Optimization with Big Data:
Network Flows

Sertac Karaman
Assistant Professor of Aeronautics and Astronautics
Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

Massachusetts Institute of Technology

MIT AEROASTRO

Recall Linear Programming

- Maximize (or minimize) an objective function with respect to constraints

$$\text{Maximize} \quad \sum_{j=1}^{n} c_j x_j$$

← objective function

$$\text{subject to:} \quad \sum_{j=1}^{n} a_{ij} x_j \le b_i \quad \text{for} \quad i = 1, 2, \ldots, m$$

$$x_j \ge 0 \quad \text{for} \quad j = 1, 2, \ldots, n$$

← constraints

Linear Programming Formulation

LP as a “Decision Making” Problem:

Maximize $Z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$

Subject to:

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \leq b_1$$

$$a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \leq b_2$$

...

$$a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n \leq b_m$$

and $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$

Value maximization subject to resource constraints

x_j Decision variable
- *we would like to determine (decide) this*

Given by the problem instance:

c_j The unit value for the *j*th decision variable

b_i The available amount for the *i*th resource

a_{ij} The unit of *i*th resource required for one unit of the *j*th decision variable

Linear Programming Formulation

Maximize $Z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$

Subject to:

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \leq b_1$$

$$a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \leq b_2$$

...

$$a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n \leq b_m$$

and $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$

- **Feasible Solution:** A set of values for the decision variables ($x_1, \ldots, x_n$) that satisfies all of the constraints.
- **Optimal Solution:** A feasible solution that gives the best value ($\sum_{j=1}^{n} c_i x_i$) among all feasible solutions.

MIT AEROASTRO

Linear Programming Formulation

Maximize $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

Subject to:

$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1$

$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2$

...

$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m$

and $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$

Minimize $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

Subject to:

$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1$

$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2$

...

$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m$

and $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$

- The problem can be cast either as a maximization or a minimization problem. (Just multiply all the “value” constants c_i by -1.)

MIT AEROASTRO

Linear Programming Formulation

Cost minimization subject to resource constraints

x_j Decision variable

Given by the problem instance:

c_j The unit **cost** for the *j*th decision variable

b_i The available amount for the *i*th resource

a_{ij} The unit of *i*th resource required for one unit of the *j*th decision variable

Minimize $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

Subject to:

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1$$

$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2$$

...

$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m$$

and $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$

- The problem can be cast either as a maximization or a minimization problem. (Just multiply all the “value” constants c_i by -1.)
- In the new problem, c_i can be interpreted as cost.

LP Example: Maximizing Capacity with Constrained Crew and Vehicle Supply

- Freight needs to be carried with trucks.
- Two types of trucks are available in limited numbers, each with different capacity and crew requirements:

	Capacity	Crew required	Number available
Anadol	300	3	40
BMC	500	2	60

- We have a limited amount of crew: **Exactly 180 number of personnel available.** All personnel can operate either truck.

