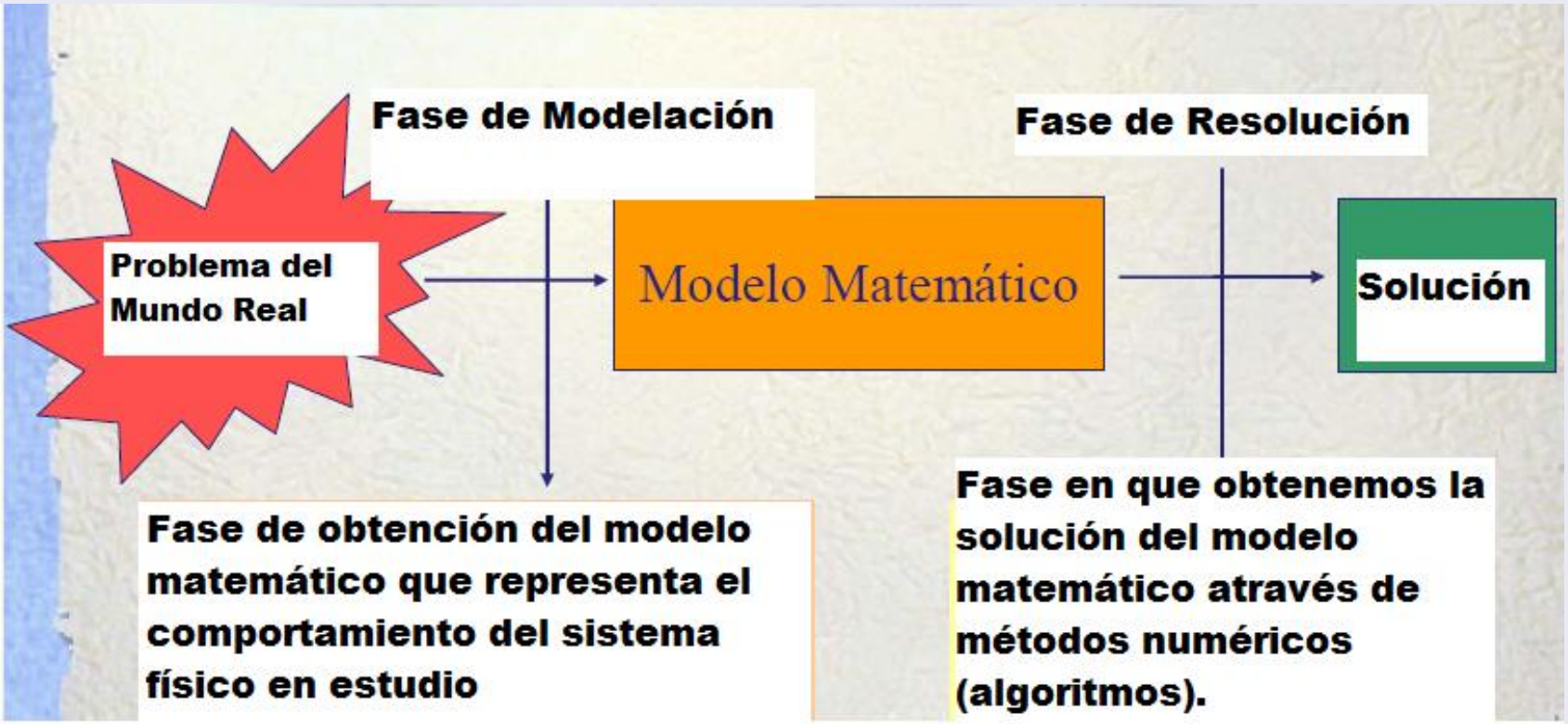
Facultad de Ingeniería Industrial
Universidad Nacional Mayor de San Marcos

Métodos Computacionales

# Contenido

1. Introducción
2. Teoria de Errores
3. Aritmetica del computador

# Introducción al estudio de métodos computacionales



# Aproximación y Errores

- Los cálculos númericos **inevitablemente** conducen a errores
- Estos son de dos clases principales:
  1. Errores de Redondeo
     - Errores asociados con la representación inexacta de números reales por la computadora.
     - Errores asociados con la maquina.
  2. Errores de Truncamiento
     - Errores asociados con el uso de un procedimiento numérico aproximado para reemplazar una expresión matemática exacta.
     - Error asociados con el algoritmos matemático.
- Ambos conducen al error total.

