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

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.

Previous Works – General Convex Optimization

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

Generalized Convex Sizing (GCS)

$$\begin{aligned} \text{Minimize} \quad & C(\mathbf{x}) \\ \text{s.t.} \quad & t_i + d_{i,j}(\mathbf{x}) - t_j \leq 0, \forall (i,j) \in E, \\ & \mathbf{x} \in \Omega. \end{aligned}$$

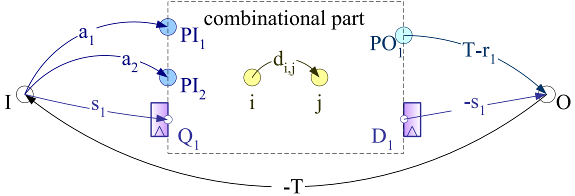
- $G = (V, E)$: a directed graph, the system structure.
- $\mathbf{x}$: the system parameters belonging to

$$\Omega \triangleq \{\mathbf{x} : l_k \leq x_k \leq u_k, \forall 1 \leq k \leq n\}.$$

- $\mathbf{t}$: (relative) arrival times.
- $C(\mathbf{x})$, $d_{i,j}(\mathbf{x})$: objective function and delays, twice differentiable and convex.

The GCS Formulation

- Timing specification in sequential circuits.



Feasible timing ⇔ 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:

$$\forall \mathbf{x} \in \Omega, \forall \mathbf{z} \neq 0, \mathbf{z}^\top H_C(\mathbf{x})\mathbf{z} \neq 0.$$

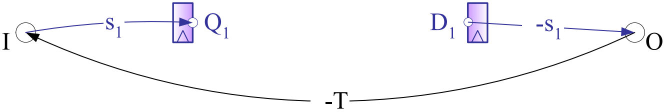
- With differentiable dual functions (shown later).
  - $d_{i,j}(\mathbf{x})$ are irrelevant.
- Optimizing total positive weighted area is proper.

Simultaneous Sizing and Clock Skew Optimization

- 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: $\frac{\partial C}{\partial s} = 0, \forall$clock skew variable s.
- Cancel s. Transform to a proper one.

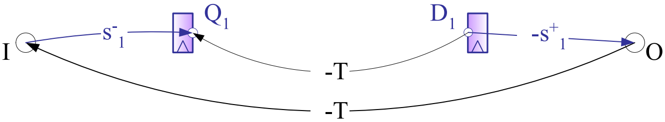
- Constraints concerning clock skew s_k:

$$(t_{\mathrm{I}} + s_k \leq t_{\mathrm{Q}_k}) \wedge (t_{\mathrm{D}_k} - s_k \leq t_{\mathrm{O}}) \wedge (s_k^- \leq s_k \leq s_k^+)$$



- Cancel s_k: $[t_{\mathrm{D}_k} - t_{\mathrm{O}}, t_{\mathrm{Q}_k} - t_{\mathrm{I}}] \cap [s_k^-, s_k^+] \neq \emptyset$. Equivalently:

$$(t_{\mathrm{D}_k} - s_k^+ \leq t_{\mathrm{O}}) \wedge (t_{\mathrm{I}} + s_k^- \leq t_{\mathrm{Q}_k}) \wedge (t_{\mathrm{D}_k} - t_{\mathrm{Q}_k} \leq T)$$



Outline

Introduction to Gate Sizing

Generalized Convex Sizing

The Dual Problems

The DualFD Algorithm

Experiments

Conclusions

Lagrangian Relaxation Overview

- Lagrangian function:

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

- Lagrangian subproblem:

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

- Lagrangian dual problem (D-GCS):

$$\begin{aligned} \text{Maximize} \quad & L(\mathbf{f}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{N}. \end{aligned}$$

$\mathcal{N}$: non-negative $\mathbf{f}$.

Duality Gap

$$P \triangleq \inf\{C(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, D \triangleq \sup\{L(\mathbf{f}) : \mathbf{f} \in \mathcal{N}\}.$$

$\mathcal{X}$: feasible $\mathbf{x}$.

