Numerical Analysis – Linear Equations(II)

Hanyang University

Jong-Il Park

Singular Value Decomposition(SVD)

- Why SVD?
  - Gaussian Elim. and LU Decomposition fail to give satisfactory results for singular or numerically near singular matrices
  - SVD can cope with over- or under-determined problems
  - SVD constructs orthonormal basis vectors

What is SVD?

Any ***MxN*** matrix ***A*** whose number of rows ***M*** is greater than or equal to its number of columns ***N***, can be written as the product of a ***MxN*** column-orthogonal matrix ***U***, an ***NxN*** diagonal matrix ***W*** with positive or zero elements(the singular values), and the transpose of an ***NxN*** orthogonal matrix ***V***:

$$\boldsymbol{A = U\,W\,V^T}$$

$$\begin{pmatrix} \\ \\ \mathbf{A} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \\ \mathbf{U} \\ \\ \\ \end{pmatrix} \cdot \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \cdots & \\ & & \cdots & \\ & & & w_N \end{pmatrix} \cdot \begin{pmatrix} & \mathbf{V}^T & \end{pmatrix}$$

Properties of SVD

- Orthonormality
  - $U^T U = I \quad \Rightarrow \quad U^{-1} = U^T$
  - $V^T V = I \quad \Rightarrow \quad V^{-1} = V^T$
- Uniqueness
  - The decomposition can always be done, no matter how singular the matrix is, and it is almost unique.

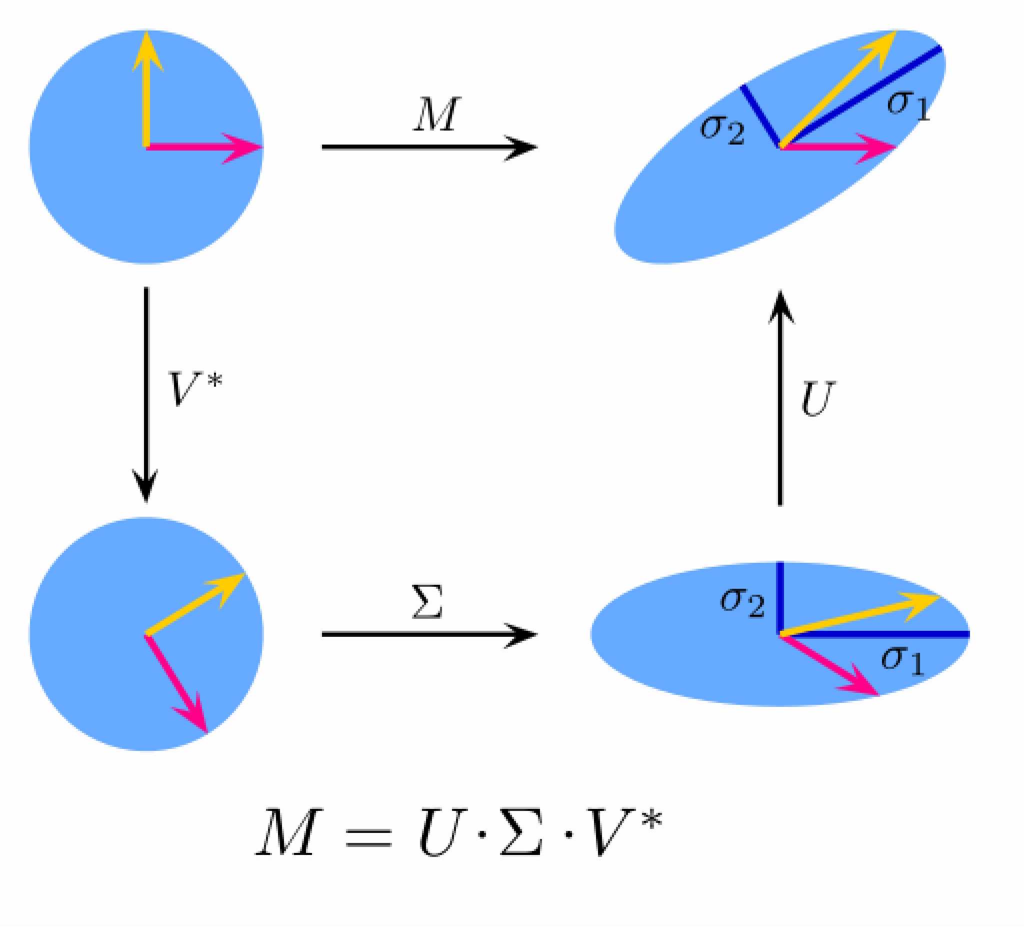
SVD of a square matrix

- SVD of a Square Matrix

$$\mathbf{A}^{-1} = \mathbf{V} \cdot [\mathrm{diag}\,(1/w_j)] \cdot \mathbf{U}^T$$

  - columns of ***U***
    - an orthonormal set of basis vectors
  - columns of ***V*** whose corresponding w_j's are zero
    - an orthonormal basis for the nullspace

Interpretation of SVD



Reminding basic concept in linear algebra

Important Concept in $Ax=b$ of order N

- Range: $\{y \mid y=Ax\}$ (subspace of b that can be reached by A)
- Rank: dimension of range
- Nullspace: $\{x \mid Ax=0\}$
- Nullity: dimension of nullspace
- N = (rank) + (nullity)

Homogeneous equation

- Homegeneous equations (***b***=0) + ***A*** is singular
  - Any column of ***V*** whose corresponding w_j is zero yields a solution

$$\underbrace{\begin{pmatrix} U \end{pmatrix}\begin{pmatrix} W \end{pmatrix}\begin{pmatrix} V^T \end{pmatrix}}_{A}\begin{pmatrix} V & \overset{x}{\Big\|} \end{pmatrix} = 0$$

Nonhomogeneous equation

- Nonhomegeneous eq. with singular ***A***

$$\mathbf{x} = \mathbf{V} \cdot [\mathrm{diag}\,(1/w_j)] \cdot (\mathbf{U}^T \cdot \mathbf{b})$$

where we replace $1/w_j$ by zero if w_j=0

※ The solution ***x*** obtained by this method is the solution vector of the smallest length.

