FASTER MINIMIZATION OF LINEAR WIRELENGTH FOR GLOBAL PLACEMENT

Charles J. Alpert, IBM Austin
Tony F. Chan, UCLA Mathematics
Dennis J.-H. Huang, UCLA CS / Synopsys
Andrew B. Kahng, UCLA CS / Cadence
Igor L. Markov, UCLA Mathematics
Pep Mulet, UCLA Mathematics
Kenneth Yan, UCLA CS

Outline of the Talk

1. Background
   - Linear and quadratic placement
   - GORDIAN-L

2. Contributions
   - β-regularizing a linear objective
   - Weiszfeld approach to linear placement generalizes GORDIAN-L
   - Achieving faster (quadratic) convergence with Newton methods
   - Primal-Dual Newton method

3. Experimental Validation

Placement

Given:

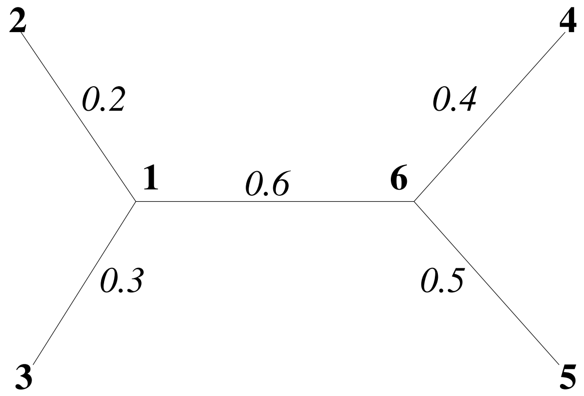
- weighted graph
- constraints on vertex locations (e.g., pad vertices fixed)
- objective function of vertex locations and edge weights

Find: locations minimizing objective function
subject to constraints

Additional considerations:

- row-based (standard-cell, gate-array) layout
- timing- and routability-driven placement
- problem sizes of interest: 10^5 and above

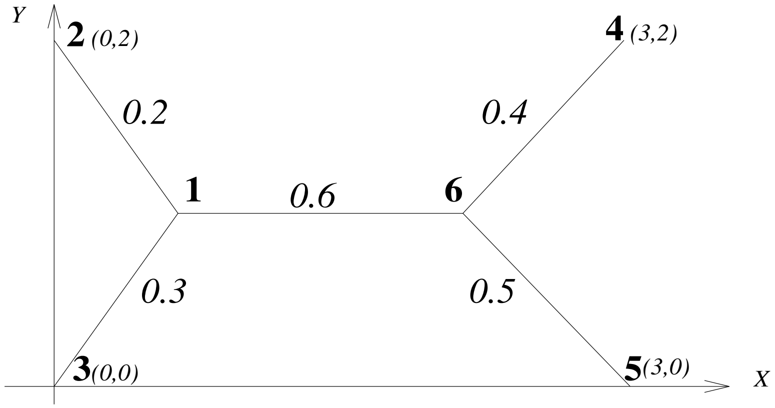
Laplacian and The Incidence Matrix



$$\textbf{Laplacian } \mathbf{Q} \;=\; \begin{pmatrix} 1.1 & -0.2 & -0.3 & & & -0.6 \\ -0.2 & 0.2 & & & & \\ -0.3 & & 0.3 & & & \\ & & & 0.4 & & -0.4 \\ & & & & 0.5 & -0.5 \\ -0.6 & & & -0.4 & -0.5 & 1.5 \end{pmatrix}$$

$$\textbf{Incidence matrix } \mathbf{C} = \begin{pmatrix} 0.2 & -0.2 & & & & \\ 0.3 & & -0.3 & & & \\ & & & 0.4 & & -0.4 \\ & & & & 0.5 & -0.5 \\ 0.6 & & & & & -0.6 \end{pmatrix}$$

