(Conic) Linear Optimization: Problem Instances I

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

LY 5th, Chapter 1, Chapter 2.1-2.2

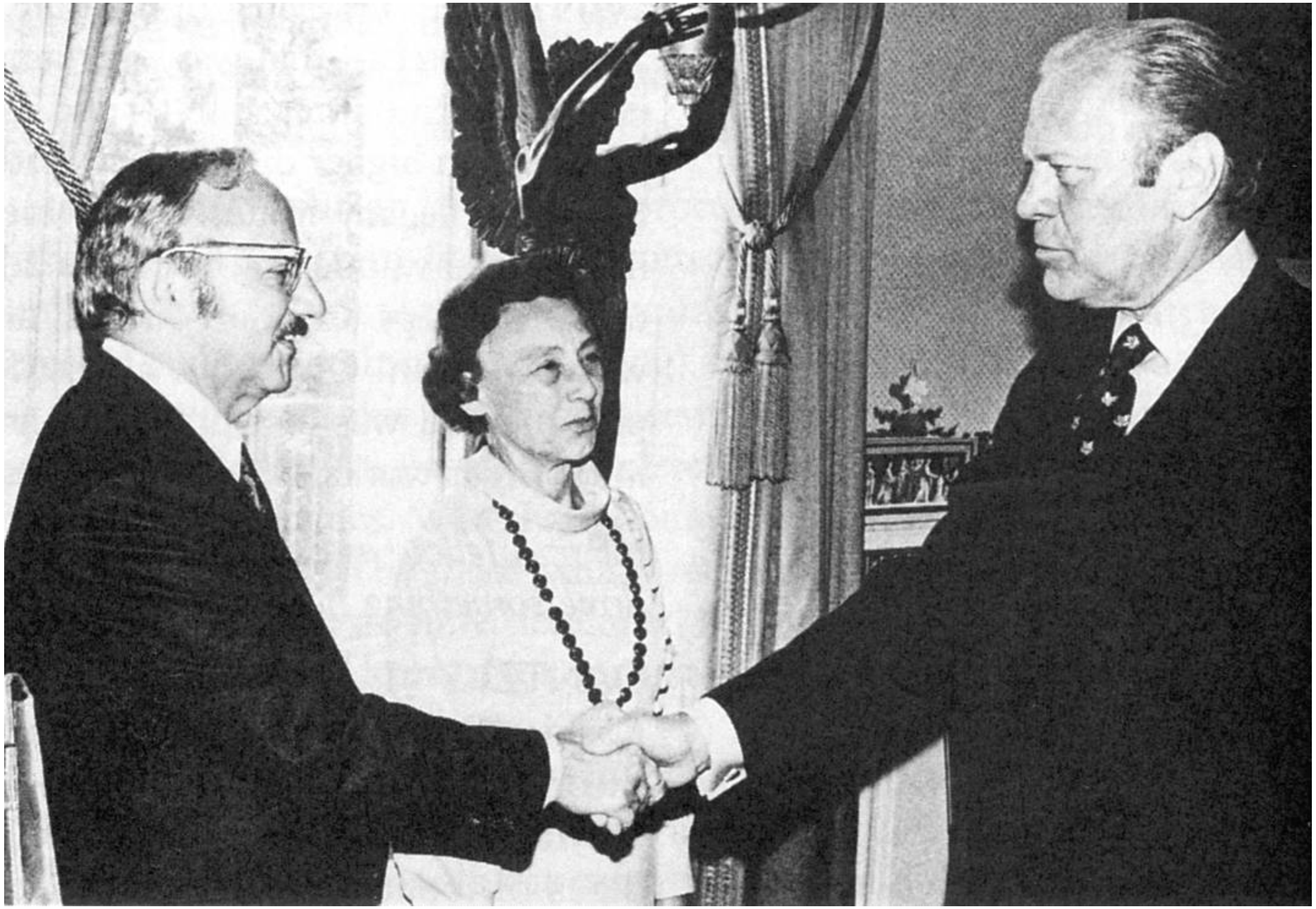
Introduction to Conic Linear Optimization

The field of optimization is concerned with the study of maximization and minimization of mathematical functions. Very often the arguments of (i.e., variables or unknowns in) these functions are subject to side conditions or constraints. By virtue of its great utility in such diverse areas as applied science, engineering, economics, finance, medicine, and statistics, optimization holds an important place in the practical world and the scientific world. Indeed, as far back as the Eighteenth Century, the famous Swiss mathematician and physicist Leonhard Euler (1707-1783) proclaimed[a] that . . . nothing at all takes place in the Universe in which some rule of maximum or minimum does not appear.

[a]See Leonhardo Eulero, *Methodus Inviendi Lineas Curvas Maximi Minimive Proprietate Gaudentes*, Lausanne & Geneva, 1744, p. 245.