※ If ***b*** is not in the range of ***A***

➔ **SVD find the solution *x* in the least-square sense**, i.e.

x which minimize ***r = |Ax-b|***

SVD solution - concept

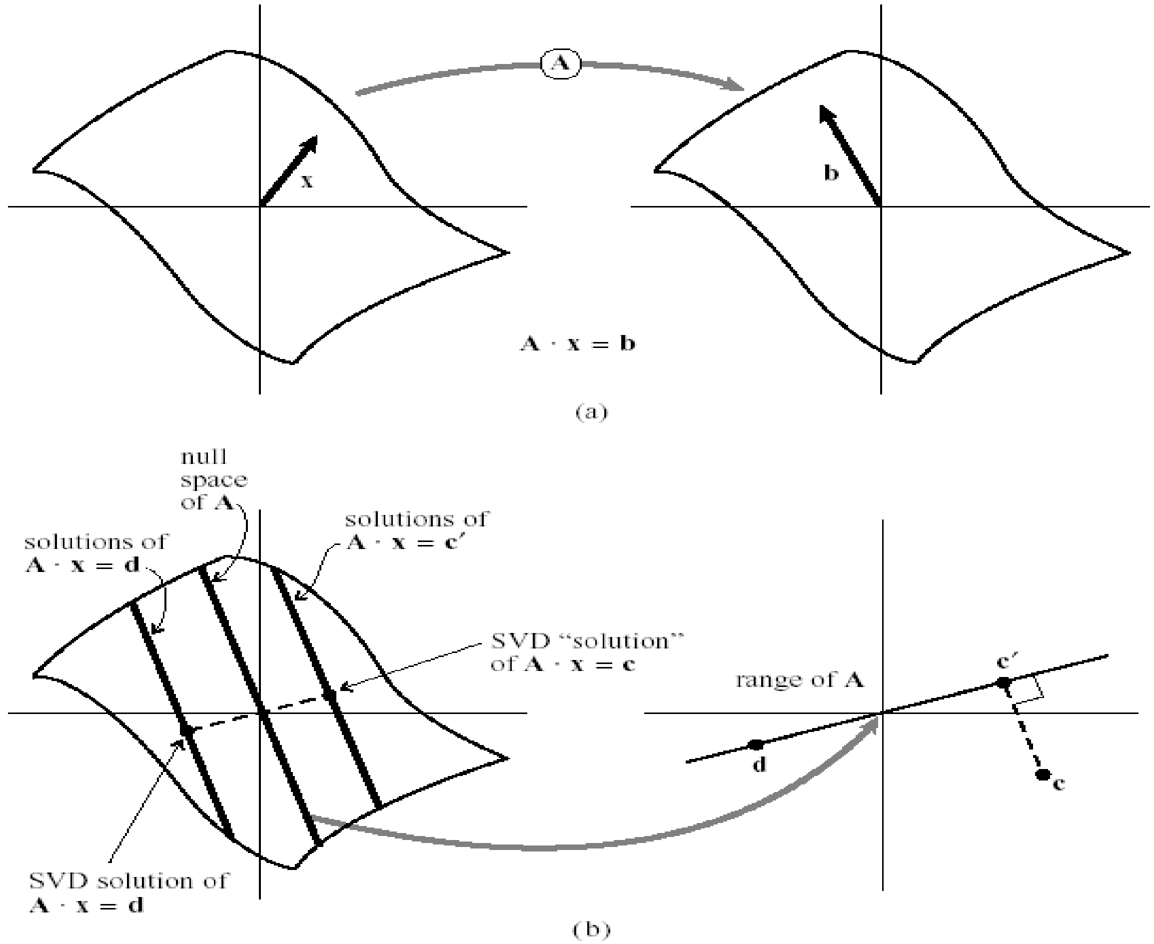
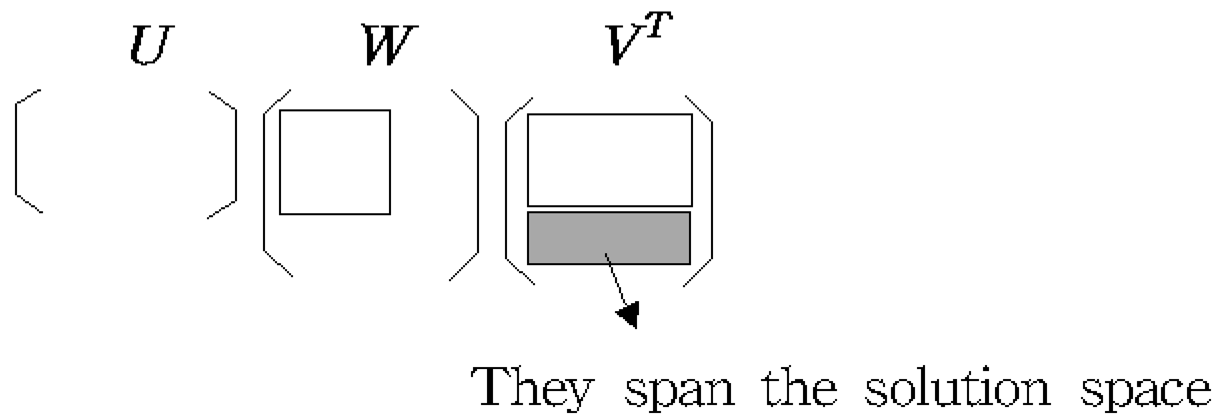


Figure 2.6.1. (a) A nonsingular matrix $\mathbf{A}$ maps a vector space into one of the same dimension. The vector $\mathbf{x}$ is mapped into $\mathbf{b}$, so that $\mathbf{x}$ satisfies the equation $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$. (b) A singular matrix $\mathbf{A}$ maps a vector space into one of lower dimensionality, here a plane into a line, called the "range" of $\mathbf{A}$. The "nullspace" of $\mathbf{A}$ is mapped to zero. The solutions of $\mathbf{A} \cdot \mathbf{x} = \mathbf{d}$ consist of any one particular solution plus any vector in the nullspace, here forming a line parallel to the nullspace. Singular value decomposition (SVD) selects the particular solution closest to zero, as shown. The point $\mathbf{c}$ lies outside of the range of $\mathbf{A}$, so $\mathbf{A} \cdot \mathbf{x} = \mathbf{c}$ has no solution. SVD finds the least-squares best compromise solution, namely a solution of $\mathbf{A} \cdot \mathbf{x} = \mathbf{c}'$, as shown.