# Fuentes de errores

Los errores en el cálculo matemático tienen varia fuentes:

- El modelado que da origen al problema matemático.
- Incertidumbre de los datos del problema.
- La codificación del modelo.
- Errores de redondeo.

# Error absoluto, relativo y precisión

Consideremos "A" el valor exacto de la medida de cierta magnitud (en general desconocida) y sea "a" un valor conocido que se llamará aproximación de "A". Evidentemente la buena cualidad de la aproximación es de acuerdo a cuan próximo está "a" de "A".

## Error Absoluto

Llamamos error absoluto del número aproximado "a" al valor:

$$\xi_a = |A - a|$$

y todo número $\xi_a^* \geq \xi_a$, se denominará cota del error absoluto.

# Error absoluto, relativo y precisión

## Error Relativo

Llamamos error relativo del número aproximado "a" al valor:

$$\delta_a = \frac{|A - a|}{|A|}, \quad A \neq 0$$

y todo número $\delta_a^* \geq \delta_a$, se denominará cota del error relativo.

## Precisión

Dado un $\epsilon > 0$ (pequeño) decimos que el valor "a" aproxima a "A" con una precisión $\epsilon$ si:

$$\xi_a = |A - a| \leq \epsilon$$

# Definiciones

Sean "A" y "a" dos números reales. Se dice que "a" es una aproximación de "A" con "n" cifras decimales exactas (o que "A" y "a" coinciden en "n" cifras decimales), si "n" es el mayor entero no negativo tal que

$$\xi_a \leq 0{,}5 \times 10^{-n}$$

Sean "A" y "a" dos números reales, con A$\neq 0$. Se dice que "a" es una aproximación de "A" con "n" cifras decimales significativas exactas (o que "A" y "a" coinciden en "n" cifras decimales significativas), si "n" es el mayor entero no negativo tal que

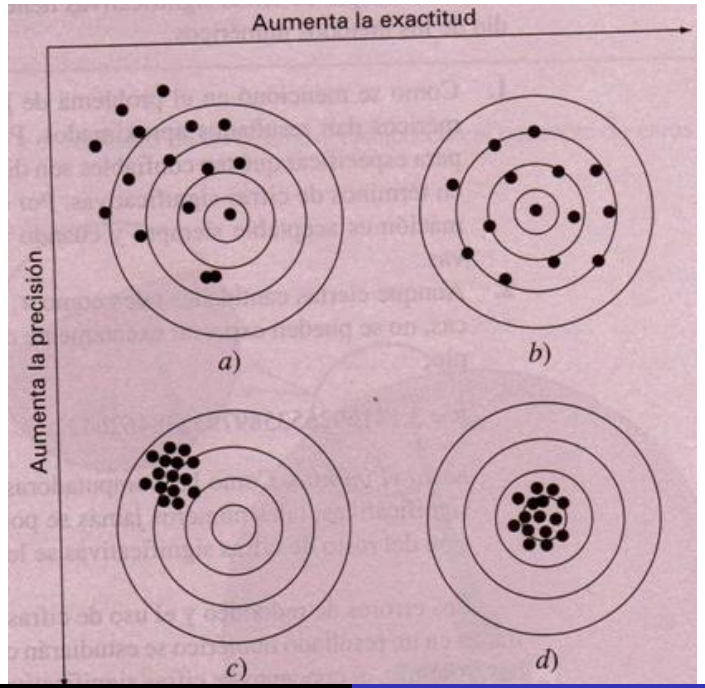
$$\delta_a \leq 5 \times 10^{-n}$$

# Exactitud vs Precisión

**Exactitud :** Se refiere a qué tan cercano está el valor calculado o medido del valor verdadero.

**Precisión:** Se refiere a qué tan cercanos se encuentran, uno de otros, diversos valores calculados o medidos.

a)Inexacto e impreciso b)Exacto e impreciso c)Inexacto y preciso d) Exacto y Preciso



# Propagación de errores

Al resolver un problema utilizando métodos numéricos, el error que se genera será consecuencia de un cúmulo de errores ocurridos en pasos sucesivos, se debe estudiar la mecánica de "propagación" de los mismos a lo largo del cálculo.

