Supplementary Material

A Proof of Prop. 2

Reflexivity ($\boldsymbol{v} \preceq_G \boldsymbol{v}, \forall \boldsymbol{v} \in V$) comes from the fact that the identity element 1_G leaves any $\boldsymbol{v}$ unchanged, and therefore $\boldsymbol{v} \in G\boldsymbol{v} \subseteq \mathcal{O}_G(\boldsymbol{v})$. To prove transitivity, it suffices to show that

$$\mathcal{O}_G(\boldsymbol{w}) \subseteq \mathcal{O}_G(\boldsymbol{v}) \quad \Leftrightarrow \quad \boldsymbol{w} \preceq_G \boldsymbol{v}. \tag{15}$$

The direct statement ($\Rightarrow$) follows from reflexivity: since $\boldsymbol{w} \in \mathcal{O}_G(\boldsymbol{w}) \subseteq \mathcal{O}_G(\boldsymbol{v})$, this implies $\boldsymbol{w} \in \mathcal{O}_G(\boldsymbol{v})$. For the converse statement ($\Leftarrow$), note that $\boldsymbol{w} \in \mathcal{O}_G(\boldsymbol{v})$ implies that $\boldsymbol{w} = \sum_i c_i h_i \boldsymbol{v}$ for $\{h_i\} \subseteq G$ and non-negative scalars $\{c_i\}$ that sum to one; using the linearity of the action, we then have that $g\boldsymbol{w} = \sum_i c_i g h_i \boldsymbol{v} \in \mathcal{O}_G(\boldsymbol{v})$ for any $g \in G$, which implies $G\boldsymbol{w} \subseteq \mathcal{O}_G(\boldsymbol{v})$ and $\mathcal{O}_G(\boldsymbol{w}) \subseteq \mathcal{O}_G(\boldsymbol{v})$ (due to convexity of $\mathcal{O}_G(\boldsymbol{v})$).

B Proof of Prop. 14

Let us start by noting that, for arbitrary $h \in G$,

$$\begin{aligned} \min_{h\in G} \frac{1}{2}\|h\boldsymbol{w} - \boldsymbol{a}\|^2 &= \min_{h\in G} \frac{1}{2}\|h\boldsymbol{w}\|^2 - \langle h\boldsymbol{w}, \boldsymbol{a}\rangle + \frac{1}{2}\|\boldsymbol{a}\|^2 \\ &= \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\|\boldsymbol{a}\|^2 - m(\boldsymbol{w}, \boldsymbol{a}) \\ &= \frac{1}{2}\|\boldsymbol{w} - \tilde{\boldsymbol{a}}\|^2, \end{aligned} \tag{16}$$

where $\tilde{\boldsymbol{a}} \in G\boldsymbol{a}$ is such that $m(\boldsymbol{w}, \boldsymbol{a}) = \langle \boldsymbol{w}, \tilde{\boldsymbol{a}}\rangle$; the optimal h satisfies $\tilde{\boldsymbol{a}} = h^{-1}\boldsymbol{a}$; and the second step is justified by the fact that G is a subgroup of $\mathsf{O}(d)$, hence its action is norm-preserving. Due to Moreau's decomposition theorem [34], we have that the projection in line 5 can be computed via proximal operator associated with $I^\star_{O_G(\boldsymbol{v})} = m_G(., \boldsymbol{v})$; namely we have that the (unique) minimizer $\boldsymbol{w}^*$ in line 5 satisfies $\boldsymbol{w}^* = \boldsymbol{a} - \mathrm{prox}_{m_G(.,\boldsymbol{v})}(\boldsymbol{a})$. Evaluating the proximal operator boils down to solving the following problem:

$$\begin{aligned} \min_{\boldsymbol{u}\in V} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{a}\|^2 + m_G(\boldsymbol{u}, \boldsymbol{v}) &= \min_{\boldsymbol{u}\in K_G(\boldsymbol{a})} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{a}\|^2 + m_G(\boldsymbol{u}, g\boldsymbol{v}) \\ &= \min_{\boldsymbol{u}\in K_G(g\boldsymbol{v})} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{a}\|^2 + \langle \boldsymbol{u}, g\boldsymbol{v}\rangle \\ &= \min_{\boldsymbol{u}\in K_G(g\boldsymbol{v})} \frac{1}{2}\|\boldsymbol{u} - (\boldsymbol{a} - g\boldsymbol{v})\|^2 + \text{constant}, \end{aligned} \tag{17}$$

where we used Eq. 16 and the fact that $g\boldsymbol{v} \in K_G(\boldsymbol{a})$. This leads to the result.

C Proof of Convergence of the Continuation Algorithm

We show that for any $\epsilon > 0$, the sequence $(L(\boldsymbol{w}_1), L(\boldsymbol{w}_2), \ldots)$ is strictly decreasing. Convergence follows from the fact that this sequence is lower bounded by the unregularized objective value $\min_{\boldsymbol{w}} L(\boldsymbol{w})$, assumed finite. The proof consists of two steps:

1. Showing that, for any $\epsilon > 0$, $\boldsymbol{w}_t$ lies in the interior of $\mathcal{O}_G(\boldsymbol{v}_{t+1})$. This follows from the fact that $\boldsymbol{v}'_t$, $\boldsymbol{w}_t$, and their convex combination all belong to the region cone $K_G(\boldsymbol{w}_t)$; in this region the pre-order induced by G is a cone ordering w.r.t. the polar cone of K_G, from which we can derive $\boldsymbol{w}_t \in O_G(\alpha\boldsymbol{v}'_t + (1-\alpha)\boldsymbol{w}_t)$, leading to the desired statement.
2. Showing that $(L(\boldsymbol{w}_1), L(\boldsymbol{w}_2), \ldots)$ strictly decreases before the algorithm terminates. This is a simple consequence of the previous fact. Since $\boldsymbol{w}_t \in O_G(\boldsymbol{v}_{t+1})$, we must have $L(\boldsymbol{w}_{t+1}) \leq L(\boldsymbol{w}_t)$. If this holds with equality, then $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$ is an optimal solution at the $(t+1)$th iteration, but since it lies in the interior of $O_G(\boldsymbol{v}_{t+1})$, we have $\|\boldsymbol{w}_{t+1}\|_{G\boldsymbol{v}_{t+1}} < 1$ and the algorithm will terminate. Therefore we must have $L(\boldsymbol{w}_{t+1}) < L(\boldsymbol{w}_t)$ for the algorithm to proceed.