Linear Programming: Nobel Prize

Linear Programming: National Metal of Science



Linear Programs and Extensions (in Standard Form)

Linear Programming

$$
\begin{array}{lll}
(LP) & \text{minimize} & \mathbf{c}^T\mathbf{x} \\
& \text{subject to} & A\mathbf{x} = \mathbf{b}, \\
& & \mathbf{x} \geq \mathbf{0}.
\end{array}
$$

Linearly Constrained Optimization Problem

$$
\begin{array}{lll}
(LCOP) & \text{minimize} & f(\mathbf{x}) \\
& \text{subject to} & A\mathbf{x} = \mathbf{b}, \\
& & \mathbf{x} \geq \mathbf{0}.
\end{array}
$$

Linear Complementarity Problem Find nonnegative vectors $\mathbf{x} \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}$ such that

$$(LCP) \qquad \begin{aligned} \mathbf{s} &= M\mathbf{x} + \mathbf{q}, \\ \mathbf{s}^T\mathbf{x} &= 0. \end{aligned}$$

Rectified Linear Unit Operation: $s = \max\{0, ay + b\} \Rightarrow sx = 0,\ s = x + ay + b, (s, x) \geq 0.$

Conic Linear Programming: Structured Convex Optimization

$$\begin{array}{lll} (CLP) & \text{minimize} & \mathbf{c} \bullet \mathbf{x} \\ & \text{subject to} & \mathbf{a}_i \bullet \mathbf{x} = b_i,\ i = 1, ..., m, \\ & & \mathbf{x} \in K, \end{array}$$

where K is a closed convex cone.

CLP: LP, SOCP, and SDP Examples

$$\begin{array}{ll} \text{minimize} & 2x_1 + x_2 + x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 1, \\ & (x_1; x_2; x_3) \geq \mathbf{0}; \end{array}$$

$$\begin{array}{ll} \text{minimize} & 2x_1 + x_2 + x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 1, \\ & \sqrt{x_2^2 + x_3^2} \leq x_1. \end{array}$$

$$\begin{array}{ll} \text{minimize} & 2x_1 + x_2 + x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 1, \\ & \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq \mathbf{0}, \end{array}$$

Facility Location Problem

Let $\mathbf{c}_j$ be the location of client $j = 1, 2, ..., m$, and $\mathbf{y}$ be the location decision of a facility to be built. Then we solve

$$\text{minimize}_{\mathbf{y}} \quad \sum_j \|\mathbf{y} - \mathbf{c}_j\|_p.$$

Or equivalently (?)

$$\begin{array}{ll} \text{minimize} & \sum_j \delta_j \\ \text{subject to} & \mathbf{y} + \mathbf{x}_j = \mathbf{c}_j, \ \|\mathbf{x}_j\|_p \le \delta_j, \ \forall j. \end{array}$$

This is a p-order conic linear program (POCP) for $p \ge 1$.

In particular, if $p = 1$, it is an LP problem; if $p = 2$, it is an SOCP problem.

For simplicity, consider $m = 3$.