- Zero duality gap is necessary: $P = D$.

Duality Gap

$$P \triangleq \inf\{C(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, D \triangleq \sup\{L(\mathbf{f}) : \mathbf{f} \in \mathcal{N}\}.$$

$\mathcal{X}$: feasible $\mathbf{x}$.

- Zero duality gap is necessary: $P = D$.
- Establish $P = D$ through Strong Duality Theorem.
  - Strictly feasible solution $\Rightarrow$ exists saddle point $(\mathbf{x}, \mathbf{t}, \mathbf{f})$:

$$D = L(\mathbf{f}) = L^*(\mathbf{x}, \mathbf{t}, \mathbf{f}) = C(\mathbf{x}) = P.$$

Duality Gap

$$P \triangleq \inf\{C(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, D \triangleq \sup\{L(\mathbf{f}) : \mathbf{f} \in \mathcal{N}\}.$$

$\mathcal{X}$: feasible $\mathbf{x}$.

- Zero duality gap is necessary: $P = D$.
- Establish $P = D$ through Strong Duality Theorem.
  - Strictly feasible solution $\Rightarrow$ exists saddle point $(\mathbf{x}, \mathbf{t}, \mathbf{f})$:

$$D = L(\mathbf{f}) = L^*(\mathbf{x}, \mathbf{t}, \mathbf{f}) = C(\mathbf{x}) = P.$$

- Misunderstanding: previous work applied Strong Duality Theorem to original non-convex formulation (transformation was necessary for convexity).
- GCS is convex without transformation.

Duality Gap

- What if there is no strictly feasible solution?

Duality Gap

- What if there is no strictly feasible solution?
- Regularity condition $\Rightarrow$ zero duality gap. [Rockafellear 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:

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

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

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

- Simplify D-GCS into FD-GCS:

$$\begin{aligned} \text{Maximize} \quad & L(\mathbf{f}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{F} \cap \mathcal{N}. \end{aligned}$$

Simplify the Dual Problem

- Flow conservation on G:

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

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

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

- Simplify D-GCS into FD-GCS:

$$\begin{aligned} \text{Maximize} \quad & L(\mathbf{f}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{F} \cap \mathcal{N}. \end{aligned}$$

- 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:

$$\begin{array}{ll} \text{Minimize} & e^x \\ \text{s.t.} & t_1 + e^x \le t_2,\ t_2 + e^{-x} \le t_3,\ t_3 \le t_1 + 2, \\ & -\ln 2 \le x \le \ln 2. \end{array}$$

- Single feasible solution $x = 0$.
  Optimal but not strictly feasible.
- KKT conditions cannot be satisfied by $x = 0$:

$$\begin{aligned} 0 = \frac{\partial L^*}{\partial x} &= e^x + f_{1,2} e^x - f_{2,3} e^{-x} = 1 + f_{1,2} - f_{2,3}, \\ 0 = \frac{\partial L^*}{\partial t_2} &= f_{2,3} - f_{1,2}. \end{aligned}$$

A trivial example that is not trivial.

- Since $\mathbf{f} \in \mathcal{F} \cap \mathcal{N}$, assume $\beta = f_{1,2} = f_{2,3} = f_{3,1} \geq 0$.
- Compute $L(\mathbf{f})$:

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

- $q(\beta)$ increases from $\frac{1}{2}$ to 1 when β increases from 0 to $+\infty$.
- Therefore,

$$\forall \mathbf{f} \in \mathcal{F} \cap \mathcal{N}, L(\mathbf{f}) < 1 = D.$$

Further Simplification of the Dual Problem

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

$$P_{\mathbf{f}}(\mathbf{x}) \triangleq C(\mathbf{x}) + \sum_{(i,j)\in E} f_{i,j} d_{i,j}(\mathbf{x}),$$

$$Q(\mathbf{f}) \triangleq \inf\{P_{\mathbf{f}}(\mathbf{x}) : \mathbf{x} \in \Omega\}.$$

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

- SD-GCS:

$$\begin{aligned} \text{Maximize} \quad & Q(\mathbf{f}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{F} \cap \mathcal{N}. \end{aligned}$$

- SD-GCS is different from FD-GCS.
- D-GCS, FD-GCS, SD-GCS are equivalent.

