# Detailed Interpretation of Phase-Rotation Experiments for a Sign-Aware Peak-to-Peak Jaccard Framework

Compiled Notes

September 25, 2025

**Abstract**

We present a comprehensive analysis of phase-rotation experiments evaluating a sign-aware, peak-to-peak Jaccard similarity metric for real-valued signals. This framework provides interpretable similarity measures that are sensitive to polarity, amplitude, and phase relationships. The analysis demonstrates superior performance over traditional metrics in scenarios where sign semantics matter, while maintaining mathematical rigor through positive semidefinite kernel properties and valid distance metrics.

## Contents

# 1 The Banana View: Why the Math Works (Plainly)

## 1.1 One line (the whole idea)

*Our score is just the fraction of the combined basket that is truly shared.* If $A$ and $B$ are two nonnegative banana piles split into bins (they could be time steps or categories),

$$J(A,B) \;=\; \frac{\sum_i \min\{A_i, B_i\}}{\sum_i \max\{A_i, B_i\}}.$$

- The top (numerator) adds up the bananas we **both** have in each bin: "shared bananas."
- The bottom (denominator) adds up the bananas that **at least one** of us has in each bin: "total available bananas."

**Tiny example (one bin).** I have 2 bananas, you have 4: shared $= \min(2,4) = 2$, union $= \max(2,4) = 4$. So $J = 2/4 = 0.5$ (that's 50%). If one person holds extra bananas that the other doesn't, the denominator grows while the numerator doesn't, so the fraction drops.

**Worked example (many bins).** Suppose across three bins we have

$$A = [2,\ 1,\ 0], \qquad B = [4,\ 0,\ 3].$$

Binwise mins and maxes are

$$\min = [2,\ 0,\ 0], \qquad \max = [4,\ 1,\ 3].$$

Totals: shared $= 2+0+0 = 2$, union $= 4+1+3 = 8$, so $J = 2/8 = 0.25$ (25%). Interpretation: only one quarter of the combined basket is truly held by both of us.

## 1.2 Signed signals = Apples and Oranges

Sometimes values can be positive or negative. We treat these like two different types of fruit that can't be compared: **apples** (positive values) and **oranges** (negative values).

$$A_i^+ = \max(A_i, 0) \quad \text{(the apple pile)}, \qquad A_i^- = \max(-A_i, 0) \quad \text{(the orange pile)}.$$

(and the same for $B$). Then our **sign-aware** similarity is

$$J_{\text{peak}}(A,B) \;=\; \frac{\sum_i \big(\min\{A_i^+, B_i^+\} + \min\{A_i^-, B_i^-\}\big)}{\sum_i \big(\max\{A_i^+, B_i^+\} + \max\{A_i^-, B_i^-\}\big)}, \qquad d_{\text{peak}} = 1 - J_{\text{peak}}.$$

**Plain meaning:** We only count shared fruit when the *type* matches (apples with apples, oranges with oranges). If one person has apples and the other has oranges in the same bin, there is no shared fruit in that bin. *(In practice you can treat very small values as zero with a tiny threshold $\varepsilon_0$ to avoid noisy sign flips.)*

*(Note: For a financial context, this is the same as comparing **Assets** (positive) and **Liabilities** (negative). You only find shared ground by comparing assets to assets and liabilities to liabilities.)*

**Mini signed example.** One bin: $A = +3$, $B = -5$. Then $A$ has 3 apples and 0 oranges; $B$ has 0 apples and 5 oranges. Shared apples $= \min(3,0) = 0$, shared oranges $= \min(0,5) = 0$, so total shared $= 0$. Union of apples $= \max(3,0) = 3$, union of oranges $= \max(0,5) = 5$, total union $= 8$. Thus $J_{\text{peak}} = 0/8 = 0$. Different fruit types $\Rightarrow$ no share.

### 1.3 Probability view (this ratio really is a probability)

Imagine all the fruit (apples and oranges) from both people are poured into one big basket (the *union*). Now pick *one* piece of fruit uniformly at random from that basket. Then

$$\Pr(\text{“picked fruit is shared by both”}) \;=\; J_{\text{peak}}(A,B)\,.$$

This matches our intuition: if $J_{\text{peak}} = 0.25$, then one out of every four pieces of fruit in the basket is a “shared” piece.

