On Continuous Piecewise Linear Approximating Functions

Lingxun Kong

Maravelias Group

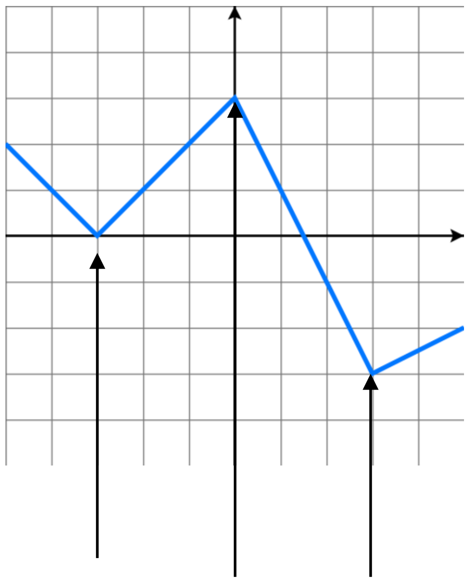
Department of Chemical and Biological Engineering
University of Wisconsin-Madison

Computing Seminar
3/1/2019

What are PWL Functions?

Definition of piecewise linear functions

- A piecewise linear or segmented function is a real-valued function defined on the real numbers or a segment thereof, whose graph is composed of straight-line sections
- The PWL function can be continuous or discontinuous



Break points

$$f(x) = \begin{cases} -x-3 & \text{if } x \leq -3 \\ x+3 & \text{if } -3 < x < 0 \\ -2x+3 & \text{if } 0 \leq x < 3 \\ 0.5x-4.5 & \text{if } x \geq 3 \end{cases}$$

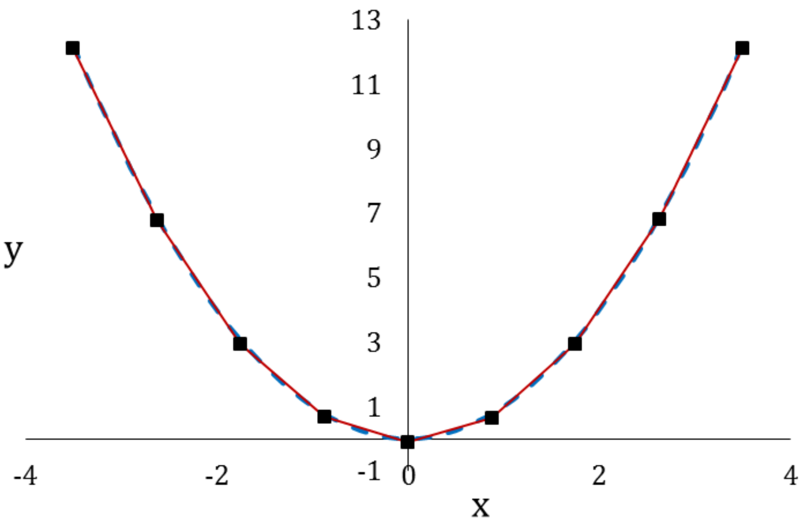
Break points

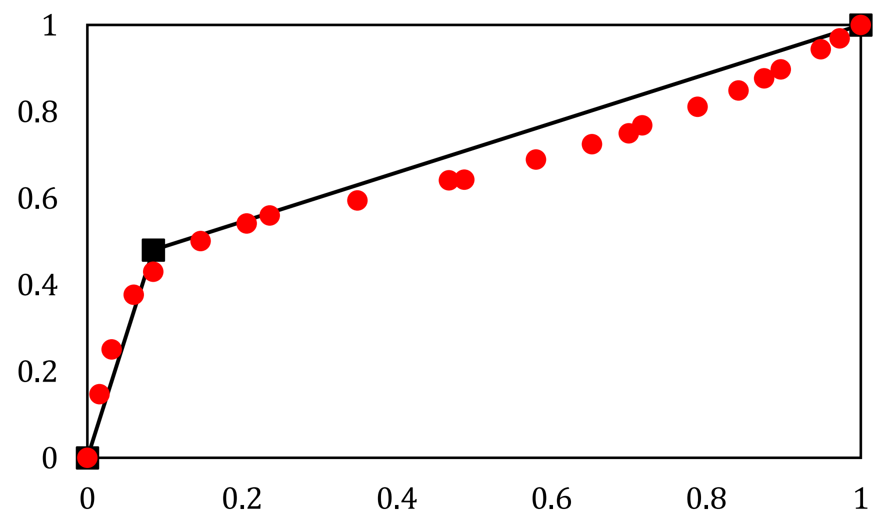
Two representations

- Parametric
- Nonparametric

What are PWLA Functions?

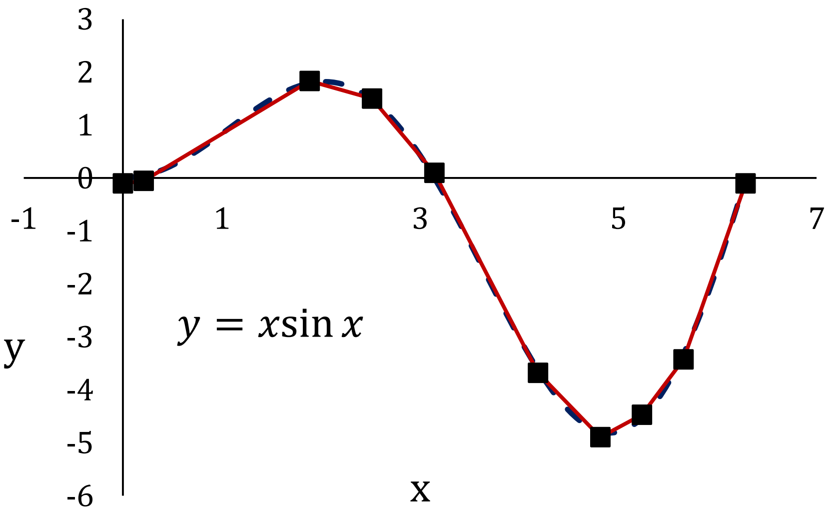
- A PWL function that is used to approximate
  - A continuous (nonlinear) function, or
  - A set of discrete data points

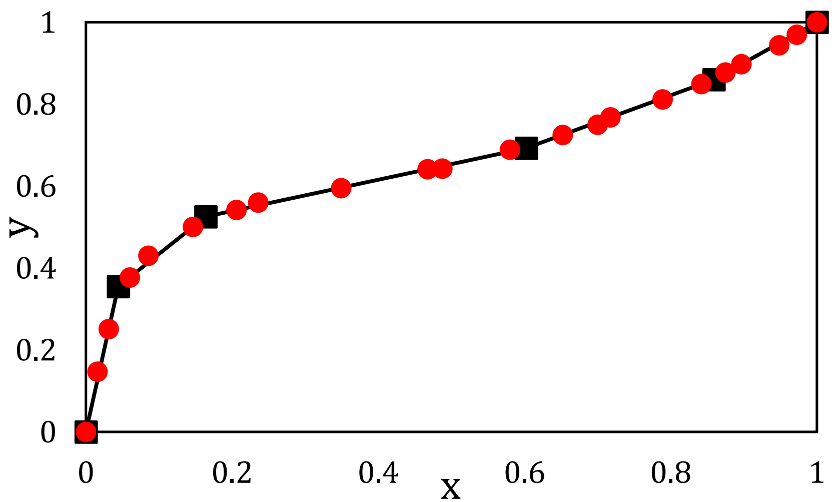




Why PWLA Functions?