SVD – under/over-determined problems

- SVD for Fewer Equations than Unknowns



- SVD for More Equations than Unknowns
  - SVD yields the least-square solution
  - In general, non-singular

Applications of SVD

- Applications

1. **Constructing an orthonormal basis**
   - *M*-dimensional vector space
   - Problem: Given N vectors,

     find an orthonormal basis
   - Solution:

     Columns of the matrix ***U*** are the desired orthonormal basis
2. **Approximation of Matrices**

$$A_{ij} = \sum_{k=1}^{N} w_k U_{ik} V_{jk}$$

Vector norm

A vector norm on $\mathcal{R}^n$ is a function, $\|\cdot\|$, from $\mathcal{R}^n$ into $\mathcal{R}$ with the following properties:

(i) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in \mathcal{R}^n$,

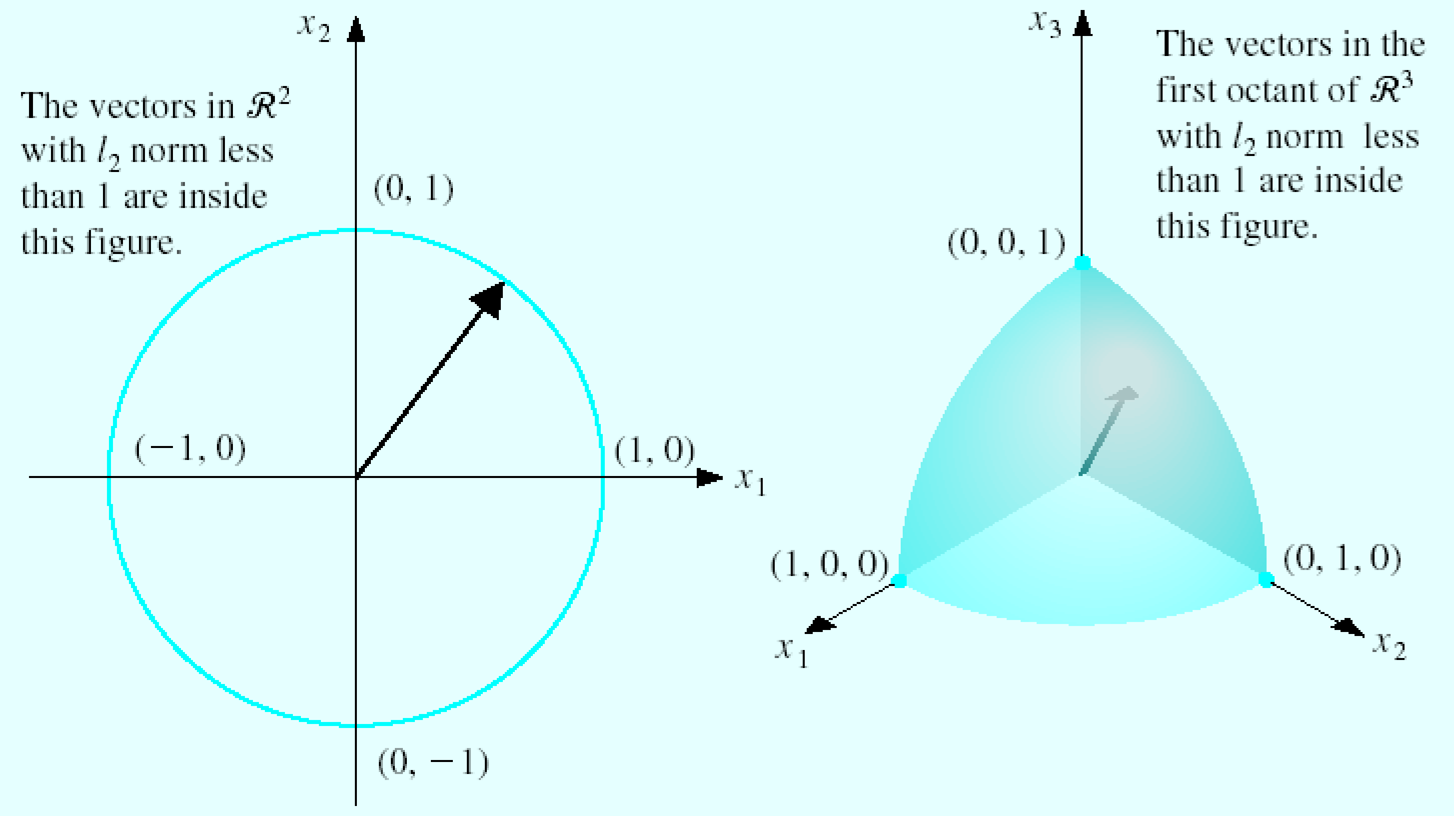
(ii) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = (0, 0, \ldots, 0)^t \equiv \mathbf{0}$,

(iii) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathcal{R}$ and $\mathbf{x} \in \mathcal{R}^n$,

(iv) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$.

