Mathematics of Heterogeneous Distance Metrics

Spatial model and geometry domain

We model voters and candidates as points in a d-dimensional hypercube:

$$x \in [-1,1]^d.$$

This choice is important for cosine distance: allowing negative coordinates permits directions spanning the full angular range $[0,\pi]$ (i.e. 180°). In contrast, sampling from $[0,1]^d$ restricts all vectors to the first orthant, which limits angles largely to $[0,\frac{\pi}{2}]$ (i.e. 90°), preventing cosine from expressing full opposition.

Cosine distance and its range

For nonzero vectors $u, v \in \mathbb{R}^d$, cosine similarity is

$$\cos(u,v) \;=\; \frac{u \cdot v}{\|u\|_2 \, \|v\|_2},$$

and cosine distance is

$$d_{\cos}(u,v) \;=\; 1 - \cos(u,v).$$

Because $\cos(u,v) \in [-1,1]$, cosine distance satisfies

$$d_{\cos}(u,v) \in [0,2],$$

achieving 0 when the vectors are aligned ($\theta = 0$) and 2 when the vectors are opposite ($\theta = \pi$).

Note on the zero vector. If $u = 0$ or $v = 0$, cosine similarity is undefined; implementations typically add a small ε to norms or define a fallback behavior.

Center/extreme classification (normalized radius)

To classify voters by how "central" they are, we use distance from the geometric center of the hypercube:

$$c \;=\; \frac{(-1)+1}{2}\mathbf{1} \;=\; 0.$$

For a voter at position x, define the (unnormalized) radius

$$r(x) \;=\; \|x - c\|_2.$$

The maximum possible radius in $[-1,1]^d$ is achieved at a corner $(\pm 1, \ldots, \pm 1)$:

$$r_{\max} \;=\; \sqrt{d}.$$

We normalize centrality to $[0, 1]$ via

$$\rho(x) \;=\; \frac{r(x)}{r_{\max}} \in [0, 1].$$

Then a threshold $\tau \in [0, 1]$ can be used to switch distance metrics:

$$\rho(x) \le \tau \Rightarrow \text{center metric}, \qquad \rho(x) > \tau \Rightarrow \text{extreme metric}.$$

1 Simulator Specification

1.1 Introduction

This document provides a complete mathematical specification of the spatial voting simulator with heterogeneous distance metrics. The simulator models elections where voters evaluate candidates based on spatial positions in a policy space, but different voters may use different distance metrics to measure "closeness" to candidates.

Methodology Note. Homogeneous comparisons are defined relative to a baseline metric chosen from the metric pair used in the heterogeneous assignment. By default, the baseline is the center metric; for completeness, results can also be computed relative to the extreme metric baseline.

1.2 Spatial Voting Model

1.3 Basic Setup

Let $\mathcal{P} = \{1, 2, \ldots, N\}$ be a set of N election profiles. For each profile $p \in \mathcal{P}$, we have:

- A set of voters $\mathcal{V}_p = \{1, 2, \ldots, n_v\}$ with n_v voters.
- A set of candidates $\mathcal{C}_p = \{1, 2, \ldots, n_c\}$ with n_c candidates.
- A d-dimensional policy space $\mathbb{R}^d$ (typically $d \in \{1, 2, 3, 4, 5, 7, 10\}$).

Each voter $v \in \mathcal{V}_p$ has a position $\mathbf{x}_v \in [-1, 1]^d$ representing their ideal policy point. Each candidate $c \in \mathcal{C}_p$ has a position $\mathbf{y}_c \in [-1, 1]^d$ representing their policy platform.

1.4 Position Generation

For uniform distribution (the primary method used in this research), positions are generated as:

$$\mathbf{x}_v \sim \text{Uniform}([-1, 1]^d) \quad \forall v \in \mathcal{V}_p \tag{1}$$

$$\mathbf{y}_c \sim \text{Uniform}([-1, 1]^d) \quad \forall c \in \mathcal{C}_p \tag{2}$$

All positions are independently and identically distributed within the hypercube $[-1, 1]^d$.