- The nonlinear function is too complex to model (in a math program)
- We need a simple function to approximate the data set
  - The underlying function of the data set might **not** be available





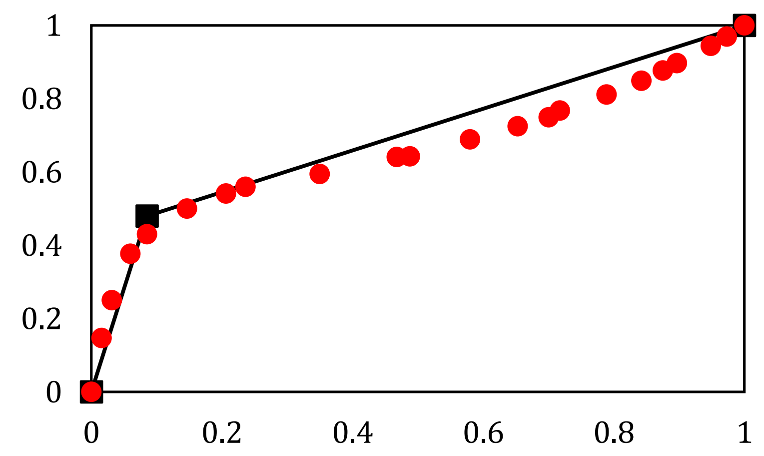
How to Find a PWLA Function?

Given

- A finite set of discrete data points, or
- A continuous function over a domain

Solve a constrained optimization problem to

- Minimize some errors of interest
- Find the slopes, intercepts, and break points



$$\min \sum fitting\ errors$$

Objective

$$s.t. \quad f(x, y) = 0$$

$$g(x, y) \geq 0$$

$$x \in \mathbb{R}^n, y \in \{0,1\}^m$$

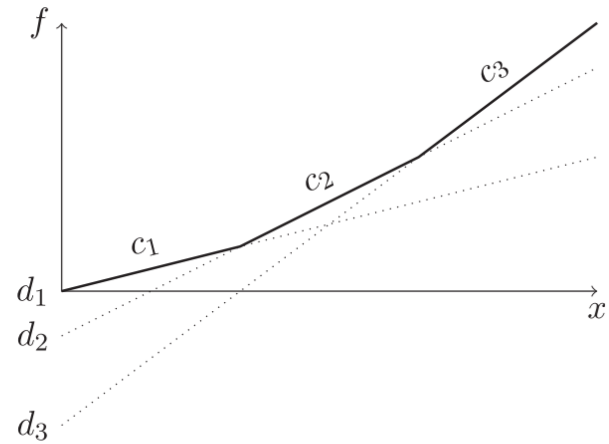
Relationship between PWLA function and data set & PWLA function is continuous

How to Model a PWLA Function?

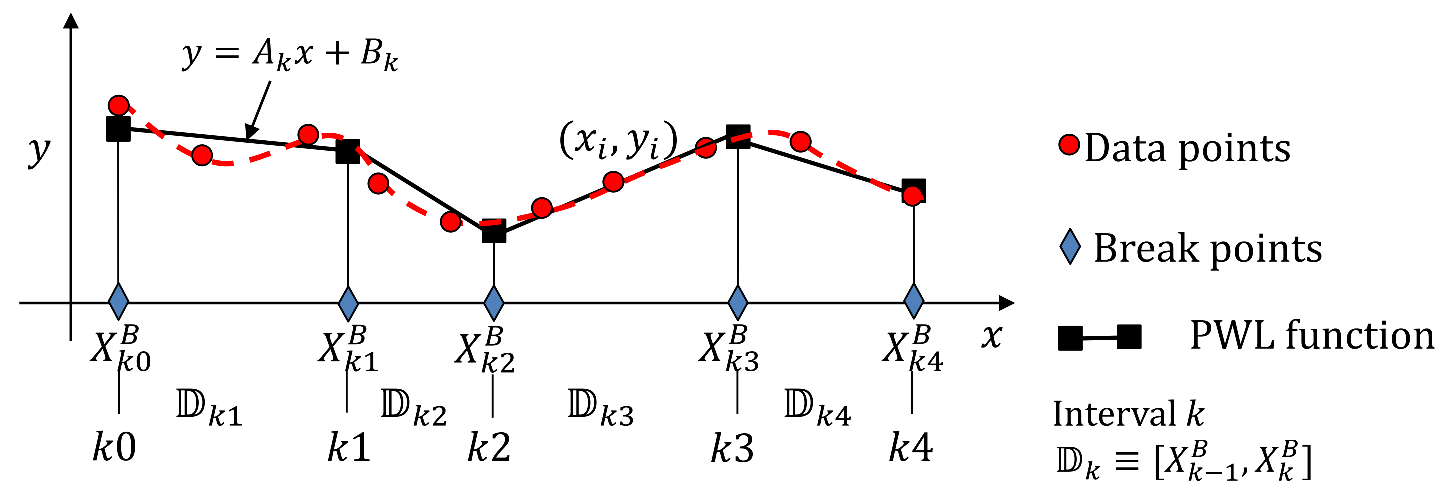
- A typical **mixed-integer** formulation to model PWLA functions

$$Z_k = \{0,1\} \qquad X = \sum_k X_k^D$$

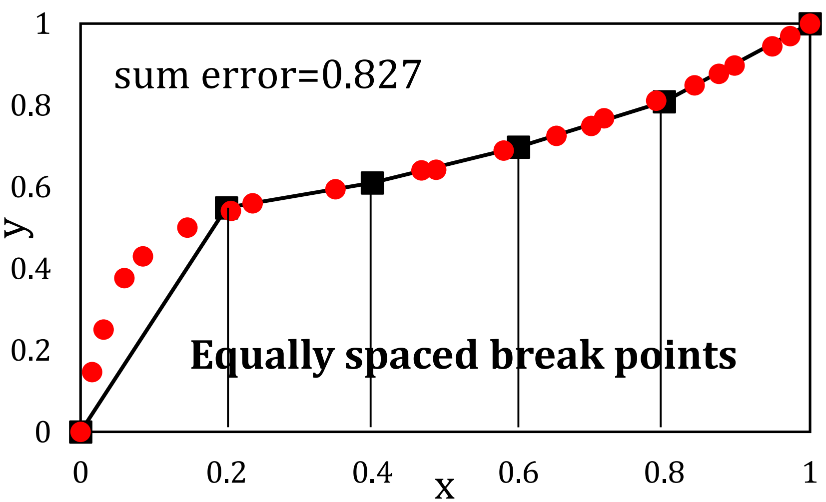
$$X_{k-1}^B Z_k \le X_k^D \le X_k^B Z_k \qquad Y = \sum_k (A_k X_k^D + B_k Z_k)$$

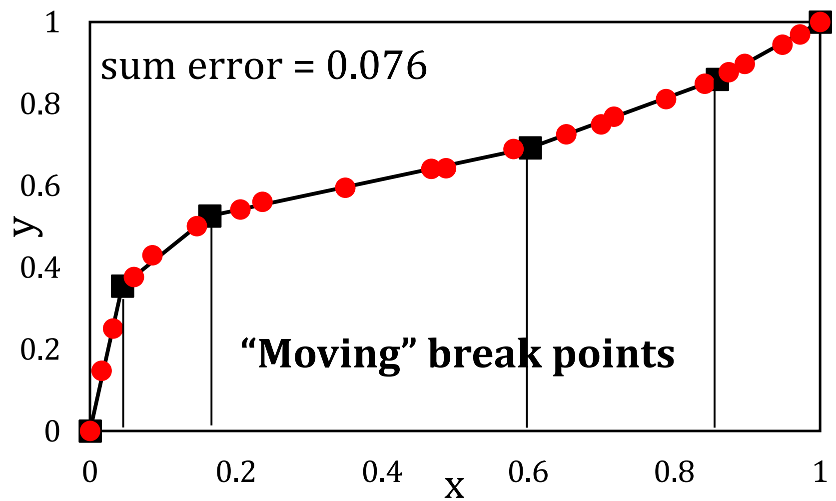


- How many break points do we need? (more break points = more difficult models)



Location and Number of Break Points





Observations

- The more break points, the better the approximation
- Location of the break points can be optimized

Questions

- Given # of break points, what are the locations of these break points? Or
- Given an error tolerance (e.g SSE), how many break points do we need?
- Can we solve this fast?
- Can we approximate nonlinear functions?