## Reglas de Propagación

a) Suma: Propaga los errores absolutos
$|\xi_{x+y}| \approx |\xi_x| + |\xi_y|$

b) Producto y cociente(Si $\delta_x$ y $\delta_y$ son pequeños): propaga los errores relativos
$|\delta_{x.y}| \approx |\delta_x| + |\delta_y|$
$|\delta_{x/y}| \approx |\delta_x| + |\delta_y|$

# Propagación de errores

c) Funciones de una variable: $y = f(x)$

$$\xi_y \approx |y'|\xi_x$$

d) Funciones de varias variables: $y = f(x_1, x_2, \ldots, x_n)$

$$\xi_y \approx \sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}|\xi_{x_i}$$

# Aritmetica del Computador

Las operaciones de suma, resta, multiplicación y división en el sistema de punto flotante (F), se denota por $\oplus, \ominus, \otimes, \oslash$ respectivamente. Estas operaciones están definidas por:

$x \oplus y = fl(fl(x) + fl(y))$

$x \ominus y = fl(fl(x) - fl(y))$

$x \otimes y = fl(fl(x) \times fl(y))$

$x \oslash y = fl(fl(x) \div fl(y)), \quad fl(y) \neq 0, \;\; y \neq 0$

Estas operaciones no son cerradas sobre F, pues en algunos casos se genera underflow u overflow;

# Representación de números del computador

Los computadores trabajan con aritmética real usando un sistema denominado de "punto flotante". Suponen un número real que tiene la expansión binaria:

**Número Normalizado**

**Notación científica en base 2**

Consiste en expresar los números no nulos en la forma:

$$\pm 1.d_1 d_2 \ldots . d_s \ldots . \bullet 2^e$$

mantisa — exponente

con: $d_i \in \{0,1\}$ y $e \in \mathbb{Z}$ (expresado también en base 2)

# Notación Normalizada

## *Aproximación de números reales en base 2 expresados en coma flotante normalizada*

Técnicas de aproximación de mantisas de la forma:

$$\pm 1.d_1 d_2 \ldots d_s d_{s+1} \ldots .$$

por mantisas con **s** dígitos decimales:

***Truncado:*** $\pm 1.d_1 d_2 \ldots d_s d_{s+1} \ldots . \approx \pm 1.d_1 d_2 \ldots d_s$

***Redondeo:***

$$\pm 1.d_1 d_2 \ldots d_s d_{s+1} \ldots . \approx \begin{cases} \pm 1.d_1 d_2 \ldots d_s & \text{si } d_{s+1} = 0 \\ \left\{ \pm (1.d_1 d_2 \ldots d_s + \right. \\ \left. +0.0\ 0\ \ldots 1\ ) \right\} & \text{si } d_{s+1} = 1 \end{cases}$$

*"Suma binaria"*

# Ejemplo

Redondeo a mantisas con 6 dígitos binarios decimales:

$z = 1.0110111010....$

7º Dígito decimal

$$\begin{array}{r} 0\ 0\ 0\ 0\ 1\ 1 \\ 1.\ 0\ 1\ 1\ 0\ 1\ 1 \\ +\ 0.\ 0\ 0\ 0\ 0\ 0\ 1 \\ \hline 1.\ 0\ 1\ 1\ 1\ 0\ 0 \end{array} = z^*$$

