CO327 Deterministic OR Models
Piecewise-linear function and robust curve fitting

Instructor: Andersen Ang
Combinatorics and Optimization, U.Waterloo, Canada
msxang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: Janurary 5, 2021 Last update: June 22, 2021

Linear, affine and nonlinear function

- Linear: $f(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n = \mathbf{a}^\top \mathbf{x}$.
- Affine: $\mathbf{a}^\top \mathbf{x} + b$.
- Linear vs affine
  - pass through zero or not.
  - Affine $=$ linear shifted.
- Nonlinear: you have polynomial in x_i (degree higher than 1).
- Geometry
  - Linear/affine: a straight line/ hyperplane.
  - Nonlinear: a curve/ surfece

Be formal and precise

A function f

- is linear if
$$f(ax + by) = af(x) + bf(y)$$
for all x, y, a, b.

- is affine if
$$f(ax + (1-a)y) = af(x) + (1-a)f(y)$$
for all x, y, a, b.

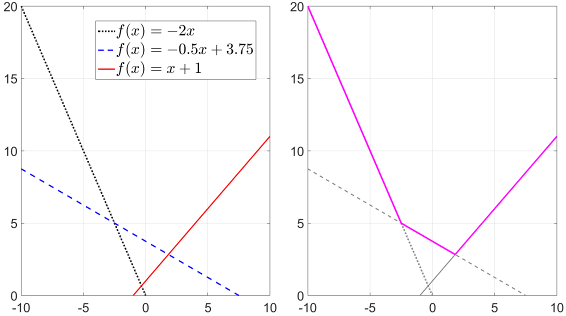
A function f is affine if and only if $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$.

Piecewise-linear (piecewise-affine) function

- A function f is piecewise-linear if it can be expressed as

$$f(\mathbf{x}) = \max_{i=1,\dots,m} \left(\mathbf{a}_i^\top \mathbf{x} + b_i \right).$$

- f is convex and is parametrized by m vectors $\mathbf{a}_i$ and m scalar b_i
- Example in 1-dimension and $m = 3$



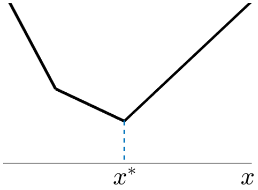
People mix up piecewise-linear as piecewise-affine, just like people mixing up linear and affine. In fact piecewise-linear means $\mathbf{b} = \mathbf{0}$, the graph will be V-shape. Basically the convention now is that people take piecewise-linear as piecewise-affine.

Minimizing piecewise-linear function

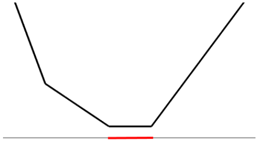
- The minimization problem

$$\min f(\mathbf{x}) = \max_{i=1,\ldots,m} \left(\mathbf{a}_i^\top \mathbf{x} + \mathbf{b}_i \right).$$

  is to find the minimizer point $\mathbf{x}^*$



- There can be ∞ many minimizers: piecewise-linear function is convex but not necessarily strictly convex



LP formulation ... (1/2)

- The minimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) = \max_{i=1,...,m} \left(\mathbf{a}_i^\top \mathbf{x} + b_i \right)$$

  means you want to minimize the max of all $\mathbf{a}_i^\top \mathbf{x} + b_i$.

- How: put a upper bound t on top of all $\mathbf{a}_i^\top \mathbf{x} + b_i$, minimizing this t will minimize the max among all $\mathbf{a}_i^\top \mathbf{x} + b_i$.

- Equivalent LP

$$\begin{aligned} \min_{t} \quad & t \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} + \mathbf{b}_i \leq t, \quad i = 1, ..., m \end{aligned}$$

  we can see the max operator is gone in the LP formulation.