1.5 Distance Metrics

1.6 Homogeneous Distance Metrics

In the standard spatial voting model, all voters use the same distance metric. We define four distance metrics:

1.6.1 Euclidean (L2) Distance

For voter position $\mathbf{x}_v$ and candidate position $\mathbf{y}_c$:

$$d_{\mathrm{L2}}(\mathbf{x}_v, \mathbf{y}_c) = \|\mathbf{x}_v - \mathbf{y}_c\|_2 = \sqrt{\sum_{k=1}^{d}(x_{v,k} - y_{c,k})^2} \tag{3}$$

1.6.2 Manhattan (L1) Distance

$$d_{\mathrm{L1}}(\mathbf{x}_v, \mathbf{y}_c) = \|\mathbf{x}_v - \mathbf{y}_c\|_1 = \sum_{k=1}^{d} |x_{v,k} - y_{c,k}| \tag{4}$$

1.6.3 Chebyshev (L∞) Distance

$$d_{\mathrm{Chebyshev}}(\mathbf{x}_v, \mathbf{y}_c) = \|\mathbf{x}_v - \mathbf{y}_c\|_\infty = \max_{k=1,\ldots,d} |x_{v,k} - y_{c,k}| \tag{5}$$

1.6.4 Cosine Distance

The cosine distance measures angular separation rather than spatial distance:

$$d_{\mathrm{cosine}}(\mathbf{x}_v, \mathbf{y}_c) = 1 - \frac{\mathbf{x}_v \cdot \mathbf{y}_c}{\|\mathbf{x}_v\|_2 \|\mathbf{y}_c\|_2} = 1 - \frac{\sum_{k=1}^{d} x_{v,k} y_{c,k}}{\sqrt{\sum_{k=1}^{d} x_{v,k}^2}\sqrt{\sum_{k=1}^{d} y_{c,k}^2}} \tag{6}$$

Note: If $\|\mathbf{x}_v\|_2 = 0$ or $\|\mathbf{y}_c\|_2 = 0$, we define $d_{\mathrm{cosine}}(\mathbf{x}_v, \mathbf{y}_c) = 0$.

1.7 Heterogeneous Distance Assignment

1.8 Voter Centrality Measure

To assign different metrics to different voters, we first compute a *centrality measure* for each voter. The centrality $\gamma_v \in [0, 1]$ measures how “extreme” voter v is relative to the center of the policy space.

Let $\mathbf{c} = (0, 0, \ldots, 0) \in \mathbb{R}^d$ be the geometric center of the hypercube $[-1, 1]^d$. The centrality is computed as:

$$\gamma_v = \frac{\|\mathbf{x}_v - \mathbf{c}\|_2}{d_{\max}^{(\mathrm{center})}} \tag{7}$$

where $d_{\max}^{(\mathrm{center})} = \sqrt{d}$ is the maximum possible distance from the center to a corner of $[-1, 1]^d$. The centrality is then clipped to $[0, 1]$:

$$\gamma_v = \min\left(1, \max\left(0, \frac{\|\mathbf{x}_v - \mathbf{c}\|_2}{d_{\max}^{(\mathrm{center})}}\right)\right) \tag{8}$$

Thus, $\gamma_v = 0$ for voters at the center, and $\gamma_v = 1$ for voters at the extreme corners.

1.9 Center-Extreme Strategy

The center-extreme strategy assigns one metric to central voters and a different metric to extreme voters based on a threshold parameter $\theta \in [0, 1]$. For a given threshold θ and metric pair $(m_{\text{center}}, m_{\text{extreme}})$, the metric assignment function is:

$$m_v = \begin{cases} m_{\text{center}} & \text{if } \gamma_v \leq \theta \\ m_{\text{extreme}} & \text{if } \gamma_v > \theta \end{cases} \tag{9}$$

The distance from voter v to candidate c is then:

$$d_v(\mathbf{x}_v, \mathbf{y}_c) = d_{m_v}(\mathbf{x}_v, \mathbf{y}_c) \tag{10}$$

where d_{m_v} is the distance function corresponding to metric m_v.

1.10 Heterogeneous Distance Matrix

For a profile p with heterogeneous metrics, we compute a distance matrix $\mathbf{D}_p \in \mathbb{R}^{n_v \times n_c}$ where:

$$[\mathbf{D}_p]_{v,c} = d_v(\mathbf{x}_v, \mathbf{y}_c) = d_{m_v}(\mathbf{x}_v, \mathbf{y}_c) \tag{11}$$

Each row v uses the metric m_v assigned to voter v.

