Effective Optimization Strategies For Large-Scale VLSI Placement

Andrew B. Kahng, UCLA

Andrew Kennings, Cypress, OR

Igor L. Markov, UCLA

Presented by Alex Zelikovsky, GSU

Outline of the Talk

- VLSI analytical placement
- Traditional approach
- New approach
  - main difficulties in ***objective functions***,
    not in *minimization algorithms*
  - ***smoothness*** and ***convexity*** bring powerful
    optimization machinery (Newton-type methods etc)
- Applications to timing-driven placement
- Conclusions

Abstractions

- Circuits as hypergraphs
- Circuit placement ≈ hypergraph placement
  - generalizes Weber problem that seeks a location to minimize sum of distances to fixed points
  - first analytical algorithm due to Weiszfeld 1937
- In general: some vertices fixed, others movable
  - minimize a function of vertex locations
  - modern algorithms exploit Newton's method

VLSI Placement

- Large circuit hypergraphs, limited CPU time
- "*Circuit delay*" to be minimized
- *Interconnect delay* dominates device delay
- Wirelength
  - represents interconnect delay and area
  - *x*- and *y*- components often independent
  - minimization formulated as convex optimization

VLSI Analytical Placement

- Analytical algos useful with non-linear terms present
  - more versatile than min-cut placement
  - faster than Simulated Annealing
- BBox: common objective function (wirelength est.)
  - for one hyperedge: half-perimeter of the bounding box of incident vertices
  - sum over all hyperedges
  - not everywhere differentiable
  - can be complemented by other, e.g., non-linear terms
  - typically dominates other terms

Traditional Approach

- ***Main features:***
  - ***conversion*** of hypergraphs to graphs
    - hyperedges $\rightarrow$ cliques or stars
  - hyperedge-based objective functions substituted by edge-length variants
    - e.g., “quadratic”, “linear” wirelength etc
- Drawbacks:
  - optimal solution not practical; heuristics applied
  - not as fast as leading optimization techniques
  - optimal solutions are not feasible

Quadratic vs Linear Wirelength Minimization

- $\min_{\mathbf{x}} \Sigma_{i>j} a_{ij}(x_i - x_j)^2$ subject to $\mathbf{Hx} = \mathbf{b}$
  - **x** = unknown node positions, **H** = linear constraints
  - **Benefits**: objective function is differentiable and convex
    - Fast unique solution (PROUD [Tsay et al. `88])
  - **Drawback**: questionable relevancy
- $\min_{\mathbf{x}} \Sigma_{i>j} a_{ij}|x_i - x_j|$ subject to $\mathbf{Hx} = \mathbf{b}$
  - **Benefits**: better model of routed wirelength
    - *Mahmoud et al.* `94
  - **Drawbacks**: not differentiable and not strictly convex
    - typically *many* minimizers
    - minimized by slow linear programming or heuristically by *GORDIAN-L* (*Sigl* et al, *DAC '91*)

