Rethinking Warm-Starts with Predictions: Learning Predictions Close to Sets of Optimal Solutions for Faster L-/$\mathrm{L}^\natural$-Convex Function Minimization

Shinsaku Sakaue, Taihei Oki

The University of Tokyo

Warm-Starting Algorithms with Predictions

Warm-start algorithms with predictions learned from past instances.

Time complexity is typically written as

prediction error

$$O(\text{single-step time} \times \|p^* - \hat{p}\|)$$

Weighted perfect bipartite matching (Dinitz et al. '21, Chen et al. '22, Khodak et al. '22 , S. & Oki '22)

- Single step is done by the $O(m\sqrt{n})$-time Hopcroft—Karp algorithm
- p^* is an optimal dual solution
- $\hat{p}$ is a dual prediction (learned from past p^*)

If $\hat{p} \approx p^*$, faster than the $O(mn)$-time Hungarian method 😊

Warm-Starting Algorithms with Predictions

Warm-start algorithms with predictions learned from past instances.

Time complexity is typically written as

prediction error

$$O(\text{single-step time} \times \|p^* - \hat{p}\|)$$

Weighted perfect bipartite matching (Dinitz et al. '21, Chen et al. '22, Khodak et al. '22 , S. & Oki '22)

- Single step is done by the $O(m\sqrt{n})$-time Hopcroft—Karp algorithm
- p^* is an optimal dual solution
- $\hat{p}$ is a dual prediction (learned from past p^*)

If $\hat{p} \approx p^*$, faster than the $O(mn)$-time Hungarian method 😊

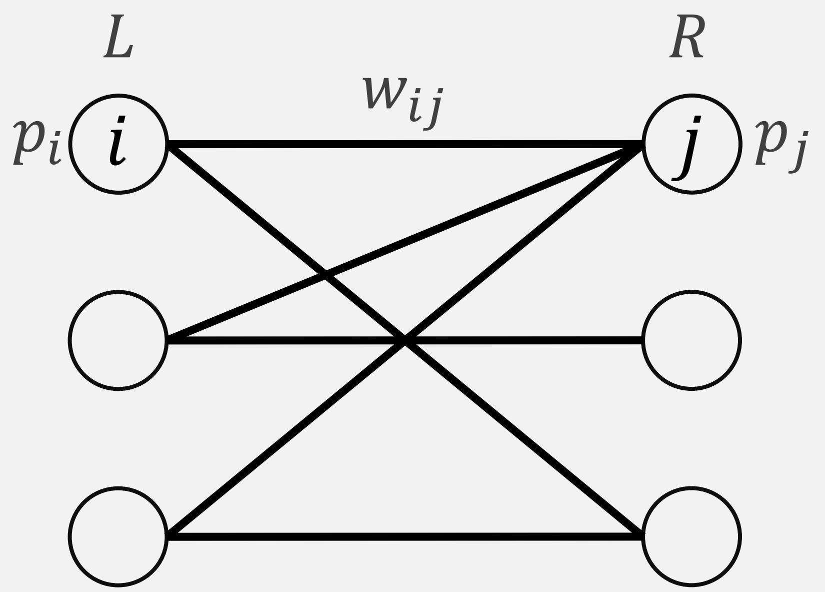
…assuming $\|p^* - \hat{p}\|$ is well-defined (or optimal p^* is unique) 🤔

Multiple Optimal Solutions Always Exist!

Dual LP of weighted perfect bipartite matching

minimize $\sum_{i \in L} p_i - \sum_{i \in R} p_j$

subject to $p_i - p_j \geq w_{ij}$ for all edges (i, j)



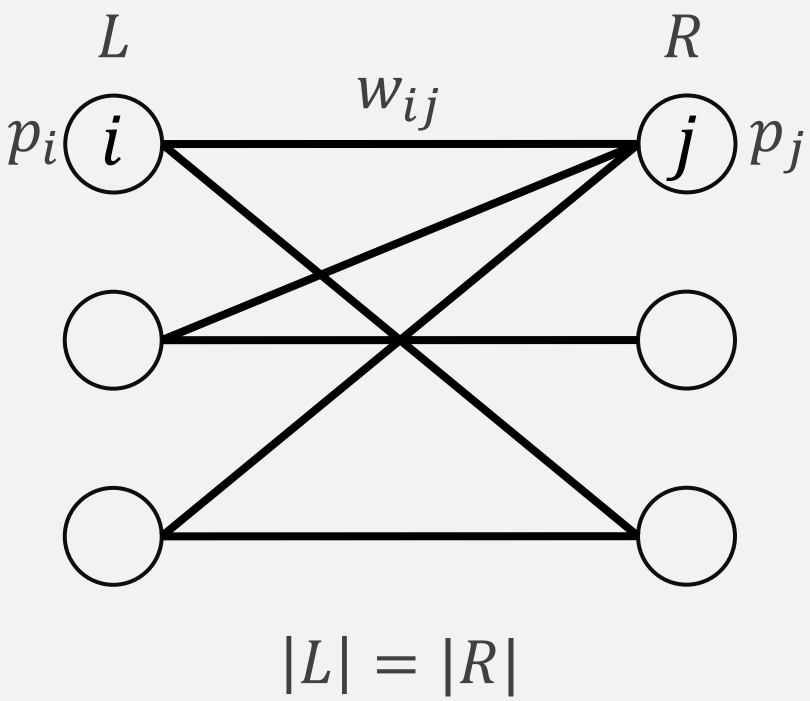
$|L| = |R|$

Multiple Optimal Solutions Always Exist!