1.11 Utility Computation

1.12 Utility Functions

Voter utilities are computed from distances using a utility function. The primary function used in this research is the linear utility function:

$$u_v(c) = \max\left(0, 1 - \frac{d_v(\mathbf{x}_v, \mathbf{y}_c)}{d_{\max}^{(\text{diag})}}\right) \tag{12}$$

where, under the geometry $[-1, 1]^d$, we use $d_{\max}^{(\text{diag})} = 2\sqrt{d}$ as the maximum possible Euclidean distance in the hypercube (the diagonal length). This ensures utilities are normalized to $[0, 1]$ and are comparable across different voter positions.

1.13 Utility Matrix

For profile p, we construct a utility matrix $\mathbf{U}_p \in \mathbb{R}^{n_v \times n_c}$:

$$[\mathbf{U}_p]_{v,c} = u_v(c) \tag{13}$$

1.14 Preference Rankings

From utilities, we derive strict preference rankings. For voter v, the ranking π_v is a permutation of $\mathcal{C}_p$ such that:

$$\pi_v(1) \succ_v \pi_v(2) \succ_v \cdots \succ_v \pi_v(n_c) \tag{14}$$

where $\pi_v(1)$ is the most preferred candidate. The ranking is computed as:

$$\pi_v = \text{argsort}(-\mathbf{u}_v) \tag{15}$$

where $\mathbf{u}_v = [u_v(1), u_v(2), \ldots, u_v(n_c)]$ is the utility vector for voter v, and argsort returns indices sorted in descending utility order. Ties (within numerical tolerance $\epsilon = 10^{-9}$) are broken arbitrarily by index order.

1.15 Voting Rules

We implement three ordinal voting rules that operate on preference rankings:

1.16 Plurality Rule

Each voter casts one vote for their top-ranked candidate. The winner is the candidate with the most first-place votes:

$$\text{score}_p(c) = \sum_{v \in \mathcal{V}_p} \mathbf{1}[\pi_v(1) = c] \tag{16}$$

$$w_p = \arg\max_{c \in \mathcal{C}_p} \text{score}_p(c) \tag{17}$$

Ties are broken randomly.

1.17 Borda Count

Each voter assigns points: $n_c - 1$ points to their first choice, $n_c - 2$ to their second, ..., 0 to their last choice. The winner is the candidate with the highest total:

$$\text{score}_p(c) = \sum_{v \in \mathcal{V}_p} (n_c - \text{rank}_v(c) - 1) \tag{18}$$

where $\text{rank}_v(c)$ is the position of candidate c in voter v's ranking (0-indexed).

$$w_p = \arg\max_{c \in \mathcal{C}_p} \text{score}_p(c) \tag{19}$$

1.18 Instant Runoff Voting (IRV)

IRV proceeds in rounds. In each round:

1. Count first-place votes: $\text{votes}_p(c) = \sum_{v \in \mathcal{V}_p} \mathbf{1}[\pi_v(1) = c]$.
2. If any candidate has $> n_v/2$ votes, that candidate wins.
3. Otherwise, eliminate the candidate with fewest first-place votes.
4. Redistribute votes from the eliminated candidate to each voter's next-ranked remaining candidate.
5. Repeat until a candidate has a majority.

Formally, let $\mathcal{C}_p^{(r)}$ be the set of remaining candidates in round r, with $\mathcal{C}_p^{(0)} = \mathcal{C}_p$. In round r:

$$\text{votes}_p^{(r)}(c) = \sum_{v \in \mathcal{V}_p} \mathbf{1}\left[\pi_v\Big(\min\{k : \pi_v(k) \in \mathcal{C}_p^{(r)}\}\Big) = c\right] \tag{20}$$

The elimination rule is:

$$c_{\text{elim}}^{(r)} = \arg\min_{c \in \mathcal{C}_p^{(r)}} \text{votes}_p^{(r)}(c) \tag{21}$$

and $\mathcal{C}_p^{(r+1)} = \mathcal{C}_p^{(r)} \setminus \{c_{\text{elim}}^{(r)}\}$. The process terminates when $\exists c \in \mathcal{C}_p^{(r)} : \text{votes}_p^{(r)}(c) > n_v/2$, and $w_p = c$.