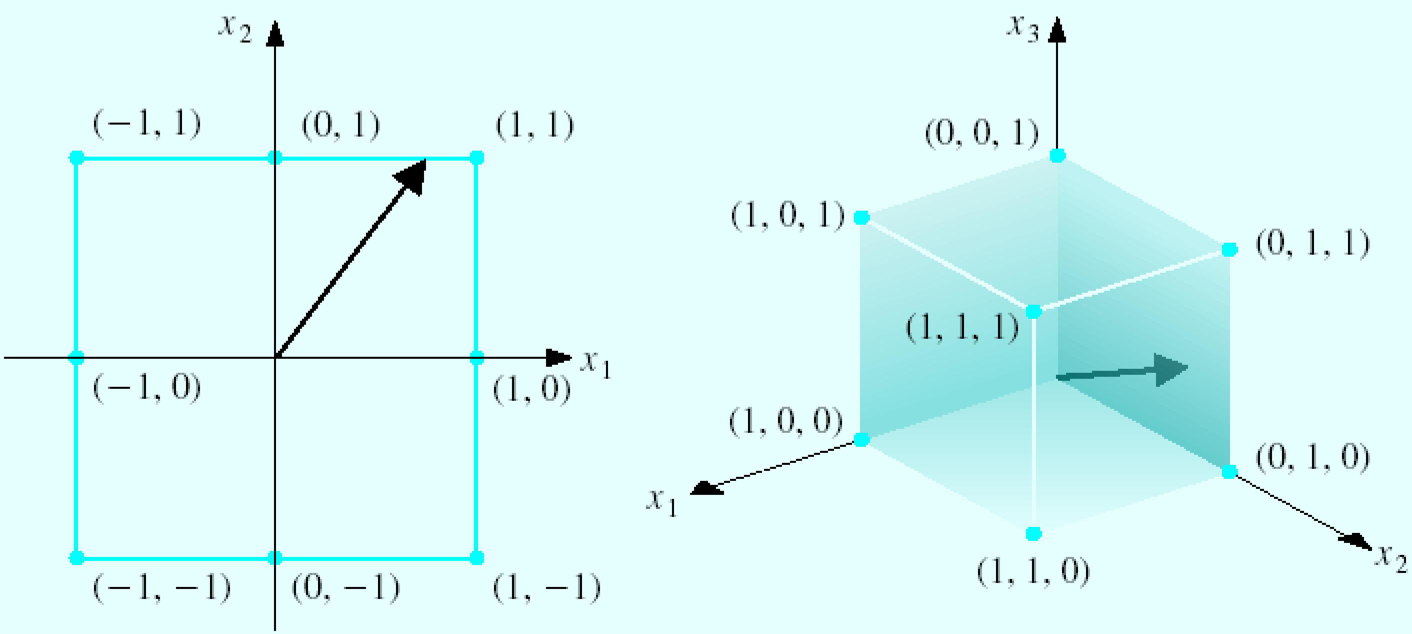
l_2 norm

$$\|\mathbf{x}\|_2 = \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{1/2}$$



l_∞ norm

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|.$$



The vectors in $\mathcal{R}^2$ with l_∞ norm less than 1 are inside this figure.

The vectors in the first octant of $\mathcal{R}^3$ with l_∞ norm less than 1 are inside this figure.

Distance between vectors

If $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ are vectors in $\mathcal{R}^n$, the l_2 and l_∞ distances between $\mathbf{x}$ and $\mathbf{y}$ are defined by

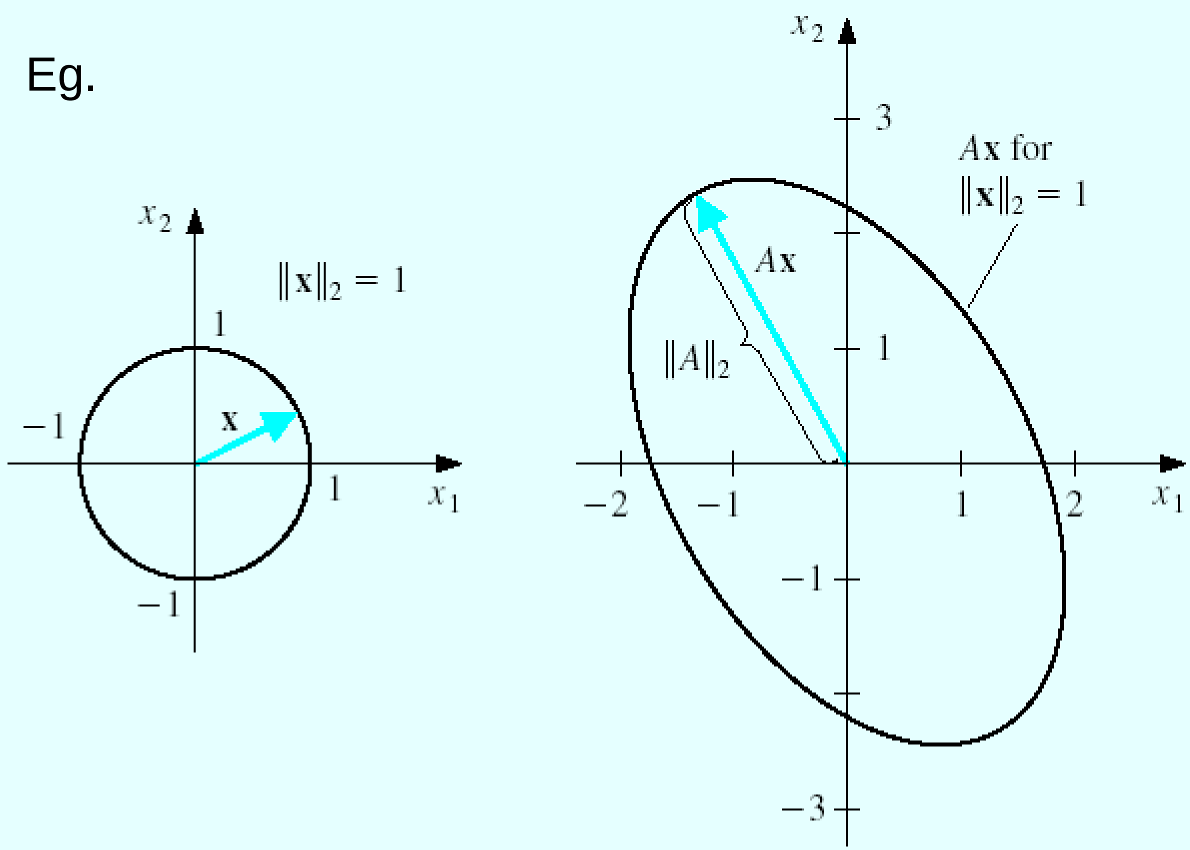
$$\|\mathbf{x} - \mathbf{y}\|_2 = \left\{ \sum_{i=1}^{n} (x_i - y_i)^2 \right\}^{1/2} \quad \text{and} \quad \|\mathbf{x} - \mathbf{y}\|_\infty = \max_{1 \le i \le n} |x_i - y_i|.$$