LP formulation ... (2/2)

- The LP

$$\begin{aligned} \min_{t} \quad & t \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} + b_i \leq t, \quad i = 1, ..., m \end{aligned}$$

- Rewrite

$$\begin{aligned} \min_{\mathbf{x}, t} \quad & \mathbf{0}^\top \mathbf{x} + t \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} - t \leq -b_i, \quad i = 1, ..., m \end{aligned}$$

- Canonical form LP

$$\begin{aligned} \min_{\hat{\mathbf{x}}} \quad & \hat{\mathbf{c}}^\top \hat{\mathbf{x}} \\ \text{s.t.} \quad & \hat{\mathbf{A}} \hat{\mathbf{x}} \leq \hat{\mathbf{b}} \end{aligned}$$

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_1^\top & -1 \\ \vdots & \vdots \\ \mathbf{a}_m^\top & -1 \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \end{bmatrix}.$$

Application: curve fitting

- Given m points (t_i, y_i), fit the affine function $f(t) = a + bt_i$.
- Goal: estimate (a, b) from these m points.
- Data are noisy: there will be fitting error $a + bt_i = y_i + \epsilon_i$.
- Matrix expression

$$\begin{aligned} a + bt_1 &= y_1 + \epsilon_1 \\ a + bt_2 &= y_2 + \epsilon_2 \\ a + bt_3 &= y_3 + \epsilon_3 \end{aligned} \rightarrow \underbrace{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\mathbf{e}}$$

  i.e., $\mathbf{Ax} = \mathbf{b} + \mathbf{e}$, or $\mathbf{e} = \mathbf{Ax} - \mathbf{b}$, where $(\mathbf{Ax} - \mathbf{b})_i = i$th element of $\mathbf{Ax} - \mathbf{b}$, which equal to ϵ_i.
- Good fit: the size of $\mathbf{e}$ is minimized.
- Optimization problem

$$\min_{\mathbf{x}} \text{size of } (\mathbf{Ax} - \mathbf{b}).$$

Least squares and least absolute value

- Least squares: minimize the sum of squares of ϵ_i

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 = \sum_i (\mathbf{A}\mathbf{x}-\mathbf{b})_i^2$$

  this is also called least L_2-norm.

- Least L_1 norm: minimize the sum of absolute value of ϵ_i

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 = \sum_i \left|(\mathbf{A}\mathbf{x}-\mathbf{b})_i\right|.$$

- Using $|x| = \max\{x, -x\}$, the equivalent problem

$$\begin{aligned}\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 &= \sum_i \max\left\{(\mathbf{A}\mathbf{x}-\mathbf{b})_i,\ -(\mathbf{A}\mathbf{x}-\mathbf{b})_i\right\}\\ &= \sum_i \max\left\{\mathbf{a}_i^\top\mathbf{x}-b_i,\ -(\mathbf{a}_i^\top\mathbf{x}-b_i)\right\}.\end{aligned}$$

LP of least L_1 norm ... (1/3)

Few slides ago, we saw that

$$
\begin{aligned}
& \min f(\mathbf{x}) = \max_{i=1,\ldots,m} \left(\mathbf{a}_i^\top \mathbf{x} + b_i\right) \\
\iff \quad & \min_{t,\mathbf{x}} t \;\text{ s.t. } \mathbf{a}_i^\top \mathbf{x} + b_i \leq t, \quad i = 1, \ldots, m \\
\iff \quad & \min_{\hat{\mathbf{x}}} \hat{\mathbf{c}}^\top \hat{\mathbf{x}} \text{ s.t. } \hat{\mathbf{A}} \hat{\mathbf{x}} \leq \hat{\mathbf{b}}
\end{aligned}
$$

with

$$
\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_1^\top & -1 \\ \vdots & \vdots \\ \mathbf{a}_m^\top & -1 \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \end{bmatrix}.
$$

Hence we derive a similar LP for

$$
\begin{aligned}
\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 \quad & = \quad \sum_i \max \left\{ (\mathbf{A}\mathbf{x} - \mathbf{b})_i, \; -(\mathbf{A}\mathbf{x} - \mathbf{b})_i \right\} \\
& = \quad \sum_i \max \left\{ \mathbf{a}_i^\top \mathbf{x} - b_i, \; -(\mathbf{a}_i^\top \mathbf{x} - b_i) \right\}.
\end{aligned}
$$