- ***How many trucks of each type would you utilize?***

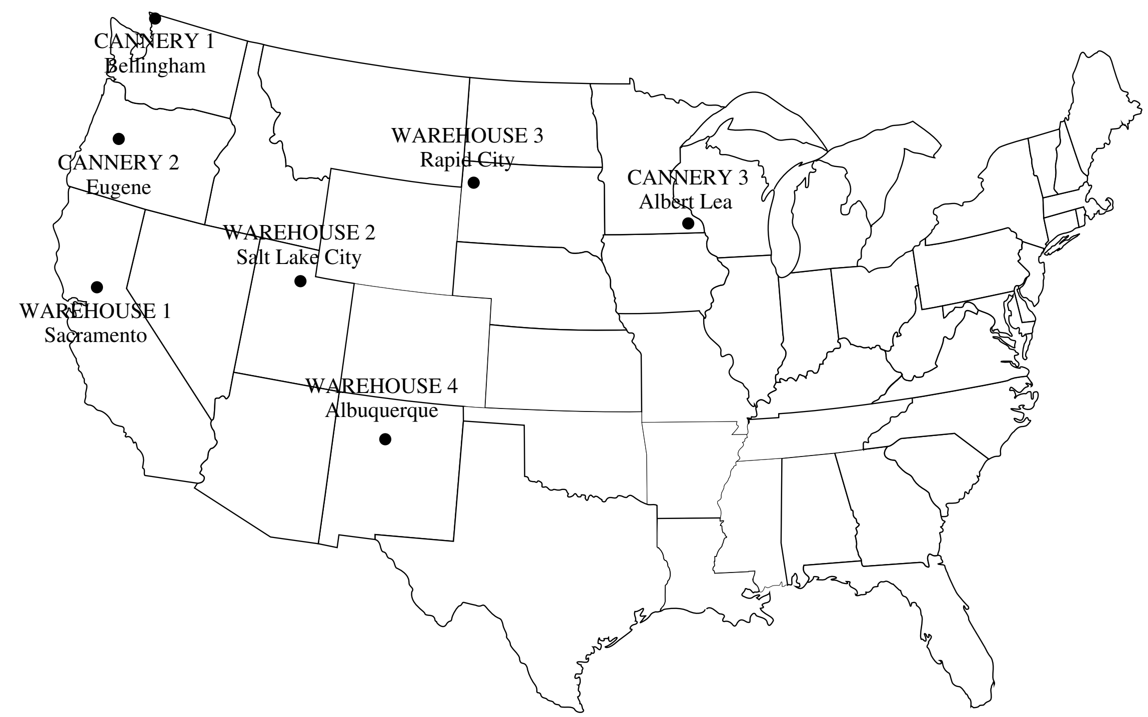
MIT AEROASTRO

Some Important Linear Programming Problems

- Transportation problem
- Assignment problem
- Maximum flow problem on a network
- Minimum-cost flow problem on a network

The Transportation Problem

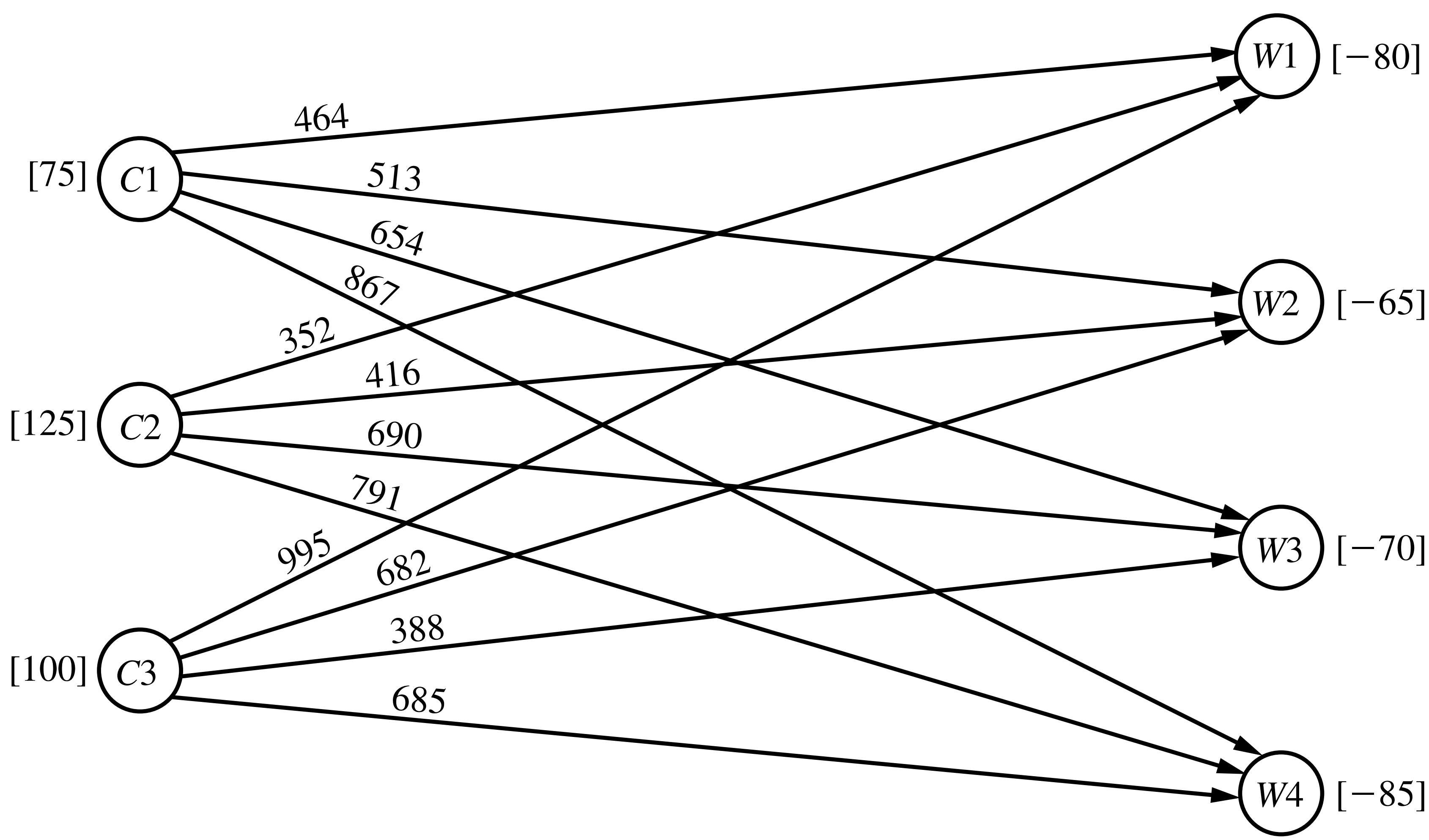
- Need to ship goods peas from canneries to warehouses.
- How can you formulate a linear program for this shipment problem?