Natural matrix norm

If $\| \cdot \|$ is a vector norm on $\mathcal{R}^n$, the natural matrix norm on the set of $n \times n$ matrices given by $\| \cdot \|$ is defined by

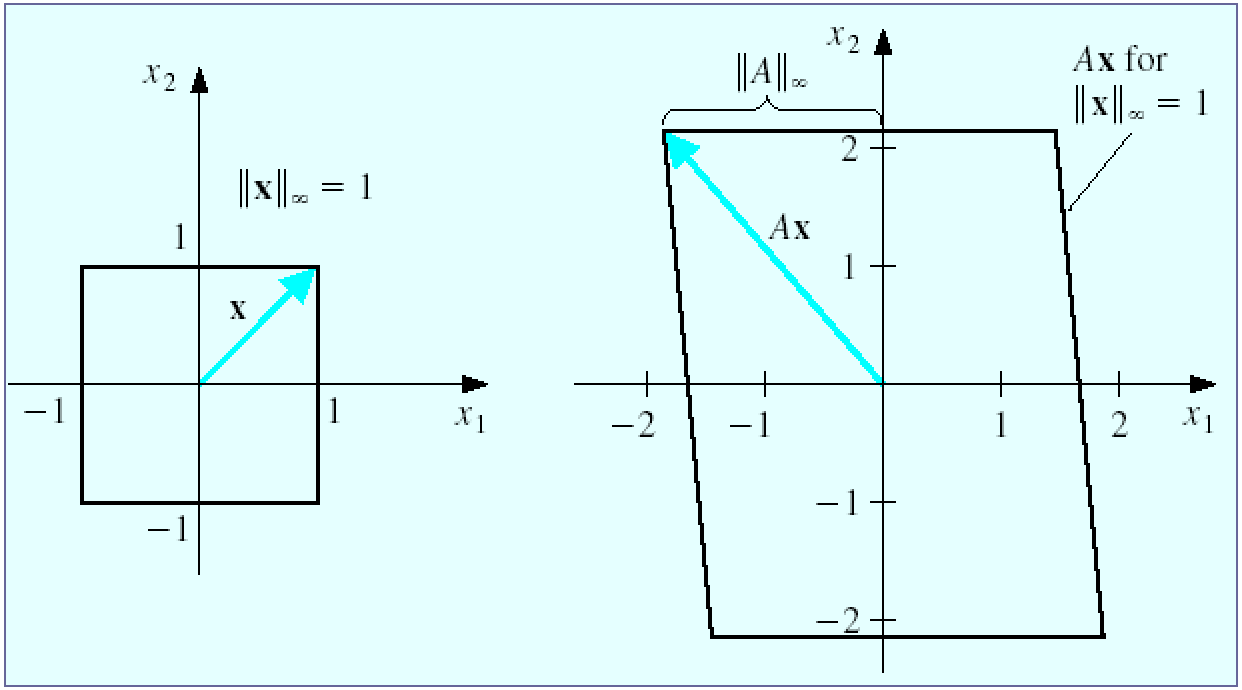
$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

Eg.

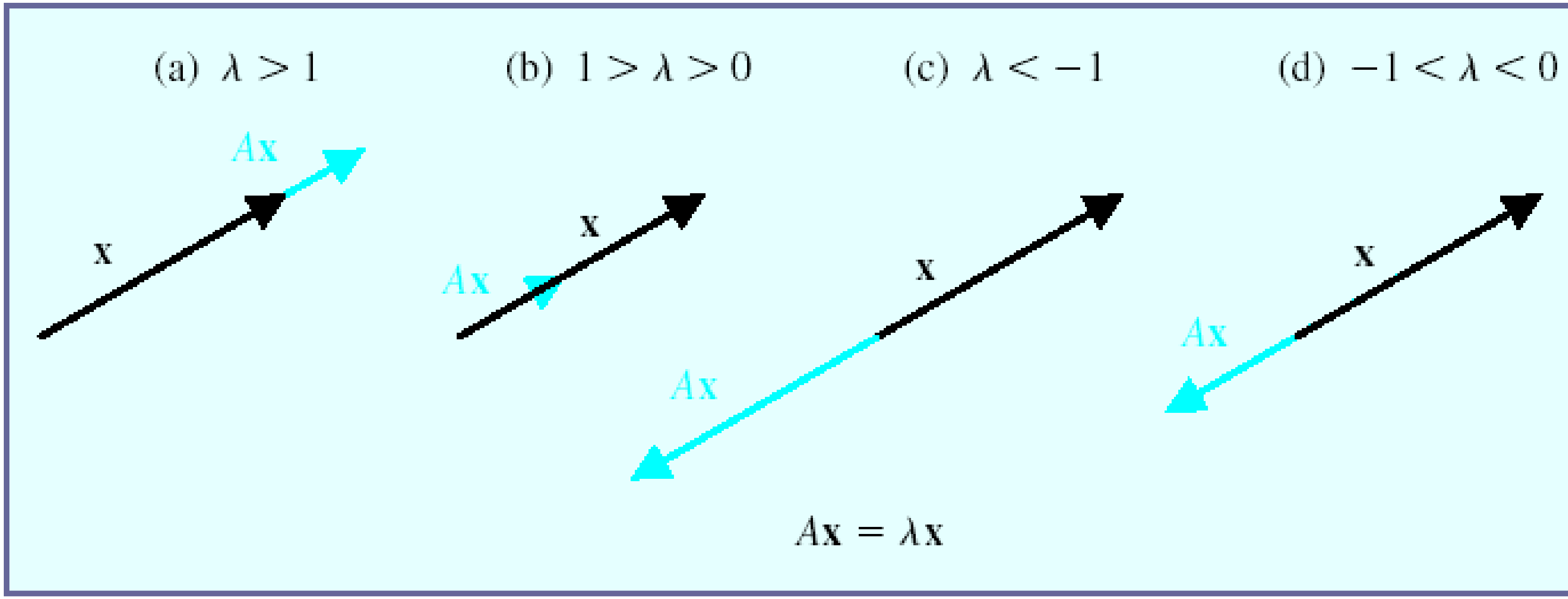


l_∞ norm of a matrix

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|.$$



Eigenvalues and eigenvectors



* To be discussed later in detail.