Infeasibility for GCS

- GCS can be infeasible: clock period is too small.
- $C(\mathbf{x})$ is continuous and Ω is compact:

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

- Assume zero duality gap, GCS is feasible iff

$$\forall \mathbf{f} \in \mathcal{F} \cap \mathcal{N}, Q(\mathbf{f}) \leq U.$$

Outline

Differentiable Dual Objective $Q(\mathbf{f})$

- Sufficient condition (from textbook):

$$\exists \mathbf{x_f} \in \Omega, \forall \mathbf{x} \in \Omega, Q(\mathbf{f}) = P_{\mathbf{f}}(\mathbf{x_f}) < P_{\mathbf{f}}(\mathbf{x}).$$

- Assume the condition is NOT satisfied.

$$\begin{aligned} & \exists \mathbf{x}' \neq \mathbf{x}'' \in \Omega, Q(\mathbf{f}) = P_{\mathbf{f}}(\mathbf{x}') = P_{\mathbf{f}}(\mathbf{x}'') \\ \Rightarrow\ & \forall 0 \leq \gamma \leq 1, Q(\mathbf{f}) = P_{\mathbf{f}}((1-\gamma)\mathbf{x}' + \gamma \mathbf{x}'') \\ \Rightarrow\ & (\mathbf{x}'' - \mathbf{x}')^\top H_{P_{\mathbf{f}}}(\frac{\mathbf{x}'' + \mathbf{x}'}{2})(\mathbf{x}'' - \mathbf{x}') = 0 \\ \Rightarrow\ & (\mathbf{x}'' - \mathbf{x}')^\top H_C(\frac{\mathbf{x}'' + \mathbf{x}'}{2})(\mathbf{x}'' - \mathbf{x}') = 0 \end{aligned}$$

GCS is NOT proper!

- For proper GCS problems, $Q(\mathbf{f})$ is differentiable, and

$$\frac{\partial Q(\mathbf{f})}{\partial f_{i,j}} = d_{i,j}(\mathbf{x_f}).$$

Improving Feasible Direction

Method of feasible directions:
Find $\Delta\mathbf{f}$, an ascent direction that is also feasible (w.r.t. SD-GCS).

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

How to find?

Improving Feasible Direction

Method of feasible directions:
Find $\Delta\mathbf{f}$, an ascent direction that is also feasible (w.r.t. SD-GCS).

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

How to find?

- $\mathbf{d}(\mathbf{x_f})$ is the gradient of $Q(\mathbf{f})$:

$$\Delta\mathbf{f}^\top \mathbf{d}(\mathbf{x_f}) > 0 \Rightarrow \exists\lambda > 0, Q(\mathbf{f} + \lambda\Delta\mathbf{f}) > Q(\mathbf{f}).$$

- Flow conservation:

$$\mathbf{f} + \Delta\mathbf{f} \in \mathcal{F} \cap \mathcal{N} \Rightarrow \forall 0 \leq \lambda \leq \min_{\Delta f_{i,j} < 0} -\frac{f_{i,j}}{\Delta f_{i,j}}, \mathbf{f} + \lambda\Delta\mathbf{f} \in \mathcal{F} \cap \mathcal{N}.$$

Improving Feasible Direction

- The direction finding (DF) problem:

$$\begin{aligned} \text{Maximize} \quad & \Delta\mathbf{f}^\top \mathbf{d}(\mathbf{x_f}) \\ \text{s.t.} \quad & \mathbf{f} + \Delta\mathbf{f} \in \mathcal{F} \cap \mathcal{N}, \\ & -u \le \Delta f_{i,j} \le u, \forall (i,j) \in E. \end{aligned}$$