M0: Optimize Break Point Locations

$$\min \sum \textit{error at each data point}$$

Objective =Minimize error

$s.t.$

$$f(x, y) = 0$$

$$g(x, y) \geq 0$$

$$x \in \mathbb{R}, y \in \{0,1\}$$

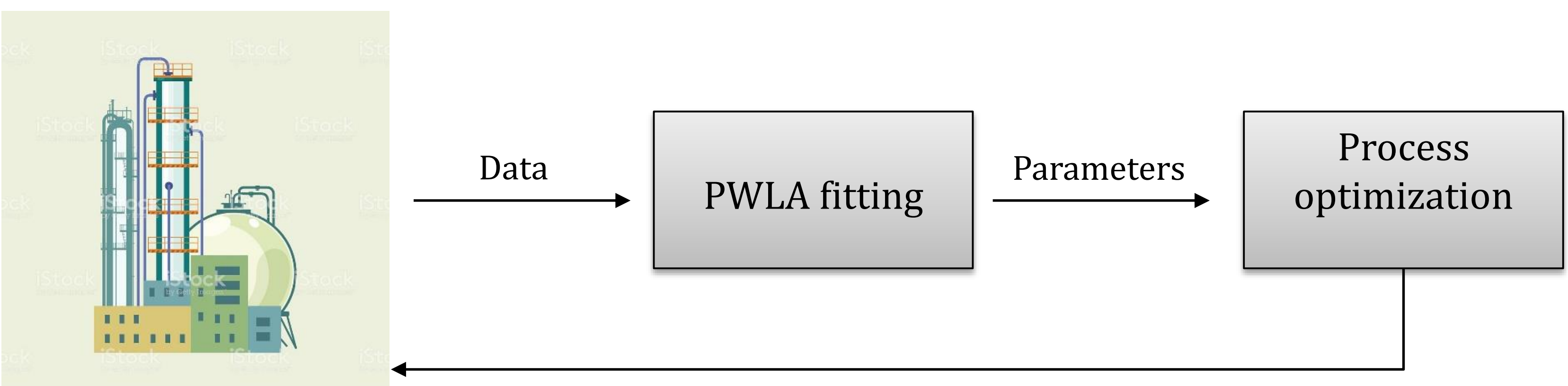
To which segment k data point i belongs

$$h(x) = 0$$

The function is continuous

Mixed Integer Non-Linear Programming = VERY DIFFICULT PROBLEM!

=TAKE VERY LONG TIME TO SOLVE



M0: An MINLP Benchmark

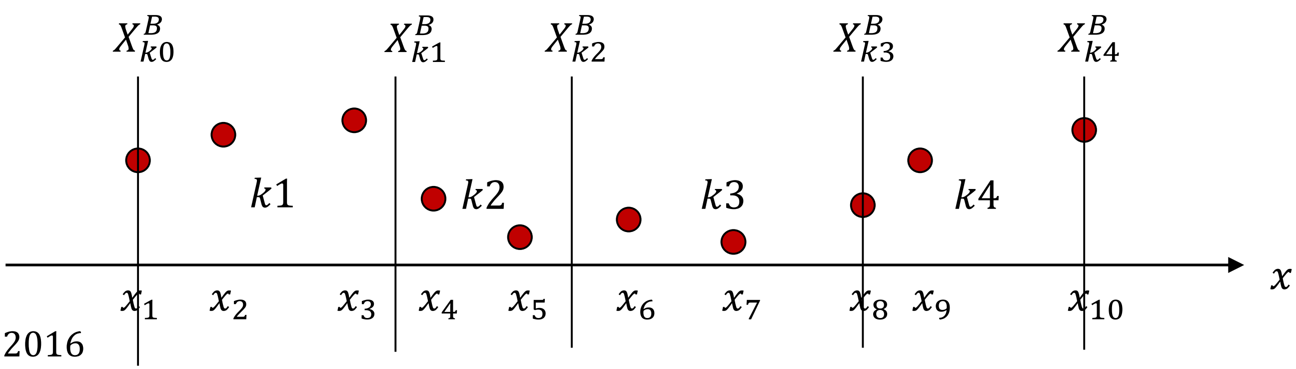
$i \in \mathbf{I}$	data points
$k \in \mathbf{K}$	break points
$\mathbf{K}^I$	$\mathbf{K} \backslash \{0\}$, segments/intervals
$\mathbf{K}^{IM}$	$\mathbf{K} \backslash \{0, \lvert\mathbf{K}\rvert\}$, all but the last segment
X_k^B	x-coordinate of break point k
$Z_{i,k}$	1 if point i belongs to k, 0 otherwise

$$\textstyle\sum_k Z_{i,k} = 1 \qquad i \in \mathbf{I}$$

$$X_{k-1}^B - \Delta(1 - Z_{i,k}) \le x_i \le X_k^B + \Delta(1 - Z_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^I$$

To which segment k data point i belongs

$$X_{k-1}^B \le X_k^B \qquad k \in \mathbf{K}^I$$

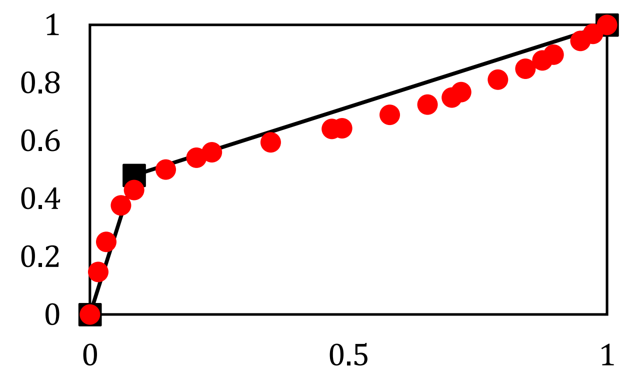


Yang et al., 2016

M0: An MINLP Benchmark

Continuous Variables

A_k	slope of the piecewise function in segment k
B_k	intercept of the piecewise function in segment k
X_k^B	x-coordinate of break point
E_i	absolute error at point i
$Y_{i,k}$	disaggregated approximation of point i



$$Y_{i,k} = x_i A_k + B_k \qquad i \in \mathbf{I}, k \in \mathbf{K}^I$$

Piecewise linear function

$$E_i \geq Y_{i,k} - y_i - bigM(1 - Z_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^I$$

$$E_i \geq y_i - Y_{i,k} - bigM(1 - Z_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^I$$

