Gate Sizing by Lagrangian Relaxation Revisited

Jia Wang, Debasish Das, and Hai Zhou
Electrical Engineering and Computer Science
Northwestern University
Evanston, Illinois, United States

October 17, 2007
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Gate Sizing

- A mathematical programming formulation.
  - Timing constraints: delays, arrival times.
  - Objective function: clock period, total weighted area, etc.
- Trade off performance and cost.
  - Optimize performance.
  - Optimize cost under performance constraint.
TILOS [Fishburn & Dunlop, ’85], [Sapatnekar et al., ’93], etc.

- Transform sizing into a convex programming problem.
  - Assume convex delays (after variable transformation).
  - Convexity through geometric programming commonly.
- Apply general convex optimization techniques.
  - Studied for decades.
  - High running time.
Previous Works – Lagrangian Relaxation

[Chen, Chu, & Wong, ’99], [Tennakoon & Sechen, ’02, ’05]

- Special structure:
  - timing constraints are system of difference inequalities.
- The Lagrangian dual problem is simplified.
  - Objective function (Lagrangian dual function):
    - available through Lagrangian subproblems.
  - Constraints: flow conservation on Lagrangian multipliers.
Previous Works – Lagrangian Relaxation

[Chen, Chu, & Wong, ’99], [Tennakoon & Sechen, ’02, ’05]

- Special structure:
  - timing constraints are system of difference inequalities.
- The Lagrangian dual problem is simplified.
  - Objective function (Lagrangian dual function): available through Lagrangian subproblems.
  - Constraints: flow conservation on Lagrangian multipliers.
- Lagrangian dual function is not differentiable in general.
  - Apply subgradient optimizations.
- Difficult to choose good initial solution and step sizes.
  - Pre-processing can help choosing initial solutions for some delay models.
  - No universal pre-processing method for all convex delays.
Contribution

- Combine gate sizing with sequential optimization.
- Revisit Lagrangian relaxation.
  - Correct misunderstandings.
  - Check primal feasibility in dual problems.
- Identify a class of problems with differentiable dual functions.
  - The DualFD algorithm: use gradient and network flow.
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Generalized Convex Sizing (GCS)

Minimize \[ C(x) \]
\[ \text{s.t.} \quad t_i + d_{i,j}(x) - t_j \leq 0, \forall (i,j) \in E, \]
\[ x \in \Omega. \]

- \( G = (V, E) \): a directed graph, the system structure.
- \( x \): the system parameters belonging to
  \[ \Omega \overset{\Delta}{=} \{ x : l_k \leq x_k \leq u_k, \forall 1 \leq k \leq n \}. \]
- \( t \): (relative) arrival times.
- \( C(x), d_{i,j}(x) \): objective function and delays, twice differentiable and convex.
The GCS Formulation

- Timing specification in sequential circuits.

Feasible timing $\iff$ No positive cycle

- Prefer edge delays to vertex delays for accuracy.
  - Timing arcs in one gate/cell.
  - Rise/fall delay and slew.

- Flexibility in choosing $x$.
  - Logarithms of sizes (gate, wire, transistor, etc.) for Elmore and posynomial delays.
  - Clock skews and clock period.
GCS is Convex

- GCS formulation is convex.
  - Objective function and feasible set are convex.
  - Not necessary to establish convexity through geometric programming.
GCS is Convex

- GCS formulation is convex.
  - Objective function and feasible set are convex.
  - Not necessary to establish convexity through geometric programming.
- Non-convex formulations can be transformed into convex ones.
  - Properties of convex formulations are not necessarily hold for equivalent non-convex ones.
Proper GCS Problems

Definition
A GCS is proper iff:

\[ \forall x \in \Omega, \forall z \neq 0, z^T H_C(x)z \neq 0. \]

- With differentiable dual functions (shown later).
  - \( d_{i,j}(x) \) are irrelevant.
- Optimizing total positive weighted area is proper.
Minimize total area under performance bound while allowing clock skew optimization.

