Function Smoothing with Applications to VLSI Layout

R. Baldick, A. B. Kahng, A. Kennings and I. L. Markov

baldick@ece.utexas.edu, abk@cs.ucla.edu,

akenning@ee.ryerson.ca, imarkov@cs.ucla.edu

ASP-DAC ’99

Outline of the Talk

- VLSI analytical placement: quadratic and linear wirelength
- GORDIAN-L and Weiszfeld methods
- our contribution
- approximating piecewise linear functions
- symbolic regularization and multivariate functions
- applications to analytical placement
- novel convex delay estimate
- conclusions and ongoing work

VLSI Analytical Placement

- large instances, limited CPU time
- interconnect delay dominates device delay
- wirelength represents interconnect delay and area
- wirelength minimization formulated as **convex** optimization

Quadratic Wirelength Minimization

$$\min_{\mathbf{x}}\{\textstyle\sum_{i>j} a_{ij}(x_i - x_j)^2 : \mathbf{Hx} = \mathbf{b}\}$$

- $\mathbf{x} \equiv$ unknown node positions, $\mathbf{H} \equiv$ linear constraints
- objective function strictly convex
  if all free vertices are reachable from fixed vertices
- objective is differentiable
- unique solution is very easy to find
  - need solve one linear system
- used in `PROUD` (*IEEE Design & Test 1988 — Tsay et al.*)

Linear Wirelength Minimization

$$\min_{\mathbf{x}}\{\textstyle\sum_{i>j} a_{ij}|x_i - x_j| : \mathbf{Hx} = \mathbf{b}\}$$

- better represents net wirelength (*Mahmoud et al. — 1994*)
- **objective is convex, but not strictly convex**
  - multiple solutions
- objective not differentiable
  - classic smooth optimization methods do not apply

GORDIAN-L

$$\min_{\mathbf{x}^\nu}\{\sum_{i>j} \frac{a_{ij}}{|x_i^{\nu-1}-x_j^{\nu-1}|}(x_i^\nu - x_j^\nu)^2 : \mathbf{H}\mathbf{x}^\nu = \mathbf{b}\}$$

- *DAC-91 — Sigl et al.*
- avoids singularities of linear objective
  - approximates it by a series of quadratic objectives
  - uses repetitive quadratic minimization with adjusted weights
- achieves better placements than quadratic minimization
- has been successful in VLSI placement tools

β-regularization and Weiszfeld Method

$$\min_{\mathbf{x}}\{\textstyle\sum_{i>j} a_{ij}\sqrt{(x_i - x_j)^2 + \beta} : \mathbf{Hx} = \mathbf{b}\}$$

- *ISPD-97 — Alpert et al.*
- *regularization:* $|x|$ changed to $\sqrt{x^2 + \beta}$
  - $\beta > 0$ gauges trade-off between quality and run time
- GORDIAN-L is special case $\beta = 0$ of Weiszfeld iteration (*Eckhardt '80*)
- regularization allows for faster numerical methods (see *Proc. ISPD-97*)

Our Contribution

- general technique to approximate continuous but not smooth objectives with good (smooth) functions
- allows application of classic optimization techniques, e.g., Newton method to wide range of objectives
- limited to convex functions with common types of singularities

Elements of Our Approach

- start with symbolic representation of objective functions
- find symbolic fragments responsible for singularities
  they typically are
  - univariate
  - absolute value
  - more general case analysis
- approximate the fragments
- produce new symbolic representation by substituting in approximations of fragments

Univariate Piecewise Linear Functions

- Original function

$$f(x) = \begin{cases} \alpha_1(x - x_0) + C, & \text{if } x \geq x_0, \\ \alpha_2(x - x_0) + C, & \text{if } x < x_0 \end{cases}$$