Error calculation

$$X_k^B A_k + B_k = X_k^B A_{k+1} + B_{k+1} \qquad k \in \mathbf{K}^{IM}$$

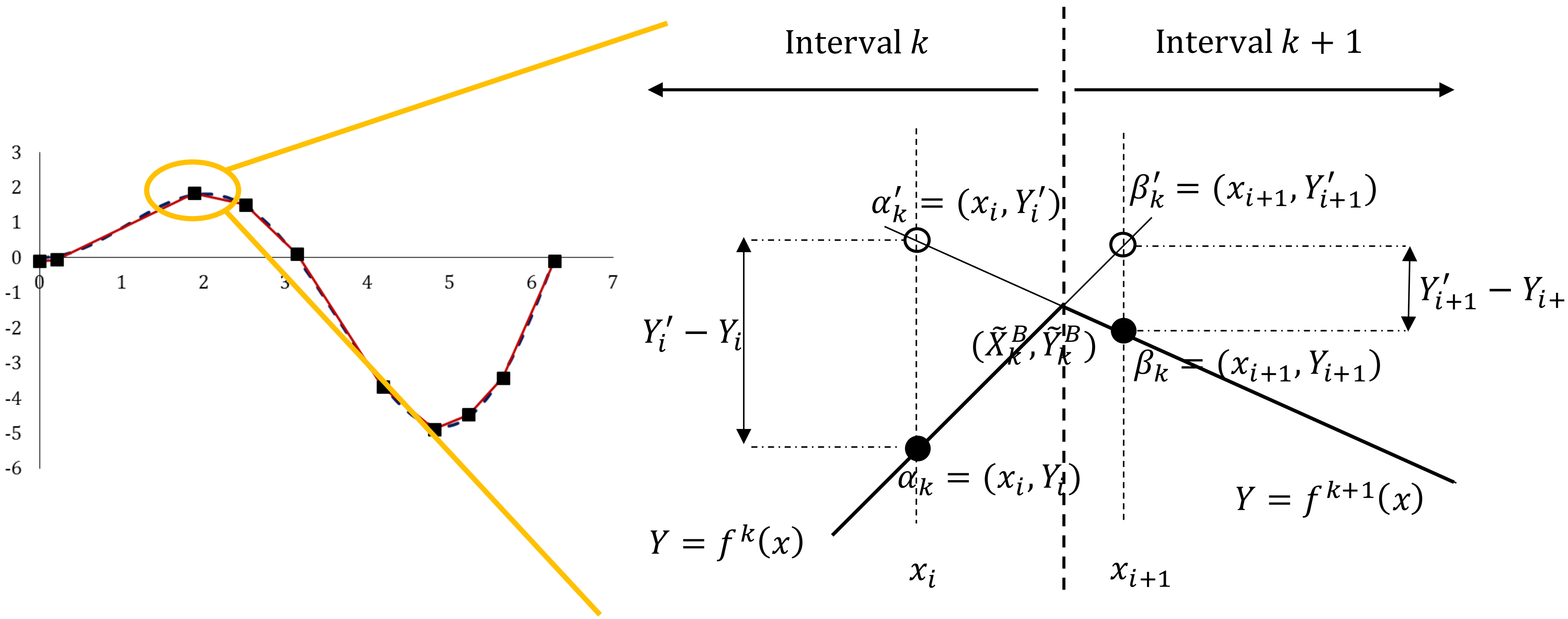
Continuity condition

Bilinear   Bilinear

Proposition 1

Define $i \in \mathbf{I}_k^L$ as the last data point in segment/interval k

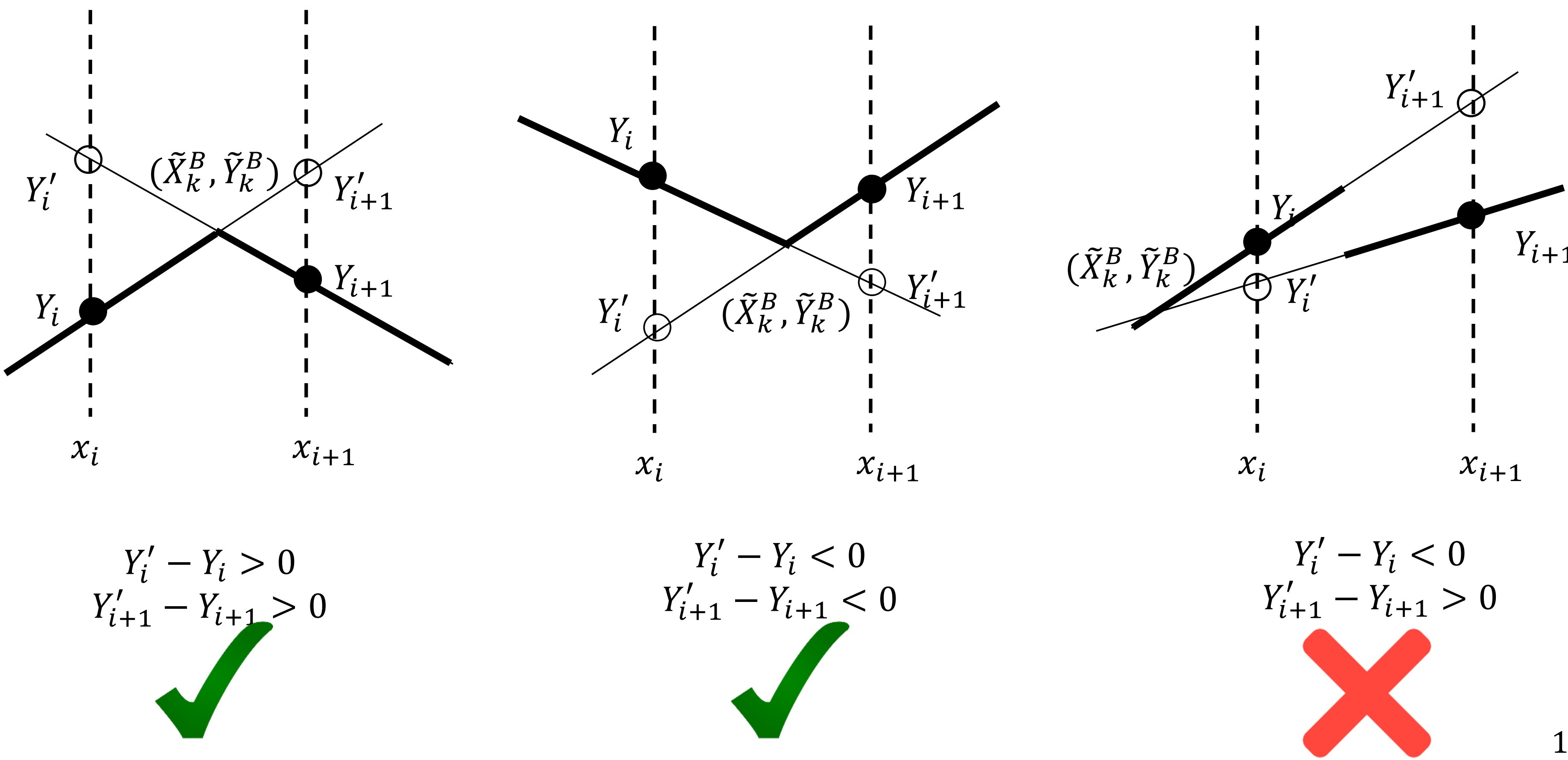
Proposition 1. *For $k \in \mathbf{K}^{IM}$ and $i \in \mathbf{I}_k^L$, when f^k and f^{k+1} have a unique point of intersection $(\tilde{X}_k^B, \tilde{Y}_k^B)$, $(Y_i' - Y_i)(Y_{i+1}' - Y_{i+1}) \geq 0$ if and only if $x_i \leq \tilde{X}_k^B \leq x_{i+1}$.*



Proposition 1

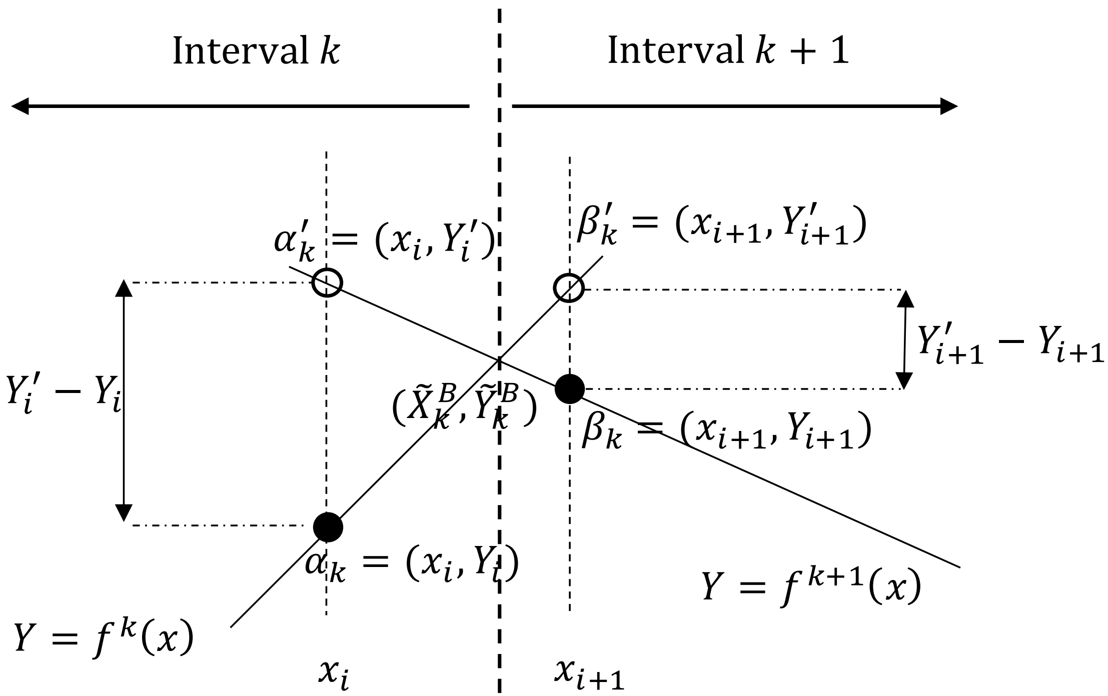
Define $i \in \mathbf{I}_k^L$ as the last data point in segment/interval k