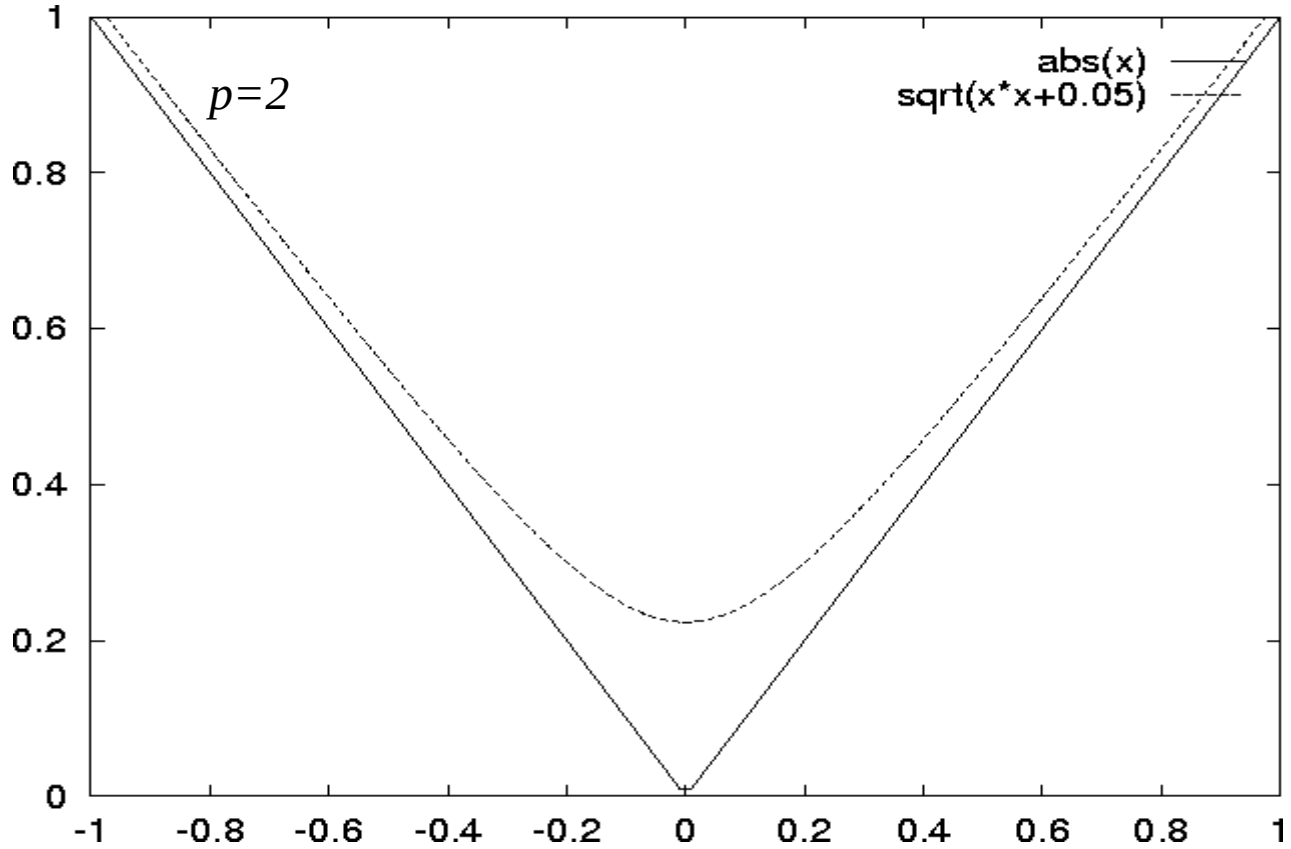
Smooth Approximations

- **Problem**: combine benefits of both objectives
- **Solution**: smooth approximations
  - high accuracy
    - *minimizers* must be “very close” to original minimizers
  - quickly computable (similar to the original function)
  - free of numerical problems
    - partials not too large
    - twice continuously differentiable
- **Problem**: combine accuracy and speed
- **Solution:** parameterized approximations:
  - trade-offs between approximation quality and runtime

β-regularization and "Weiszfeld method"

- *Alpert, Chan, Kahng, Markov et al., ISPD-97*
- Regularization: $|x| \rightarrow (x^2+\beta)^{1/2}$
- $\beta > 0$ gauges trade-off: quality *vs* run time
- GORDIAN-L a special case $\beta = 0$ of Weiszfeld iteration (*Eckhardt '80*)
- Regularization allows for faster numerical methods (see *Proc. ISPD-97)*

Simple Regularization



Symbolic Regularization

- Look at symbolic representation of objective function
- Find symbolic fragments responsible for singularities
- Relevant fragments often are
  - **univariate** functions
  - absolute value or more general **case analysis**
- Hence our interest in **piece-wise linear functions**
- Approximate ("**regularize**") the fragments
  - e.g., send $|x|$ into $(|x|^p+\beta)^{1/p}$
- Produce new symbolic representation