### 1.4 Many people (multi-way) and “budget closure”

With multiple people, we say the fruit *everyone* shares in a bin is the per-bin minimum across all of them (done separately for apples and oranges). The “budget closure” property simply means that if you account for all the fruit—what's unique to each person, what's shared by exactly two, etc.—no fruit is ever lost or double-counted. The accounting always balances perfectly.

Möbius inversion is just systematic bookkeeping to assign each piece to exactly one exclusive pile. Summing all exclusives yields the union (no fruit is lost or double-counted), and summing the exclusives that *contain* a person yields that person's total.

### 1.5 When you only know sizes: The “Basket Overflow” Bound or Union Bound

Suppose three people have $|A| = 2$, $|B| = 3$, $|C| = 4$ pieces of fruit, but we don't know the overlaps. We want to find the **guaranteed minimum overlap** between A and B.

Think of the total basket $\Omega = A \cup B \cup C$.

1. Imagine this basket has a fixed capacity of $|\Omega|$ items.
2. Now, try to dump all of A's fruit ($|A|$) and all of B's fruit ($|B|$) into it. The total you're trying to fit is $|A| + |B|$.
3. If $|A| + |B|$ is **greater than** the basket's capacity $|\Omega|$, then some fruit *must* have been double-counted. There wasn't enough space for it all to be unique.
4. The number of items forced to overlap is the **“overflow”**: $\max(0, |A| + |B| - |\Omega|)$.

**Example:**

- **Worst case (no overlap):** The basket must be huge to hold everything separately, so its capacity is $|\Omega| = 2 + 3 + 4 = 9$. The overflow for A and B is $(2 + 3) - 9 = -4$, so the guaranteed minimum overlap is **0**.
- **Best case (perfect nesting, $A \subseteq B \subseteq C$):** The total basket only needs a capacity of $|\Omega| = 4$. The overflow for A and B is $(2+3)-4 = 1$. This tells us that **at least 1** piece of fruit *must* be shared between A and B (the true overlap is $\geq 1$, and here it's actually 2).

### 1.6 Metric axioms made intuitive (banana version)

Let $d_{\text{peak}}(A,B) = 1 - J_{\text{peak}}(A,B)$ with the sign-split $A \mapsto (A^+, A^-)$ and $J_{\text{peak}}$ defined by per-bin mins/maxes on $(A^+, A^-)$.

- **Non-negativity.** You can't have negative fruit: shared amounts are $\sum_i \min\{\cdot,\cdot\} \geq 0$ and totals are $\sum_i \max\{\cdot,\cdot\} \geq 0$, so $0 \leq J_{\text{peak}} \leq 1$ and $d_{\text{peak}} \geq 0$.

- **Symmetry.** "My sharing with you" equals "your sharing with me": $\min\{A_i, B_i\} = \min\{B_i, A_i\}$ and $\max\{A_i, B_i\} = \max\{B_i, A_i\}$, hence $J_{\text{peak}}(A, B) = J_{\text{peak}}(B, A)$ and $d_{\text{peak}}(A, B) = d_{\text{peak}}(B, A)$.
- **Identity of indiscernibles.** If our fruit collections are identical *bin-by-bin and by sign* (i.e., $A^+ = B^+$ and $A^- = B^-$), then $J_{\text{peak}} = 1$ and $d_{\text{peak}} = 0$. Conversely, if $d_{\text{peak}} = 0$, then the mins equal the maxes in every bin for both signs, forcing $A^+ = B^+$ and $A^- = B^-$.
- **Triangle inequality (detour penalty).** Going $A \to B \to C$ cannot be shorter than the direct $A \to C$ when we measure "unshared portions." Formally, $d_{\text{peak}}$ is the (Ruzicka/Jaccard) distance on the *nonnegative* vector obtained by concatenating the two baskets $[A^+; A^-]$; this distance is known to satisfy the triangle inequality. Intuitively, any banana that fails to match from $A$ to $C$ cannot be *more* matched by detouring through $B$.

*Why the sign-split preserves these:* We just apply the same fruit-sharing logic to two *independent* baskets (apples = positives, oranges = negatives) and add them. Since each basket individually satisfies the axioms, their sum does too.