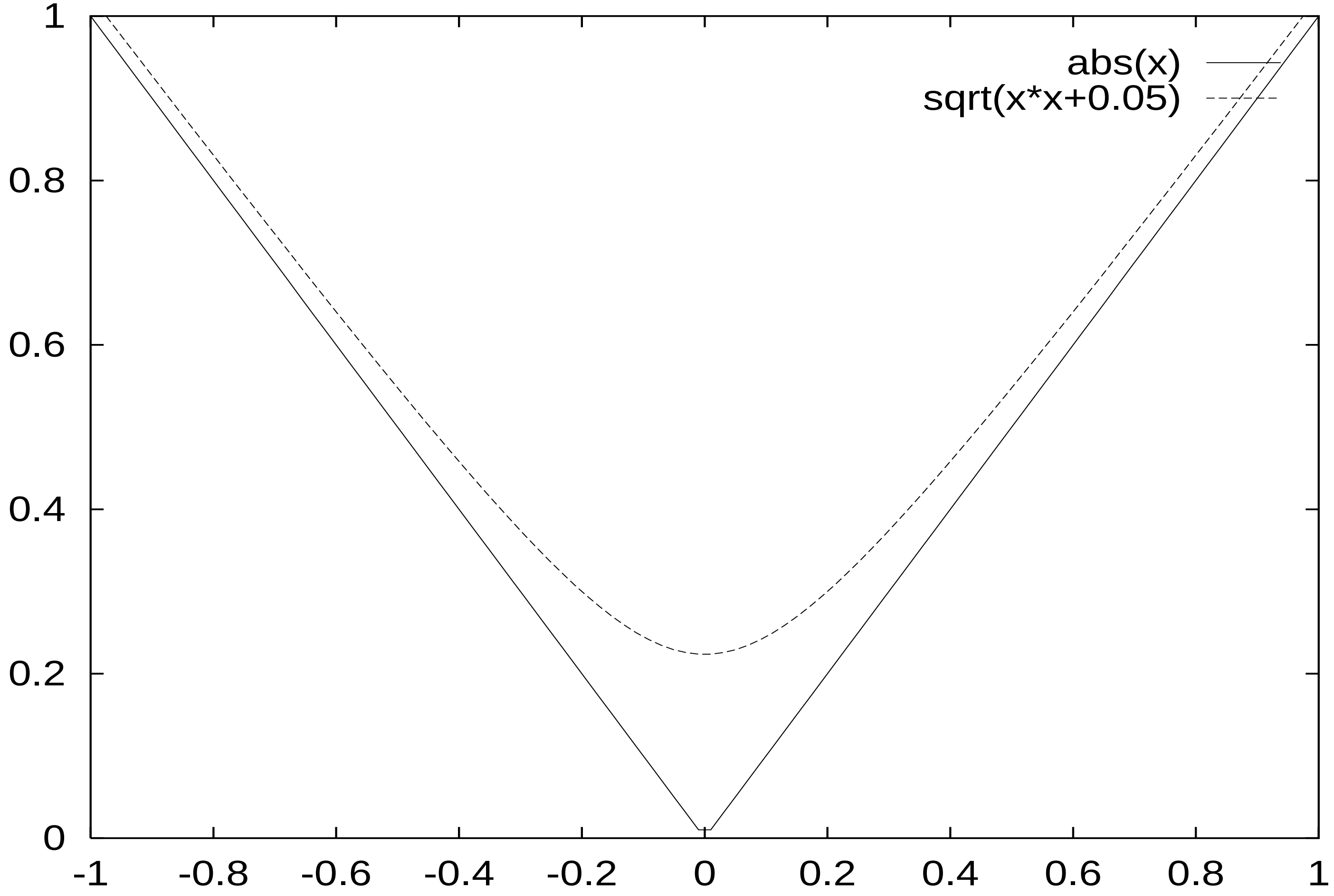
where $\alpha_1 > 0$, $\alpha_2 < 0$, and C is arbitrary

- For $p \geq 2$, the β-regularization of f is defined by:

$$\forall x, f_\beta(x) = C + (|f(x) - C|^p + \beta)^{\frac{1}{p}}$$

Example

For $f(x) = |x|$ and $p = 2$, $f_\beta(x) = \sqrt{x^2 + \beta}$



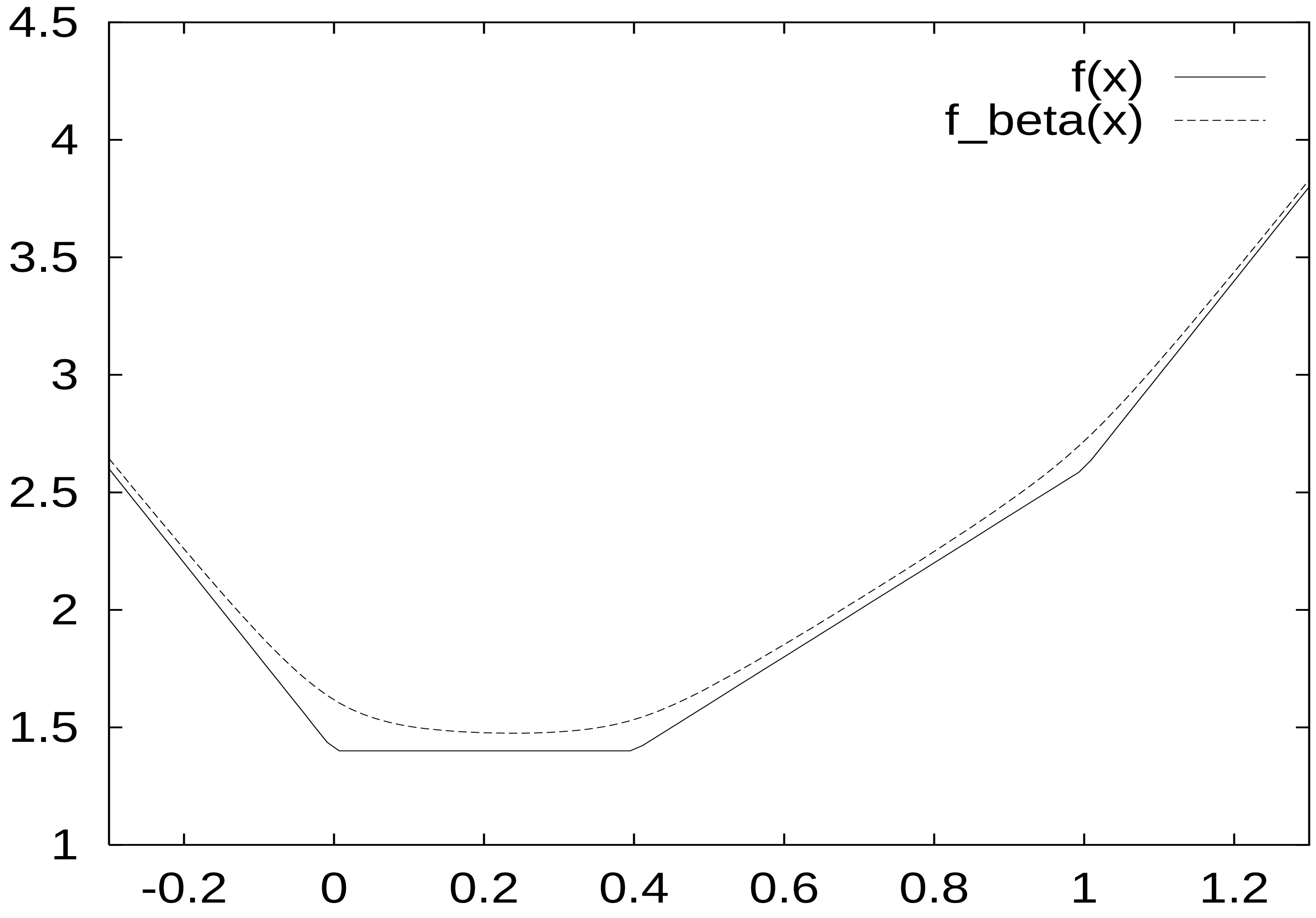
Univariate Piecewise Linear Functions

- **Theorem:** regularized function is smooth and *strictly* convex
- **Theorem:** overestimates original function by at most $\beta^{1/p}$
- **Theorem:** minimizers are almost the same

$$\lim_{\beta \to 0} \inf_{x \in R^n} f_\beta(x) = \inf_{x \in R^n} f(x)$$

- **Corollary:**
  all convex piecewise linear functions can be regularized

Example $p = 2, \beta = 0.01$



Symbolic Regularization

Original function

$$F(x) = \begin{cases} F_1(x - x_0) + C, & \text{if } x \geq x_0, \\ F_2(x - x_0) + C, & \text{if } x < x_0, \end{cases}$$