Handle long path conditions (setup condition).

Assume post-processing for repairing of violated short path conditions (hold condition).

Not proper: \( \frac{\partial C}{\partial s} = 0, \forall \text{clock skew variable } s \).

Cancel \( s \). Transform to a proper one.
Simultaneous Sizing and Clock Skew Optimization
Make a Non-Proper Problem Proper

- Constraints concerning clock skew $s_k$:

\[
(t_I + s_k \leq t_{Q_k}) \land (t_{D_k} - s_k \leq t_O) \land (s_k^- \leq s_k \leq s_k^+) \]

- Cancel $s_k$: $[t_{D_k} - t_O, t_{Q_k} - t_I] \cap [s_k^-, s_k^+] \neq \emptyset$. Equivalently:

\[
(t_{D_k} - s_k^+ \leq t_O) \land (t_I + s_k^- \leq t_{Q_k}) \land (t_{D_k} - t_{Q_k} \leq T) \]
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Lagrangian Relaxation Overview

- Lagrangian function:

\[ L^*(x, t, f) = C(x) + \sum_{(i,j) \in E} f_{i,j}(t_i + d_{i,j}(x) - t_j) \]

- Lagrangian subproblem:

\[ L(f) = \inf \{ L^*(x, t, f) : x \in \Omega, t \in \mathbb{R}^{|V|} \} \]

- Lagrangian dual problem (D-GCS):

Maximize \[ L(f) \]

s.t. \[ f \in \mathcal{N}. \]

\[ \mathcal{N} \text{: non-negative } f. \]
Duality Gap

\[ P \overset{\Delta}{=} \inf \{ C(x) : x \in \mathcal{X} \}, \quad D \overset{\Delta}{=} \sup \{ L(f) : f \in \mathcal{N} \}. \]

\( \mathcal{X} \): feasible \( x \).

- Zero duality gap is necessary: \( P = D \).
Duality Gap

\[ P \overset{\Delta}{=} \inf \{ C(x) : x \in \mathcal{X} \}, \quad D \overset{\Delta}{=} \sup \{ L(f) : f \in \mathcal{N} \}. \]

\( \mathcal{X} \): feasible \( x \).

- Zero duality gap is necessary: \( P = D \).
- Establish \( P = D \) through Strong Duality Theorem.
  - Strictly feasible solution \( \Rightarrow \) exists saddle point \( (x, t, f) \):
    \[ D = L(f) = L^*(x, t, f) = C(x) = P. \]
\[
P \triangleq \inf \{ C(x) : x \in \mathcal{X} \}, \quad D \triangleq \sup \{ L(f) : f \in \mathcal{N} \}.
\]

\( \mathcal{X} \): feasible \( x \).

- Zero duality gap is necessary: \( P = D \).
- Establish \( P = D \) through Strong Duality Theorem.
  - Strictly feasible solution \( \Rightarrow \) exists saddle point \((x, t, f)\):
    \[
    D = L(f) = L^*(x, t, f) = C(x) = P.
    \]
- Misunderstanding: previous work applied Strong Duality Theorem to original non-convex formulation (transformation was necessary for convexity).
- GCS is convex without transformation.
What if there is no strictly feasible solution?
What if there is no strictly feasible solution?

Regularity condition $\Rightarrow$ zero duality gap. [Rockafellar 1971]

More general than Strong Duality Theorem for zero duality gap.

No guarantee for saddle points as Strong Duality Theorem.

$$\forall \mathbf{f} \in \mathcal{N}, L(\mathbf{f}) < D.$$
Simplify the Dual Problem

▶ Flow conservation on $G$:

$$\mathcal{F} \triangleq \{ f : \sum_{(i,k) \in E} f_{i,k} = \sum_{(k,j) \in E} f_{k,j}, \forall k \in V \}. $$

▶ $L(f) = -\infty$ for $f \not\in \mathcal{F}$, since

$$\forall f \not\in \mathcal{F}, M \in R, x \in \Omega, \exists t \in R^{|V|}, L^*(x, t, f) < M.$$