**Detour penalty (triangle inequality) — tiny example.** Let $A = [2, 0]$, $B = [1, 1]$, $C = [0, 2]$ be two-bin banana piles (nonnegative case). Compute Jaccard-style overlap $J(X, Y) = \frac{\sum_i \min\{X_i, Y_i\}}{\sum_i \max\{X_i, Y_i\}}$ and distance $d = 1 - J$:

$$
\begin{aligned}
J(A, B) &= \tfrac{\min(2,1)+\min(0,1)}{\max(2,1)+\max(0,1)} = \tfrac{1}{3}, & d(A, B) &= \tfrac{2}{3}, \\
J(B, C) &= \tfrac{\min(1,0)+\min(1,2)}{\max(1,0)+\max(1,2)} = \tfrac{1}{3}, & d(B, C) &= \tfrac{2}{3}, \\
J(A, C) &= \tfrac{\min(2,0)+\min(0,2)}{\max(2,0)+\max(0,2)} = 0, & d(A, C) &= 1.
\end{aligned}
$$

Triangle inequality holds:

$$
d(A, C) = 1 \ \le \ d(A, B) + d(B, C) = \tfrac{2}{3} + \tfrac{2}{3} = \tfrac{4}{3}.
$$

*Intuition:* $A$ has only bin 1 bananas, $C$ only bin 2 bananas. The "detour" via $B = [1, 1]$ shares a little with each leg, but the total unshared mass cannot be less than going directly from $A$ to $C$.

# 2 Synopsis

This section provides a thorough, panel-by-panel interpretation of two composite figures produced by phase-rotation experiments that evaluate a sign-aware, peak-to-peak Jaccard similarity for real-valued signals.

## 2.1 Key Contributions

The analysis accomplishes several important objectives:

- **Methodological validation**: Explains how the metric responds to phase, sign, and amplitude variations
- **Comparative analysis**: Contrasts performance with Pearson correlation and Euclidean distance
- **Multi-way extensions**: Interprets ensemble behavior and consensus structures

- **Practical guidance**: Provides implementation recipes, strengths assessment, and caveat identification
- **Theoretical foundations**: Establishes positive semidefinite kernel properties and Euclidean embedding capabilities
- **Intuitive exposition**: Presents core concepts through accessible "Banana View" metaphors

# 3 Mathematical Foundations

## 3.1 Core Definitions

Let $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$ denote positive and signed-negative magnitudes respectively. For two signals $A, B$ sampled at indices $i = 1, \ldots, n$:

$$N(A,B) = \sum_{i=1}^{n} \left(\min\{A_i^+, B_i^+\} + \min\{A_i^-, B_i^-\}\right) \tag{1}$$

$$U_{\text{peak}}(A,B) = \sum_{i=1}^{n} \left(\max\{A_i^+, B_i^+\} + \max\{A_i^-, B_i^-\}\right) \tag{2}$$

$$J_{\text{peak}}(A,B) = \begin{cases} \dfrac{N(A,B)}{U_{\text{peak}}(A,B)}, & \text{if } U_{\text{peak}}(A,B) > 0 \\ 1, & \text{if } U_{\text{peak}}(A,B) = 0 \end{cases} \tag{3}$$

The associated distance metric is defined as:

$$d_{\text{peak}} = 1 - J_{\text{peak}} \tag{4}$$

## 3.2 Experimental Parameters

Throughout these experiments, we employ a fixed zero-tolerance $\varepsilon_0 = 10^{-12}$ for sign tests to avoid spurious sign flips from numerical noise.

The test signals are defined as:

$$A(t) = \sin t \tag{5}$$

$$B(t) = 2\sin(t - \phi) \tag{6}$$

$$C(t) = 0.5\cos t \tag{7}$$

where $\phi \in [0, 360)$ is swept in equal increments.

**Sampling specifications:**

- Duration: Two complete periods
- Sample points: $n = 500$
- Phase divisions: $M = 36$ ($10°$ increments)
- Numerical tolerance: $\varepsilon_0 = 10^{-12}$