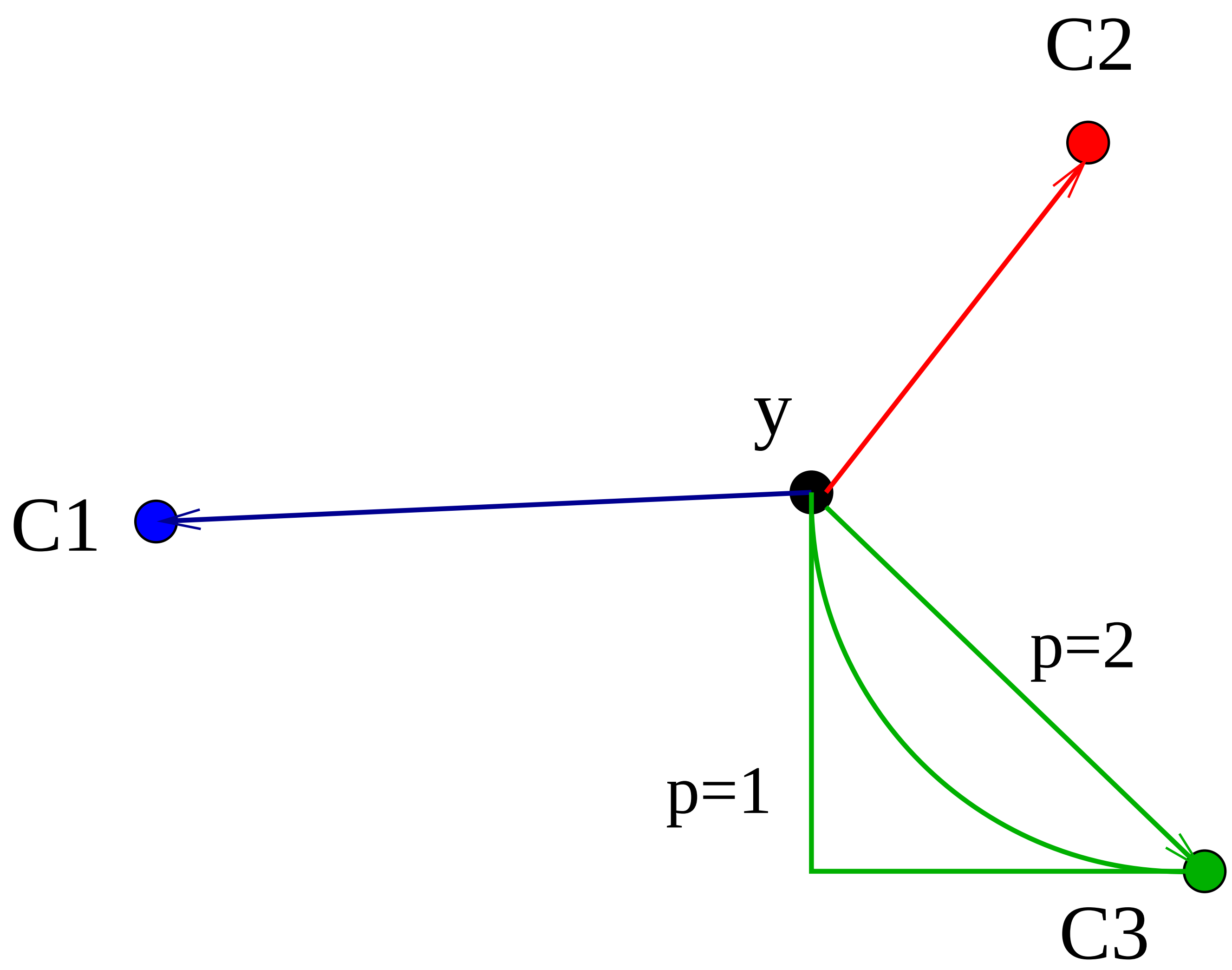


Figure 1: Facility Location at Point $\mathbf{y}$.

CLP Terminology

- decision variable/activity, data/parameter
- objective/goal/target, coefficient vector
- constraint/limitation/requirement, satisfied/violated
- equality/inequality constraint, direction of inequality, non-negativity
- constraint matrix/tensor/the right-hand side vector
- feasible/infeasible solution, interior feasible solution
- optimizers and optimum values
- active constraint (binding constraint), inactive constraint, redundant constraint

We start with LP...

Linear Programming Facts

- The feasible region is a convex polyhedron.
- Every linear program is either feasible/bounded, feasible/unbounded, or infeasible.
- If feasible/bounded, every local optimizer is global and all optimizers form a convex polyhedron set.
- All optimizers are on the boundary of the feasible region.
- If the feasible region has an extreme point, then there must be an extreme optimizer.
- LP possesses efficient algorithms in both practice and theory (polynomial-time).

Linear Optimization Model and Formulation

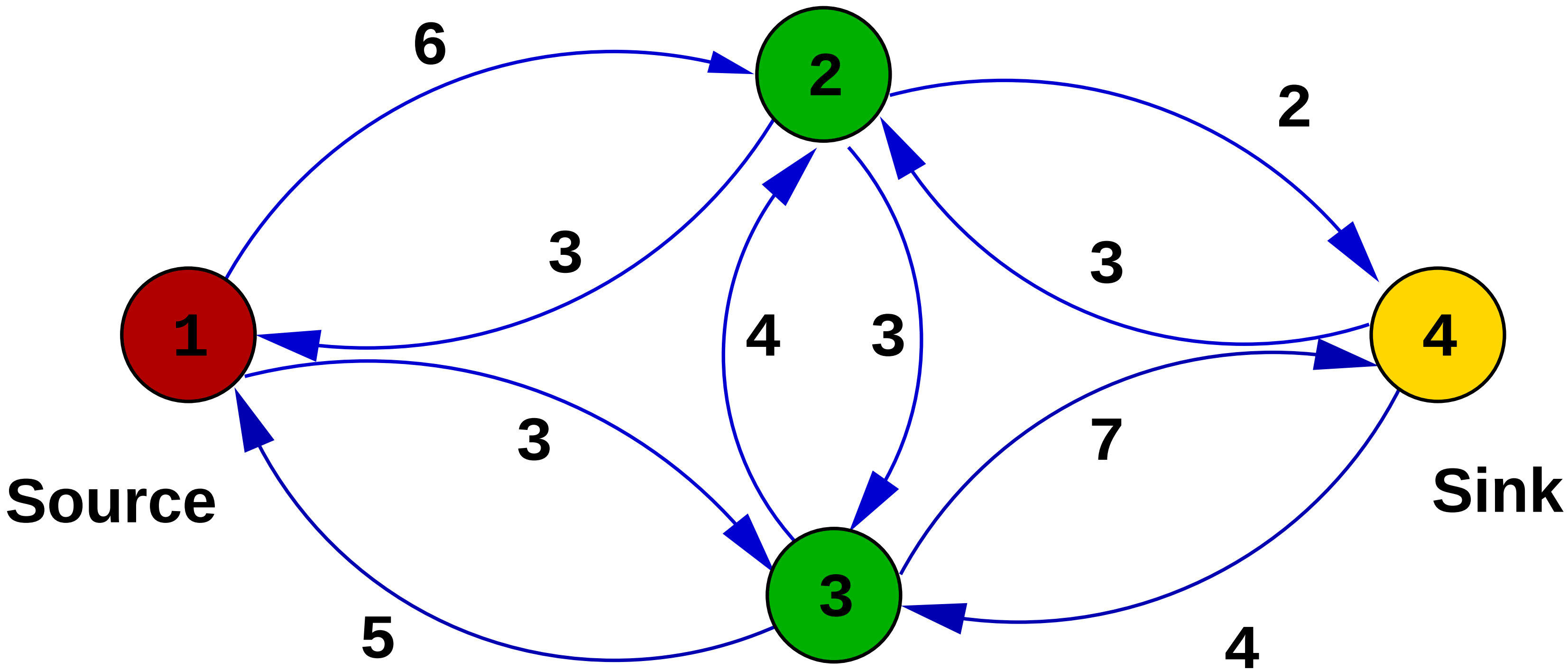
- Sort out data and parameters from the verbal description
- Define the set of decision variables
- Formulate the linear objective function of data and decision variables
- Set up linear equality and inequality constraints