▶ Simplify D-GCS into FD-GCS:

Maximize $L(f)$

s.t. $f \in \mathcal{F} \cap \mathcal{N}$. 
Simplify the Dual Problem

Flow conservation on $G$:

$\mathcal{F} \triangleq \{ f : \sum_{(i,k) \in E} f_{i,k} = \sum_{(k,j) \in E} f_{k,j}, \forall k \in V \}.$

$L(f) = -\infty$ for $f \notin \mathcal{F}$, since

$\forall f \notin \mathcal{F}, M \in R, x \in \Omega, \exists t \in R^{|V|}, L^*(x, t, f) < M.$

Simplify D-GCS into FD-GCS:

Maximize $L(f)$

s.t. $f \in \mathcal{F} \cap \mathcal{N}.$

Misunderstanding: previous works obtained FD-GCS by the Karush-Kuhn-Tucker (KKT) conditions $\frac{\partial L^*}{\partial t_k} = 0$. However, KKT conditions are not necessary conditions.
A trivial example that is not trivial.

Will the extreme cases happen for sizing?

- Optimal solutions do NOT satisfy KKT conditions.
- No saddle point.

\[ \forall \mathbf{f} \in \mathcal{F} \cap \mathcal{N}, L(\mathbf{f}) < D. \]
A trivial example that is not trivial.

Optimize a single inverter with size $e^x$ and fixed driver/load:

\[
\begin{align*}
\text{Minimize} & \quad e^x \\
\text{s.t.} & \quad t_1 + e^x \leq t_2, \quad t_2 + e^{-x} \leq t_3, \quad t_3 \leq t_1 + 2, \\
& \quad -\ln 2 \leq x \leq \ln 2.
\end{align*}
\]

▶ Single feasible solution $x = 0$.
Optimal but not strictly feasible.

▶ KKT conditions cannot be satisfied by $x = 0$:

\[
0 = \frac{\partial L^*}{\partial x} = e^x + f_{1,2}e^x - f_{2,3}e^{-x} = 1 + f_{1,2} - f_{2,3},
\]
\[
0 = \frac{\partial L^*}{\partial t_2} = f_{2,3} - f_{1,2}.
\]
A trivial example that is not trivial.

Since \( f \in \mathcal{F} \cap \mathcal{N} \), assume \( \beta = f_{1,2} = f_{2,3} = f_{3,1} \geq 0 \).

Compute \( L(f) \):

\[
L(f) = q(\beta) = \begin{cases} 
\frac{1+\beta}{2}, & \text{if } 0 \leq \beta < \frac{1}{3} \\
\frac{2}{\sqrt{1+1/\beta + 1}}, & \text{if } \beta \geq \frac{1}{3}
\end{cases}
\]

\( q(\beta) \) increases from \( \frac{1}{2} \) to 1 when \( \beta \) increases from 0 to \( +\infty \).

Therefore,

\[
\forall f \in \mathcal{F} \cap \mathcal{N}, \quad L(f) < 1 = D.
\]
Further Simplification of the Dual Problem

- Simplify \( L(f) \) given \( f \in \mathcal{F} \):

\[
P_f(x) \triangleq C(x) + \sum_{(i,j) \in E} f_{i,j} d_{i,j}(x),
\]

\[
Q(f) \triangleq \inf \{ P_f(x) : x \in \Omega \}.
\]

\[
\Rightarrow L(f) = Q(f), \forall f \in \mathcal{F}.
\]

- **SD-GCS:**

Maximize \( Q(f) \)

s.t. \( f \in \mathcal{F} \cap \mathcal{N} \).

- SD-GCS is different from FD-GCS.

- D-GCS, FD-GCS, SD-GCS are equivalent.
GCS can be infeasible: clock period is too small.