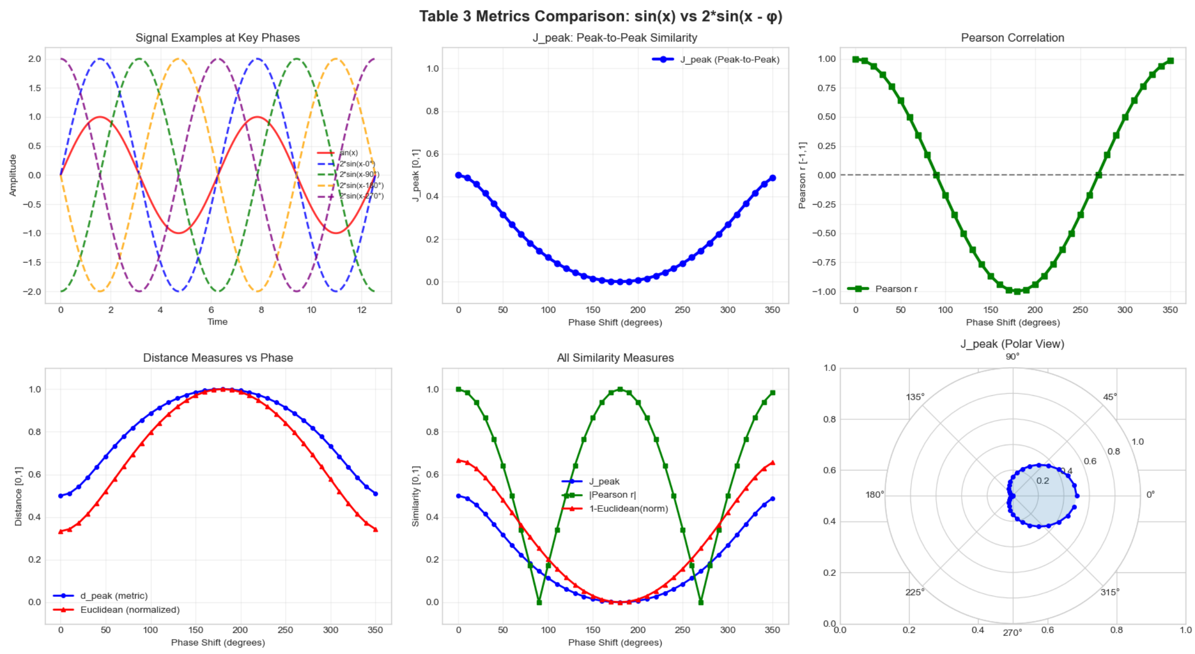


Figure 1: Pairwise experiment results showing six analytical perspectives on the phase-rotation relationship between unit sine and scaled sine signals.

# 4 Figure 1 Analysis: Pairwise Rotation $A(t) = \sin t$ vs $B(t) = 2\sin(t - \phi)$

## 4.1 Panel-by-Panel Interpretation

### 4.1.1 Top-Left: Signal Examples at Key Phases

This panel displays $A$ (unit sine) alongside several instances of $B$ at critical phase points: $\phi \in \{0, 90, 180, 270\}$ with amplitude factor 2.

**Key observations:**

(a) **Perfect alignment ($\phi = 0$):** Waves maintain identical sign patterns with 2:1 amplitude ratio

(b) **Perfect opposition ($\phi = 180$):** Complete sign disagreement—when $A$ is positive, $B$ is negative (and vice versa)

### 4.1.2 Top-Middle: $J_{\text{peak}}$ vs Phase (Peak-to-Peak Similarity)

The similarity curve exhibits a characteristic symmetric U-shape with distinct behavioral regions:

- **Maxima at $\phi \approx 0, 360$ ($J_{\text{peak}} \approx 0.5$):** For in-phase sinusoids with scale ratio $k \geq 1$, the min/max comparison yields $J_{\text{peak}} = 1/k$. With $k = 2$, we obtain $J_{\text{peak}} \approx 0.5$.
- **Minimum near $\phi \approx 180$ ($J_{\text{peak}} \approx 0$):** Anti-phase configuration results in systematic sign disagreement, collapsing the sign-aware intersection while maintaining large union.
- **Intermediate values at $\phi \approx 90, 270$:** Orthogonal phases still share some same-sign regions, producing small but non-zero similarity.

#### 4.1.3 Top-Right: Pearson Correlation vs Phase

As expected for sinusoidal pairs, the correlation follows $r(\phi) = \cos\phi$:

- $r = 1$ at $\phi = 0$
- $r = -1$ at $\phi = 180$
- $r = 0$ at $\phi = 90, 270$

**Critical limitation:** The absolute correlation $|r|$ cannot distinguish between 0 and 180 (both yield $|r| = 1$), whereas $J_{\text{peak}}$ correctly identifies 180 as near-zero similarity due to sign disagreement.