  $\Delta\mathbf{f}$ are decision variables, u is positive constant.
- $\Delta\mathbf{f}^\top \mathbf{d}(\mathbf{x_f})$: first order approx. of $Q(\mathbf{f} + \Delta\mathbf{f}) - Q(\mathbf{f})$.
- DF is a min-cost network flow problem.
- For optimal $\Delta\mathbf{f}$, $\Delta\mathbf{f}^\top \mathbf{d}(\mathbf{x_f}) \ge 0$. And

$$\Delta\mathbf{f}^\top \mathbf{d}(\mathbf{x_f}) = 0 \Rightarrow \mathbf{x_f} \text{ is optimal for GCS.}$$

The DualFD Algorithm

- Iterative algorithm. $Q(\mathbf{f})$ is increasing every iteration.
  - $Q(\mathbf{f})$ is convex $\Rightarrow$ local maximal is global maximum.
  - For subgradient optimizations, $Q(\mathbf{f})$ may decrease.
- Each iteration,
  - Solve DF for $\Delta\mathbf{f}$.
  - Perform a line search to find $Q(\mathbf{f} + \lambda\Delta\mathbf{f}) > Q(\mathbf{f})$
  - Check GCS infeasibility when computing Q.
  - Claim optimality if the changes are marginal.

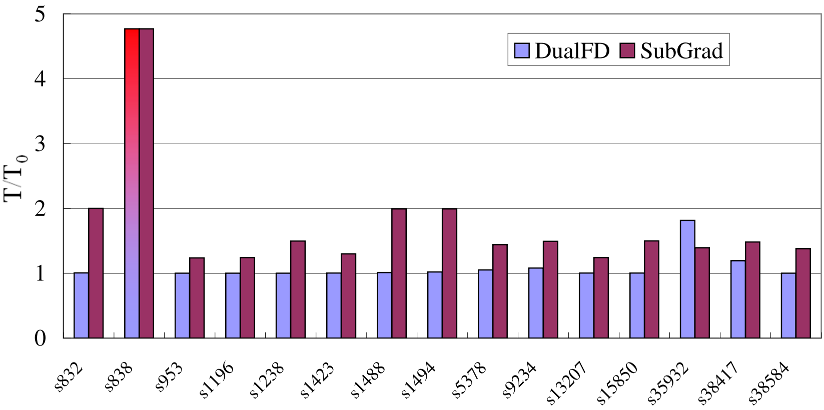
Outline

Experimental Setup

- Minimize total area under performance bound with Elmore delay model.
- ISCAS89 sequential circuits, 29 totally.
  Largest: $\sim$55000 vertices, $\sim$70000 edges, $\sim$21000 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.