Spectral radius

$$\boxed{\rho(A) = \max |\lambda|,}$$

If A is an $n \times n$ matrix, then

(i) $\|A\|_2 = [\rho(A^t A)]^{1/2}$;

(ii) $\rho(A) \leq \|A\|$ for any natural norm.

Convergent matrix equivalences

The following are equivalent statements:

(i) A is a convergent matrix.

(ii) $\lim_{n\to\infty} \|A^n\| = 0$, for some natural norm.

(iii) $\lim_{n\to\infty} \|A^n\| = 0$, for all natural norms.

(iv) $\rho(A) < 1$.

(v) $\lim_{n\to\infty} A^n\mathbf{x} = \mathbf{0}$, for every $\mathbf{x}$.

Convergence of a sequence

- An important connection between the eigen values of the matrix T and the expectation that the iterative method will converge

  ➔ spectral radius

The sequence

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$$

converges to the unique solution of $\mathbf{x} = T\mathbf{x} + \mathbf{c}$ for any $\mathbf{x}^{(0)}$ in $\mathcal{R}^n$ if and only if $\rho(T) < 1$.

Iterative Methods - Jacobi Iteration

$$\boldsymbol{Ax}=\boldsymbol{b} \quad => \quad \sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (i=1,2,\cdots,n)$$

If $a_{ii} \neq 0$,

$$x_i = \frac{1}{a_{ii}}\left\{b_i - \left(\sum_{j=1}^{i-1} a_{ij}x_j + \sum_{j=i+1}^{n} a_{ij}x_j\right)\right\}$$

- Jacobi Iteration

$$\begin{aligned}
x_i^{(k)} &= \frac{1}{a_{ii}}\left\{b_i - \left(\sum_{j=1}^{i-1} a_{ij}x_j^{(k-1)} + \sum_{j=i+1}^{n} a_{ij}x_j^{(k-1)}\right)\right\} \\
&= x_i^{(k-1)} + \frac{1}{a_{ii}}\left\{b_i - \sum_{j=1}^{n} a_{ij}x_j^{(k-1)}\right\} \\
&= x_i^{(k-1)} + \Delta x_i^{(k-1)}
\end{aligned}$$

Jacobi Iteration

- Initial guess

$$\boldsymbol{x}^{(0)} = [x_1^{(0)},\ \ x_2^{(0)}, \cdots, x_n^{(0)}]$$

- Convergence Condition

The Jacobi iteration is convergent, irrespective of an initial guess, if the matrix ***A*** is diagonal-dominant:

$$|a_{ii}| \geq \sum_{j=1,\, j\neq i}^{n} |a_{ij}|$$

Eg. Jacobi iteration

The linear system $A\mathbf{x} = \mathbf{b}$ given by

$$\begin{aligned}
E_1:&\quad 10x_1 - x_2 + 2x_3 = 6,\\
E_2:&\quad -x_1 + 11x_2 - x_3 + 3x_4 = 25,\\
E_3:&\quad 2x_1 - x_2 + 10x_3 - x_4 = -11,\\
E_4:&\quad 3x_2 - x_3 + 8x_4 = 15
\end{aligned}$$

has solution $\mathbf{x} = (1, 2, -1, 1)^t$. To convert $A\mathbf{x} = \mathbf{b}$ to the form $\mathbf{x} = T\mathbf{x} + \mathbf{c}$, solve equation E_i for x_i obtaining

$$\begin{aligned}
x_1 &= \frac{1}{10}x_2 - \frac{1}{5}x_3 + \frac{3}{5},\\
x_2 &= \frac{1}{11}x_1 + \frac{1}{11}x_3 - \frac{3}{11}x_4 + \frac{25}{11},\\
x_3 &= -\frac{1}{5}x_1 + \frac{1}{10}x_2 + \frac{1}{10}x_4 - \frac{11}{10},\\
x_4 &= -\frac{3}{8}x_2 + \frac{1}{8}x_3 + \frac{15}{8}.
\end{aligned}$$

Then $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{x} = T\mathbf{x} + \mathbf{c}$, with

$$T = \begin{bmatrix} 0 & \frac{1}{10} & -\frac{1}{5} & 0 \\ \frac{1}{11} & 0 & \frac{1}{11} & -\frac{3}{11} \\ -\frac{1}{5} & \frac{1}{10} & 0 & \frac{1}{10} \\ 0 & -\frac{3}{8} & \frac{1}{8} & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \frac{3}{5} \\ \frac{25}{11} \\ -\frac{11}{10} \\ \frac{15}{8} \end{bmatrix}.$$

For an initial approximation, suppose $\mathbf{x}^{(0)} = (0, 0, 0, 0)^t$. Then $\mathbf{x}^{(1)}$ is given by

$$\begin{aligned}
x_1^{(1)} &= \phantom{-\frac{1}{5}x_1^{(0)} +} \frac{1}{10}x_2^{(0)} - \frac{1}{5}x_3^{(0)} \phantom{+ \frac{1}{10}x_4^{(0)}} + \frac{3}{5} = 0.6000, \\
x_2^{(1)} &= \frac{1}{11}x_1^{(0)} \phantom{+ \frac{1}{10}x_2^{(0)}} + \frac{1}{11}x_3^{(0)} - \frac{3}{11}x_4^{(0)} + \frac{25}{11} = 2.2727, \\
x_3^{(1)} &= -\frac{1}{5}x_1^{(0)} + \frac{1}{10}x_2^{(0)} \phantom{+ \frac{1}{11}x_3^{(0)}} + \frac{1}{10}x_4^{(0)} - \frac{11}{10} = -1.1000, \\
x_4^{(1)} &= \phantom{-\frac{1}{5}x_1^{(0)}} - \frac{3}{8}x_2^{(0)} + \frac{1}{8}x_3^{(0)} \phantom{+ \frac{1}{10}x_4^{(0)}} + \frac{15}{8} = 1.8750.
\end{aligned}$$