#### 4.1.4 Bottom-Left: Distance Comparison

- $d_{\mathbf{peak}} = 1 - J_{\mathbf{peak}}$**:** Mirrors the U-shape pattern—minimum at 0 ($\approx 0.5$ due to 2:1 amplitude ratio) and maximum near 180 (approaching 1.0)
- **Normalized Euclidean distance:** Also peaks near 180 due to maximum pointwise differences, but conflates amplitude and phase effects without clean sign semantics

#### 4.1.5 Bottom-Middle: Similarity Overlay

The overlay of $J_{\text{peak}}$, $|r|$, and (1 – Euclidean normalized) reveals:

- **At** 0**:** $J_{\text{peak}} \approx 0.5$ accurately reflects the 2:1 amplitude ratio, while $|r| = 1$ overstates similarity when amplitudes differ
- **At** 180**:** $|r| = 1$ suggests perfect similarity, but $J_{\text{peak}} \approx 0$ correctly identifies near-complete dissimilarity
- **At** 90**:** All measures decrease, but only $J_{\text{peak}}$ maintains interpretable ”same-sign overlap” semantics

#### 4.1.6 Bottom-Right: Polar Visualization

The polar representation emphasizes the periodic nature and pronounced minimum around 180. The constrained radius range ($\sim 0$ to $\sim 0.5$) reflects the fixed amplitude ratio limitation.

### 4.2 Figure 1 Summary

The $J_{\text{peak}}$ metric demonstrates three crucial properties:

1. **Sign awareness:** Anti-phase relationships yield near-zero similarity
2. **Scale awareness:** Best-case similarity equals the ratio of smaller to larger amplitude
3. **Phase resolution:** Successfully distinguishes 0 from 180, unlike $|r|$

## 5 Figure 2 Analysis: Three-Signal Ensemble $(A, B, C)$ with Rotating $B$

### 5.1 Panel-by-Panel Interpretation

#### 5.1.1 Top-Left: Component Signals

The configuration maintains $A(t) = \sin t$ and $C(t) = 0.5\cos t$ as fixed references (90° phase separation), while rotating $B(t) = 2\sin(t - \phi)$.

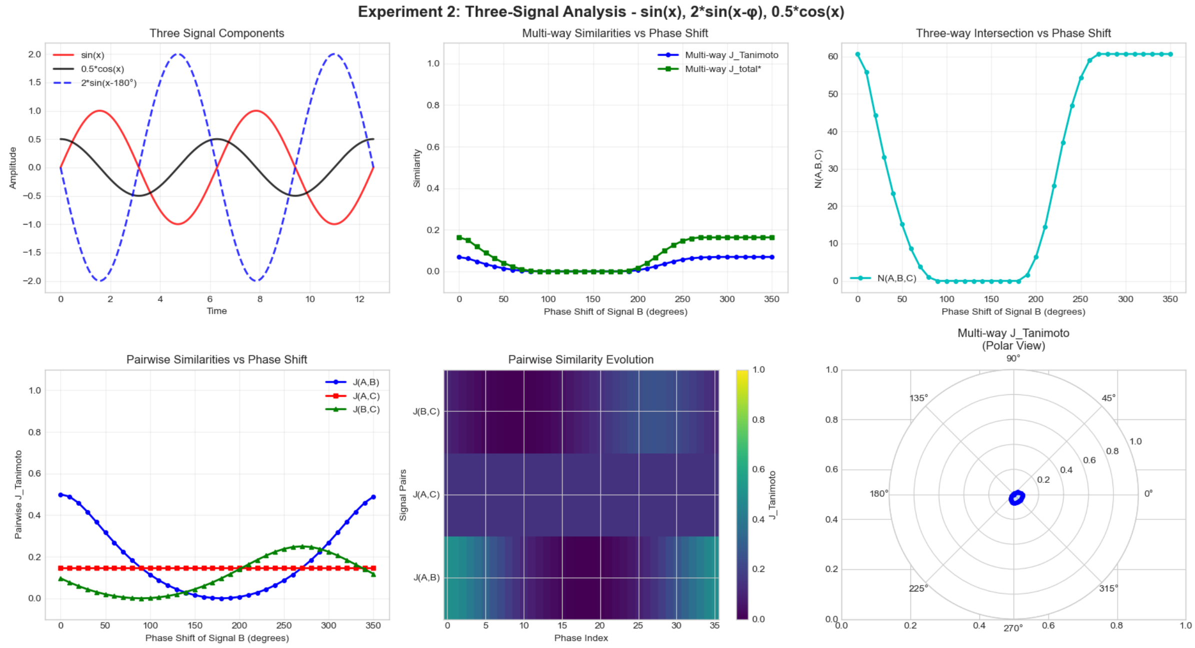


Figure 2: Three-signal experiment demonstrating multi-way consensus behavior with fixed signals $A$ and $C$, and rotating signal $B$.