  by substituting in approximations of fragments

– e.g., $min(a,b)=(a+b-|a-b|)/2$ by $(a+b-((a-b)^p+\beta)^{1/p})/2$

Symbolic Regularization

Original function

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

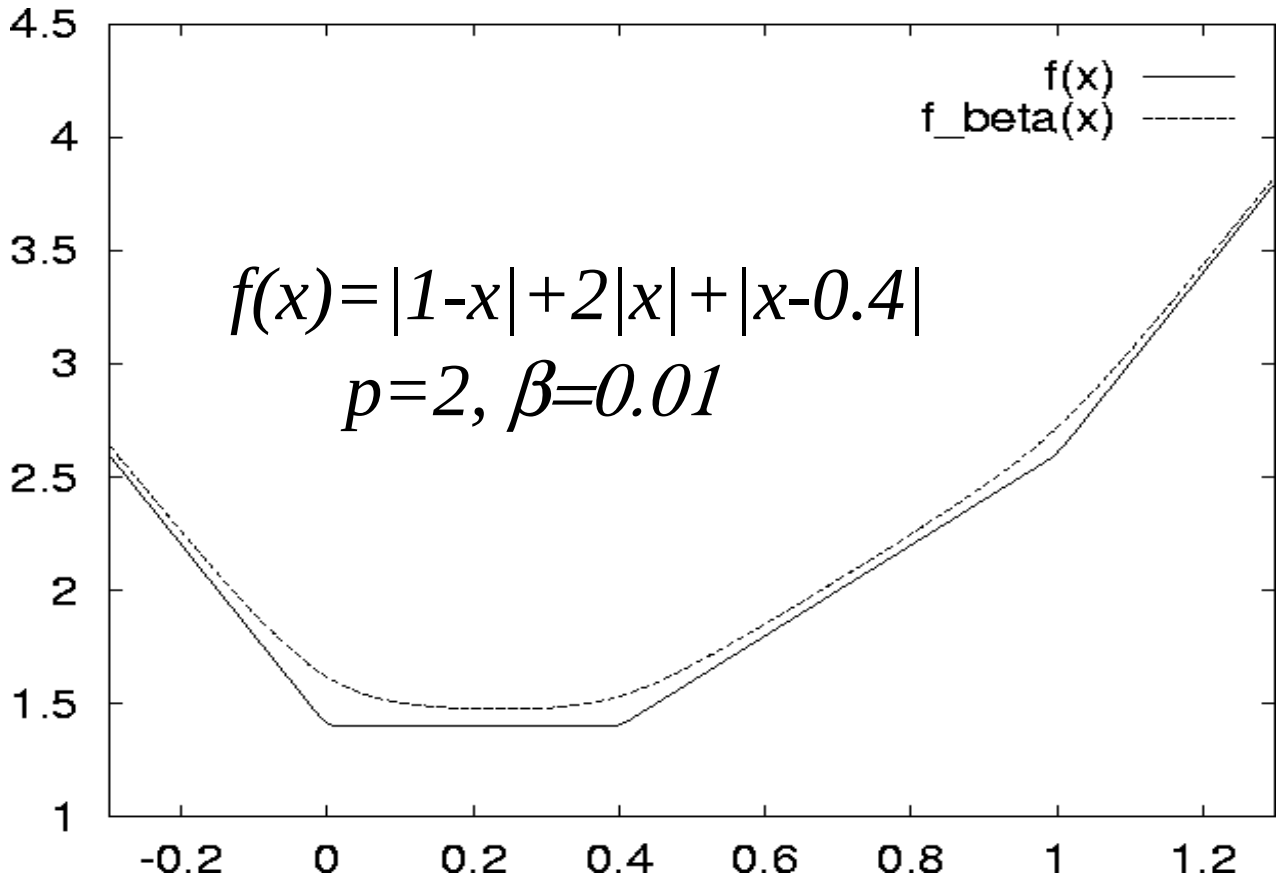
- $F_1(t), t \leq 0$ and $F_2(t), t \geq 0$
- $F_1(t)$ and $F_1(t)$ continuously differentiable and convex
- $F_1(0) = F_2(0) = 0$, but not necessarily $F_1'(0^+) = F_2'(0^-)$.
- arbitrary C

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

Univariate Piece-Wise Linear Functions

- Assume orig. function convex (not nec. *strictly* convex!)
- Present original function as sum of *convex PWL* funcs having at most one cusp at $x=x_0$ (simple functions)
- Regularize simple functions one by one
  - send $f(x)$ into $f_\beta(x)=f(x_0)+((f(x)-f(x_0))^p+\beta)^{1/p}$
  - slightly different regularization if $f(x)$ has plateaux
- **Theorem:** regularized func smooth and *strictly* convex
- **Theorem:** overestimates original func by at most $\beta^{1/p}$
- **Theorem:** minimizers converge $lim_{\beta\to 0}\ inf_x\ f_\beta(x) = inf_x\ f(x)$

