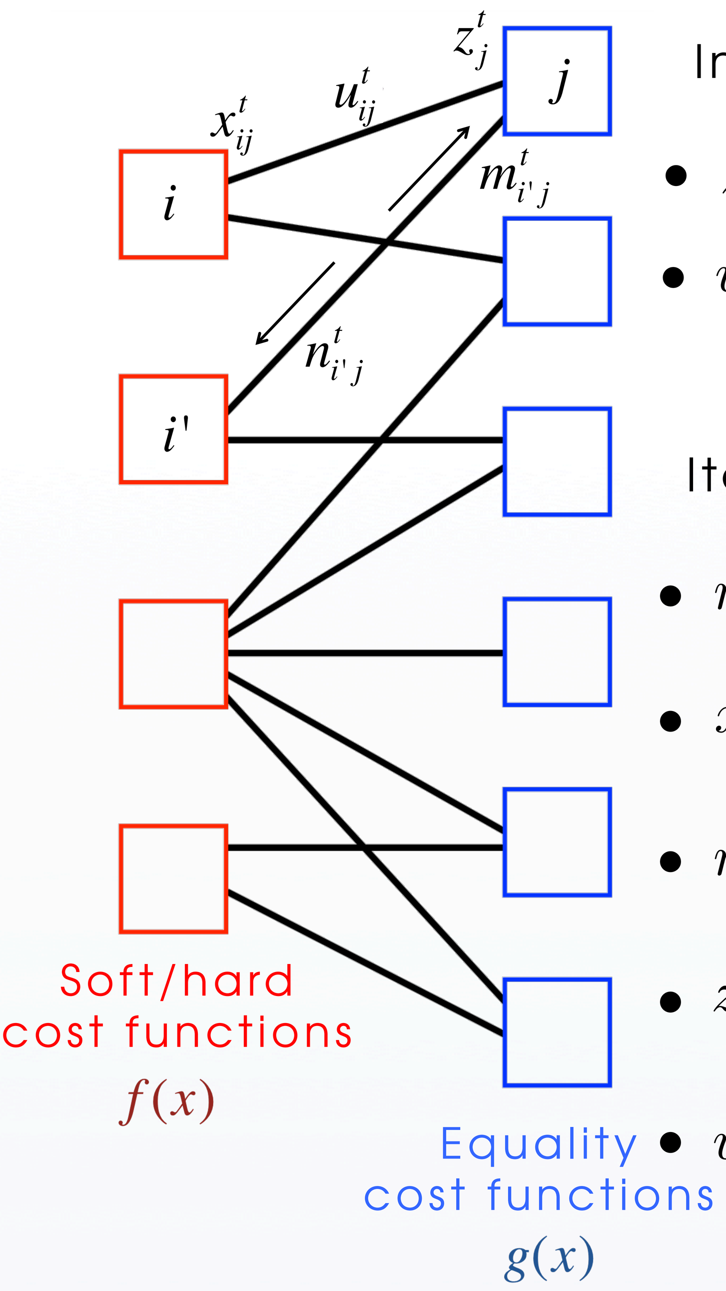
Improved message-passing algorithm incorporating certainty information

Nate Derbinsky José Bento Veit Elser Jonathan Yedidia

Summary Message-passing alg. based on ADMM | Distinct certainty weights: "no opinion"; "certain"; "standard" | Improved performance on non-convex opt. problems

Message-passing ADMM



Initialization:

- z^0 ← some value
- $u^0 = 0$

Iterations:

- $n^t = z^t - u^t$
- $x^t = \underset{x}{\text{argmin}} \left[f(x) + (\rho/2)(x - n^t)^2 \right]$
- $m^t = x^t + u^t$
- $z^{t+1} = \underset{z}{\text{argmin}} \left[g(z) + (\rho/2)(z - m^t)^2 \right]$
- $u^{t+1} = u^t + \frac{\alpha}{\rho}(x^t - z^{t+1})$