Proposition 1. *For $k \in \mathbf{K}^{IM}$ and $i \in \mathbf{I}_k^L$, when f^k and f^{k+1} have a unique point of intersection $(\tilde{X}_k^B, \tilde{Y}_k^B)$, $(Y_i' - Y_i)(Y_{i+1}' - Y_{i+1}) \geq 0$ if and only if $x_i \leq \tilde{X}_k^B \leq x_{i+1}$.*



Theorem 1

Theorem 1. *If* $(Y_i' - Y_i)(Y_{i+1}' - Y_{i+1})$ *is non-negative for all* $k \in \boldsymbol{K}^{IM}$ *and* $i \in \boldsymbol{I}_k^L$, *then the piecewise linear function is continuous.*



Consider three possible cases:

(1) $f^k(x)$ and $f^{k+1}(x)$ have **one** intersecting point

(2) $f^k(x)$ and $f^{k+1}(x)$ have **infinitely many** intersecting points (identical)

(3) $f^k(x)$ and $f^{k+1}(x)$ have **no** intersecting point (parallel)

Implementing Theorem 1

Binary variables

$Z_{i,k}$	1 if point i is in k
$Z_{i,k}^F$	1 if point i is the **first** point in k
$Z_{i,k}^L$	1 if point i is the **last** point in k

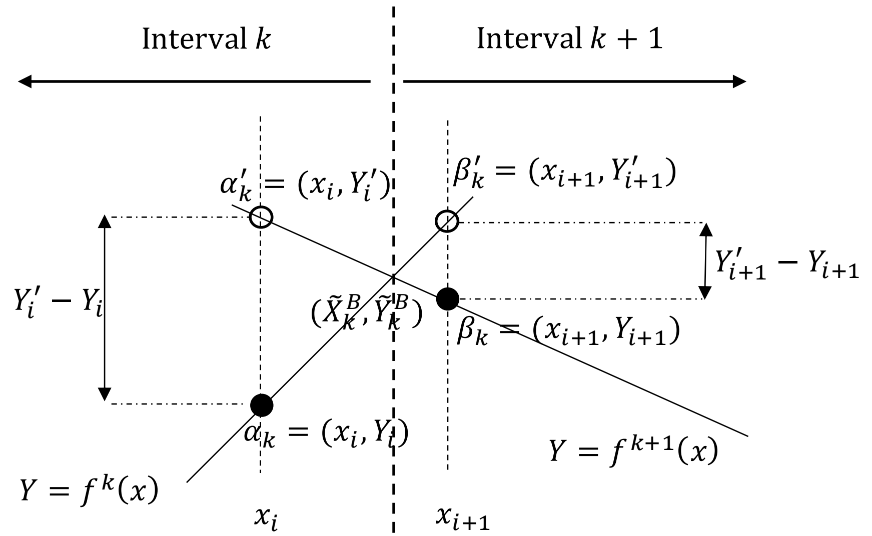
$$Z_{i,k} = Z_{i-1,k} + Z_{i,k}^F - Z_{i-1,k}^L \qquad i \in \mathbf{I}, k \in \mathbf{K}^I$$

$$\textstyle\sum_i Z_{i,k}^F = 1 \qquad k \in \mathbf{K}^I$$

$$\textstyle\sum_i Z_{i,k}^L = 1 \qquad k \in \mathbf{K}^I$$

Theorem 1 can be expressed with the following disjunction

$$\left[\begin{array}{c} Z_{i,k}^L \\ \begin{bmatrix} Y_i' - Y_i \geq 0 \\ Y_{i+1}' - Y_{i+1} \geq 0 \end{bmatrix} \vee \begin{bmatrix} Y_i' - Y_i \leq 0 \\ Y_{i+1}' - Y_{i+1} \leq 0 \end{bmatrix} \end{array}\right] \vee \neg Z_{i,k}^L$$



Enforcing Continuity – Linear Approach

$$x_i A_{k+1} + B_{k+1} - (x_i A_k + B_k) = P^+_{i,k} - P^-_{i,k} \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$x_{i+1} A_k + B_k - (x_{i+1} A_{k+1} + B_{k+1}) = Q^+_{i+1,k+1} - Q^-_{i+1,k+1} \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$P^+_{i,k} \leq bigM(1 - Z^+_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$Q^+_{i+1,k+1} \leq bigM(1 - Z^+_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

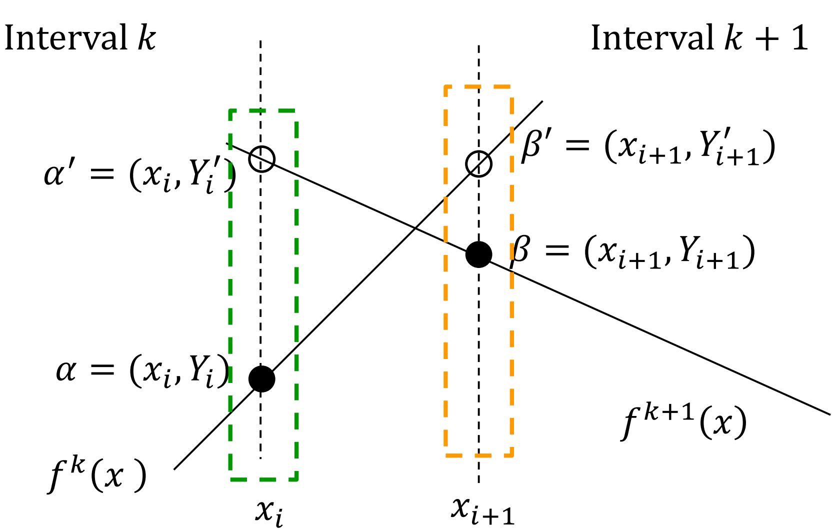
$$P^-_{i,k} \leq bigM(1 - Z^-_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$Q^-_{i+1,k+1} \leq bigM(1 - Z^-_{i,k}) \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$Z^+_{i,k} + Z^-_{i,k} = Z^L_{i,k} \qquad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$$

$$P^+_{i,k} /\ P^-_{i,k} /\ Q^+_{i,k} /\ Q^-_{i,k} \in \mathbb{R}^+$$

$$Z^+_{i,k} / Z^-_{i,k} \in \{0,1\}$$



$$\left[\begin{array}{c} Z^L_{i,k} \\ \left[\begin{array}{c} Y_i' - Y_i \geq 0 \\ Y_{i+1}' - Y_{i+1} \geq 0 \end{array}\right] \vee \left[\begin{array}{c} Y_i' - Y_i \leq 0 \\ Y_{i+1}' - Y_{i+1} \leq 0 \end{array}\right] \end{array}\right] \vee \neg Z^L_{i,k}$$

M1*: MILP/MIQCP Formulation

$$\min \sum_{i\in\mathbf{I}} (E_i)^q$$ ← convex and/or linear

s.t.