Constraints



$$\mathbf{H_x x = b_x} \qquad \begin{pmatrix} 0 & 1 & & & & 0 \\ & & 1 & & & \\ & & & 1 & & \\ 0 & & & & 1 & 0 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 3 \end{bmatrix}$$

$$\mathbf{H_y y = b_y} \qquad \begin{pmatrix} 0 & 1 & & & & 0 \\ & & 1 & & & \\ & & & 1 & & \\ 0 & & & & 1 & 0 \end{pmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 0 \end{bmatrix}$$

Objective Functions

- Classic 1-dim **linear** (Manhattan) objective

$$\Phi_L(\mathbf{x}) \quad = \quad \| \mathbf{C}\mathbf{x} \|_1 \quad = \quad \sum_{i>j}^{n} w_{ij}|x_i - x_j|$$

- Classic 1-dim **quadratic** objective

$$\Phi_Q(\mathbf{x}) \quad = \quad \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} \quad = \quad \sum_{i>j}^{n} w_{ij}(x_i - x_j)^2$$

$\mathbf{Q}$ — the Laplacian
$\mathbf{C}$ — the incidence matrix
x_i — locations of vertices (modules)
w_{ij} — edge weights

What does *linear* mean ?

1. One-dimensional (not used)

2. Algebraically linear: $\Psi(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\Psi(\mathbf{x}) + \beta\Psi(\mathbf{y})$
   (e.g., constraints)

3. Linear growth: $O(\| \mathbf{x} \|), \mathbf{x} \to \infty$
   (e.g., objective functions)

4. Piecewise linear ("broken" straight line)
   (e.g., objective functions)

5. Logarithmically linear: $\Theta(exp(-N)), N \to \infty$
   (e.g., error in *linear convergence*)

What does *quadratic* mean ?

1. Quadratic assignment
   (e.g., discrete placement problem, not used)

2. 2-homogeneous: $\Psi(\alpha\mathbf{x}) = \alpha^2\Psi(\mathbf{x})$
   (e.g., objective functions)

3. Quadratic growth: $O(\| \mathbf{x} \|^2), \mathbf{x} \to \infty$
   (e.g., objective functions)

4. Logarithmically quadratic: $\Theta(exp(-N^2)), N \to \infty$
   (e.g., error in *quadratic convergence*)

Quadratic Objective versus Linear

Quadratic objective

- Has unique minimizer
- "Enough" to solve $\frac{d\Phi_Q}{dx} = 0$
- Can be combined with top-down approach (GORDIAN)
- Many implementations available

Linear objective

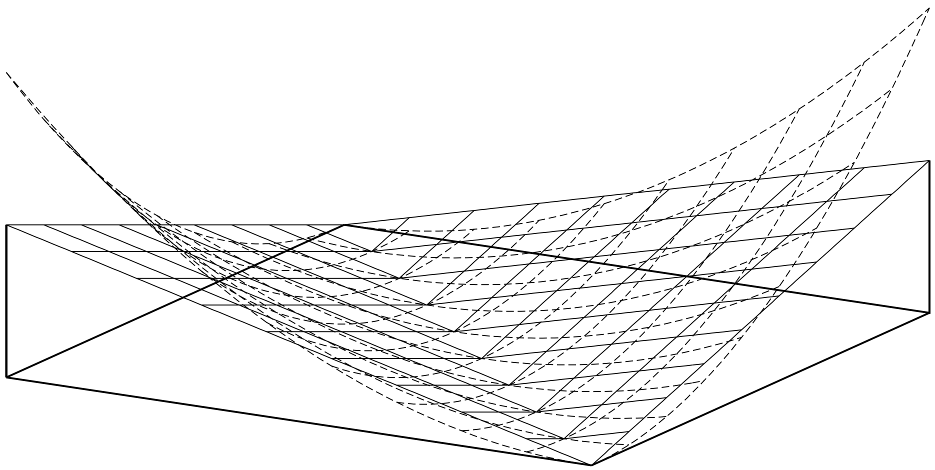
- Better models timing and congestion
- Distributes modules more evenly
- Harder to minimize
- Can be approximated by a sequence of quadratic objectives (GORDIAN-L)

GORDIAN-L (Sigl et al. '91)