Max-Flow Problem

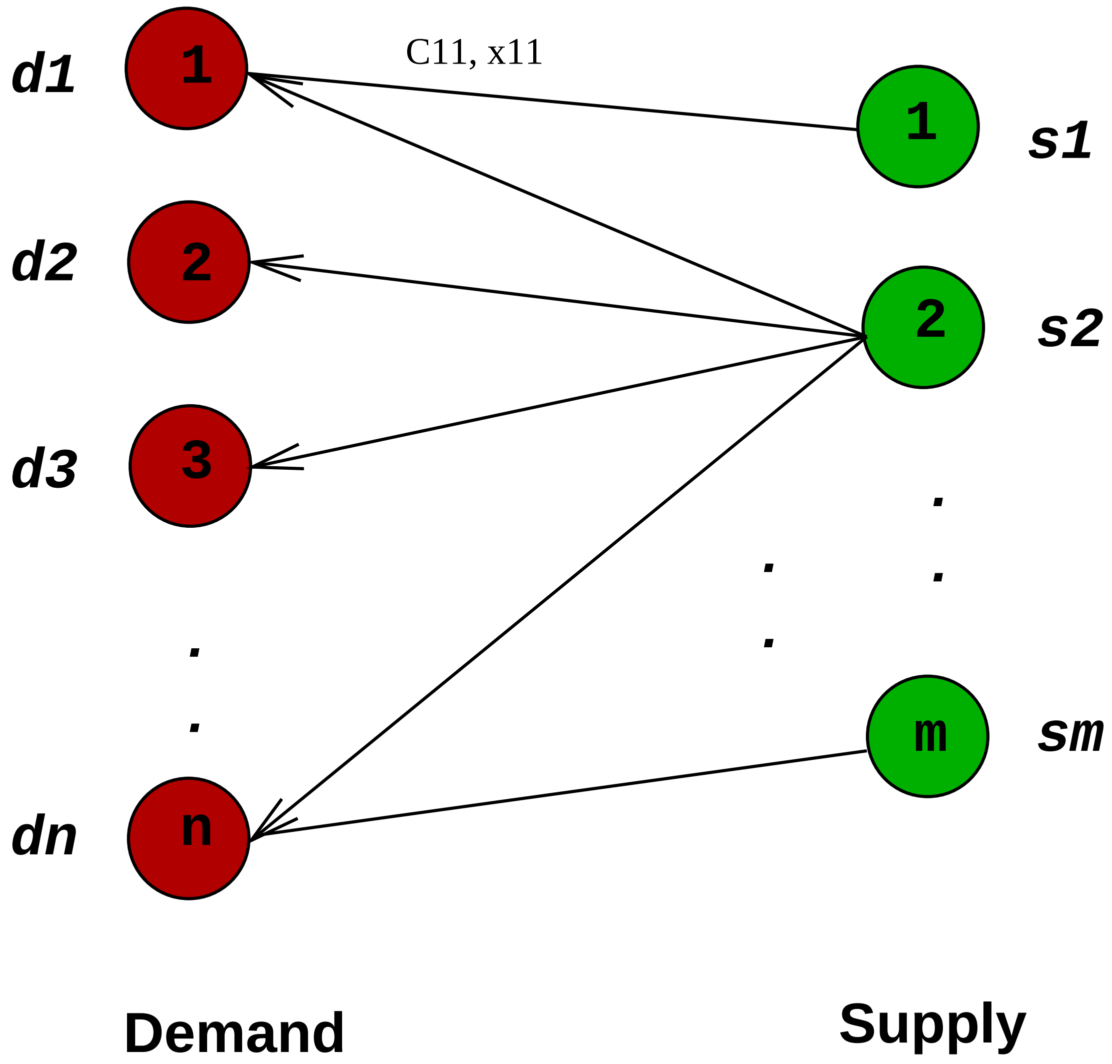
Given a directed graph with nodes $1, ..., m$ and edges $\mathcal{A}$, where node 1 is called source and node m is called the sink, and each edge (i, j) has a flow rate capacity k_{ij}. The Max-Flow problem is to find the largest possible flow rate from source to sink.

Let x_{ij} be the flow rate from node i to node j. Then the problem can be formulated as

$$
\begin{array}{rl}
\text{maximize} & x_{m1} \\
\text{s.t.} & \sum_{j:(j,1)\in\mathcal{A}} x_{j1} - \sum_{j:(1,j)\in A} x_{1j} + x_{m1} = 0, \\
& \sum_{j:(j,i)\in\mathcal{A}} x_{ji} - \sum_{j:(i,j)\in\mathcal{A}} x_{ij} = 0, \forall i = 2, ..., m-1, \\
& \sum_{j:(j,m)\in\mathcal{A}} x_{jm} - \sum_{j:(m,j)\in\mathcal{A}} x_{mj} - x_{m1} = 0, \\
& 0 \leq x_{ij} \leq k_{ij}, \; \forall (i,j) \in \mathcal{A}.
\end{array}
$$


6
2
3
2
3
1
4
3
4
3
7
Source
Sink
3
5
4

The Transportation Problem



Mathematical Optimization Model:

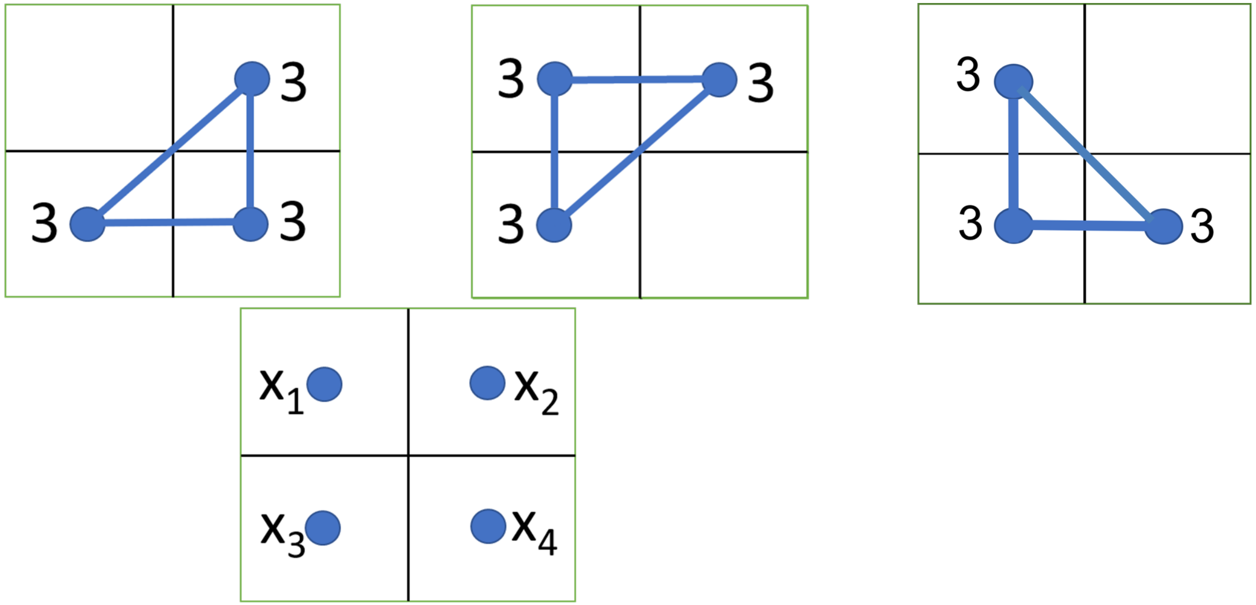
$$
\begin{array}{lll}
\min & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} & \\
\text{s.t.} & \sum_{j=1}^{n} x_{ij} & = s_i, \ \forall i = 1, ..., m \\
& \sum_{i=1}^{m} x_{ij} & = d_j, \ \forall j = 1, ..., n \\
& x_{ij} & \geq 0, \ \forall i, j.
\end{array}
$$

The minimal transportation cost is called the Wasserstein Distance (WD) between supply distribution $\mathbf{s}$ and demand distribution $\mathbf{d}$ (can be interpreted as two probability distributions after normalization). This is a linear program!

What happen if supplies $\mathbf{s}$ are also decision variables?

The Wasserstein Barycenter Problem is to find a distribution such that the sum of its Wasserstein Distance to each of a set of distributions would be minimized.

A Wassestein Barycenter Application: Stochastic Approach



Find distribution of $x_i, i = 1, 2, 3, 4$ to minimize

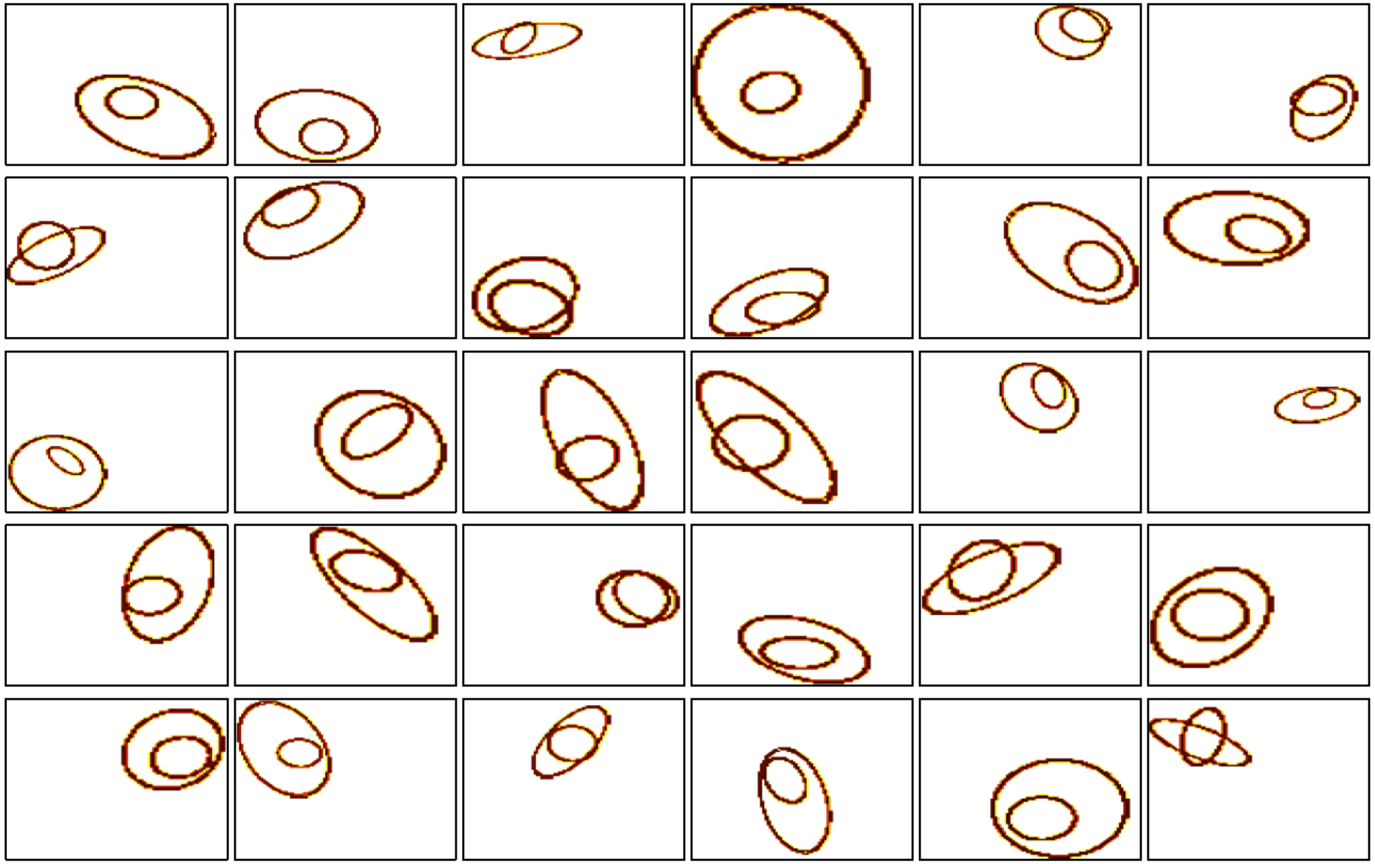
$$\begin{array}{ll} \min & WD(\mathbf{x}, \mathbf{d}_l) + WD(\mathbf{x}, \mathbf{d}_m) + WD(\mathbf{x}, \mathbf{d}_r) \\ \text{s.t.} & x_1 + x_2 + x_3 + x_4 = 9, \qquad x_i \geq 0,\ i = 1, 2, 3, 4; \end{array}$$