- Above algorithm is well-defined even for non-convex functions
- For convex functions global minimum is attained for weight parameter > 0 (see Boyd et al. 2011)
- Global convergence rate of $O(1/t)$ under convexity and linear rate under strict convexity. Dependency on ρ is significant and poorly understood (see Deng and Yin 2012)

Novel 3-weight algorithm

Different weights per iteration (t), edge (i, j) and direction $i \rightarrow j$

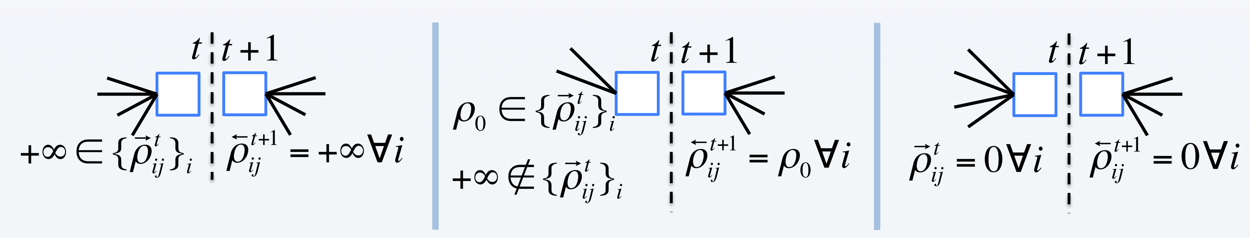
$\overleftarrow{\rho}_{ij}^t, \overrightarrow{\rho}_{ij}^t \in \{0, +\infty, \rho_0\}$ 'no opinion', 'certain', 'standard'

Initialization:

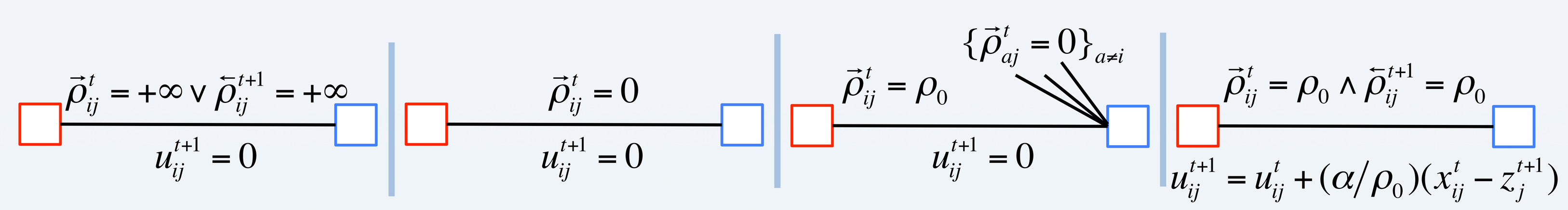
- z^0
- $u^0 = 0 \quad \overleftarrow{\rho}_{ij}^0 = 0$

Iterations:

- $n^t = z^t - u^t$
- $x^t = \underset{x}{\text{argmin}} \left[f(x) + (\overleftarrow{\rho}^t/2)(x - n^t)^2 \right]$
- $\overrightarrow{\rho}^t$ ← Update logic depends on individual cost function
- $m^t = x^t + u^t$
- $z_j^{t+1} = \sum_i \overrightarrow{\rho}_{ij}^t m_{ij}^t / \sum_i \overrightarrow{\rho}_{ij}^t$ (Exception: All-zeros incoming weights treated as all-standard incoming weights)
- $\overleftarrow{\rho}^{t+1}$ ← Update logic depends on 3 scenarios