Additional iterates, $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)})^t$, are generated in a similar manner and are presented in Table 7.1. The decision to stop after 10 iterations was based on the criterion

$$\|\mathbf{x}^{(10)} - \mathbf{x}^{(9)}\|_\infty = 8.0 \times 10^{-4} < 10^{-3}.$$

Since we know that $\mathbf{x} = (1, 2, -1, 1)^t$, we have $\|\mathbf{x}^{(10)} - \mathbf{x}\|_\infty \approx 0.0002$. ■

Gauss-Seidel Iteration

- Idea
  - Utilize recently updated x_i
- Iteration formula

$$
\begin{aligned}
x_i^{(k)} &= \frac{1}{a_{ii}}\left\{b_i - \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k)} + \sum_{j=i+1}^{n} a_{ij} x_j^{(k-1)}\right)\right\} \\
&= x_i^{(k-1)} + \frac{1}{a_{ii}}\left\{b_i - \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k)} + \sum_{j=i}^{n} a_{ij} x_j^{(k-1)}\right)\right\} \\
&= x_i^{(k-1)} + \Delta x_i^{(k-1)}
\end{aligned}
$$

- Convergence Condition
  - The same as the Jacobi iteration
- Advantage over Jacobi iteration
  - Fast convergence

Eg. Gauss-Seidel iteration

In Example 1 we used the Jacobi method to solve the linear system

$$\begin{array}{rcrcrcrcl} 10x_1 & - & x_2 & + & 2x_3 & & & = & 6, \\ -x_1 & + & 11x_2 & - & x_3 & + & 3x_4 & = & 25, \\ 2x_1 & - & x_2 & + & 10x_3 & - & x_4 & = & -11, \\ & & 3x_2 & - & x_3 & + & 8x_4 & = & 15. \end{array}$$

Using the Gauss-Seidel method as described in Eq. (7.2) gives the equations

$$\begin{aligned} x_1^{(k)} &= \frac{1}{10}x_2^{(k-1)} - \frac{1}{5}x_3^{(k-1)} + \frac{3}{5}, \\ x_2^{(k)} &= \frac{1}{11}x_1^{(k)} + \frac{1}{11}x_3^{(k-1)} - \frac{3}{11}x_4^{(k-1)} + \frac{25}{11}, \\ x_3^{(k)} &= -\frac{1}{5}x_1^{(k)} + \frac{1}{10}x_2^{(k)} + \frac{1}{10}x_4^{(k-1)} - \frac{11}{10}, \\ x_4^{(k)} &= -\frac{3}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} + \frac{15}{8}. \end{aligned}$$

Letting $\mathbf{x}^{(0)} = (0, 0, 0, 0)^t$, we generate the Gauss-Seidel iterates in Table 7.2. Since

$$\|\mathbf{x}^{(5)} - \mathbf{x}^{(4)}\|_\infty = 0.0008 < 10^{-3},$$

$\mathbf{x}^{(5)}$ is accepted as a reasonable approximation to the solution. Note that Jacobi's method in Example 1 required twice as many iterations for the same accuracy. ■

Jacobi vs. Gauss-Seidel

- Comparison: Eg. 1 vs. Eg. 2

Eg. 1 Jacobi

k	0	1	2	3	4	5	6	7	8	9	10
$x_1^{(k)}$	0.000	0.6000	1.0473	0.9326	1.0152	0.9890	1.0032	0.9981	1.0006	0.9997	1.0001
$x_2^{(k)}$	0.0000	2.2727	1.7159	2.053	1.9537	2.0114	1.9922	2.0023	1.9987	2.0004	1.9998
$x_3^{(k)}$	0.0000	−1.1000	−0.8052	−1.0493	−0.9681	−1.0103	−0.9945	−1.0020	−0.9990	−1.0004	−0.9998
$x_4^{(k)}$	0.0000	1.8750	0.8852	1.1309	0.9739	1.0214	0.9944	1.0036	0.9989	1.0006	0.9998

Faster convergence

Eg. 2 Gauss-Seidel

k	0	1	2	3	4	5
$x_1^{(k)}$	0.0000	0.6000	1.030	1.0065	1.0009	1.0001
$x_2^{(k)}$	0.0000	2.3272	2.037	2.0036	2.0003	2.0000
$x_3^{(k)}$	0.0000	−0.9873	−1.014	−1.0025	−1.0003	−1.0000
$x_4^{(k)}$	0.0000	0.8789	0.9844	0.9983	0.9999	1.0000

Variation of Gauss-Seidel Iteration

$$\boxed{x_i^{(k)} = x_i^{(k-1)} + w \, \Delta x_i^{(k-1)}}$$

- Successive Over Relaxation(SOR)
  - $1 < w < 2$
  - fast convergence
  - Well-suited for linear problem
- Successive Under Relaxation(SUR)
  - $0 < w < 1$
  - slow convergence
  - stable
  - Well-suited for nonlinear problem

Eg. Gauss-Seidel vs. SOR

The linear system $A\mathbf{x} = \mathbf{b}$ given by

$$\begin{aligned} 4x_1 + 3x_2 \phantom{{}-x_3} &= 24, \\ 3x_1 + 4x_2 - x_3 &= 30, \\ -x_2 + 4x_3 &= -24 \end{aligned}$$

has the solution $(3, 4, -5)^t$. The Gauss-Seidel method and the SOR method with $\omega = 1.25$ will be used to solve this system, using $\mathbf{x}^{(0)} = (1, 1, 1)^t$ for both methods. For each $k = 1, 2, \ldots,$ the equations for the Gauss-Seidel method are

$$\begin{aligned} x_1^{(k)} &= -0.75x_2^{(k-1)} + 6, \\ x_2^{(k)} &= -0.75x_1^{(k)} + 0.25x_3^{(k-1)} + 7.5, \\ x_3^{(k)} &= 0.25x_2^{(k)} - 6, \end{aligned}$$

and the equations for the SOR method with $\omega = 1.25$ are

$$\begin{aligned} x_1^{(k)} &= -0.25x_1^{(k-1)} - 0.9375x_2^{(k-1)} + 7.5, \\ x_2^{(k)} &= -0.9375x_1^{(k)} - 0.25x_2^{(k-1)} + 0.3125x_3^{(k-1)} + 9.375, \\ x_3^{(k)} &= 0.3125x_2^{(k)} - 0.25x_3^{(k-1)} - 7.5. \end{aligned}$$

The first seven iterates for each method are listed in Tables 7.3 and 7.4. To be accurate to seven decimal places, the Gauss-Seidel method required 34 iterations, as opposed to only 14 iterations for the SOR method with $\omega = 1.25$. ■

(cont.)

