Taiwan-Japan Joint Workshop on Nonlinear Analysis and Optimization (TJWNAO2026)
held in Nagaoka Univ. on Aug.3 and Niigata Univ. on Aug.4–5, 2026

CONE-ORDER REGRESSION FROM PREFERENCE DATA AND EXISTENCE OF ITS SOLUTIONS

YUTAKA SAITO
NIIGATA UNIVERSITY IN JAPAN

In machine learning and optimization, the choice of a reward function or an evaluation function has a substantial influence on the obtained solutions and generated outputs. Recent methods such as reinforcement learning from human feedback (RLHF) and direct preference optimization (DPO) use human pairwise preference data to learn reward models, policies, or policy objectives. In many such approaches, preference relations are often represented through scalar rewards or probabilistic pairwise-comparison models [5]. On the other hand, in multiobjective and vector optimization, there are frameworks in which multidimensional objects are compared directly by means of cone orders or Pareto orders, without necessarily scalarizing them [6, 1].

The aim of this study is to estimate an order structure from human pairwise preference data in a regression-like manner. The distinctive point of our approach is the combination of a circular cone model, text-pair embeddings of the type used in DPO datasets, and regression-like estimation of the induced order structure. We consider a set of preference observations on a vector space,

$$\{(x, y) \in X \times X \mid x \succ y\},$$

and approximate the relation $\succ$ by a cone order $\leq_K$ induced by a circular cone K. In many order-theoretic formulations, preference data are interpreted as observations of an underlying consistent preference relation. The present study is motivated by situations in which such an interpretation is difficult, as in human pairwise comparison data, and in which the data may contain contradictions or the unconstrained cone model may admit only a degenerate solution, such as an excessively wide cone that reduces empirical loss but yields a less informative order structure. We therefore introduce penalty terms so that a compromise solution can be obtained even in such cases.

For a sample $x_i \succ y_i$ of pairwise comparison data, we define the difference vector

$$z_i = x_i - y_i \quad (z_i \neq 0).$$

We interpret the observation as being better explained when z_i belongs to an unknown cone K. A basic estimation problem is formulated as

$$\widehat{K} \in \operatorname*{arg\,min}_{K \in \mathcal{K}} \left[\sum_{i=1}^{N} d(z_i, K) + \lambda \Omega(K) \right],$$

where d is a loss function associated with the cone, Ω is a regularization term controlling the complexity or width of the cone, and $\lambda > 0$ is a regularization parameter.

As a basic model, this talk focuses on a circular cone determined by a central direction $c \in \mathbb{R}^n$ and an angular parameter $\alpha = \cos\theta$:

$$K(c, \alpha) = \left\{ z \in \mathbb{R}^n \mid \langle z, c \rangle \leq \alpha \|z\|,\ \|c\| = 1,\ 0 < \alpha \leq 1 \right\}.$$

In this model, the object to be estimated is a finite-dimensional parameter consisting of the central direction and the angle. Compared with directly estimating a general closed convex cone, this setting is more tractable for discussing existence, stability, and visualization. As background, this type of conic constraint is also related to standard frameworks of conic optimization [3]. For example, using the angular margin loss

$$\ell(z; c, \alpha) = \left[\max\left\{ 0, \alpha - \frac{\langle z, c \rangle}{\|z\|} \right\} \right]^2,$$

we can construct a cone-order regression problem in which the loss becomes smaller as the observed preference difference approaches the central region of the cone.

*Presenting author.

We establish a basic existence result under compact parameter restrictions and continuous losses, and then discuss possible extensions toward stability, algorithmic convergence, and non-uniqueness. For instance, one may restrict the parameter space of (c, α) to

$$S^{n-1} \times [\varepsilon, 1] \quad (\varepsilon > 0)$$

and assume continuity of the loss function and the regularization term. Since uniqueness does not generally hold, we do not initially aim to prove global uniqueness. Instead, we discuss uniqueness from the viewpoints of isolated local solutions and uniqueness of the limit point for an initial region of an iterative algorithm.

In a validation phase, we plan to use DPO-style text-pair preference data as an example of pairwise preference observations, embed the texts into a vector space, estimate the circular cone parameters, and visualize the obtained central direction and angle. This visualization is expected to help explain latent evaluation axes in the dataset and possible biases of labelers. In this sense, the present study connects human preference learning represented by RLHF and DPO with multiple criteria decision-making research based on convex preference cones and cone contraction algorithms [2, 4].

References

[1] A. Dehnokhalaji, P. J. Korhonen, M. Koksalan, N. Nasrabadi, and J. Wallenius, *Convex cone-based partial order for multiple criteria alternatives*, Decision Support Systems, **51** (2011), 256–261.

[2] M. Kadziński, M. K. Tomczyk, and R. Słowiński, *Preference-based cone contraction algorithms for interactive evolutionary multiple objective optimization*, Swarm and Evolutionary Computation, **52** (2020), 100602.

[3] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Applications of second-order cone programming*, Linear Algebra and its Applications, **284** (1998), 193–228.

[4] N. Nasrabadi, A. Dehnokhalaji, P. Korhonen, and J. Wallenius, *Using convex preference cones in multiple criteria decision making and related fields*, Journal of Business Economics, **89** (2019), 699–717.

[5] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, *Direct preference optimization: Your language model is secretly a reward model*, Advances in Neural Information Processing Systems, **36** (2023).

[6] R. Vetschera, *Estimating preference cones from discrete choices: computational techniques and experiences*, Discussion Papers, Series I, No. 259, University of Konstanz, 1992.

(Y.Saito) Niigata University, Niigata 950–2181, Japan.
Email address: yutaka_saito@ge.niigata-u.ac.jp