Dueling Convex Optimization with General Preferences

Aadirupa Saha (UIC), Tomer Koren (Tel Aviv University, Google), Yishay Mansour (Tel Aviv University, Google)

International Conference on Machine Learning, 2025

DCO with General Preferences

Motivation

•Many applications (recommender systems, search, ranking) rely on pairwise preferences rather than absolute scores.
•Classical convex optimization assumes access to gradients or function values.
•How to optimize if you only get noisy *comparisons* of pairs of decisions?

Problem Setting

- Minimize a convex function $f : \mathbb{R}^d \to \mathbb{R}$
- Feedback: only a **noisy 1-bit preference** between two points
- Formally, at each query of (w, w'):

$$\mathbb{E}[o \mid w, w'] = \rho\big(f(w) - f(w')\big)$$

where

- $o \in \{\pm 1\}$
- ρ is a transfer function mapping function differences to preferences
- The optimization algorithm never observes actual function values.

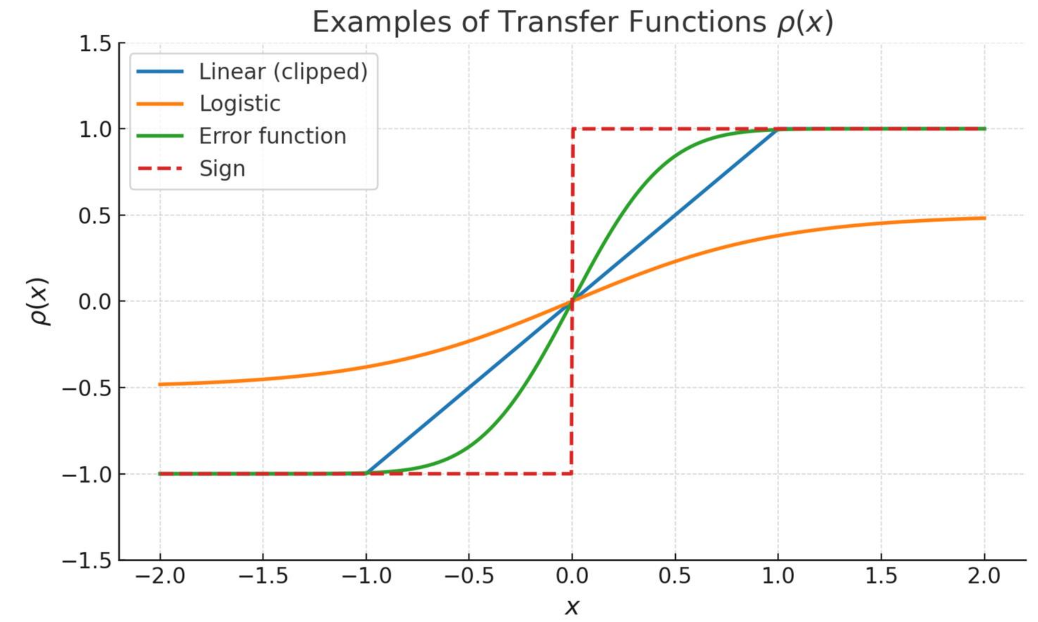
Why Hard?

•Pairwise preferences are even weaker than zero-order oracles.
•The transfer function ρ can distort magnitude information about f.
•Existing methods typically assume simple transfers (sign, polynomial) — this paper tackles arbitrary well-behaved ρ.

A No-Regret Algorithm for General DCO

Structure of ρ

- Differentiable, monotonic
- $\rho(0) = 0$
- Series expansion about 0
  $\rho(x) = \sum_{n=p}^{\infty} a_n x^n$
- $p \geq 1,\ a_p > 0$
- Bounded coefficients



Projected Dueling Descent (PDD)

- At round t:
  1. Sample a random direction u_t
  2. Compare two perturbed points around current iterate:
  $$x_t = w_t + \gamma u_t, \quad y_t = w_t - \gamma u_t$$
  3. Receive 1-bit feedback o_t
  4. Update in direction of "relative gradient" estimate:
  $$g_t = o_t \cdot u_t$$
  5. Project back onto feasible domain
- Crucially: PDD is **agnostic** to the form of ρ.

Why Relative Gradient Descent?

- You cannot estimate the true gradient from preferences
- Instead, you estimate a directional signal proportional to the gradient, up to a polynomial distortion
- PDD carefully controls the bias from this distortion.

Theoretical Performance Analysis

Convergence of Convex+β-smooth functions

Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex, β-smooth, G-Lipschitz over D with diameter D.
For any $0 < \epsilon < \bar{\epsilon} = 5\beta \min\left\{dD^2, \sqrt{drD/G}\right\}$ and $\delta \in (0,1)$,
run Algorithm 1 with $\gamma = \dfrac{\epsilon}{10\beta D\sqrt{d}}$ and $0 < \eta \leq \dfrac{c_\rho p \epsilon^{2p}}{\beta^p (80D)^{2p-1} d^{p+1/2}}$
for $T > 2\left(\dfrac{D^2}{\eta^2} + 1\right)\log\dfrac{1}{\delta}$ steps, then with probability at least $1-\delta$,
$\exists 1 \leq t \leq T$ such that $f(w_t) \leq f(w^*) + \epsilon$, with sample complexity $O(\epsilon^{-4p})$.

Algorithm for smooth & strongly convex functions

- Runs PDD in successive **epochs** with decreasing error tolerance
- Warm-starts each epoch from the previous solution
- Shrinks the domain radius across epochs
- Adapts learning rate and perturbation each epoch

Convergence with smooth & strongly convex functions

Let $f : \mathbb{R}^d \to \mathbb{R}$ be α-strongly convex, β-smooth, and G-Lipschitz over D.
Assume minimizer $w^* \in D$.
Then Algorithm 2 (Epoch-PDD) returns w such that $f(w) \leq f(w^*) + \epsilon$
with probability at least $1-\delta$,
using at most $\tilde{O}\left(\dfrac{\beta^{2p} d^{2p+1} D^{4p}}{c_\rho^2 \alpha^{2p} \epsilon^{2p}}\right)$ pairwise queries.

References

- Jamieson, K. & Nowak, R. Active Ranking using Pairwise Comparisons. NIPS'11.
- Ailon, N., Karnin, Z., & Joachims, T. Reducing Dueling Bandits to Cardinal Bandits. ICML'14.
- Saha, A., Koren, T., & Mansour, Y. Dueling Convex Optimization. ICML'21