- Locally approximates $\Phi_L(\mathbf{x})$ by $\Phi_Q^k(\mathbf{x})$
- Minimizer of $\Phi_L(\mathbf{x})$ is found as $\lim_{k\to\infty} \mathbf{x}^k$ of quadratic minimizers (uses GORDIAN)

$$\begin{aligned}
\Phi_L(\mathbf{x}) &= \sum_{i>j} a_{ij}|x_i - x_j| = \sum_{i>j} \frac{a_{ij}(x_i - x_j)^2}{|x_i - x_j|} \\
&\approx \sum_{i>j} \frac{a_{ij}(x_i - x_j)^2}{|x_i^{k-1} - x_j^{k-1}|} = \sum_{i>j} g_{ij}^k (x_i - x_j)^2 \\
&= \Phi_Q^k(\mathbf{x})
\end{aligned}$$

$\Phi_L(\mathrm{x})$ **versus** $\Phi_Q^k(\mathrm{x})$



$|x_1 - x_2|$ versus $(x_1 - x_2)^2/|0.3 - 0.2|$

β-regularization

Linear placement is hard:

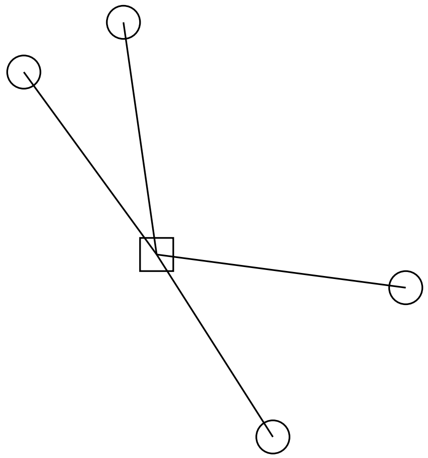
$$\Phi_L(\mathbf{x}) = \| \mathbf{C}\mathbf{x} \|_1 = \sum_{i>j}^{n} w_{ij} |x_i - x_j|$$

- $\Phi_L(\mathbf{x})$ only piece-wise differentiable
- not strictly convex (only convex)
- can have more than one minimizer

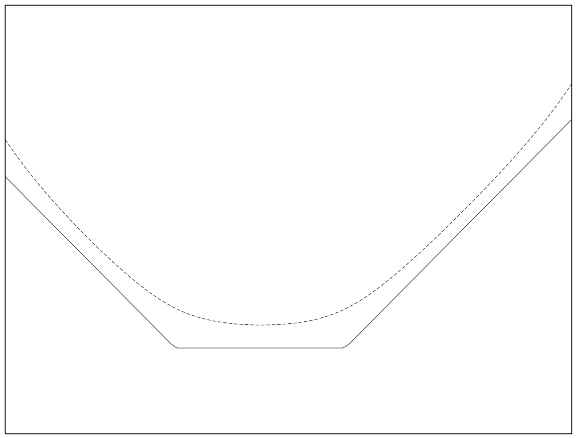
Cure: change all $|\cdot|$ in $\Phi_L(\mathbf{x})$ to $\sqrt{x^2 + \beta}$

Result: $\Phi_L^\beta(\mathbf{x})$ — the β-regularization of $\Phi_L(\mathbf{x})$