	Shipping Cost ($) per Truckload				
	Warehouse				
	1	2	3	4	Output
Cannery 1	464	513	654	867	75
2	352	416	690	791	125
3	995	682	388	685	100
Allocation	80	65	70	85	

The Transportation Problem

- Consider the network formulation:



The Transportation Problem

- Define the decision variables:
  x_{ij} is the amount shipped from cannery i to warehouse j,
  where i = 1, 2, 3 and j = 1, 2, 3, 4.

- Formulate the objective function:

$$
\begin{aligned}
\text{Minimize} \quad Z = {} & 464x_{11} + 513x_{12} + 654x_{13} + 867x_{14} + 352x_{21} + 416x_{22} \\
& + 690x_{23} + 791x_{24} + 995x_{31} + 682x_{32} + 388x_{33} + 685x_{34},
\end{aligned}
$$

- Formulate the constraints:

subject to the constraints

$$
\begin{array}{lllllllllllllr}
x_{11} + x_{12} + x_{13} + x_{14} & & & = & 75 \\
& x_{21} + x_{22} + x_{23} + x_{24} & & = & 125 \\
& & x_{31} + x_{32} + x_{33} + x_{34} & = & 100 \\
x_{11} & + x_{21} & + x_{31} & = & 80 \\
\quad x_{12} & \quad + x_{22} & \quad + x_{32} & = & 65 \\
\qquad x_{13} & \qquad + x_{23} & \qquad + x_{33} & = & 70 \\
\qquad\quad x_{14} & \qquad\quad + x_{24} & \qquad\quad + x_{34} & = & 85
\end{array}
$$

and

$$x_{ij} \geq 0 \qquad (i = 1, 2, 3;\ j = 1, 2, 3, 4).$$

MIT AEROASTRO

The General Transportation Problem

- Several similar problems are in fact "source-to-destination transportation problem."

Prototype Example	General Problem
Truckloads of canned peas	Units of a commodity
Three canneries	m sources
Four warehouses	n destinations
Output from cannery i	Supply s_i from source i
Allocation to warehouse j	Demand d_j at destination j
Shipping cost per truckload from cannery i to warehouse j	Cost c_{ij} per unit distributed from source i to destination j

MIT AEROASTRO

The General Transportation Problem

- The general transportation problem is formulated as a linear program as follows:

General Problem
Units of a commodity
m sources
n destinations
Supply s_i from source i
Demand d_j at destination j
Cost c_{ij} per unit distributed from source i to destination j

		Cost per Unit Distributed				
		Destination				
		1	2	⋯	n	Supply
Source	1	c_{11}	c_{12}	⋯	c_{1n}	s_1
	2	c_{21}	c_{22}	⋯	c_{2n}	s_2
	⋮	⋯	⋯	⋯	⋯	⋮
	m	c_{m1}	c_{m2}	⋯	c_{mn}	s_m
Demand		d_1	d_2	⋯	d_n	

$$\text{Minimize} \quad Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij},$$

subject to

$$\sum_{j=1}^{n} x_{ij} = s_i \quad \text{for } i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} = d_j \quad \text{for } j = 1, 2, \ldots, n,$$