**Structural constraint:** Since $C$ has the smallest amplitude (0.5), any triple intersection using per-index minima will be bottlenecked by $C$, even under optimal phase alignment.

#### 5.1.2 Top-Middle: Multi-Way Similarities vs Phase

Two closely related multi-way coefficients demonstrate:

- **Peak alignment:** Maximum values occur when $B$ aligns with $A$ (near $\phi \approx 0, 360$), maximizing the probability of simultaneous same-sign agreement across all three signals
- **Overall constraint:** Values remain small even at optimal phases due to $C$'s 90° phase offset from $A$ and reduced amplitude—triple same-sign agreement is structurally rare and magnitude-limited
- **Anti-phase suppression:** Deep minima around $\phi \approx 180$ result from $B$-$A$ anti-phase alignment destroying triple overlap

#### 5.1.3 Top-Right: Three-Way Intersection $N(A, B, C)$

The unnormalized common mass exhibits:

- **Sharp peaks:** Rapid increases near $\phi \approx 0$–30 and $\phi \approx 330$–360
- **Broad suppression:** Near-zero values across wide ranges ($\sim$ 90–270) where simultaneous same-sign agreement is geometrically difficult

#### 5.1.4 Bottom-Left: Pairwise Similarity Evolution

Individual pair relationships show distinct behaviors:

- $J(A, B)$**:** Strong phase dependence (replicating Figure 1 pattern)
- $J(A, C)$**:** Nearly constant and low due to fixed 90° phase separation

- $J(B, C)$**:** Phase-dependent, peaking where $B$ and $C$ achieve optimal sign alignment

**Key insight:** Strong pairwise coalitions can exist even when triple consensus approaches zero.

#### 5.1.5 Bottom-Middle: Pairwise Evolution Heatmap

Provides compact visual confirmation of bottom-left panel trends:

- AB relationship brightens near 0, dims near 180
- AC relationship remains uniformly low
- BC relationship brightens where $B$ and $C$ achieve sign alignment

#### 5.1.6 Bottom-Right: Polar Multi-Way Similarity

The small-radius ring pattern emphasizes the stringent requirements for multi-way similarity:

- Limited by rare three-way same-sign occurrences
- Constrained by amplitude bottlenecks (half-amplitude cosine component)
- Maintains clear periodicity despite restrictions

# 6 Practical Implementation Guidance

## 6.1 Optimal Use Cases

The sign-aware, peak-to-peak Jaccard framework excels in scenarios where polarity and relative amplitude carry semantic meaning:

- **Physical systems:** Sources vs sinks, positive/negative fluxes, charge/discharge cycles
- **Financial data:** Profits/losses, inflows/outflows, market movements
- **Biological signals:** Upregulation/downregulation, excitatory/inhibitory responses
- **Engineering applications:** Control signals, system responses, error measurements

## 6.2 Advantages Over Traditional Metrics

### 6.2.1 Versus Pearson Correlation

- Cannot distinguish 0 from 180 phase relationships (both yield $|r| = 1$)
- $J_{\text{peak}}$ correctly identifies anti-phase as dissimilar
- Provides interpretable amplitude-aware similarity scores

### 6.2.2 Versus Euclidean Distance

- Conflates phase and amplitude effects
- $J_{\text{peak}}$ maintains clean overlap semantics through min/max operations on signed components
- Lacks direct interpretability for signal analysis

## 6.3 Mathematical Properties

### 6.3.1 Kernel and Positive Semidefinite Structure

Defining kernel $K_{\text{peak}}(A, B) = J_{\text{peak}}(A, B)$, under the sign-split embedding that maps each real signal to nonnegative vectors of positive and negative parts:

- $J_{\text{peak}}$ corresponds to min/max Tanimoto (Jaccard) similarity in nonnegative feature space
- Resulting Gram matrix $K = [K_{\text{peak}}(A_i, A_j)]_{i,j}$ is **positive semidefinite**
- Centered kernel $K_c = HKH$ (where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$) defines valid covariance in RKHS
- Enables standard PCA, Gaussian process, and SVM implementations

### 6.3.2 Euclidean Embedding Properties

Since $d_{\text{peak}}$ constitutes a valid distance metric:

- Classical metric MDS on matrix $[d_{\text{peak}}(A_i, A_j)]$ yields Euclidean embeddings
- Preserves $d_{\text{peak}}$ relationships optimally in reduced dimensions
- Supports both mass/overlap interpretation and distance geometry computation

# 7 Quick Reference Tables

## 7.1 Figure 1 Panel Summary

| **Panel** | **Content** | **Key Takeaway** |
|---|---|---|
| Top-left | Signal examples at critical phases | 2:1 amplitude ratio and sign alignment/opposition establish $J_{\text{peak}}$ ceiling/floor |
| Top-middle | $J_{\text{peak}}$ vs phase | U-shaped curve: max $\approx 0.5$ at 0° (amplitude ratio), near 0 at 180° (sign clash) |
| Top-right | Pearson correlation vs phase | $r = \cos\phi$; $\lvert r\rvert$ treats 0° and 180° identically, missing sign semantics |
| Bottom-left | Distance comparison | $d_{\text{peak}}$ exhibits U-shape; Euclidean distance conflates phase and amplitude effects |
| Bottom-middle | Similarity overlay | Only $J_{\text{peak}}$ distinguishes aligned vs anti-aligned while maintaining interpretability |
| Bottom-right | Polar visualization | Clear periodicity with pronounced 180° minimum; radius limited by amplitude ratio |

Table 1: Panel-by-panel summary for pairwise rotation experiment

## 7.2 Figure 2 Panel Summary

# 8 Robustness and Implementation Notes

## 8.1 Numerical Considerations

- **Near-zero handling:** With finite sampling and small but nonzero $\varepsilon_0$, similarity minima at 180° approach rather than achieve exact zero

| Panel | Content | Key Takeaway |
|---|---|---|
| Top-left | Three-component system | Triple overlap constrained by smallest amplitude and phase geometry |
| Top-middle | Multi-way similarities | Peaks near 0°/360°; overall small due to C's 90° offset and reduced amplitude |
| Top-right | Three-way intersection mass | Near-zero across broad ranges where simultaneous same-sign agreement is rare |
| Bottom-left | Pairwise relationships | Strong bilateral coalitions can exist despite minimal triple consensus |
| Bottom-middle | Pairwise heatmap | Visual confirmation: AB varies strongly, AC remains low, BC shows phase dependence |
| Bottom-right | Polar multi-way view | Small radius reflects stringent triple same-sign requirements and bottlenecks |

Table 2: Panel-by-panel summary for three-signal ensemble experiment

- **Parameter sensitivity:** Variations in sample count $n$ or tolerance $\varepsilon_0$ affect only fine details near extrema; qualitative behavior patterns remain stable
- **Computational efficiency:** Implementation requires only element-wise min/max operations and summation—highly efficient for large signals

### 8.2 Experimental Reproducibility

All figures were generated using consistent parameters:

- Signal duration: Two complete periods
- Sample count: $n = 500$ points
- Phase resolution: $M = 36$ steps (10° increments)
- Numerical tolerance: $\varepsilon_0 = 10^{-12}$

## 9 Conclusion

The sign-aware, peak-to-peak Jaccard framework provides a mathematically rigorous and intuitively interpretable approach to signal similarity analysis. Its key advantages include:

1. **Semantic clarity:** Distinguishes between aligned and anti-aligned signals based on sign agreement
2. **Scale awareness:** Incorporates amplitude relationships in interpretable ratios
3. **Mathematical validity:** Guarantees positive semidefinite kernels and valid distance metrics
4. **Multi-way extension:** Naturally extends to ensemble analysis with budget closure properties
5. **Practical applicability:** Excels in domains where polarity carries semantic meaning

The experimental validation through systematic phase rotation demonstrates superior performance compared to traditional correlation and distance measures in scenarios requiring sign-sensitive similarity assessment. The framework's combination of intuitive interpretation ("banana sharing") and mathematical rigor positions it as a valuable tool for signal analysis applications where semantic relationships between positive and negative values are crucial.