which is an LP problem with three sets of local variables and constraints.

Or it can be viewed as nonlinear minimization: what is the gradient vector $\nabla_x WD(\mathbf{x}, \mathbf{d}_.)$?

The Wasserstein Barycenter (Mean) Problem in Data Science

What is the "mean or consensus" image from a set of images/distributions:



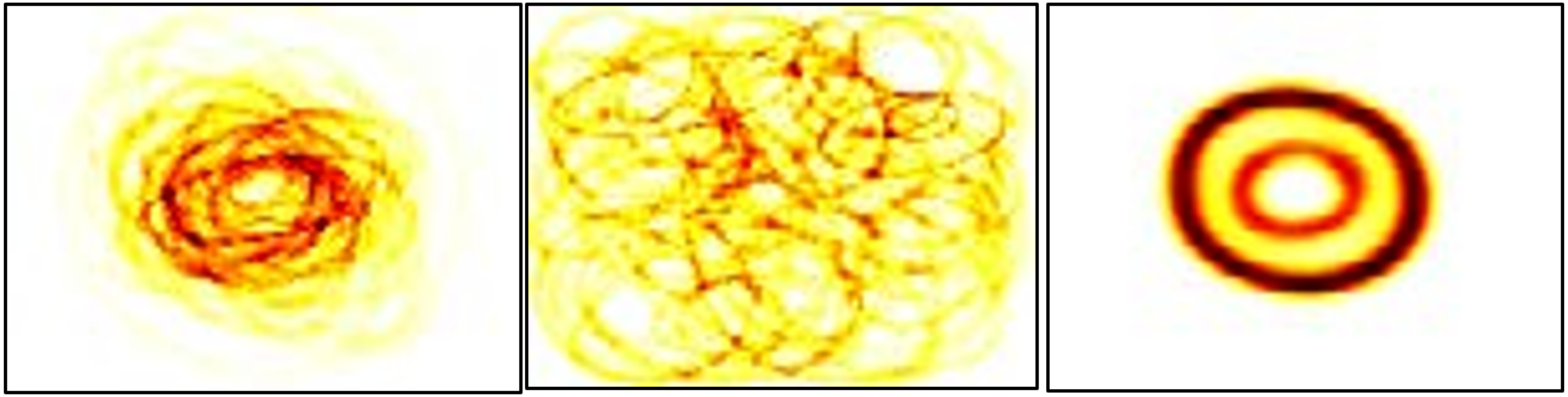


Figure 2: Mean picture constructed from the (a) Euclidean mean after re-centering images (b) Euclidean mean (c) Wasserstein Barycenter (self recenter, resize and rotate)

Euclidean Mean/Center:

$$\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{a}_i, \quad \text{or} \quad \min_{\mathbf{x}} \sum_{i=1}^{n}\|\mathbf{x}-\mathbf{a}_i\|_2^2,$$

which is an unconstrained optimization, or least-squares, problem

Linear Classification: Support Vector Machine I

There are two set of points $\mathbf{a}_i,\ i \in A$ and $\mathbf{b}_j,\ j \in B$, is there a(hyper)plane/line to separate the two sets of points?