$E_i \geq Y_{i,k} - y_i - bigM(1 - Z_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$\sum_k Z_{i,k} = 1 \quad i \in \mathbf{I}$

$E_i \geq y_i - Y_{i,k} - bigM(1 - Z_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$Y_{i,k} = x_i A_k + B_k \quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$X^B_{k-1} \leq X^B_k \quad k \in \mathbf{K}^I$

$Z_{i,k} = Z_{i-1,k} + Z^F_{i,k} - Z^L_{i-1,k} \quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$\sum_i Z^F_{i,k} = 1 \quad k \in \mathbf{K}^I$

$\sum_i Z^L_{i,k} = 1 \quad k \in \mathbf{K}^I$

$x_i A_{k+1} + B_{k+1} - (x_i A_k + B_k) = P^+_{i,k} - P^-_{i,k} \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$x_{i+1} A_k + B_k - (x_{i+1} A_{k+1} + B_{k+1}) = Q^+_{i+1,k+1} - Q^-_{i+1,k+1} \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$P^+_{i,k} \leq bigM(1 - Z^+_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$Q^+_{i+1,k+1} \leq bigM(1 - Z^+_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$P^-_{i,k} \leq bigM(1 - Z^-_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$Q^-_{i+1,k+1} \leq bigM(1 - Z^-_{i,k}) \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$Z^+_{i,k} + Z^-_{i,k} = Z^L_{i,k} \quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

} linear

M1: MILP/MIQCP with Tight Bounds

$$\min \sum_{i \in \mathbf{I}} (E_i)^q$$ ← convex and/or linear

s.t.

$E_i \geq Y_{i,k} - y_i - \mu_i^L (1 - Z_{i,k})$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$\sum_k Z_{i,k} = 1$ $\quad i \in \mathbf{I}$

$E_i \geq y_i - Y_{i,k} - \mu_i^U (1 - Z_{i,k})$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$Y_{i,k} = x_i A_k + B_k$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$X_{k-1}^B \leq X_k^B$ $\quad k \in \mathbf{K}^I$

$Z_{i,k} = Z_{i-1,k} + Z_{i,k}^F - Z_{i-1,k}^L$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^I$

$\sum_i Z_{i,k}^F = 1$ $\quad k \in \mathbf{K}^I$

$\sum_i Z_{i,k}^L = 1$ $\quad k \in \mathbf{K}^I$

$x_i A_{k+1} + B_{k+1} - (x_i A_k + B_k) = P_{i,k}^+ - P_{i,k}^-$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$x_{i+1} A_k + B_k - (x_{i+1} A_{k+1} + B_{k+1}) = Q_{i+1,k+1}^+ - Q_{i+1,k+1}^-$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$P_{i,k}^+ \leq \nu_i^U (1 - Z_{i,k}^+)$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$Q_{i+1,k+1}^+ \leq \nu_i^L (1 - Z_{i,k}^+)$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$P_{i,k}^- \leq \nu_i^U (1 - Z_{i,k}^-)$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

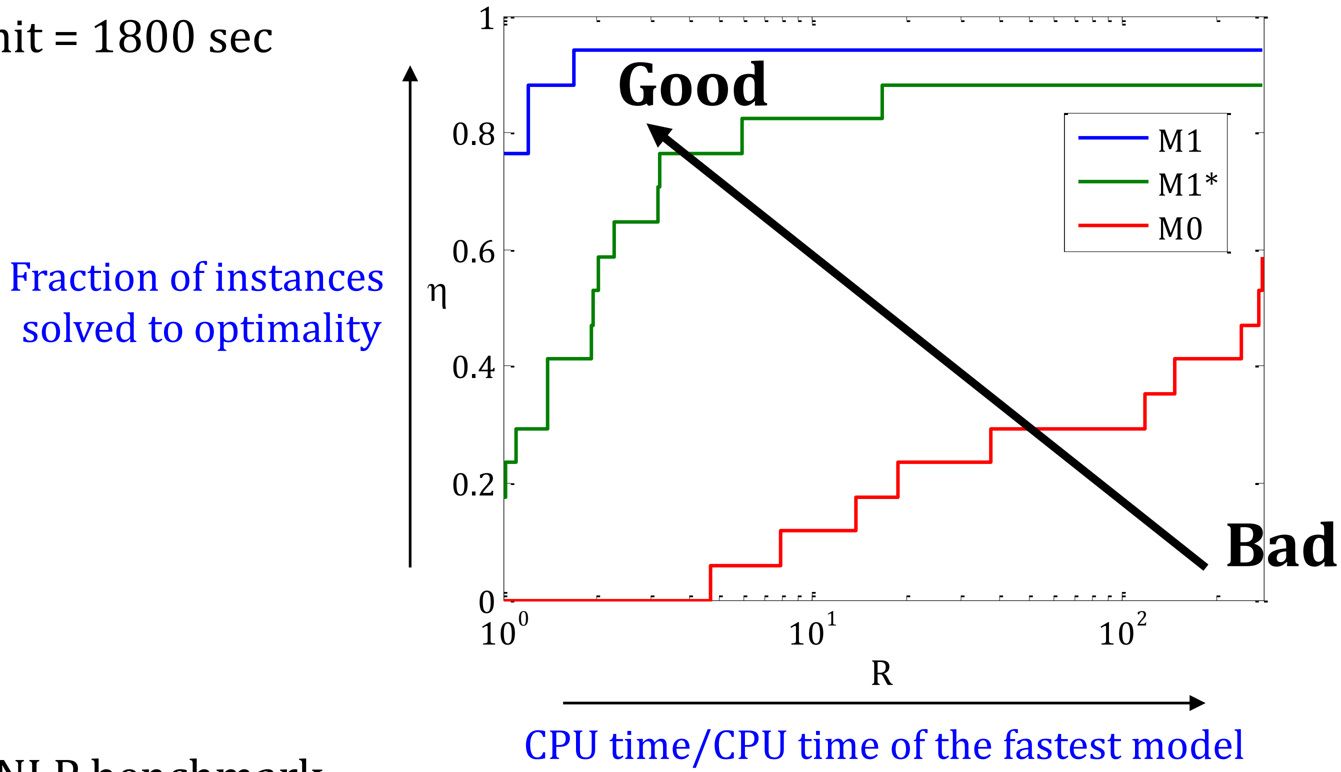
$Q_{i+1,k+1}^- \leq \nu_i^L (1 - Z_{i,k}^-)$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

$Z_{i,k}^+ + Z_{i,k}^- = Z_{i,k}^L$ $\quad i \in \mathbf{I}, k \in \mathbf{K}^{IM}$

} linear

Computational Studies

- 6 data sets from physical science, engineering and social science
- 17 different instances (data sets + break points)
- Quadratic objective function ($q = 2$, SSE)
- CPU limit = 1800 sec



- M0: MINLP benchmark
  - Solved using BARON, ANTIGONE, and SCIP
- M1*: proposed linear model with bigM parameters
- M1: proposed linear model with tight bounds/parameters
  - Solved with CPLEX