LP of least L_1 norm ... (2/3)

$$\begin{aligned}\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 &= \sum_i \max\left\{(\mathbf{A}\mathbf{x}-\mathbf{b})_i,\ \ -(\mathbf{A}\mathbf{x}-\mathbf{b})_i\right\} \\ &= \sum_i \max\left\{\mathbf{a}_i^\top\mathbf{x}-b_i,\ \ -(\mathbf{a}_i^\top\mathbf{x}-b_i)\right\}\end{aligned}$$

is equivalent to LP

$$\begin{aligned}\min_{u_i} \quad & \sum_i u_i \\ \text{s.t.} \quad & -u_i \leq \mathbf{a}_i^\top\mathbf{x}-b_i \leq u_i, \quad i=1,...,m\end{aligned}$$

In matrix form

$$\begin{aligned}\min_{\mathbf{u},\mathbf{x}} \quad & \mathbf{1}^\top\mathbf{u}+\mathbf{0}^\top\mathbf{x} \\ \text{s.t.} \quad & -\mathbf{u} \leq \mathbf{A}\mathbf{x}-\mathbf{b} \leq \mathbf{u}\end{aligned}$$

LP of least L_1 norm ... (3/3)

$$\begin{array}{ll}\min_{\mathbf{u},\mathbf{x}} & \mathbf{1}^\top \mathbf{u} + \mathbf{0}^\top \mathbf{x} \\ \text{s.t.} & -\mathbf{u} \leq \mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{u}\end{array} \iff \begin{array}{ll}\min_{\mathbf{u},\mathbf{x}} & \mathbf{1}^\top \mathbf{u} + \mathbf{0}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{u} \\ & -\mathbf{A}\mathbf{x} + \mathbf{b} \leq \mathbf{u}\end{array}$$

$$\iff \begin{array}{ll}\min_{\mathbf{u},\mathbf{x}} & \mathbf{1}^\top \mathbf{u} + \mathbf{0}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} - \mathbf{u} \leq \mathbf{b} \\ & -\mathbf{A}\mathbf{x} - \mathbf{u} \leq -\mathbf{b}\end{array} \iff \begin{array}{ll}\min_{\hat{\mathbf{x}}} & \hat{\mathbf{c}}^\top \hat{\mathbf{x}} \\ \text{s.t.} & \hat{\mathbf{A}}\hat{\mathbf{x}} \leq \hat{\mathbf{b}}\end{array}$$

with

$$\hat{\mathbf{x}} = \begin{bmatrix}\mathbf{x} \\ \mathbf{u}\end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix}\mathbf{0} \\ \mathbf{1}\end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix}\mathbf{A} & -\mathbf{I} \\ -\mathbf{A} & -\mathbf{I}\end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix}\mathbf{b} \\ -\mathbf{b}\end{bmatrix}.$$

Robust curve fitting, summary

- Given m points (t_i, y_i), fit the affine function $f(t) = a + bt_i$.

$$\begin{aligned} a + bt_1 &= y_1 + \epsilon_1 \\ &\vdots \\ a + bt_m &= y_m + \epsilon_m \end{aligned} \quad \rightarrow \quad \underbrace{\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{bmatrix}}_{\mathbf{e}}.$$