Example of β-regularization



Multivariate Functions

- can typically be represented symbolically with univariate functions
- examples of multivariate regularization
  - $\asymp$ $\max\{a,b\} = \frac{a+b+|a-b|}{2}$ regularized with $\frac{a+b+(|a-b|^p+\beta)^{1/p}}{2}$
  - $\asymp$ $\min\{a,b\} = \frac{a+b-|a-b|}{2}$ regularized with $\frac{a+b-(|a-b|^p+\beta)^{1/p}}{2}$
  - $\asymp$ $\max\{a,b,c\}$ regularized via $\max\{a,\max\{b,c\}\}$ (wirelength)
  - $\not\asymp$ $\max\{(a-b)^2,(a-b)^2\} = a^2-b^2-2|ab|$ regularized with $a^2-b^2-2\sqrt{a^2b^2-\beta}$ (delay)
  - $\asymp$ $\max\{0,(t-t_0)\}$ regularized with $\frac{(t-t_0)+\sqrt{(t-t_0)^2+\beta}}{2}$ (excessive for delay)

Practical Issues

- Minimizers of the regularization often minimize the original function (no error)
- Different trade-offs in $(x^2+\beta)^{1/2}$ with different scaling of x
  - $\beta = \beta_0 |x|^2$
  - β_0 – scale-independent component
  - 10^{-8} through 10^{-1}

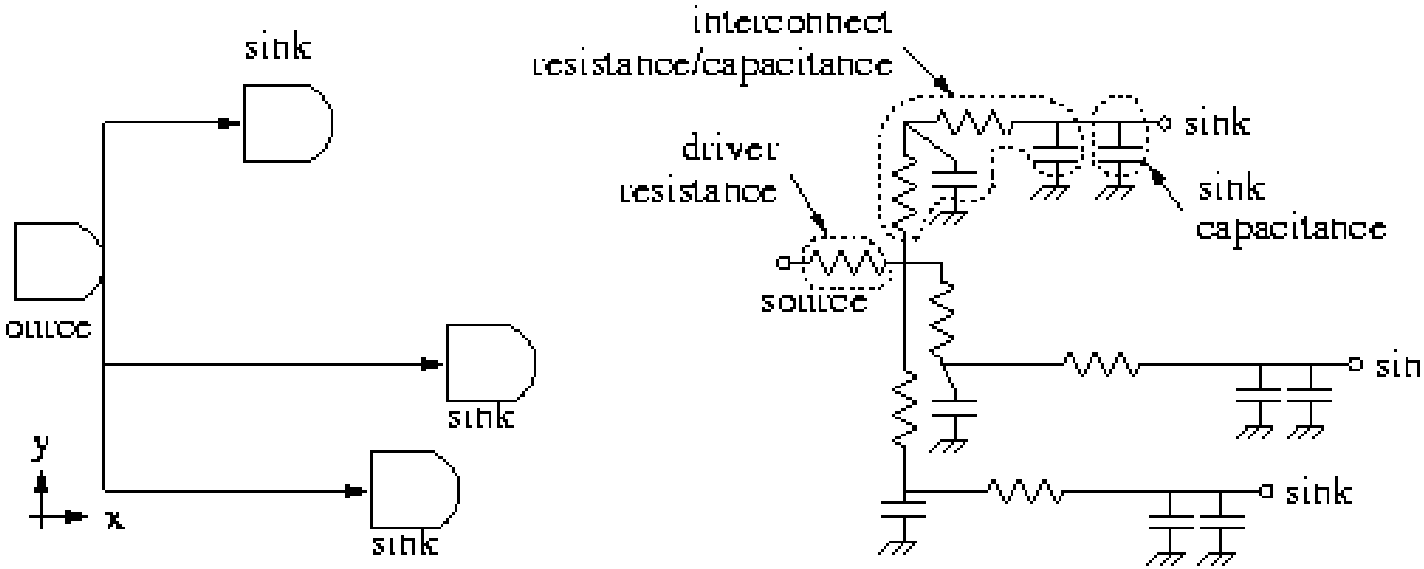
Benefits of New Approach

- Overcome difficulties in objective functions
  - lack of smoothness
  - lack of strict convexity
- Overcome difficulties in minimization algorithms
  - no need to create specialized algorithms
  - can apply Newton-type methods
    - Primal-Dual

Outline of the Talk

- VLSI analytical placement
- Traditional approach
- New approach
  - main difficulties in ***objective functions***,
    not in *minimization algorithms*
  - ***smoothness*** and ***convexity*** bring powerful
    optimization machinery (Newton-type methods etc)
- **Applications to timing-driven placement**
- Conclusions

Lumped-L Elmore Delay



Three components of Elmore delay

- source resistance times all downstream capacitance
- interconnect resistance times sink capacitance
- interconnect resistance times interconnect delay