$z = 1.1111111010..... \ 2^e$

7º Dígito decimal

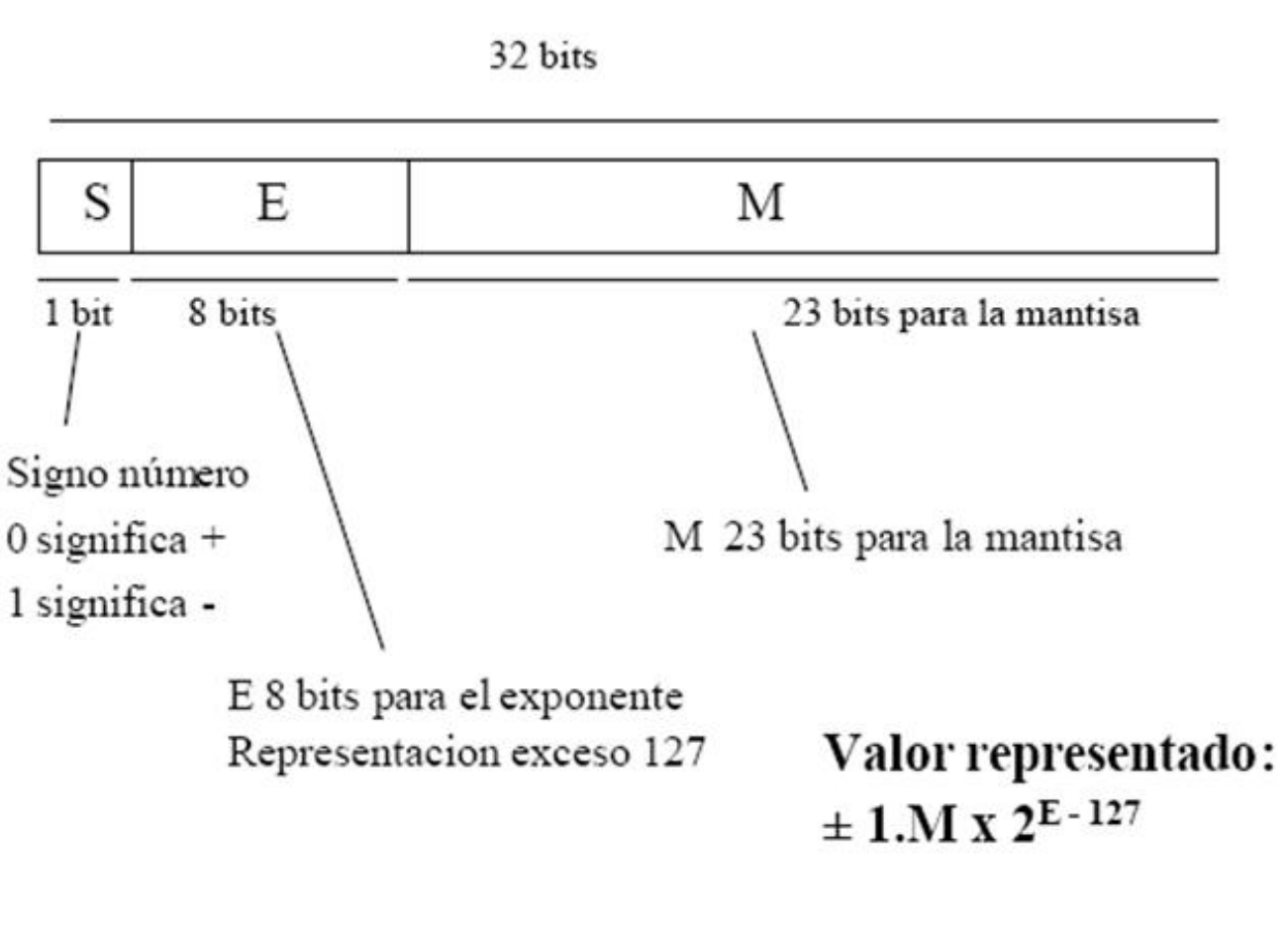
$$\begin{array}{r} 1\ 1\ 1\ 1\ 1\ 1\ 1 \\ 1.\ 1\ 1\ 1\ 1\ 1\ 1 \cdot 2^e \\ +\ 0.\ 0\ 0\ 0\ 0\ 0\ 1 \cdot 2^e \\ \hline 10.\ 0\ 0\ 0\ 0\ 0\ 0 \cdot 2^e \end{array}$$