M2: Minimize Segments with an Error Tolerance

$$\min \sum_{k \in \mathbf{K}^I} kW_k$$

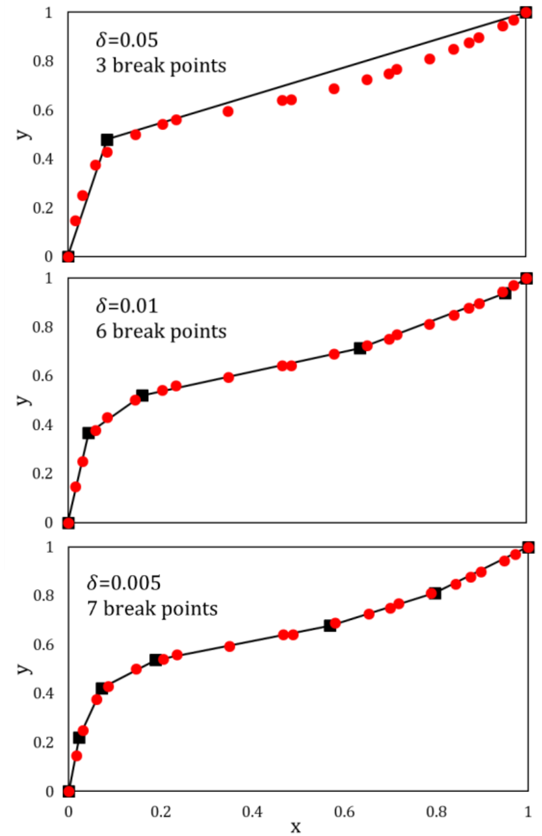
$$s.t. \quad E_i \leq \delta \qquad i \in \mathbf{I}$$

$$\textstyle\sum_i Z_{i,k}^F = W_k \qquad k \in \mathbf{K}^I$$

$$\textstyle\sum_i Z_{i,k}^L = W_k \qquad k \in \mathbf{K}^I$$

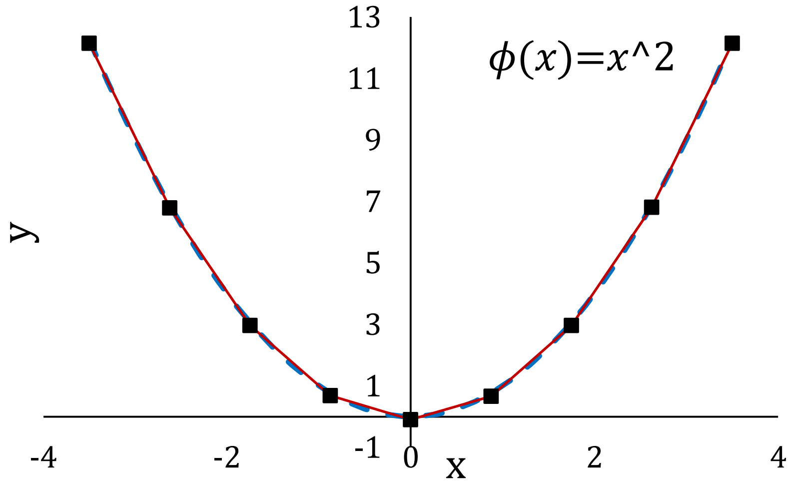
$$W_{k+1} \leq W_k \qquad k \in \mathbf{K}^I$$

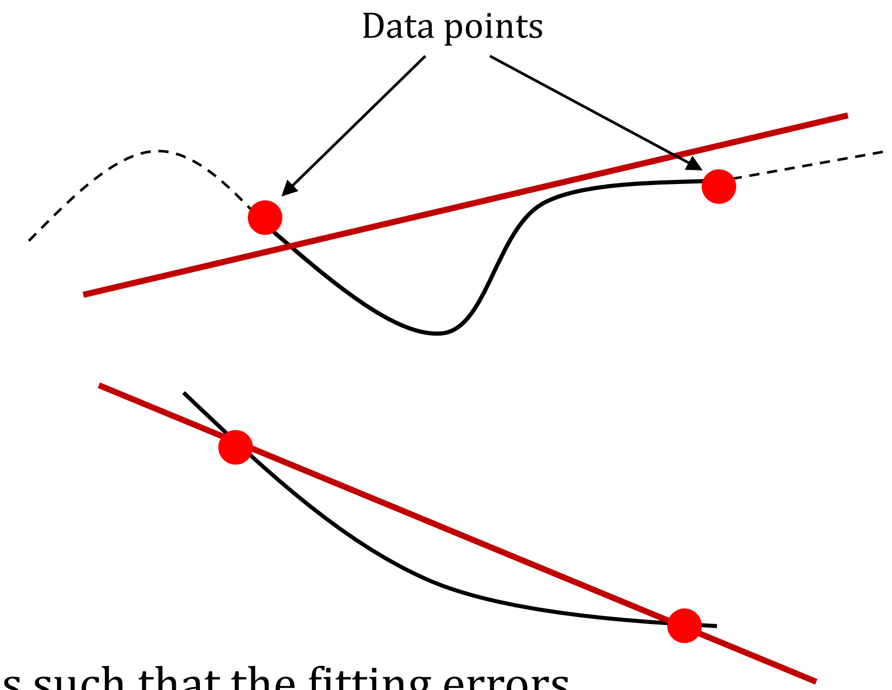
$+$ other constraints from M1



- Error at each point should not exceed specified tolerance
- $W_k \in \{0,1\} = 1$ if segment k is selected, 0 otherwise
- If a segment is **not** selected ($W_k = 0$), then both $Z_{i,k}^F$ and $Z_{i,k}^L$ are deactivated
- Minimize number of segments/break points selected

Approximate Univariate Functions





Goal: find a set of piecewise linear functions such that the fitting errors are within a tolerance δ (δ approximator).

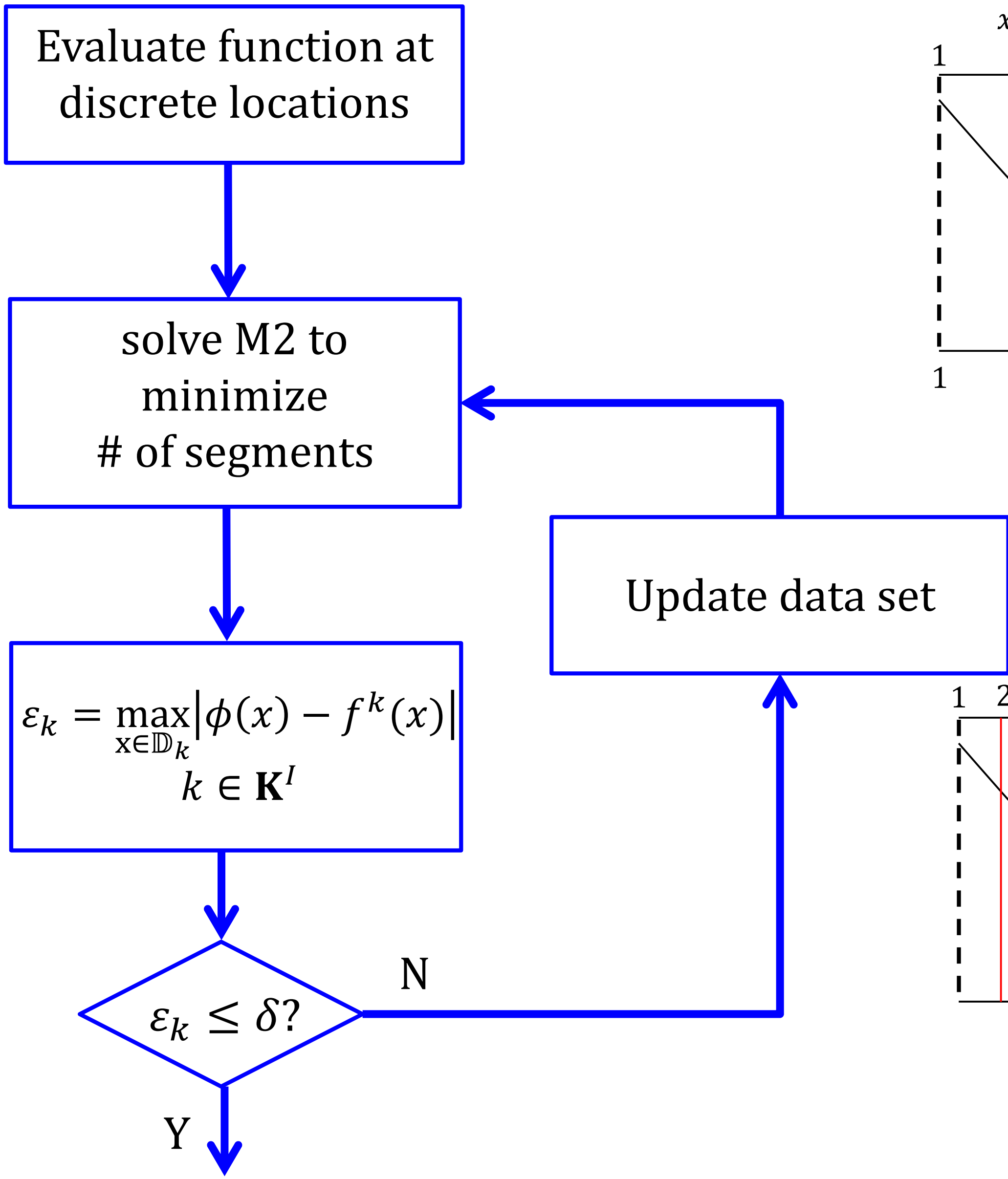
Idea: evaluate the univariate function at discretized points (i.e., data points), and solve M2 to minimize # of break points