2 Experimental Methodology

2.1 Disagreement Rate

The primary measure of heterogeneity effects is the disagreement rate, defined as the fraction of profiles where the heterogeneous winner differs from the homogeneous winner. For a metric pair $(m_{\text{center}}, m_{\text{extreme}})$ and threshold θ, let:

- w_p^{het} be the winner under heterogeneous conditions (center uses m_{center}, extreme uses m_{extreme}),
- $w_p^{\text{homo}}(m)$ be the winner under homogeneous conditions when all voters use a baseline metric m.

We report disagreement relative to a chosen baseline metric $m_{\text{base}} \in \{m_{\text{center}}, m_{\text{extreme}}\}$:

$$D(m_{\text{center}}, m_{\text{extreme}}, \theta \mid m_{\text{base}}) = \frac{1}{N} \sum_{p=1}^{N} \mathbf{1}\left[w_p^{\text{het}} \neq w_p^{\text{homo}}(m_{\text{base}})\right] \times 100\% \tag{22}$$

2.2 Asymmetry Measure

For a metric pair (A, B), we define the asymmetry as:

$$\Delta_{A,B}(\theta \mid m_{\text{base}}) = |D(A \to B, \theta \mid m_{\text{base}}) - D(B \to A, \theta \mid m_{\text{base}})| \tag{23}$$

where $D(A \to B, \theta)$ denotes disagreement when A is assigned to center voters and B to extreme voters.

2.3 Voter Scaling Analysis

To understand how effects depend on voter count, we compute disagreement rates for different values of n_v:

$$D(n_v; m_{\text{center}}, m_{\text{extreme}}, \theta) = \frac{1}{N} \sum_{p=1}^{N} \mathbf{1}[w_p^{\text{het}}(n_v) \neq w_p^{\text{homo}}(n_v)] \tag{24}$$

We test $n_v \in \{10, 25, 50, 100, 200, 300, 400, 500\}$ and fit a linear trend:

$$D(n_v) = \alpha n_v + \beta + \epsilon \tag{25}$$

where α is the slope (change per voter) and β is the intercept.

2.4 Threshold Sweep

To detect phase transitions, we sweep the threshold parameter:

$$\theta \in \{0.05, 0.10, 0.15, \ldots, 0.95\} \tag{26}$$

For each θ, we compute $D(\theta)$ and analyze:

- Inflection points: values where $\frac{d^2 D}{d\theta^2} = 0$,
- Maximum curvature: $\max_\theta \left|\frac{d^2 D}{d\theta^2}\right|$,
- Sudden jumps: $\max_\theta \left|\frac{dD}{d\theta}\right|$.

2.5 Dimensional Scaling

To understand dimensional dependence, we compute disagreement for different dimensions d:

$$D(d; m_{\text{center}}, m_{\text{extreme}}, \theta) = \frac{1}{N} \sum_{p=1}^{N} \mathbf{1}[w_p^{\text{het}}(d) \neq w_p^{\text{homo}}(d)] \tag{27}$$

We test $d \in \{1, 2, 3, 4, 5, 7, 10\}$ and fit a power law:

$$D(d) = D_0 \cdot d^{\alpha} \quad \text{for } d \leq d_{\text{peak}} \tag{28}$$

Using log-log regression:

$$\log D(d) = \log D_0 + \alpha \log d + \epsilon \tag{29}$$

For $d > d_{\text{peak}}$, we fit a linear model:

$$D(d) = D_{\text{peak}} + \beta(d - d_{\text{peak}}) + \epsilon \tag{30}$$

2.6 Condorcet Metrics

2.6.1 Condorcet Winner

A candidate c is a Condorcet winner if they beat all other candidates in pairwise majority comparisons. Formally, define the pairwise margin:

$$m_{c,c'} = \sum_{v \in \mathcal{V}_p} \mathbf{1}[\pi_v(c) < \pi_v(c')] - \sum_{v \in \mathcal{V}_p} \mathbf{1}[\pi_v(c') < \pi_v(c)] \tag{31}$$

where $\pi_v(c) < \pi_v(c')$ means voter v prefers c to c' (lower rank number = higher preference). Candidate c is a Condorcet winner if:

$$\text{CW}(c) \Leftrightarrow \forall c' \neq c : \; m_{c,c'} > 0 \tag{32}$$

Equivalently, c beats c' in a pairwise majority vote if $m_{c,c'} > 0$, which occurs when more than half of voters prefer c to c'.