Ajuste de exponentes:

$$1.\ 0\ 0\ 0\ 0\ 0\ 0 \cdot 2^{e+1} = z^*$$

# Estándar IEEE-754

## Precisión Simple: 32 bits



# Ejemplos:

**Ejemplo 1: ¿ Cuál es el valor decimal de: 1 01111100 11000000000000000000000?**

- ✓ El bit de signo es 1 : número negativo
- ✓ El exponente contiene 01111100 = 124
- ✓ La mantisa es 0.11000... = 0.75

El valor es : (- 1) x (1 + 0.75) x (2^(124-127)) = -1.75 x (2 ^(-3)) = - 0.21875

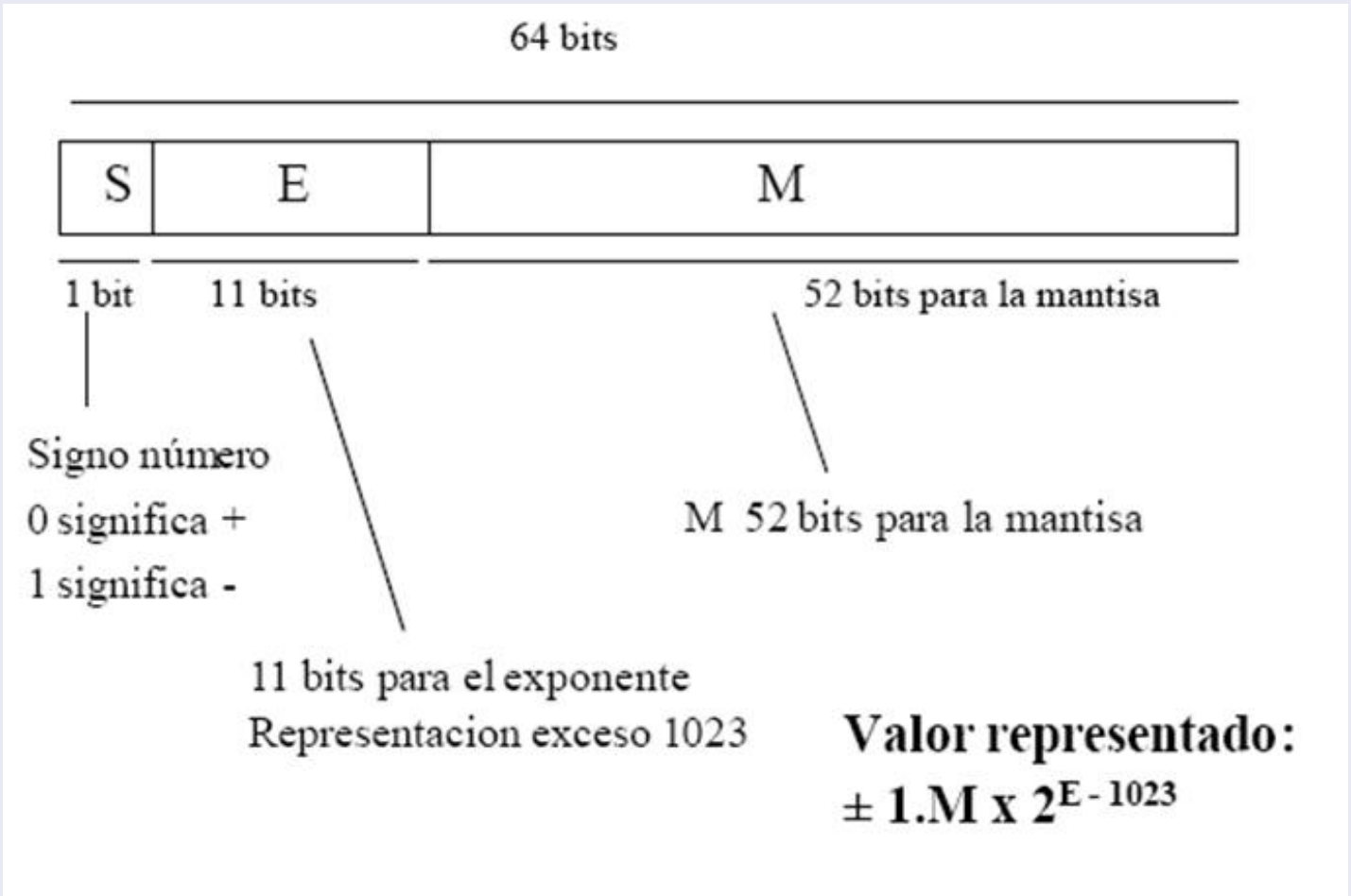
**Ejemplo 2: ¿ Cuál es la representación en simple precisión de : 347.625 ?**