Linear Objective: 5-point example



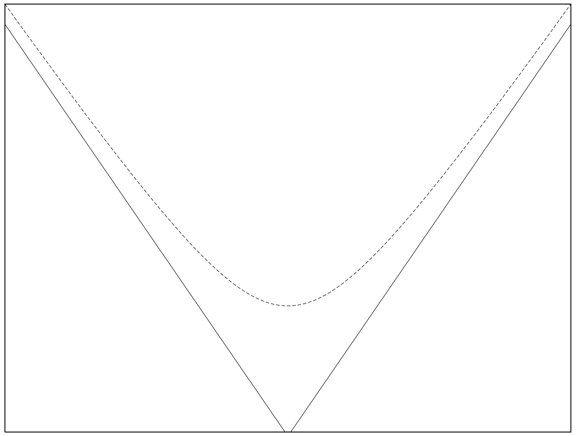
Placement problem



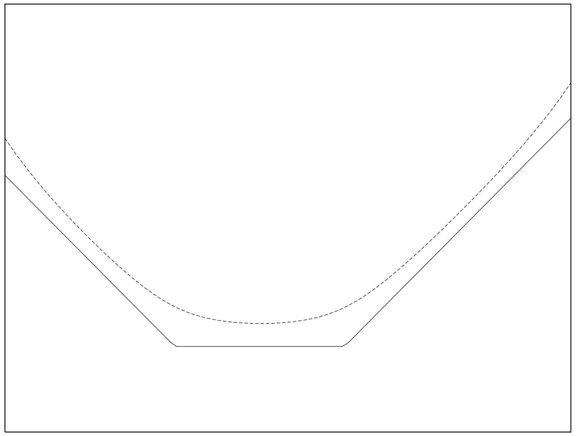
Objective function (x-component)

β-regularization (continued)

- **Theorem**
  (a) $\lim_{\beta\to 0} \Phi_L^\beta(\mathbf{x}) = \Phi_L(\mathbf{x})$ uniformly on $\mathbf{R^n}$
  (b) $\forall \mathbf{x} \in \mathbf{R^n}, \forall \beta_1 > \beta_2 \quad \Phi_L^{\beta_1}(\mathbf{x}) > \Phi_L^{\beta_2}(\mathbf{x}) > \Phi_L^0(\mathbf{x}) = \Phi_L(\mathbf{x})$
  (c) $\lim_{\beta\to 0} \min_{\mathbf{x}\in\mathbf{R^n}} \Phi_L^\beta(\mathbf{x}) = \min_{\mathbf{x}\in\mathbf{R^n}} \Phi_L(\mathbf{x})$
- consider $\Phi_L^\beta(\mathbf{x})$ instead of $\Phi_L(\mathbf{x})$



$|x|$ and $\sqrt{x^2+\beta}$



β-regularization (5-point example)

Weiszfeld (Fixed Point) Approach

- Minimize $\Phi_L^{\beta}(\mathbf{x})$ by solving the *nonlinear* equation $\frac{d\Phi_L^{\beta}(\mathbf{x})}{d\mathbf{x}} = 0$
- Find matrix $B(\mathbf{x})$, s.t. $\frac{d\Phi_L^{\beta}(\mathbf{x})}{dx} = B(\mathbf{x})\ \mathbf{x}\ =\ 0$
- Starting with initial guess $\mathbf{x^0}$, compute $\mathbf{x^k}$ from equation $B(\mathbf{x^{k-1}})\mathbf{x^k} = 0$
- *Directly handles:*
  - Arbitrary linear constraints (e.g., centers of gravity)
  - Many nonlinear constraints (with Lagrange multipliers)
  - 2-dimensional linear objectives
  - Floating pads
  - Novel objective functions (quadratic objective as well)
  - "Soft constraints"
    (penalty $BigNum \times \max\{0, -\Gamma(\mathbf{x})\}$ to prohibit $\Gamma(\mathbf{x}) < 0$)

Weiszfeld: Features and Limitations

- Proven linear convergence (Eckhardt '80)