$C(x)$ is continuous and $\Omega$ is compact:

$$\exists U, \forall x \in \Omega, C(x) \leq U.$$

Assume zero duality gap, GCS is feasible iff

$$\forall f \in F \cap N, Q(f) \leq U.$$
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Differentiable Dual Objective $Q(f)$

- Sufficient condition (from textbook):
  \[
  \exists x_f \in \Omega, \forall x \in \Omega, Q(f) = P_f(x_f) < P_f(x).
  \]

- Assume the condition is NOT satisfied.
  \[
  \exists x' \neq x'' \in \Omega, Q(f) = P_f(x') = P_f(x'') \Rightarrow \forall 0 \leq \gamma \leq 1, Q(f) = P_f((1 - \gamma)x' + \gamma x'')
  \]
  \[
  \Rightarrow (x'' - x')^\top H_P \left( \frac{x'' + x'}{2} \right)(x'' - x') = 0
  \]
  \[
  \Rightarrow (x'' - x')^\top H_C \left( \frac{x'' + x'}{2} \right)(x'' - x') = 0
  \]

  GCS is NOT proper!

- For proper GCS problems, $Q(f)$ is differentiable, and
  \[
  \frac{\partial Q(f)}{\partial f_{i,j}} = d_{i,j}(x_f).
  \]
Method of feasible directions:
Find $\Delta f$, an ascent direction that is also feasible (w.r.t. SD-GCS).

$$\exists \lambda > 0, (Q(f + \lambda \Delta f) > Q(f)) \land (f + \lambda \Delta f \in F \cap N).$$

How to find?
Method of feasible directions:
Find $\Delta f$, an ascent direction that is also feasible (w.r.t. SD-GCS).

$$\exists \lambda > 0, (Q(f + \lambda \Delta f) > Q(f)) \land (f + \lambda \Delta f \in \mathcal{F} \cap \mathcal{N}).$$

How to find?

- $d(x_f)$ is the gradient of $Q(f)$:
  $$\Delta f^\top d(x_f) > 0 \Rightarrow \exists \lambda > 0, Q(f + \lambda \Delta f) > Q(f).$$

- Flow conservation:
  $$f + \Delta f \in \mathcal{F} \cap \mathcal{N} \Rightarrow \forall 0 \leq \lambda \leq \min_{\Delta f_{i,j} < 0} -\frac{f_{i,j}}{\Delta f_{i,j}}, f + \lambda \Delta f \in \mathcal{F} \cap \mathcal{N}.$$
The direction finding (DF) problem:

Maximize \( \Delta f^\top d(x_f) \)

s.t. \( f + \Delta f \in \mathcal{F} \cap \mathcal{N} \),

\[ -u \leq \Delta f_{i,j} \leq u, \forall (i, j) \in E. \]

\( \Delta f \) are decision variables, \( u \) is positive constant.

\( \Delta f^\top d(x_f) \): first order approx. of \( Q(f + \Delta f) - Q(f) \).

DF is a min-cost network flow problem.

For optimal \( \Delta f \), \( \Delta f^\top d(x_f) \geq 0 \). And

\[ \Delta f^\top d(x_f) = 0 \Rightarrow x_f \text{ is optimal for GCS.} \]
The DualFD Algorithm

- Iterative algorithm. $Q(f)$ is increasing every iteration.
  - $Q(f)$ is convex $\Rightarrow$ local maximal is global maximum.
  - For subgradient optimizations, $Q(f)$ may decrease.

- Each iteration,
  - Solve DF for $\Delta f$.
  - Perform a line search to find $Q(f + \lambda \Delta f) > Q(f)$
  - Check GCS infeasibility when computing $Q$.
  - Claim optimality if the changes are marginal.
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Experimental Setup

- Minimize total area under performance bound with Elmore delay model.
- ISCAS89 sequential circuits, 29 totally.
  Largest: $\sim55000$ vertices, $\sim70000$ edges, $\sim21000$ variables.
- Implement DualFD algorithm in C++.
- Use the CS2 min-cost network flow solver for DF. [Goldberg]
- Compare to subgradient optimizations: SubGrad.
Experimental Results

- 15 largest benchmarks.
- Clock period achieved $T$ vs. target clock period $T_0$
  - s838 is NOT feasible.
Experimental Results

- 15 largest benchmarks.
- Compare area, dual objective, and running time.