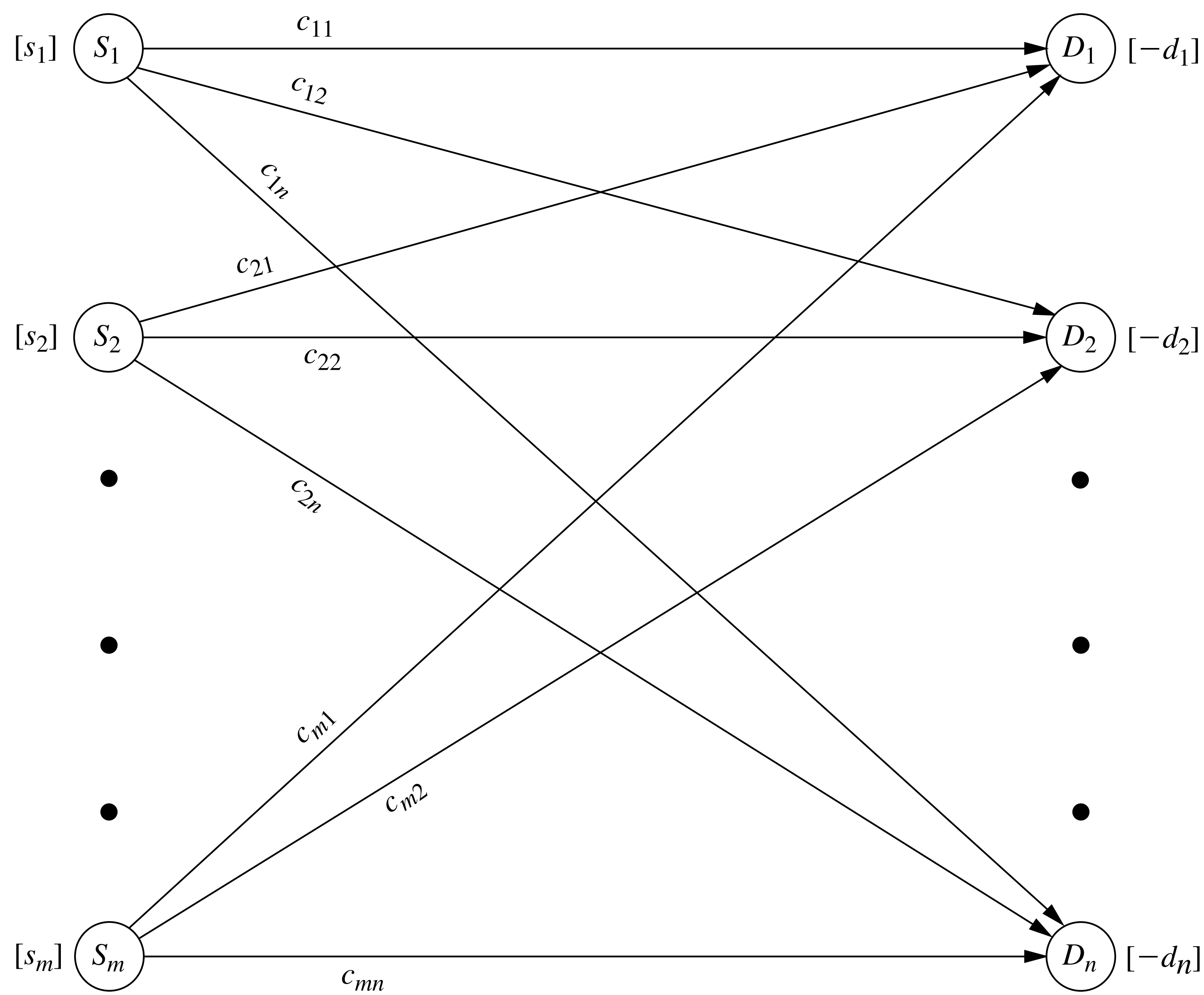
and

$$x_{ij} \geq 0, \quad \text{for all } i \text{ and } j.$$

MIT AEROASTRO

The General Transportation Problem

- The same transportation problem can be considered as a network flow problem as follows:



The General Transportation Problem

- In Excel the transportation problem can be written down as follows:

	A	B	C	D	E	F	G	H	I	J
1	P&T Co. Distribution Problem									
2										
3				Unit Cost						
4				Destination (Warehouse)						
5				Sacramento	Salt Lake City	Rapid City	Albuquerque	Supply		
6		Source	Bellingham	$464	$513	$654	$867	75		
7		(Cannery)	Eugene	$352	$416	$690	$791	125		
8			Albert Lea	$995	$682	$388	$685	100		
9		Demand		80	65	70	85			
10										
11										
12				Shipment Quantities (Truckloads)						
13				Destination (Warehouse)						
14				Sacramento	Salt Lake City	Rapid City	Albuquerque	Totals		Supply
15		Source	Bellingham	0	20	0	55	75	=	75
16		(Cannery)	Eugene	80	45	0	0	125	=	125
17			Albert Lea	0	0	70	30	100	=	100
18		Totals		80	65	70	85	$ 152,535	=	Total Cost
19				=	=	=	=			
20		Demand		80	65	70	85			

Solver Parameters

Set Target Cell: $H$18

Equal To: ○ Max ● Min ○

By Changing Cells:

$D$15:$G$17

Subject to the Constraints:

$D$18:$G$18 = $D$20:$G$20
$H$15:$H$17 = $J$15:$J$17

Solver Options

☑ Assume Linear Model
☑ Assume Non-Negative

	H
15	=SUM(D15:G15)
16	=SUM(D16:G16)
17	=SUM(D17:G17)
18	=SUMPRODUCT(D6:G8,D15:G17)

	D	E	F	G
18	=SUM(D15:D17)	=SUM(E15:E17)	=SUM(F15:F17)	=SUM(G15:G17)

MIT AEROASTRO

Another Transportation Problem in Excel

- Consider a special case, where we assign products to plants for production:

		Unit Cost ($) for Product				**Capacity Available**
		1	**2**	**3**	**4**	
	1	41	27	28	24	75
Plant	2	40	29	—	23	75
	3	37	30	27	21	45
Production rate		20	30	30	40	

Another Transportation Problem in Excel

- Consider a special case, where we assign products to plants for production:

	A	B	C	D	E	F	G	H	I	J
1	Better Products Co. Production-Planning Problem (Option 2)									
2										
3				Unit Cost						
4				Product						
5				1	2	3	4			
6			1	$41	$27	$28	$24			
7		Plant	2	$40	$29	-	$23			
8			3	$37	$30	$27	$21			
9		Required Production		20	30	30	40			
10										
11										
12				Cost ($/day)						
13				Task (Product)						
14				1	2	3	4	Supply		
15		Assignee	1	$820	$810	$840	$960	2		
16		(Plant)	2	$800	$870	-	$920	2		
17			3	$740	$900	$810	$840	1		
18		Demand		1	1	1	1			
19										
20										
21				Assignments						
22				Task (Product)						
23				1	2	3	4	Totals		Supply
24		Assignee	1	0	1	1	0	2	≤	2
25		(Plant)	2	1	0	0	0	1	≤	2
26			3	0	0	0	1	1	=	1
27		Totals		1	1	1	1	$ 3,290	=	Total Cost
28				=	=	=	=			
29		Demand		1	1	1	1			

Solver Parameters

Set Target Cell: $H$27

Equal To: ○ Max ● Min ○

By Changing Cells:

$D$24:$G$26

Subject to the Constraints:

$D$27:$G$27 = $D$29:$G$29
$F$25 = 0
$H$24:$H$25 <= $J$24:$J$25
$H$26 = $J$26

Solver Options

☑ Assume Linear Model
☑ Assume Non-Negative

	H
24	=SUM(D24:G24)
25	=SUM(D25:G25)
26	=SUM(D26:G26)
27	=SUMPRODUCT(D15:G17,D24:G26)

	D	E	F	G
15	=D6*D$9	=E6*E$9	=F6*F$9	=G6*G$9
16	=D7*D$9	=E7*E$9	–	=G7*G$9
17	=D8*D$9	=E8*E$9	=F8*F$9	=G8*G$9
27	=SUM(D24:D26)	=SUM(E24:E26)	=SUM(F24:F26)	=SUM(G24:G26)

The Assignment Problem

- **Goal:** Match assignees to tasks, so that:
  - The number of assignees and the number of tasks are the same. (This number is denoted by n.)
  - Each assignee is to be assigned to exactly one task.
  - Each task is to be performed by exactly one assignee.
  - There is a cost c_{ij} associated with assignee i (i = 1, 2, . . . , n) performing task j (j = 1, 2, . . . , n).
  - The objective is to determine how all n assignments should be made to minimize the total cost.

		Task (Location)			
		1	**2**	**3**	**4**
	1	13	16	12	11
Assignee	2	15	*M*	13	20
(Machine)	3	5	7	10	6
	4(*D*)	0	0	0	0

The Assignment Problem

- The assignment problem can be formulated as an **integer program** as follows:
  - Define the decision variables:

$$x_{ij} = \begin{cases} 1 & \text{if assignee } i \text{ performs task } j, \\ 0 & \text{if not,} \end{cases}$$

  - Write down the objective function and the constraints:

$$\text{Minimize} \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij},$$

subject to

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \text{for } i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \text{for } j = 1, 2, \ldots, n,$$