Table 7.3 Gauss-Seidel

k	0	1	2	3	4	5	6	7
$x_1^{(k)}$	1	5.250000	3.1406250	3.0878906	3.0549316	3.0343323	3.0214577	3.0134110
$x_1^{(k)}$	1	3.812500	3.8828125	3.9267578	3.9542236	3.9713898	3.9821186	3.9888241
$x_1^{(k)}$	1	−5.046875	−5.0292969	−5.0183105	−5.0114441	−5.0071526	−5.0044703	−5.0027940

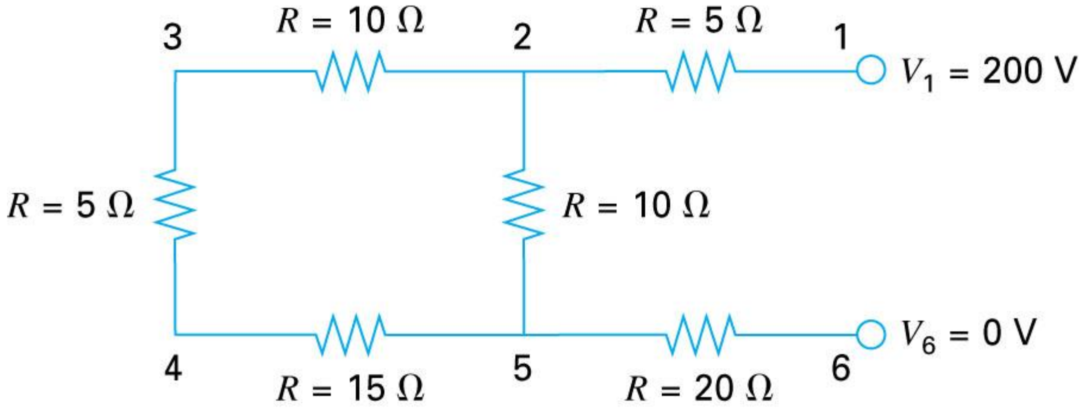
Faster convergence

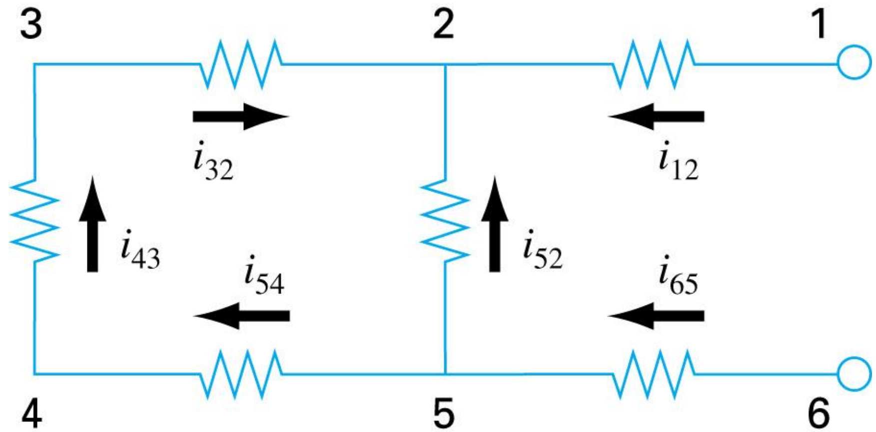
Table 7.4 SOR with $\omega = 1.25$

k	0	1	2	3	4	5	6	7
$x_1^{(k)}$	1	6.312500	2.6223145	3.1333027	2.9570512	3.0037211	2.9963276	3.0000498
$x_2^{(k)}$	1	3.5195313	3.9585266	4.0102646	4.0074838	4.0029250	4.0009262	4.0002586
$x_3^{(k)}$	1	−6.6501465	−4.6004238	−5.0966863	−4.9734897	−5.0057135	−4.9982822	−5.0003486

Application: Circuit analysis

- Kirchhoff's current and voltage law





Current rule: 4 nodes
Voltage rule: 2 meshes

6 unknowns, 6 equations

Given these assumptions, Kirchhoff's current rule is applied at each node to yield

$$i_{12} + i_{52} + i_{32} = 0$$
$$i_{65} - i_{52} - i_{54} = 0$$
$$i_{43} - i_{32} = 0$$
$$i_{54} - i_{43} = 0$$

Application of the voltage rule to each of the two loops gives

$$-i_{54}R_{54} - i_{43}R_{43} - i_{32}R_{32} + i_{52}R_{52} = 0$$
$$-i_{65}R_{65} - i_{52}R_{52} + i_{12}R_{12} - 200 = 0$$

or, substituting the resistances from Fig. 12.8 and bringing constants to the right-hand side,

$$-15i_{54} - 5i_{43} - 10i_{32} + 10i_{52} = 0$$
$$-20i_{65} - 10i_{52} + 5i_{12} = 200$$

Therefore, the problem amounts to solving the following set of six equations with six unknown currents:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 10 & -10 & 0 & -15 & -5 \\ 5 & -10 & 0 & -20 & 0 & 0 \end{bmatrix} \begin{Bmatrix} i_{12} \\ i_{52} \\ i_{32} \\ i_{65} \\ i_{54} \\ i_{43} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 200 \end{Bmatrix}$$