- $F_1(t), t \leq 0$ and $F_2(t), t \geq 0$
- $F_1(t)$ and $F_1(t)$ continuously differentiable and convex
- $F_1(0) = F_2(0) = 0$, but not necessarily $F_1'(0^+) = F_2'(0^-)$.
- arbitrary C

Regularization $F_\beta(x) = C + (|F(x) - C|^p + \beta)^{\frac{1}{p}}$

Multivariate Functions

- can typically be represented symbolically with univariate functions

- examples of multivariate regularization

  ⋆ $\max\{a,b\} = \frac{a+b+|a-b|}{2}$ regularized with $\frac{a+b+(|a-b|^p+\beta)^{1/p}}{2}$

  ⋆ $\min\{a,b\} = \frac{a+b-|a-b|}{2}$ regularized with $\frac{a+b-(|a-b|^p+\beta)^{1/p}}{2}$

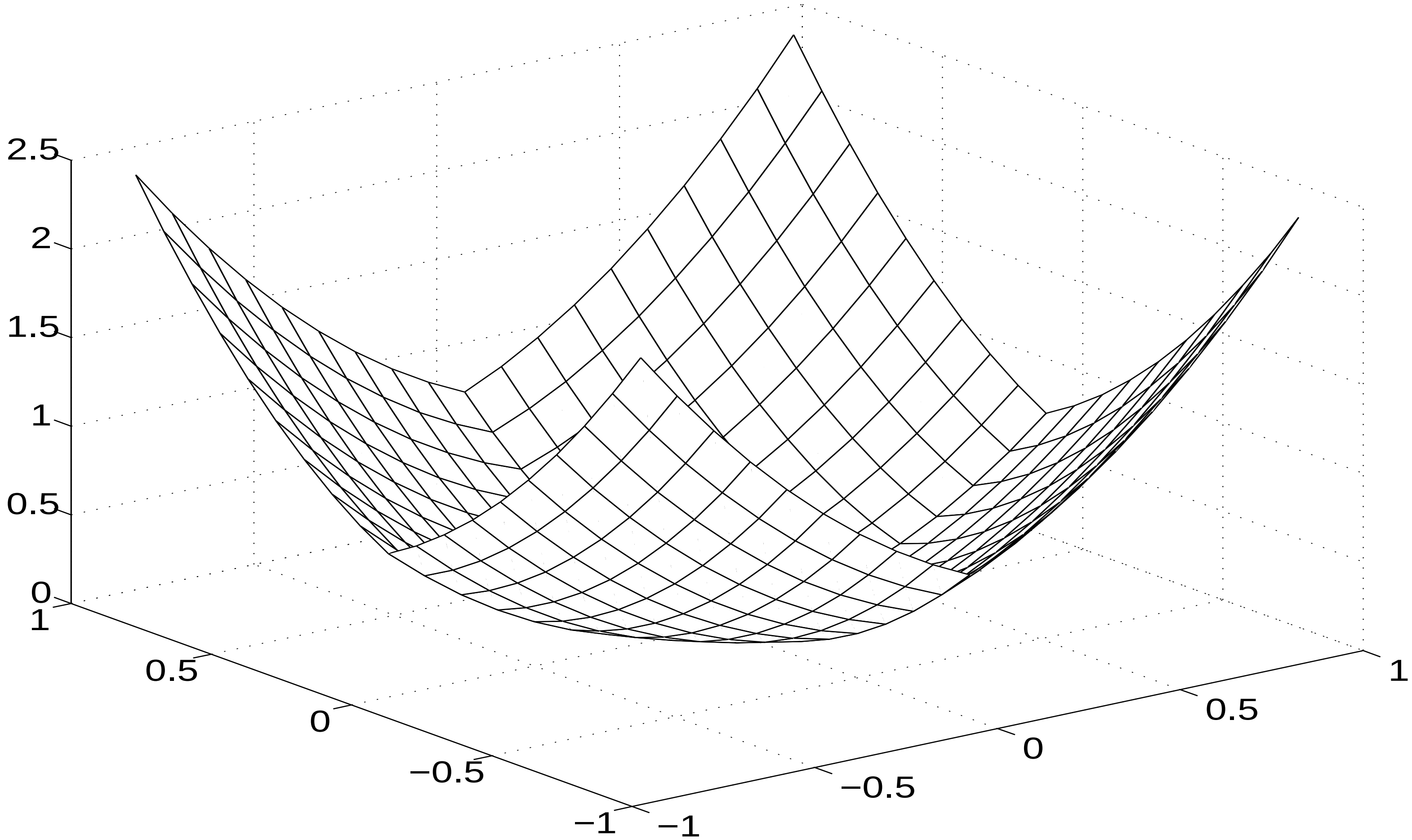
  ⋆ $\max\{a,b,c\}$ regularized via $\max\{a, \max\{b,c\}\}$ (wirelength)