A hyperplane is defined by a normal direction or slope vector $\mathbf{x}$ and an intersect scalar x_0. Therefore, if the two sets can be strictly separable, we must be able to find $(\mathbf{x}, x_0)$ such that

$$\mathbf{a}_i^T\mathbf{x} + x_0 > 0\ \forall i \in A, \quad \text{and} \quad \mathbf{b}_j^T\mathbf{x} + x_0 < 0\ \forall j \in B$$

or

$$\mathbf{a}_i^T\mathbf{x} + x_0 \geq 1\ \forall i \in A, \quad \text{and} \quad \mathbf{b}_j^T\mathbf{x} + x_0 \leq 1\ \forall j \in B.$$

This is an LP problem with the null objective function.

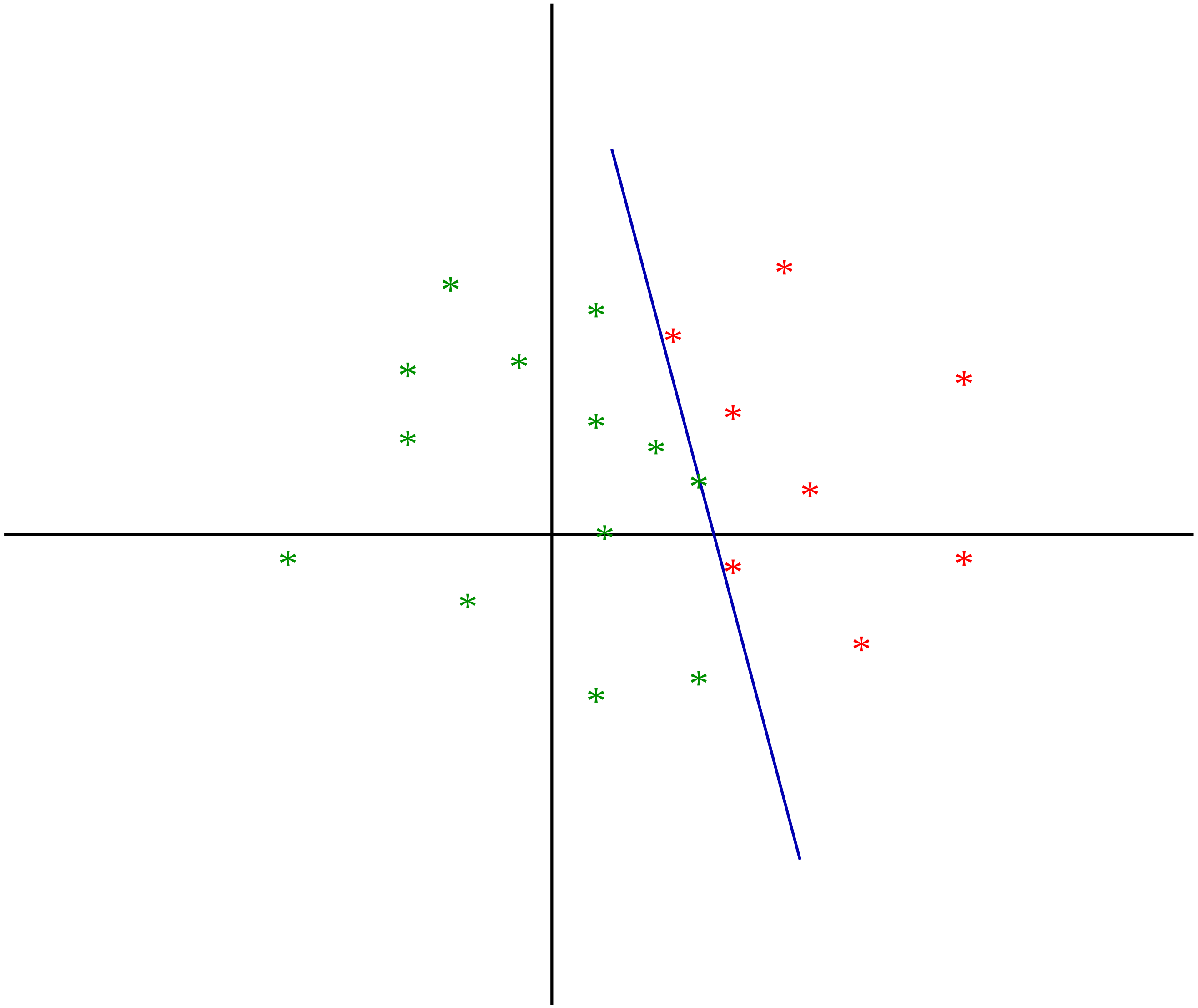


Figure 3: Linear Support Vector Machine

Linear Classification: Supporting Vector Machine II

If strict separation is impossible, we then minimize error variable β

$$\begin{array}{rl} \text{minimize} & \beta \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} + x_0 + \beta \geq 1, \ \forall i \in A, \\ & \mathbf{b}_j^T\mathbf{x} + x_0 - \beta \leq -1, \ \forall j \in B, \\ & \beta \geq 0. \end{array}$$

Frequently we add the regularization term on the slope vector

$$\begin{array}{rl} \text{minimize} & \beta + \mu\|\mathbf{x}\|_2 \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} + x_0 + \beta \geq 1, \ \forall i, \\ & \mathbf{b}_j^T\mathbf{x} + x_0 - \beta \leq -1, \ \forall j, \\ & \beta \geq 0, \end{array}$$

where μ is a fixed positive regularization parameter. This becomes a SOCP after changing objective to $\beta + \alpha$ and adding the constraint $\|\mathbf{x}\|_2 \leq \alpha$. If $\mu = 0$, then it is a pure linear program (LP)!