1- Convertir a binario: 347.625 = 101011011.101
2- Normalizar el número (mover el punto decimal hasta que haya un solo 1 a la izquierda)
101011011.101 =1.01011011101 x (2 ^(8))
3- mantisa: 01011011101
4- exponente: 8 en exceso 127 = 8 + 127 = 135 = 10000111
5- El número es positivo: bit de signo 0

Resultado : 0 10000111 01011011101000000000000

# Estándar IEEE-754

## Precisión Doble: 64 bits



# Cambio de Base

- Base 2 para Base 10

a) $1011_2 = 1.2^3 + 0.2^2 + 1.2^1 + 1.2^0 = 11_{10}$

b) $10{,}1_2 = 1.2^1 + 0.2^0 + 1.2^{-1} = 2 + 0 + 0{,}5 = 2{,}5_{10}$

c) $11{,}01_2 = 1.2^1 + 1.2^0 + 0.2^{-1} + 1.2^{-2}$
$= 2 + 1 + 0{,}25$
$= 3{,}25_{10}$

d) $0{,}000111_2 = 0.2^{-1} + 0.2^{-2} + 0.2^{-3} + 1.2^{-4} + 1.2^{-5} + 1.2^{-6}$
$= 0 + 0 + 0 + 0{,}0625 + 0{,}3125 + 0{,}015625$
$= 0{,}109375_{10}$

## Error de representación de los números reales:

Debido a la imposibilidad de manejar infinitos decimales y a la necesidad de aproximar los números por otros con un número finito de cifras.

**Ejemplo:** $x = \frac{2}{3} = 0.666666.....\widehat{6}....$

$\hat{x} = 0.666666$ (Truncando a 6 decimales)

$\hat{x} = 0.666667$ (Redondeando a 6 decimales)

*NOTA: Se denominarán errores de redondeo*

# Ejemplo:

3°. Resolver, trabajando con 3 decimales, el sistema:

$$\begin{cases} 98 \bullet x + 293.97 \bullet y = -195.97 \\ \frac{2}{3} \bullet x + 2.01 \bullet y = \frac{2}{3} - 2.01 \end{cases}$$