  ⋆ $\max\{(a+b)^2, (a-b)^2\} = a^2+b^2+2|ab|$ regularized with $a^2+b^2+2\sqrt{a^2b^2+\beta}$ (delay)

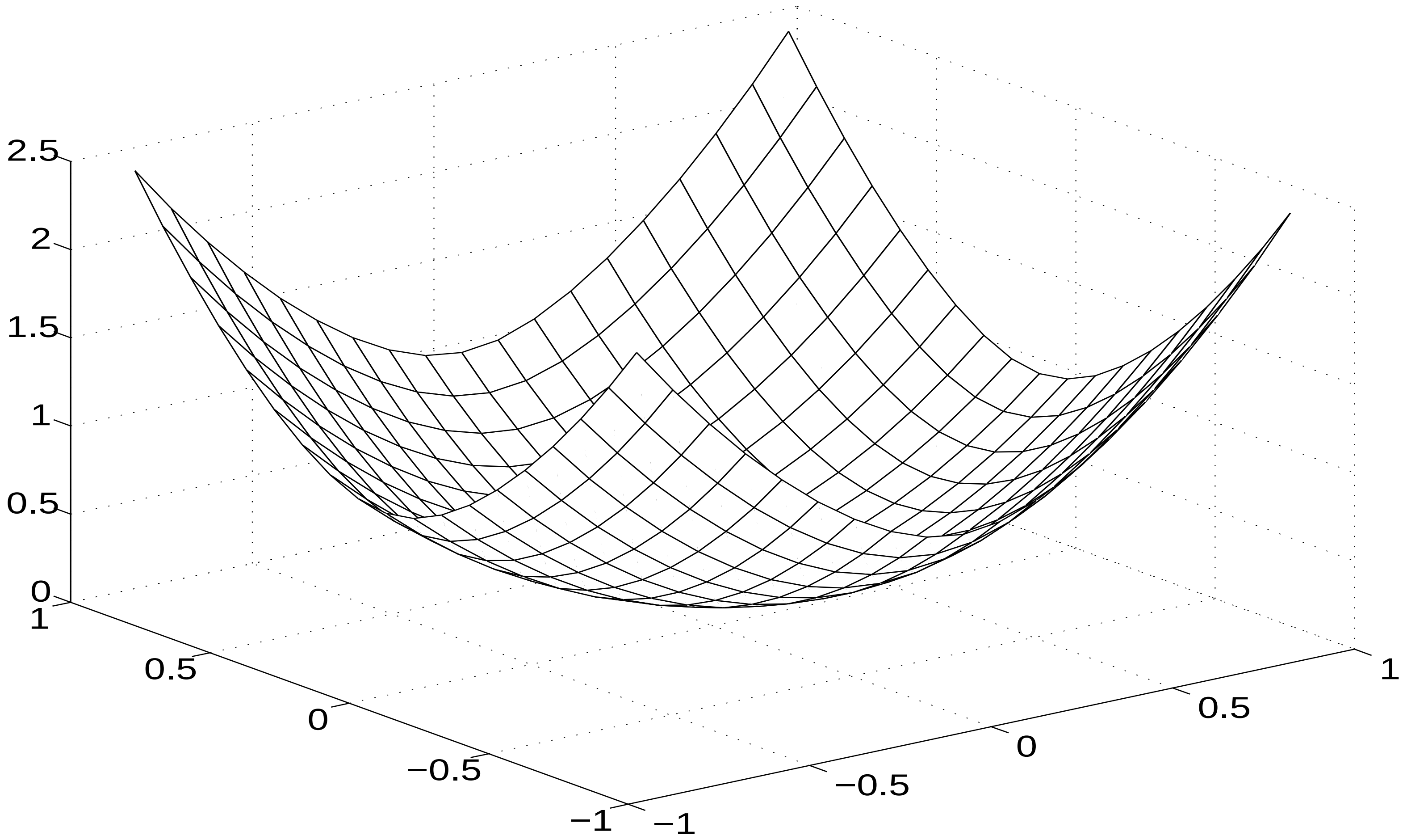
  ⋆ $\max\{0,(t-t_0)\}$ regularized with $\frac{(t-t_0)+\sqrt{(t-t_0)^2+\beta}}{2}$ (excessive for delay)