and

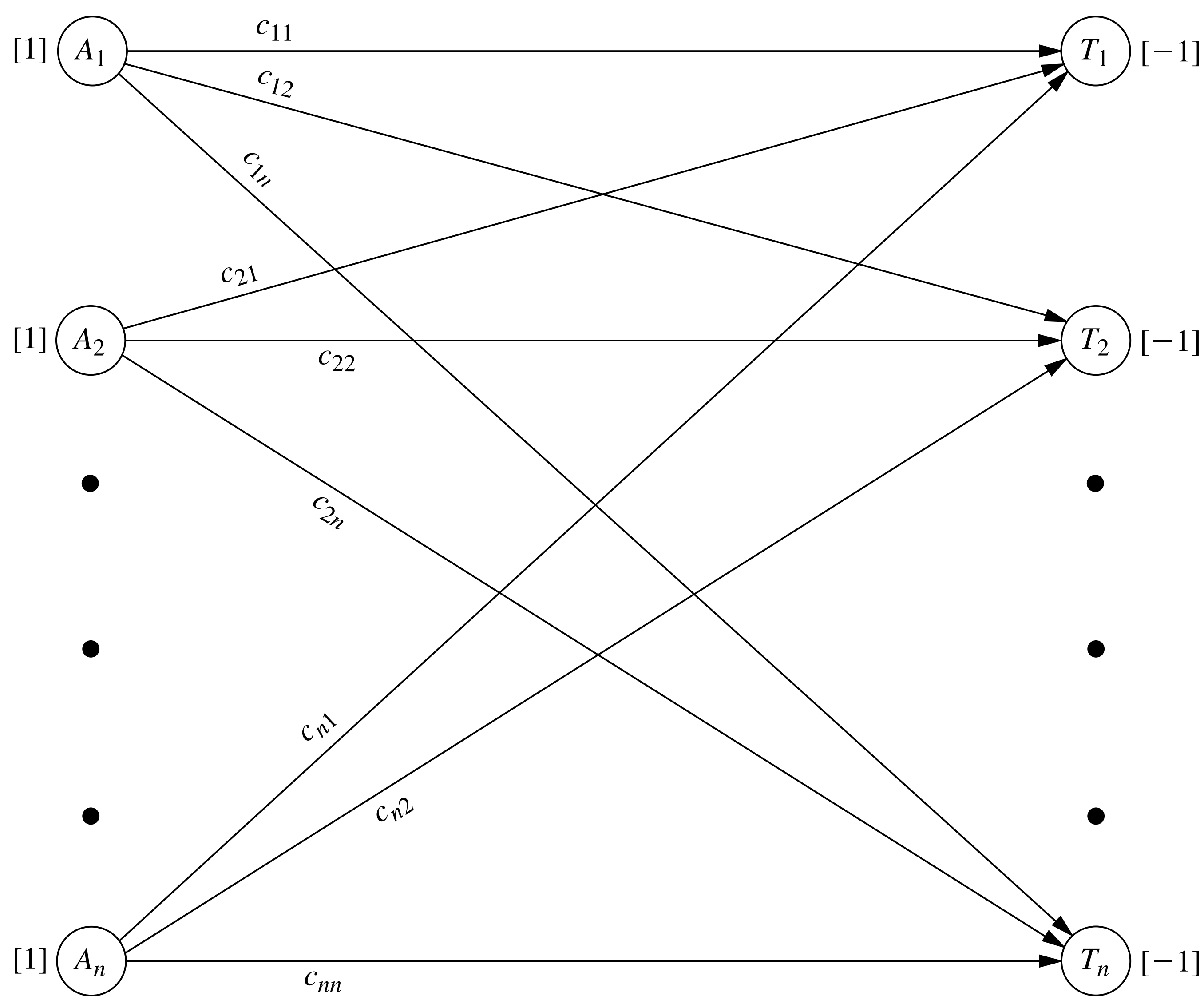
$$x_{ij} \geq 0, \qquad \text{for all } i \text{ and } j$$
$$(x_{ij} \text{ binary}, \qquad \text{for all } i \text{ and } j).$$

		Cost per Unit Distributed				
		Destination				
		1	**2**	$\cdots$	***n***	**Supply**
Source	1	c_{11}	c_{12}	$\cdots$	c_{1n}	1
	2	c_{21}	c_{22}	$\cdots$	c_{2n}	1
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$m = n$	c_{n1}	c_{n2}	$\cdots$	c_{nn}	1
Demand		1	1	$\cdots$	1	

MIT AEROASTRO

The Assignment Problem

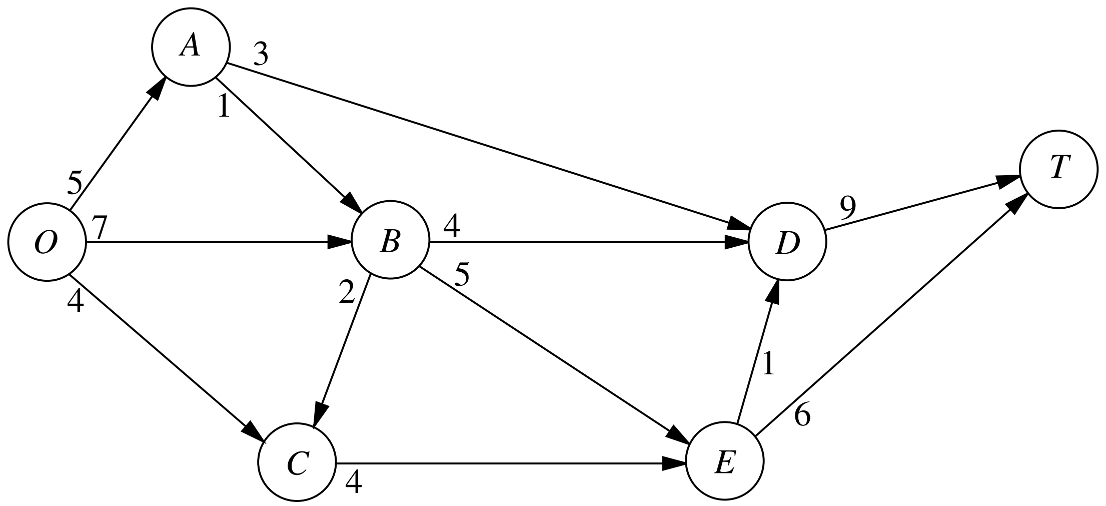
- The assignment problem can be expressed in the network flow as follows:



The Maximum Flow Problem

- All flow in a network originates from node (called the *source*) and terminates in another node (called the *destination*).
- All remaining nodes are the *transshipment* nodes.
- Flow between two nodes only allowed in the direction of the arrow, and at most at the rate of the given capacity.

- *How can we maximize the flow from the source to the destination?*



- **Applications:**
  - Maximize the flow through a company's distribution network from its factories to its customers.
  - Maximize the flow through a company's supply network from its vendors to its factories.
  - Maximize the flow of oil through a system of pipelines.
  - Maximize the flow of water through a system of aqueducts.
  - Maximize the flow of vehicles through a transportation network.

MIT AEROASTRO

The Maximum Flow Problem in Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Seervada Park Maximum Flow Problem										
2											
3		From	To	Flow		Capacity		Nodes	Net Flow		Supply/Demand
4		O	A	4	≤	5		O	14		
5		O	B	7	≤	7		A	0	=	0
6		O	C	3	≤	4		B	0	=	0
7		A	B	1	≤	1		C	0	=	0
8		A	D	3	≤	3		D	0	=	0
9		B	C	0	≤	2		E	0	=	0
10		B	D	4	≤	4		T	-14		
11		B	E	4	≤	5					
12		C	E	3	≤	4					
13		D	T	8	≤	9					
14		E	D	1	≤	1					
15		E	T	6	≤	6					
16											
17		Maximum Flow =		14							

Solver Parameters

Set Target Cell: $D$17

Equal To: (•) Max ( ) Min ( )

By Changing Cells:

$D$4:$D$15

Subject to the Constraints:

$D$4:$D$15 <= $F$4:$F$15
$I$5:$I$9 = $K$5:$K$9

Solver Options

☑ Assume Linear Model
☑ Assume Non-Negative

	D
17	=I4

	I
4	=D4+D5+D6
5	=-D4+D7+D8
6	=-D5-D7+D9+D10+D11
7	=-D6-D9+D12
8	=-D8-D10+D13-D14
9	=-D11-D12+D14+D15
10	=-D13-D15

The Minimum-Cost Flow Problem

- At least one of the nodes is a supply node.
- At least one of the other nodes is a demand node.
- All the remaining nodes are transshipment nodes.
- Flow through an arc is allowed only in the direction indicated by the arrowhead, where
- The network has enough arcs with sufficient capacity to enable all the flow generated at the supply nodes to reach all the demand nodes.
- The cost of the flow through each arc is proportional to the amount of that flow, where the cost per unit flow is known.
- The objective is to minimize the total cost of sending the available supply through the network to satisfy the given demand.

Kind of Application	Supply Nodes	Transshipment Nodes	Demand Nodes
Operation of a distribution network	Sources of goods	Intermediate storage facilities	Customers
Solid waste management	Sources of solid waste	Processing facilities	Landfill locations
Operation of a supply network	Vendors	Intermediate warehouses	Processing facilities
Coordinating product mixes at plants	Plants	Production of a specific product	Market for a specific product
Cash flow management	Sources of cash at a specific time	Short-term investment options	Needs for cash at a specific time

MIT AEROASTRO

The Minimum-Cost Flow Problem

- The minimum-cost flow problem can be formulated as follows:
  - Define the decision variables:

$$x_{ij} = \text{flow through arc } i \to j,$$

$$c_{ij} = \text{cost per unit flow through arc } i \to j,$$
$$u_{ij} = \text{arc capacity for arc } i \to j,$$
$$b_i = \text{net flow generated at node } i.$$

  - Write down the objective and constraints:

$$\text{Minimize} \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij},$$

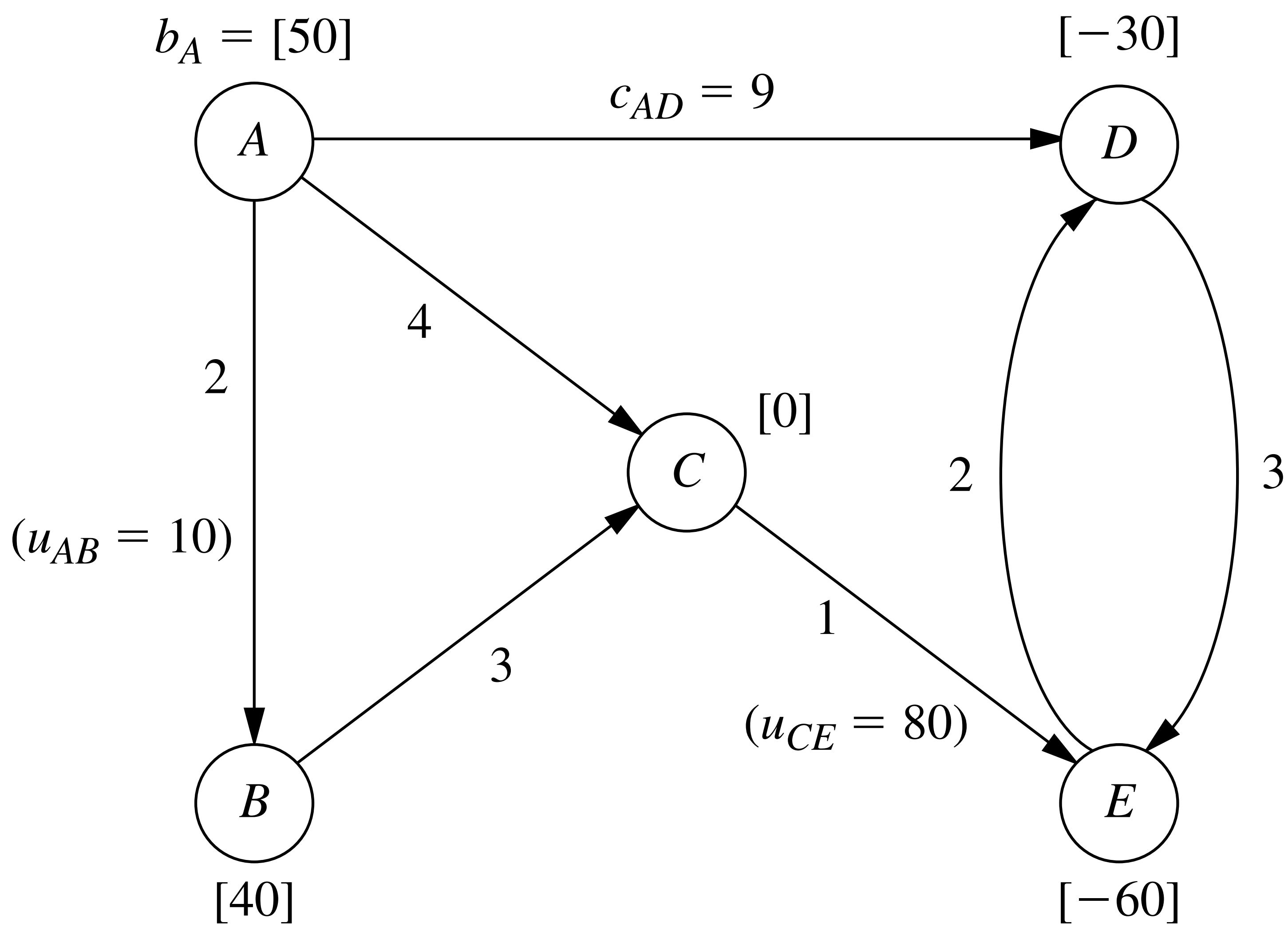
subject to

$$\sum_{j=1}^{n} x_{ij} - \sum_{j=1}^{n} x_{ji} = b_i, \qquad \text{for each node } i,$$

and

$$0 \le x_{ij} \le u_{ij}, \qquad \text{for each arc } i \to j.$$

The Minimum-Cost Problem as a Network Flow



The Minimum-Cost Problem in Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Distribution Unlimited Co. Minimum Cost Flow Problem											
2												
3		From	To	Ship		Capacity	Unit Cost		Nodes	Net Flow		Supply/Demand
4		A	B	0	≤	10	2		A	50	=	50
5		A	C	40			4		B	40	=	40
6		A	D	10			9		C	0	=	0
7		B	C	40			3		D	-30	=	-30
8		C	E	80	≤	80	1		E	-60	=	-60
9		D	E	0			3					
10		E	D	20			2					
11												
12		Total Cost =		490								

Solver Parameters

Set Target Cell: $D$12

Equal To: ( ) Max (•) Min ( )

By Changing Cells:

$D$4:$D$10

Subject to the Constraints:

$D$4 <= $F$4
$D$8 <= $F$8
$J$4:$J$8 = $L$4:$L$8

Solver Options

☑ Assume Linear Model
☑ Assume Non-Negative

	J
4	=D4+D5+D6
5	=-D4+D7
6	=-D5-D7+D8
7	=-D6+D9-D10
8	=-D8-D9+D10

	D
12	=SUMPRODUCT(D4:D10,G4:G10)