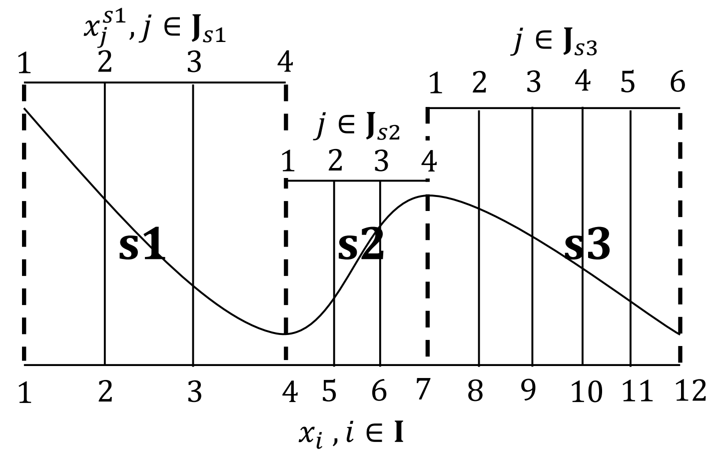
Challenge: ensure the tolerance is satisfied everywhere (semi-infinite programming)

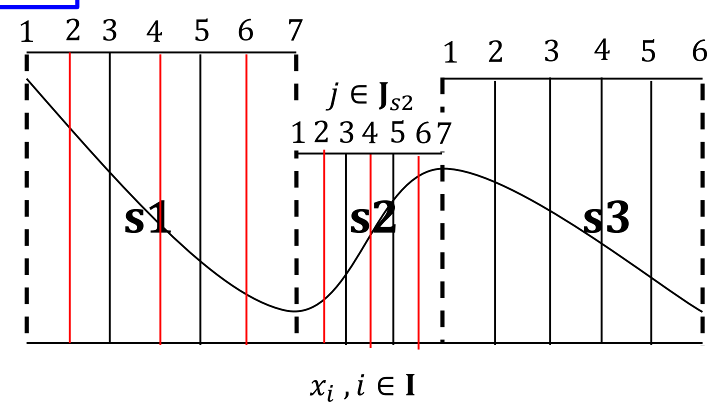
Solution: solve nonconvex NLPs to check if tolerance is indeed satisfied

$$\varepsilon_k = \max_{x \in \mathbb{D}_k} \left| \phi(x) - f^k(x) \right| \qquad k \in \mathbf{K}^I \qquad \boldsymbol{\delta}\textbf{-test}$$

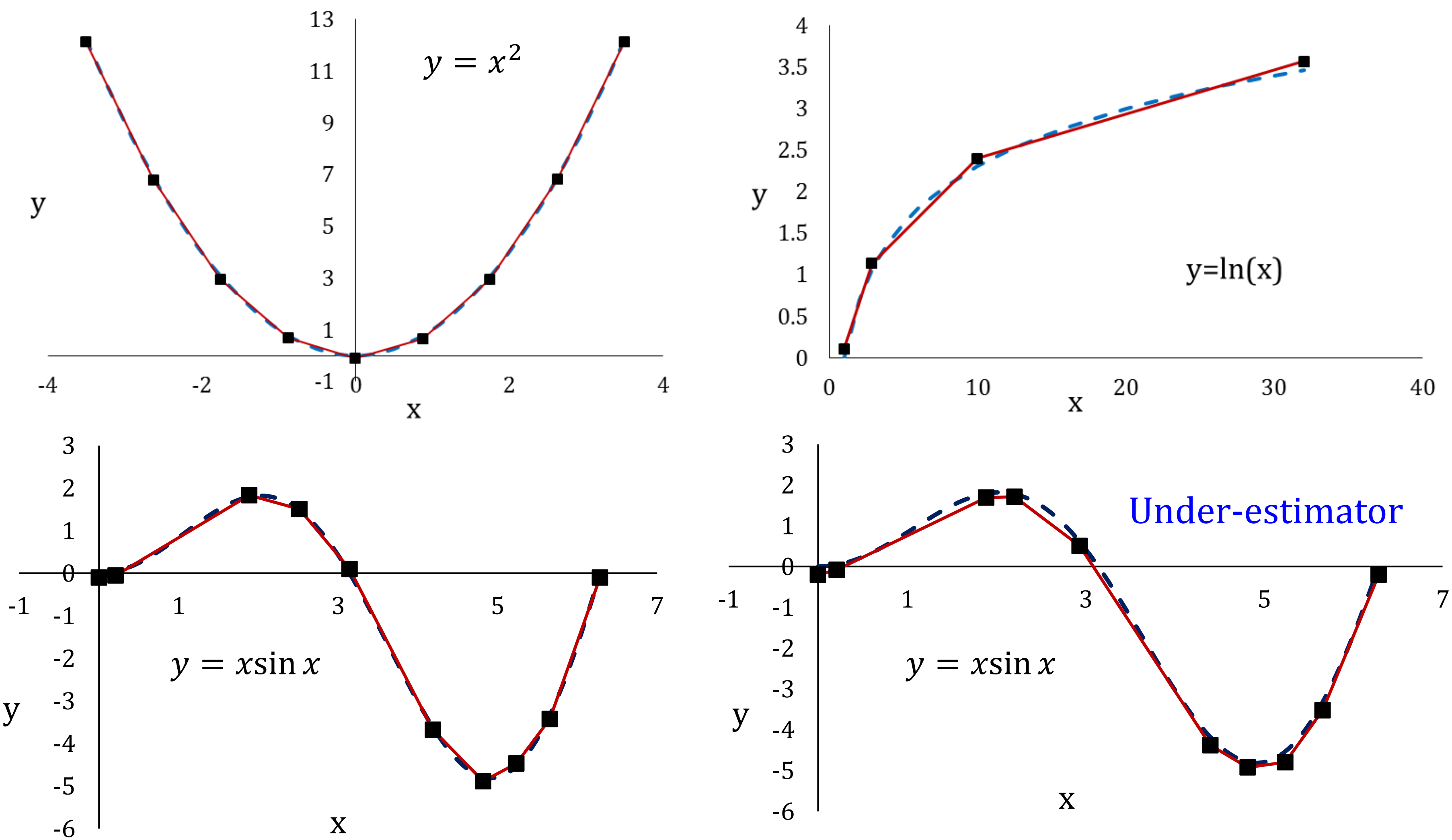
Approximating Univariate Functions







Approximating Univariate Functions



Conclusions

Proposed models

- A mixed-integer model (M1) with convex objective function and linear constraints for error minimization with given number of segments
- An MILP model (M2) that allows finding the minimum number of break points required for a given error tolerance

Continuity condition

- Enforced by a set of mixed-integer linear constraints
- Significantly faster than the nonlinear approach

Approximating functions

- Find δ approximators for univariate functions