Original Function

$$\max\{(a+b)^2, (a-b)^2\} = a^2 + b^2 + 2|ab|$$



β-regularization $a^2 + b^2 + 2\sqrt{a^2b^2 + \beta}$



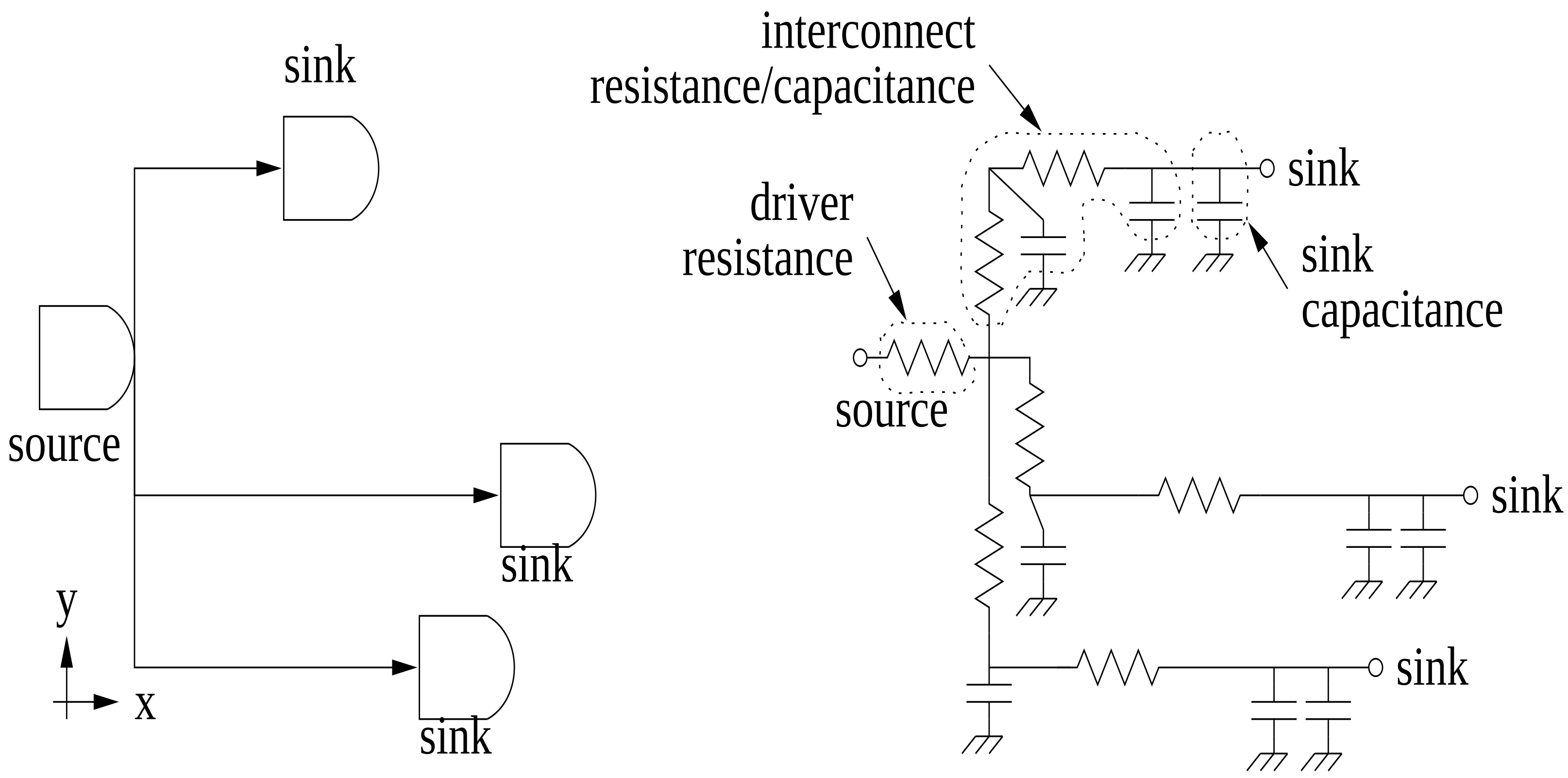
Practical Issues

- minimizers of the regularized function often minimize the original function
- β controls trade-off between speed and quality
  - larger β, easier problem
  - smaller β, better solution quality
- different trade-offs in $\sqrt{|x| + \beta}$ with different scaling of x
  - need scale-independent beta: $\beta = \beta_0 |x_{max}|^p$
  - β_0 — scale-independent component: 10^{-8} through 10^{-1}

Application to Analytical Placement

- Continuous *convex* objective functions of interest
  - estimates of delay and path timing violations
  - wirelengths
  - **many possibilities**
- regularize and apply classic optimization techniques

Lumped L-Segment Model for Multi-pin Nets



Lumped L-Segment Model (cont'd)

- reasonable approximation for 3- to 5-pin nets
- quite exact for 2-pin nets
- larger nets
  - very different Steiner topologies possible
  - good approximations unlikely before Steiner topology known
  - convexity unlikely before Steiner topology known

Components of Source-Sink Delay

(a) $R_d(\sum_s c_x|\chi| + c_y|\gamma| + C_s)$

(b) $(r_x|\chi| + r_y|\gamma|)C_s$

(c) $r_x c_x \chi^2 + r_y c_y \gamma^2 + r_y c_x |\chi||\gamma|$

- C_s — sink capacitance
- R_d — driver resistance
- r_x and r_y — per-unit interconnect resistance
- c_x and c_y — per-unit grounded capacitance
- $\chi = |x_d - x_s|$ and $\gamma = |y_d - y_s|$ —
  x- and y-distance from source to sink

Convex and Smooth Delay Model

(a) $R_d(\sum_s c_x|\chi| + c_y|\gamma| + C_s)$

(b) $(r_x|\chi| + r_y|\gamma|)C_s$