Dual LP of weighted perfect bipartite matching

minimize $\sum_{i \in L} p_i - \sum_{i \in R} p_j$

subject to $p_i - p_j \geq w_{ij}$ for all edges (i, j)



p and $p + \alpha \mathbf{1}$ have the same objective value and l.h.s. of the constraints:

- $\sum_{i \in L}(p_i + \alpha) - \sum_{i \in R}(p_j + \alpha) = \sum_{i \in L} p_i - \sum_{i \in R} p_j$
- $(p_i + \alpha) - (p_j + \alpha) = p_i - p_j$

If p^* is optimal, so is $p^* + \alpha \mathbf{1}$ for all $\alpha \in \mathbb{R}$.

$\Rightarrow$ **Infinitely many optimal solutions exist!** 😱

Multiple Optimal Solutions Always Exist!

Need to predetermine p^* uniquely by some tie-breaking rule.

However, the $\|p^* - \hat{p}\|$-dependent bound becomes poor if p^* is far from $\hat{p}$ 😢

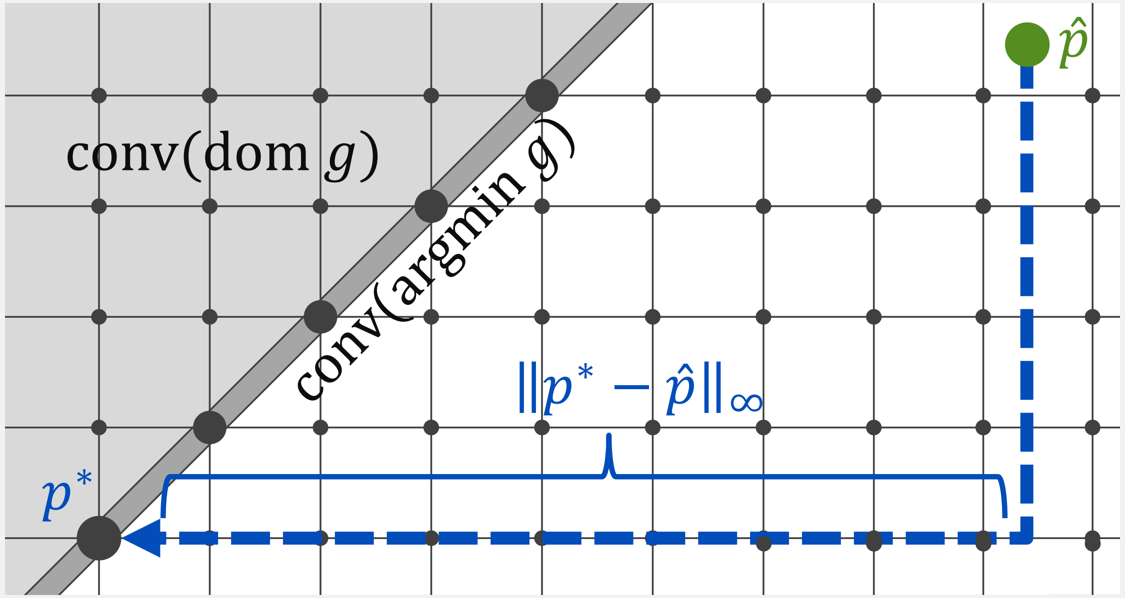
Multiple Optimal Solutions Always Exist!

Need to predetermine p^* uniquely by some tie-breaking rule.

However, the $\|p^* - \hat{p}\|$-dependent bound becomes poor if p^* is far from $\hat{p}$ 😢

Similar issues occur in more general L-/$\mathrm{L}^\natural$-*convex minimization*. (S. & Oki '22)

including bipartite matching, matroid intersection, min-cost flow, etc.



Our Idea: Learn Predictions Close to Sets of Optimal Solutions

Learn $\hat{p}$ to minimize $\bar{\mu}(\hat{p}) \coloneqq \min\{\|p^* - \hat{p}\|_\infty^\pm \mid p^* \in \mathrm{conv}(\mathrm{argmin}\, g)\}$.

distance between $\hat{p}$ and the set of optimal solutions

✓ L-/L$^\natural$-convex min. alg. takes $O(\text{single}-\text{step time} \times \bar{\mu}(\hat{p}))$ time.

based on a DCA result (Murota & Shioura '14)

✓ We can provably learn $\hat{p}$ to minimize $\bar{\mu}(\hat{p})$ approximately in polynomial time.