**Solución:**

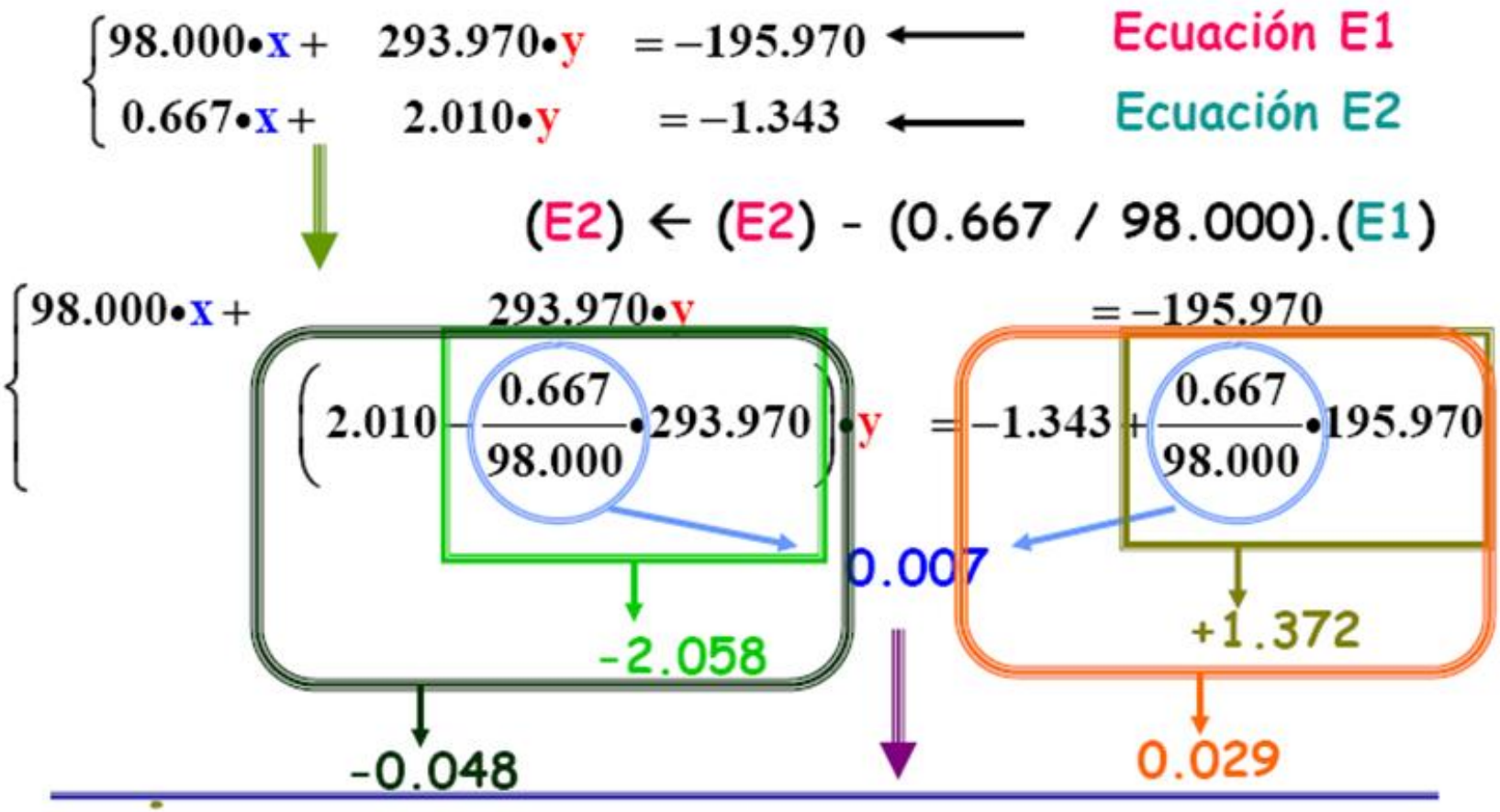
Las soluciones exactas del sistema exacto son $x = 1$ e $y = -1$.

Sistema aproximado (redondeando a la 3ª cifra decimal):

$$\begin{cases} 98.000 \bullet x + 293.970 \bullet y = -195.970 \\ 0.667 \bullet x + 2.010 \bullet y = -1.343 \end{cases}$$

Las soluciones exactas del sistema aproximado son $x = 1$ e $y = -1$.

.... PERO RESOLVÁMOSLO REDONDEANDO A 3 DECIMALES


98.000•x + 293.970•y = −195.970
Ecuación E1
0.667•x + 2.010•y = −1.343
Ecuación E2
(E2) ← (E2) - (0.667 / 98.000).(E1)
98.000•x +
293.970•y
= −195.970
2.010 − 0.667/98.000 •293.970 •y
= −1.343 + 0.667/98.000 •195.970
0.007
-2.058
+1.372
-0.048
0.029

Luego:

$$\begin{cases} 98.000 \bullet x + 293.970 \bullet y = -195.970 \\ -0.048 \bullet y = 0.029 \end{cases}$$

de donde:

$$y = \frac{0.029}{-0.048} = -0.604$$

$$x = \frac{-195.970 - 293.970 \bullet y}{98.000} = \frac{-195.970 + 177.558}{98.000} = \frac{-18.412}{98.000} = -0.188$$

... ¡ que no tienen nada que ver con las soluciones exactas !