- **For** $\beta = 0$ **get GORDIAN-L**

- $\beta > 0$ — trade-off between speed and precision

- Is robust and handles large examples

- Quickly finds a reasonable approximation

- Slow in achieving higher precision

Newton Method

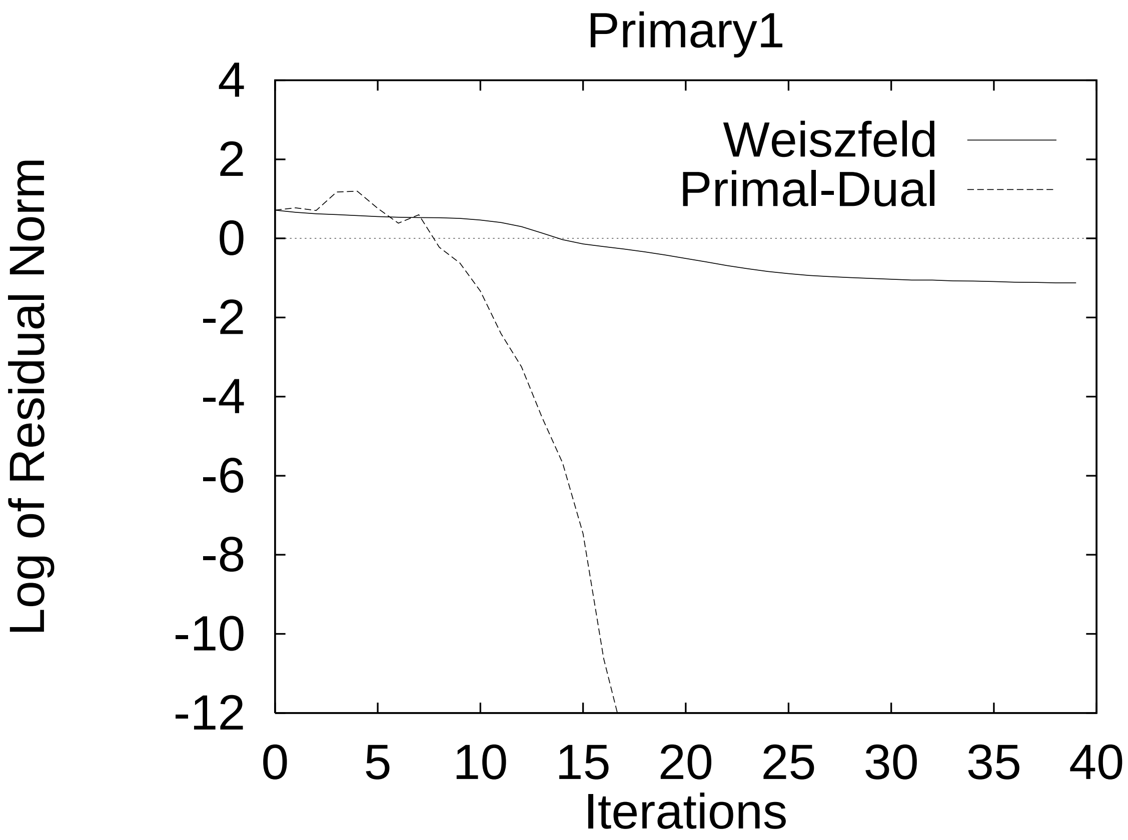
Solve the equation $F(\mathbf{x}) = \frac{d\Phi_L^\beta(\mathbf{x})}{dx} = 0$

- Start with $\mathbf{x}^0$ and update $\mathbf{x}^{k-1}$ with $\delta\mathbf{x}^k$ computed from

$$\frac{dF(\mathbf{x})}{d\mathbf{x}}\Big|_{\mathbf{x}=\mathbf{x}^{k-1}} \quad \delta\mathbf{x}^k = -F(\mathbf{x}^{k-1})$$

- Proven *local* quadratic convergence
- *Not globally convergent*
- Requires a good initial guess
  (e.g., from several iterations of Weiszfeld)
- Becomes robust when used with local search
- Requires a high-precision linear system solver