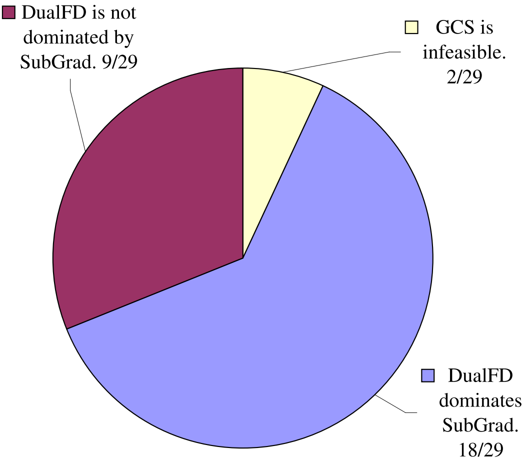
Experimental Results

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

	DualFD			SubGrad		
name	area	dual	t(s)	area	dual	t(s)
s832	1060	1074	8.28	493	493	0.11
s838	2916	80458	0.07	10189	-257K	51.51
s953	776	775	5.92	3449	-80K	55.15
s1196	1089	1088	10.90	1642	-14K	54.88
s1238	1080	1080	7.80	974	-1823	27.19
s1423	1668	1670	1.53	3346	-251K	120.76
s1488	2055	2107	28.22	1150	557	31.28
s1494	2161	2318	28.33	1140	529	32.02
s5378	5856	6083	91.49	9396	-52K	308.39
s9234	12935	15508	236.49	11517	-46K	384.86
s13207	14608	14608	111.91	15642	-121K	432.81
s15850	17766	17766	229.25	20628	-287K	600.10
s35932	33522	44344	304.61	80650	-1M	600.34
s38417	42176	44551	301.48	49126	-363K	600.35
s38584	34973	34973	149.87	35016	-23K	600.19

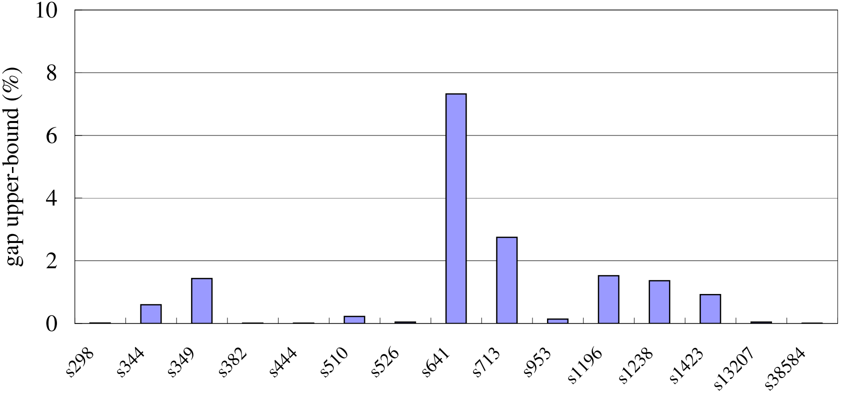
DualFD vs. SubGrad

- 29 benchmarks totally.
- SubGrad never dominates DualFD.

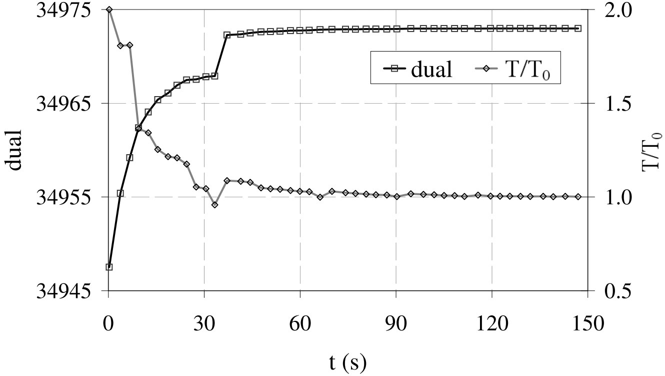


How far away are we from the optimals?

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



Convergence of s38584



Outline

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!

$L(\mathbf{f})$ vs. $Q(\mathbf{f})$

$$L(\mathbf{f}) = Q(\mathbf{f}), \quad \forall \mathbf{f} \in \mathcal{F} \cap \mathcal{N},$$

$$\begin{aligned} L(\mathbf{f}) &= -\infty, \quad \forall \mathbf{f} \in \mathcal{N} - \mathcal{F}, \\ Q(\mathbf{f}) &\neq -\infty, \quad \forall \mathbf{f} \in \mathcal{N} - \mathcal{F}. \end{aligned}$$

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

Proper GCS Problems

Posynomial objective functions of e^{x_k}:

$$C(\mathbf{x}) = \sum_{i=1}^{l} c_i \ e^{\mathbf{b}_i^\top \mathbf{x}}, c_i > 0, \forall 1 \leq i \leq l$$

Let $B = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_l)$, $\mathrm{rank}(B) = n \Leftrightarrow$ proper.
Full row rank.