2.6.2 Condorcet Cycle

A profile has a Condorcet cycle if no Condorcet winner exists:

$$\text{Cycle}_p = \mathbf{1}[\nexists c \in \mathcal{C}_p : \text{CW}(c)] \tag{33}$$

The cycle rate is:

$$C = \frac{1}{N} \sum_{p=1}^{N} \text{Cycle}_p \times 100\% \tag{34}$$

2.6.3 Condorcet Efficiency

The Condorcet efficiency of a voting rule is the fraction of profiles where the rule selects the Condorcet winner (when one exists):

$$E = \frac{1}{N_{\text{CW}}} \sum_{p:\exists \text{CW}} \mathbf{1}[w_p = \text{CW}_p] \times 100\% \tag{35}$$

where N_{CW} is the number of profiles with a Condorcet winner, and CW_p is the Condorcet winner in profile p.

2.7 Experimental Configuration

2.7.1 Base Parameters

All experiments use the following base configuration:

- Number of profiles: $N = 200$ (minimum for statistical significance).
- Number of voters: $n_v = 100$ (base case, with scaling tests from 10 to 500).
- Number of candidates: $n_c = 5$.
- Dimensions: $d \in \{1, 2, 3, 4, 5, 7, 10\}$.
- Threshold values: $\theta \in \{0.05, 0.10, \ldots, 0.95\}$ (19 points).
- Random seed: fixed at 42 for reproducibility.
- Geometry method: uniform distribution in $[-1, 1]^d$.
- Utility function: linear (as defined above).

2.7.2 Metric Pairs Tested

We test all ordered pairs of the four metrics:

$$\mathcal{M} = \{\mathrm{L1}, \mathrm{L2}, \mathrm{Cosine}, \mathrm{Chebyshev}\} \tag{36}$$

For each pair $(A, B) \in \mathcal{M} \times \mathcal{M}$ with $A \neq B$, we test both directions:

- $A \to B$: center voters use A, extreme voters use B.
- $B \to A$: center voters use B, extreme voters use A.

This yields $4 \times 3 = 12$ ordered pairs.

2.7.3 Experimental Phases

2.7.4 Phase 1: Voter Scaling

For metric pair (L1, Cosine) with $\theta = 0.5$ and $d = 2$:

$$\{D(n_v) : n_v \in \{10, 25, 50, 100, 200, 300, 400, 500\}\} \tag{37}$$

2.7.5 Phase 2: Threshold Sweep

For metric pair (L1, Cosine) with $n_v = 100$ and $d = 2$:

$$\{D(\theta) : \theta \in \{0.05, 0.10, \ldots, 0.95\}\} \tag{38}$$

2.7.6 Phase 3: Dimensional Scaling

For metric pair (L1, Cosine) with $n_v = 100$ and $\theta = 0.5$:

$$\{D(d) : d \in \{1, 2, 3, 4, 5, 7, 10\}\} \tag{39}$$

2.7.7 Phase 4: Metric Pair Interactions

For all 12 ordered pairs with $n_v = 100$, $d = 2$, and $\theta = 0.5$:

$$\{D(A \to B), D(B \to A), \Delta_{A,B} : (A, B) \in \mathcal{M} \times \mathcal{M}, A \neq B\} \tag{40}$$

2.7.8 Phase 5: Verification

Re-run Phase 4 with $n_v = 500$ to verify conclusions at higher voter counts.

2.8 Statistical Analysis

2.8.1 Coefficient of Variation

To assess stability of effects across voter counts, we compute the coefficient of variation:

$$\text{CV} = \frac{\sigma_D}{\mu_D} \tag{41}$$

where μ_D and σ_D are the mean and standard deviation of disagreement rates across voter counts.