Quadratic Convergence versus Linear



log of error is plotted against iterations

Primal-Dual (Interior point) Newton

Solves $F(\mathbf{x}) = 0$ with more complicated $F(\mathbf{x})$

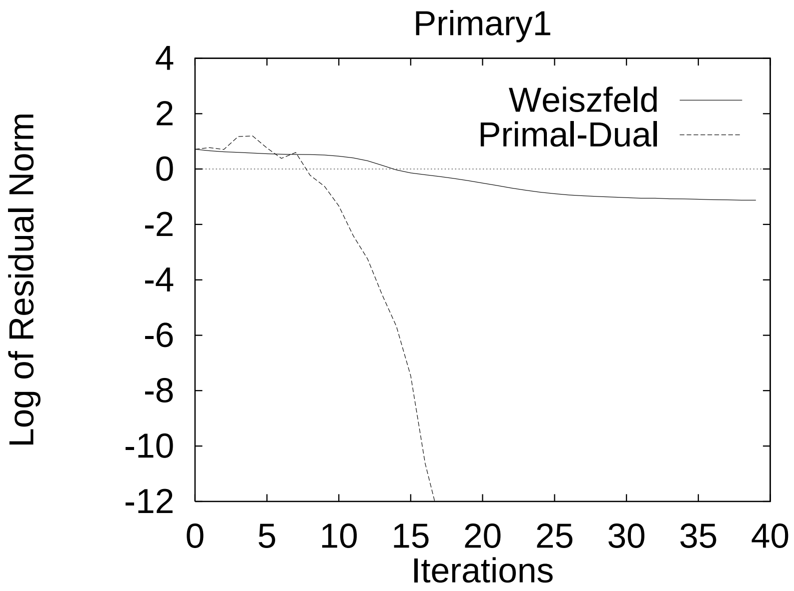
- Auxiliary (dual) variables are edge lengths
- Dual variables eliminated within Newton method (when deriving formulae, not at runtime)
- size of linear systems is # of movable modules

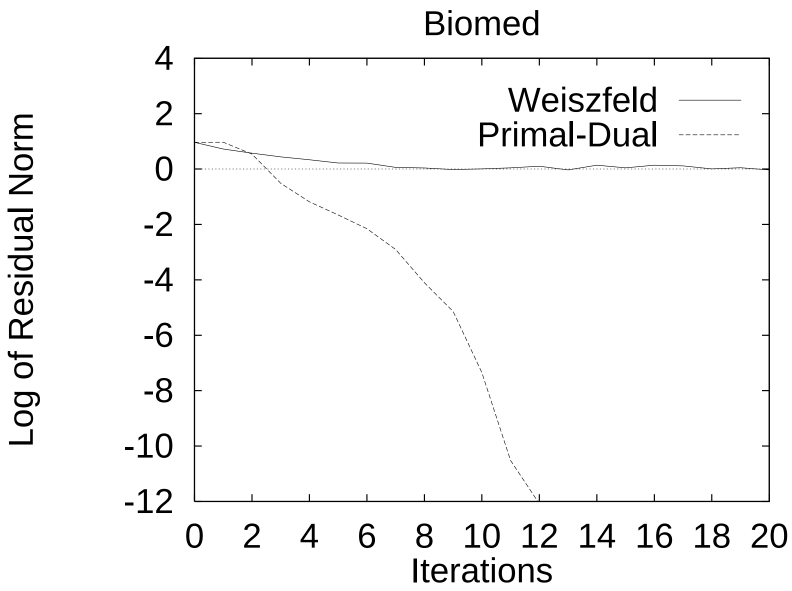
- More robust, faster than most Newton methods
- Observed (not proven) to converge globally
- Proven local quadratic convergence
- Like Weiszfeld, needs regularization and can handle nonstandard objective functions (e.g., quadratic or of any real exponent $p \geq 1$)