- (a) and (b) are convex, but not smooth
  - take small $\beta > 0$
  - change $|\cdot|$ to $\sqrt{\cdot^2 + \beta}$ in (a) and (b)

Convex and Smooth Delay Model (cont'd)

(c) $r_x c_x \chi^2 + r_y c_y \gamma^2 + r_y c_x |\chi||\gamma|$

- (c) is neither smooth nor convex
  - nonconvexity caused by cross-term $r_y c_x |\chi||\gamma|$
  - cross-term can not be ignored
  - **Proposition:** (c) is convex if $\frac{c_y}{r_y} > \frac{c_x}{4r_x}$
  - given convexity, regularize (c) as

$$\sqrt{(\chi\gamma r_y c_x)^2 + \beta} + \chi^2 r_x c_x + \gamma^2 r_y c_y$$

New Model is Easier

- Previous approaches to timing constraints
  - timing analysis results represented as *net weights* in wirelength objective — loose heuristic
  - explicit nondifferentiable path-based constraints — non-linear optimization with too many constraints ($\approx \#critical\_paths \times av\#edges\_in\_path$)
- Our approach
  - one constraint per critical path

Conclusions

- new method of function smoothing
- provably good
- enables use of classic smooth minimization techniques
- applications to VLSI analytical placement
  - handles linear wirelength estimates
  - handles path delay via new convex delay model
- fewer path-timing optimization constraints, e.g., compared to `PRIME` (*DAC-93 — Hamada, Cheng and Chau*)

Ongoing Work

- *properly multivariate* regularizations
  (i.e. not reducible to *univariate* regularizations)
- new regularizations applicable to VLSI placement
- bounds on suboptimality due to regularization
- variants of classic optimization methods
  best suited to regularized functions
- new placement engines based on regularizations
- preliminary results are encouraging