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

r_x, r_y - per-unit resistance

c_x, c_y - per-unit capacitance

χ, γ – source-to-sink distances

Delay Approximation

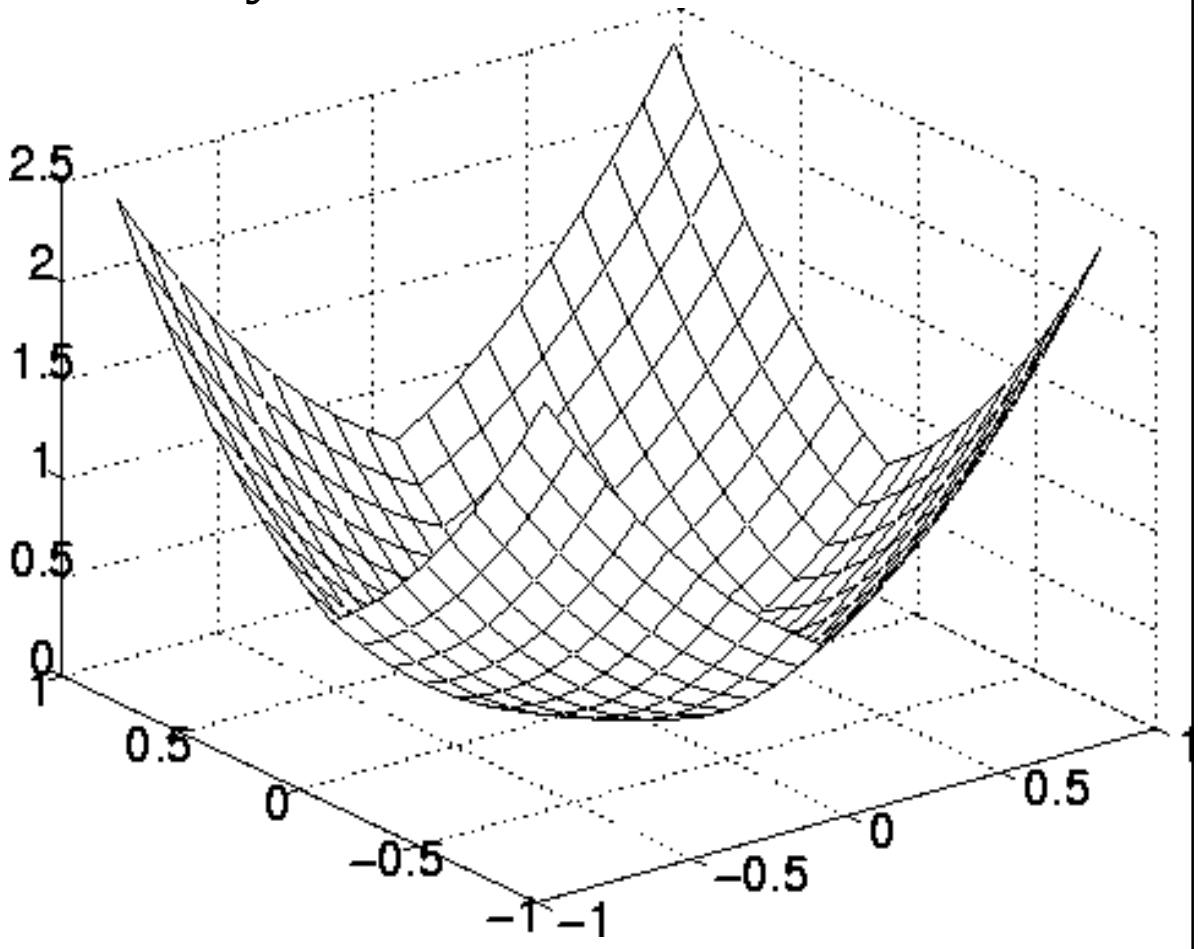
- Third term of Elmore delay neither smooth nor convex:

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

- We establish convexity when $4c_y/r_y > c_x/r_x$, which is actually the case

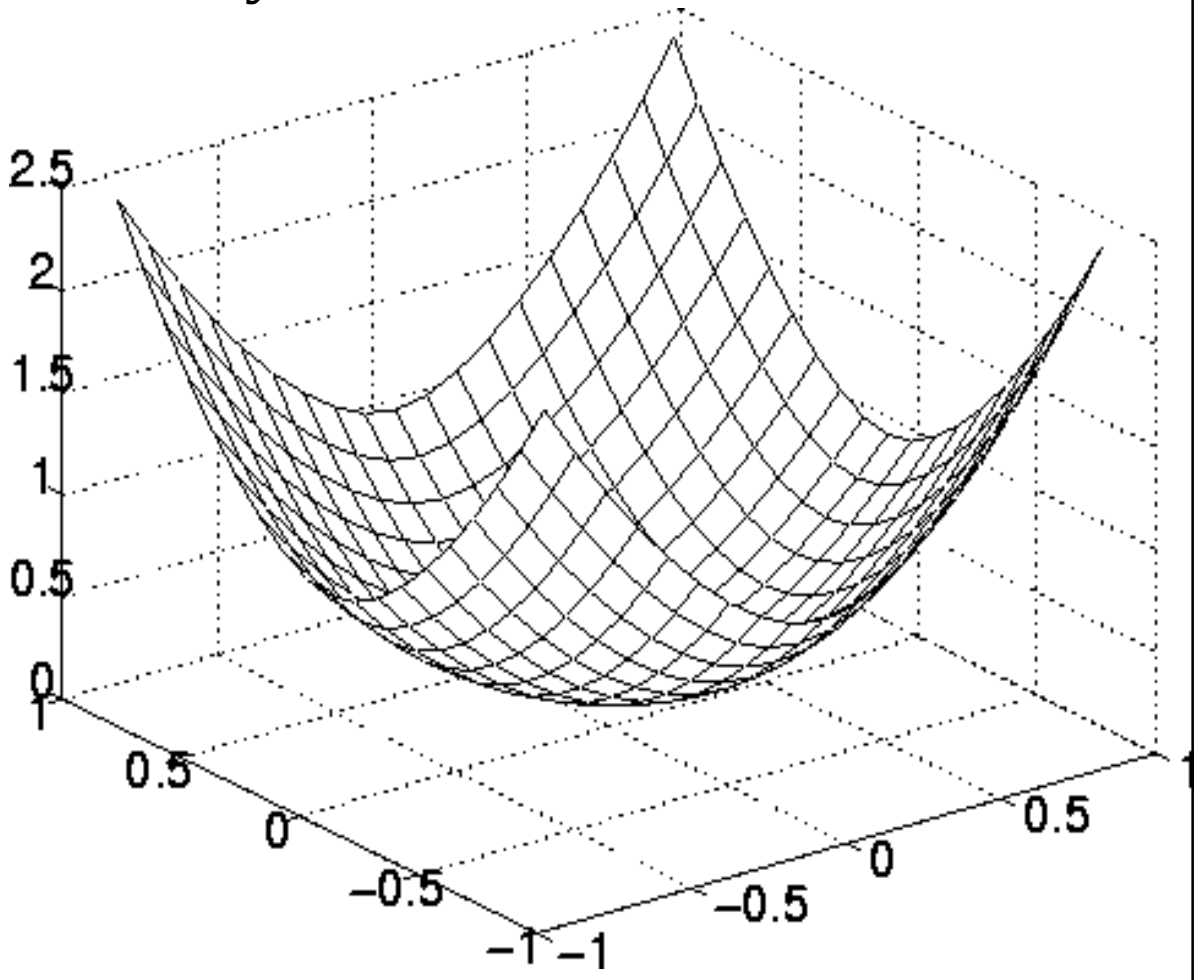
Delay Function Has Cusps

- $r_x c_x \chi^2 + r_y c_y \gamma^2 - r_y c_x |\chi\gamma|$
- Now need to regularize
- Non-linear cusps to be eliminated



Regularization Removes Cusps

- $r_x c_x \chi^2 + r_y c_y \gamma^2 - r_y c_x ((\chi\gamma)^2+\beta)^{1/2}$
- Regularized



- No cusps

Implementation Experience

- Experiments with industrial circuits
  - use different algorithms
  - evaluate the linear (BBox) WL
  - compare to the optimal WL produced by LP
- Quadratic WL minimization
  - can be 10x off, but better than random
  - takes very little time
  - can be used for initial solutions
- Minimizing regularization
  - can get within 20% in very short time
  - work underway to improve quality w/o sacrificing runtime

Conclusions

- A systematic approach to a class of (convex) circuit design problems:
  - mathematical properties of the objective more important than specialized minimization algos
  - have to use *best* general-purpose optimization algos for better results + reportable, replicable
- Uncovered *hidden convexity*, allowing to use new techniques in timing-driven placement
- Experimental confirmation