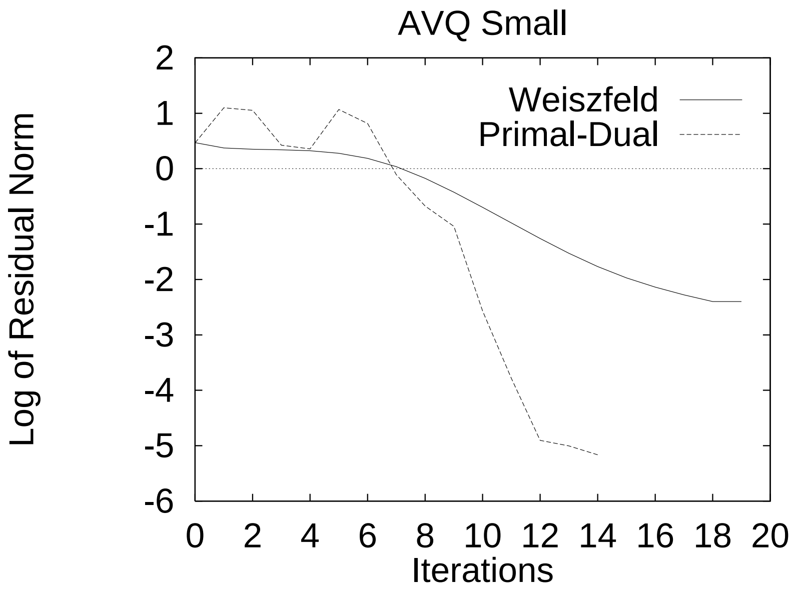
Experimental Validation

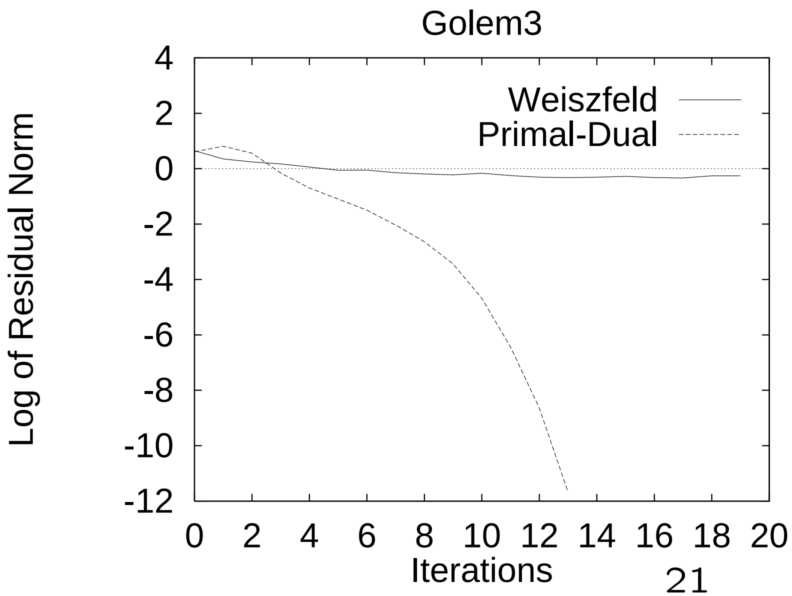
Test Case	Pads	Cells	Nets
primary1	107	752	704
biomed	97	6417	6442
avq_small	64	21854	21884
golem3	2767	100281	144949

- One Weiszfeld iteration for `avq_small` takes less than **7 seconds** on a 140 MHz Sun Ultra-1.
- Primal-Dual Newton is only $\mathbf{40-50\%}$ more expensive than Weiszfeld (but much harder to implement)
- Primal-Dual develops **quadratic** convergence and shows **overall performance improvement** over Weiszfeld


Primary1
Weiszfeld
Primal-Dual
Log of Residual Norm
4
2
0
-2
-4
-6
-8
-10
-12
0
5
10
15
20
25
30
35
40
Iterations


Biomed
Weiszfeld
Primal-Dual
Log of Residual Norm
4
2
0
-2
-4
-6
-8
-10
-12
0
2
4
6
8
10
12
14
16
18
20
Iterations


AVQ Small
Weiszfeld
Primal-Dual
Log of Residual Norm
2
1
0
-1
-2
-3
-4
-5
-6
0
2
4
6
8
10
12
14
16
18
20
Iterations


Golem3
Weiszfeld
Primal-Dual
Log of Residual Norm
4
2
0
-2
-4
-6
-8
-10
-12
0
2
4
6
8
10
12
14
16
18
20
Iterations

Conclusions

- **Weiszfeld subsumes GORDIAN-L** and allows for additional trade-off between speed and quality. **Both are linearly convergent**.

- Primal-Dual Newton achieves **quadratic convergence** on later iterations and is therefore **faster than Weiszfeld**.

- **Weiszfeld and Primal-Dual are robust and can be applied to large problems**.

- Our methods can directly handle **2-dimensional placement, nonstandard objectives and various constraints**.