Quadratic Classification: Ellipsoidal Separation?

$$
\begin{array}{ll}
\text{minimize} & \text{trace}(X) + \|\mathbf{x}\|_2 \\
\text{subject to} & \mathbf{a}_i^T X \mathbf{a}_i + \mathbf{a}_i^T \mathbf{x} + x_0 \geq 1, \ \forall i, \\
& \mathbf{b}_j^T X \mathbf{b}_j + \mathbf{b}_j^T \mathbf{x} + x_0 \leq -1, \ \forall j, \\
& X \succeq \mathbf{0}.
\end{array}
$$

This type of problems is semidefinite programming (SDP). When the problem is not separable:

$$
\begin{array}{ll}
\text{minimize} & \beta + \mu(\text{trace}(X) + \|\mathbf{x}\|_2) \\
\text{subject to} & \mathbf{a}_i^T X \mathbf{a}_i + \mathbf{a}_i^T \mathbf{x} + x_0 + \beta \geq 1, \ \forall i, \\
& \mathbf{b}_j^T X \mathbf{b}_j + \mathbf{b}_j^T \mathbf{x} + x_0 - \beta \leq -1, \ \forall j, \\
& \beta \geq 0, \\
& X \succeq \mathbf{0}.
\end{array}
$$

This is a mixed linear and SDP program.

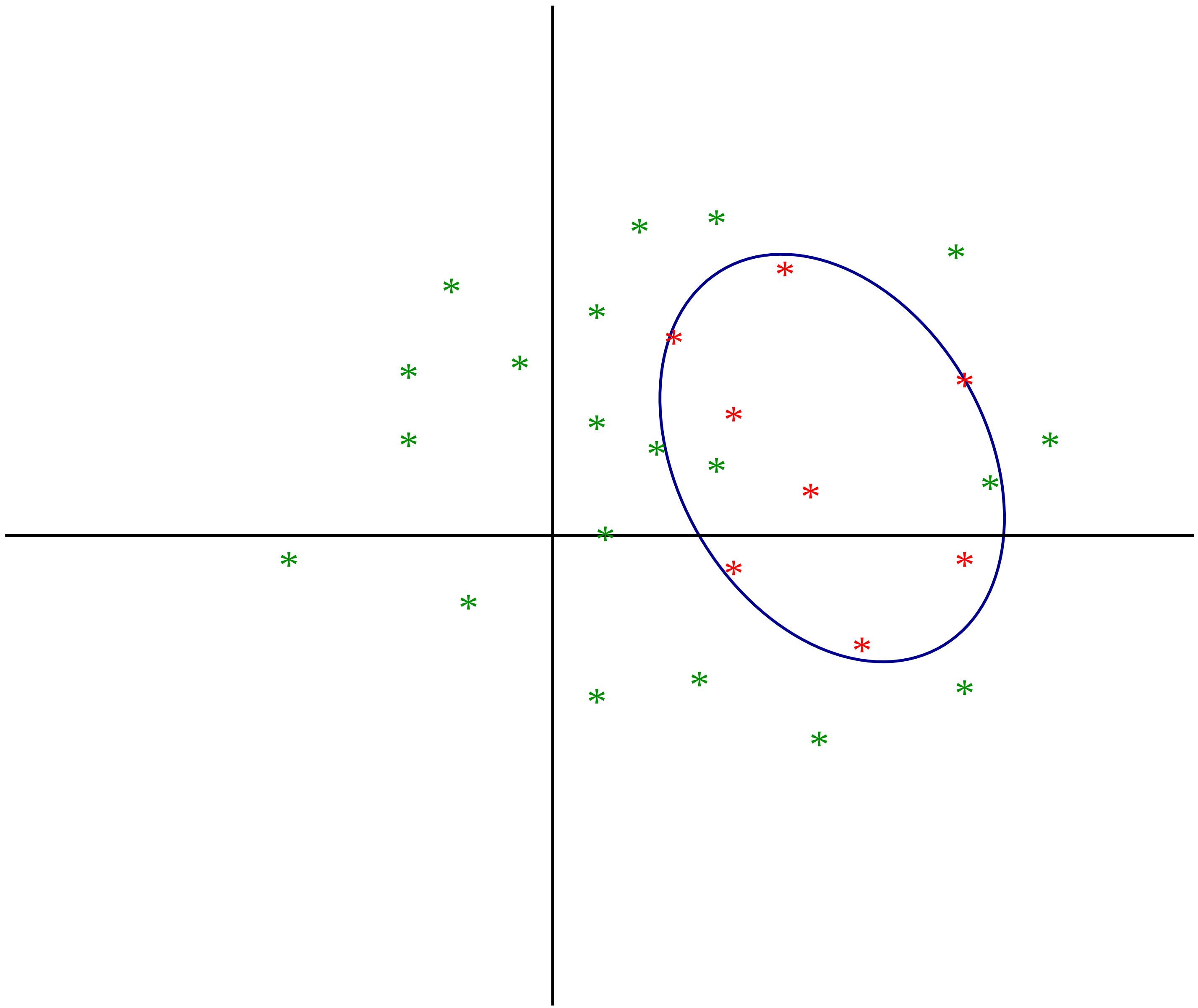


Figure 4: Quadratic Support Vector Machine