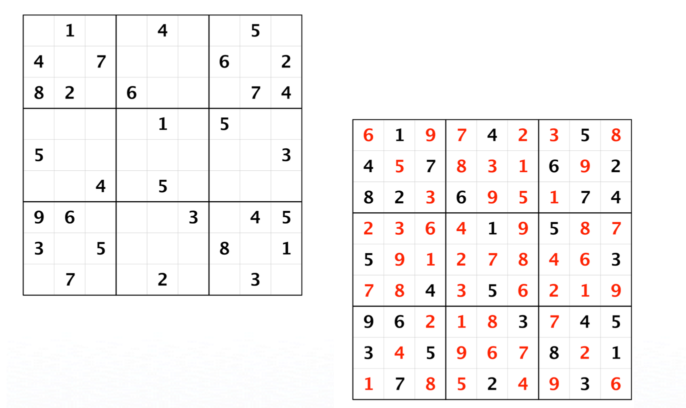


- u^{t+1} ← Update rule depends on 4 scenarios



Performance comparison

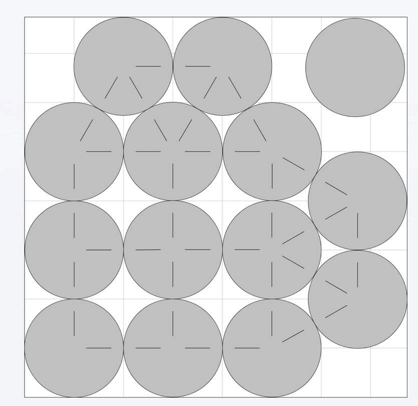
Sudoku (∞-weight intensive, non-convex problem)

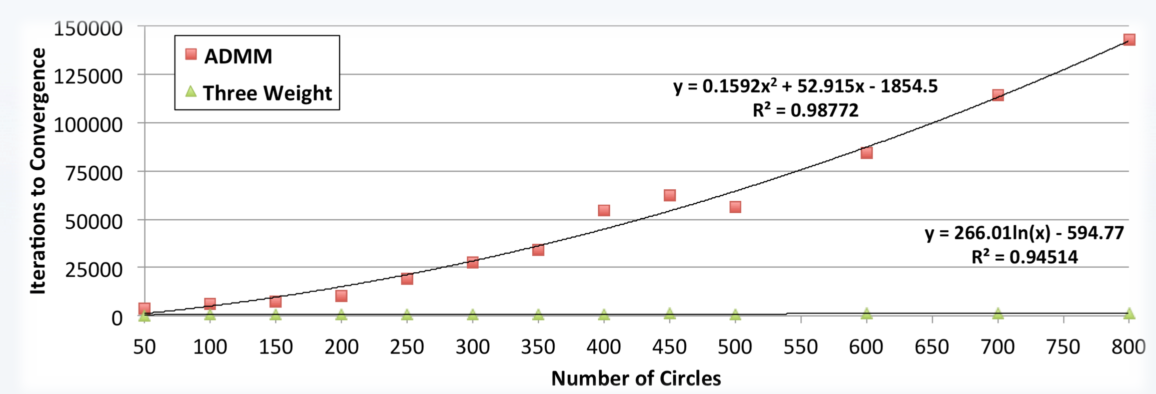


N	# Puzzles	% Improved $> 2\times$	Median Speedup
9	50	83.20%	3.35×
16	50	74.40%	3.65×
25	50	82.80%	5.58×
36	25	77.60%	5.35×
49	5	80.00%	4.44×
64	4	100.00%	6.01×
81	1	60.00%	2.03×

- N binary variables per cell (one per symbol, $O(N^3)$ total)
- One hard constraint per cell (one symbol per cell)
- One hard constraint per row, column and square (one symbol of each kind per column, per row and per square)
- ∞-weights used to propagate logical certainties from initial puzzle clues

Packing (0-weight intensive, non-convex problem)





- 2 real variables per circle (one per component)
- One hard constraint enforcing no overlap with box
- One hard constraint per pair of circles enforcing no overlap between them, $O(N^2)$ total
- 0-weights used to avoid interference from distant circles

References

- S. Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers", Foundations and Trends in Machine Learning (2011)
- W. Deng and W. Yin. "On the global and linear convergence of the generalized alternating direction method of multipliers", No. RICE-CAAM-TR12-14. Rice University (2012)