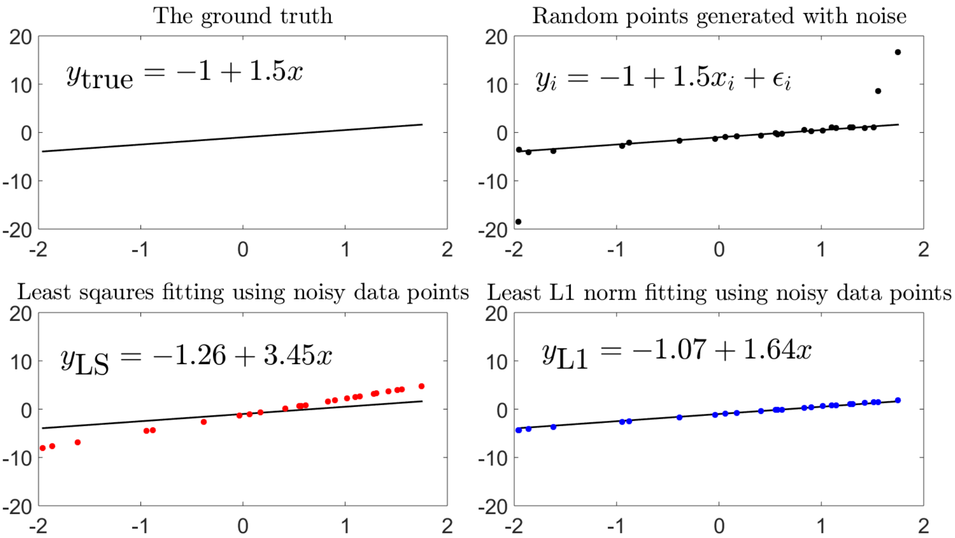
- We minimize $\|\mathbf{Ax} - \mathbf{b}\|_1$, which leads to the LP in canonical form

$$\begin{aligned} \min_{\hat{\mathbf{x}}} \quad & \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}^\top \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{A} & -\mathbf{I} \\ -\mathbf{A} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} \leq \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}. \end{aligned}$$

Least L_∞ norm / Cheybshev approximation

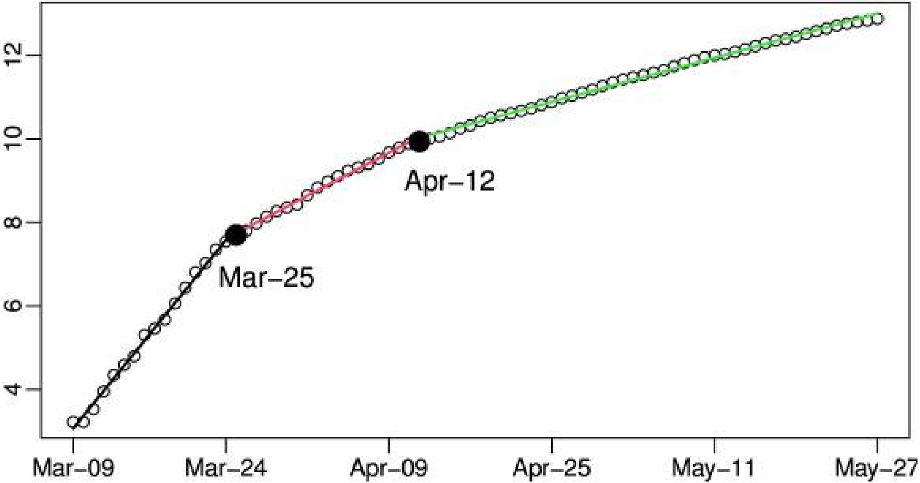
- Replacing L_1 to L_∞
$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_\infty.$$
- L_∞ norm
$$\|\mathbf{x}\|_\infty = \max_{i=1,2\ldots,m} |x_i|.$$
- The equivalent LP (in canonical form)
$$\begin{array}{ll} \min_{\hat{\mathbf{x}}} & \hat{\mathbf{c}}^\top \hat{\mathbf{x}} \\ \text{s.t.} & \hat{\mathbf{A}}\hat{\mathbf{x}} \leq \hat{\mathbf{b}} \end{array}$$
with
$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}, \quad \hat{\mathbf{c}} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & -\mathbf{1} \\ -\mathbf{A} & -\mathbf{1} \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}.$$
- Geometrically, L_∞-norm is the largest component (in magnitude) of $\mathbf{x}$.

Application: robust curve fitting



Least L_1 norm fit is more robust against outliers.

Application: Change point detection



Total number of cases of coronavirus in a country. Source: internet.

How: the data is piecewise linear, so we fit a subset of data points in each segment.

Summary

- Piecewise-linear function
- LP formulation of minimizing piecewise-linear function
- Application in Robust curve fitting
- Application in change point detection

End of document