2.8.2 Power Law Fitting

For dimensional scaling, we fit:

$$\log D(d) = \alpha \log d + \beta + \epsilon \tag{42}$$

and compute the coefficient of determination:

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} \tag{43}$$

where:

$$\text{SS}_{\text{res}} = \sum_d \left(\log D(d) - \widehat{\log D}(d)\right)^2 \tag{44}$$

$$\text{SS}_{\text{tot}} = \sum_d \left(\log D(d) - \overline{\log D}\right)^2 \tag{45}$$

2.9 Interaction Strength Hierarchy

For each voting rule R, we rank metric pairs by average disagreement:

$$S_{A,B} = \frac{D(A \to B) + D(B \to A)}{2} \tag{46}$$

Pairs are sorted in descending order of $S_{A,B}$ to form the hierarchy.

3 Implementation and Validation

3.1 Implementation Details

3.1.1 Numerical Precision

- Distance computations use double precision (64-bit floating point).

- Utility normalization uses maximum distance per voter to avoid division by zero.
- Ranking tie-breaking uses $\epsilon = 10^{-9}$ tolerance.
- All random number generation uses NumPy's Mersenne Twister with fixed seed.

3.1.2 Computational Complexity

For a single profile with n_v voters and n_c candidates in d dimensions:

- Distance computation: $O(n_v \cdot n_c \cdot d)$.
- Utility computation: $O(n_v \cdot n_c)$.
- Ranking computation: $O(n_v \cdot n_c \log n_c)$ (sorting).
- Plurality: $O(n_v)$.
- Borda: $O(n_v \cdot n_c)$.
- IRV: $O(n_v \cdot n_c^2)$ worst case.

Total per profile: $O(n_v \cdot n_c^2 \cdot d)$ in the worst case (IRV).

3.1.3 Reproducibility

All experiments use:

- Fixed random seed: 42.
- Deterministic algorithms (no parallelization that could affect order).
- Exact same geometry generation across heterogeneous and homogeneous runs.

This ensures that differences in outcomes are solely due to metric heterogeneity, not random variation.

3.2 Validation

3.2.1 Methodology Validation

Homogeneous comparisons are performed using a baseline metric selected from the heterogeneous metric pair. Results may be reported relative to the center baseline, the extreme baseline, or both.

3.2.2 Statistical Validation

- Minimum 200 profiles per configuration for statistical significance.
- Multiple independent runs with different random seeds (when testing robustness).
- Cross-validation across voting rules.
- Verification at higher voter counts (500 voters).

4 Summary of Key Findings

The mathematical framework described above enables investigation of several phenomena:

4.1 Asymmetric Metric Interactions

The order of metric assignment matters: $D(A \to B, \theta) \neq D(B \to A, \theta)$ for most metric pairs (A, B). The asymmetry magnitude is:

$$\Delta_{A,B}(\theta) = |D(A \to B, \theta) - D(B \to A, \theta)| \tag{47}$$

4.2 Dimensional Scaling

Disagreement rates scale with dimension following a power law:

$$D(d) \sim d^{\alpha} \tag{48}$$

with scaling exponent α typically in the range $[9.0, 9.5]$ for the L1–Cosine metric pair.

4.3 Voter Count Dependence

Disagreement rates decrease with voter count:

$$D(n_v) = \alpha n_v + \beta \tag{49}$$

with negative slope $\alpha < 0$, indicating stronger heterogeneity effects in smaller electorates.

4.4 Interaction Strength Hierarchy

Different metric pairs create systematically different effects. Cosine-based assignments (Cosine $\to$ any other metric) show the strongest interactions ($\sim$ 58–62% disagreement), while L2-based assignments show minimal effects due to methodology constraints.

4.5 Conclusion

This mathematical specification provides a complete description of the spatial voting simulator with heterogeneous distance metrics. The simulator enables systematic investigation of how different distance metrics affect voting outcomes, with rigorous experimental methodology and statistical analysis.

The key innovation is the heterogeneous distance assignment, which models cognitive diversity in how voters perceive policy distances. This framework allows investigation of phenomena such as asymmetric metric interactions, dimensional scaling laws, and voter count dependencies.

The mathematical framework is fully specified, enabling replication and extension of the research. All experimental parameters, statistical measures, and computational procedures are precisely defined to ensure reproducibility.