<table>
<thead>
<tr>
<th>name</th>
<th>DualFD</th>
<th></th>
<th>SubGrad</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>area</td>
<td>dual</td>
<td>t(s)</td>
<td>area</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s832</td>
<td>1060</td>
<td>1074</td>
<td>8.28</td>
<td>493</td>
</tr>
<tr>
<td>s838</td>
<td>2916</td>
<td><strong>80458</strong></td>
<td>0.07</td>
<td>10189</td>
</tr>
<tr>
<td>s953</td>
<td>776</td>
<td>775</td>
<td>5.92</td>
<td>3449</td>
</tr>
<tr>
<td>s1196</td>
<td>1089</td>
<td>1088</td>
<td>10.90</td>
<td>1642</td>
</tr>
<tr>
<td>s1238</td>
<td>1080</td>
<td>1080</td>
<td>7.80</td>
<td>974</td>
</tr>
<tr>
<td>s1423</td>
<td>1668</td>
<td>1670</td>
<td>1.53</td>
<td>3346</td>
</tr>
<tr>
<td>s1488</td>
<td>2055</td>
<td>2107</td>
<td>28.22</td>
<td>1150</td>
</tr>
<tr>
<td>s1494</td>
<td>2161</td>
<td>2318</td>
<td>28.33</td>
<td>1140</td>
</tr>
<tr>
<td>s5378</td>
<td>5856</td>
<td>6083</td>
<td>91.49</td>
<td>9396</td>
</tr>
<tr>
<td>s9234</td>
<td>12935</td>
<td>15508</td>
<td>236.49</td>
<td>11517</td>
</tr>
<tr>
<td>s13207</td>
<td>14608</td>
<td>14608</td>
<td>111.91</td>
<td>15642</td>
</tr>
<tr>
<td>s15850</td>
<td>17766</td>
<td>17766</td>
<td>229.25</td>
<td>20628</td>
</tr>
<tr>
<td>s35932</td>
<td>33522</td>
<td>44344</td>
<td>304.61</td>
<td>80650</td>
</tr>
<tr>
<td>s38417</td>
<td>42176</td>
<td>44551</td>
<td>301.48</td>
<td>49126</td>
</tr>
<tr>
<td>s38584</td>
<td>34973</td>
<td>34973</td>
<td>149.87</td>
<td>35016</td>
</tr>
</tbody>
</table>
DualFD vs. SubGrad

- 29 benchmarks totally.
- SubGrad never dominates DualFD.
How far away are we from the optimals?

- Collect feasible solutions to estimate the duality gap.
- 15 out of 29 benchmarks.
Convergence of s38584
Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions
Conclusions

- Formulate GCS problems to handle sequential optimization.
- Correct misunderstandings when applying Lagrangian relaxation.
- Show how to detect infeasibility in GCS.
- Prove gradient exists for proposed proper GCS problems.
- Design the DualFD algorithm to solve proper GCS problems.
Q & A
Thank you!
\[ L(f) = Q(f), \quad \forall f \in \mathcal{F} \cap \mathcal{N}, \]

\[ L(f) = -\infty, \quad \forall f \in \mathcal{N} - \mathcal{F}, \]

\[ Q(f) \neq -\infty, \quad \forall f \in \mathcal{N} - \mathcal{F}. \]

- \( L(f) \) is convex but not differentiable for \( f \in \mathcal{N} \).
- \( Q(f) \) is convex and differentiable for \( f \in \mathcal{N} \).
Proper GCS Problems

Posynomial objective functions of $e^{x_k}$:

$$C(x) = \sum_{i=1}^{l} c_i e^{b_i^T x}, \quad c_i > 0, \forall 1 \leq i \leq l$$

Let $B = (b_1, b_2, \ldots, b_l)$, $\text{rank}(B) = n \iff \text{proper.}$
Full row rank.