ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Polarizing Kernels: A Definite Approach to Clustering with Indefinite Similarities

Maximilian Münch[1], Peter Preinesberger[2], Manuel Röder[2,3], Frank-Michael Schleif[2*]

1 - University of Groningen, Groningen, NL
2 - Technical University of Applied Sciences Würzburg-Schweinfurt, Würzburg, DE
3 - Bielefeld University, Bielefeld, DE

Abstract. Many real-world similarity measures are indefinite, violating the assumptions of kernel-based clustering methods. We propose a principled framework based on the *polar decomposition* of the similarity matrix, yielding a positive semi-definite component that preserves relational structure while enabling consistent out-of-sample extensions also for dissimilarities. The resulting *polarized kernels* support stable and interpretable clustering across synthetic and real datasets, demonstrating that polar decomposition provides a theoretically sound and practically effective bridge between indefinite similarity learning and kernel-based methods.

1 Introduction

Machine learning models rely on comparing data representations. For structured or non-vectorial inputs such as graphs, sequences, or time series, transformer-based deep models perform well when large datasets or pretrained foundations are available.An alternative are proximity-based methods, particularly kernel approaches employing a similarity function $s(\cdot,\cdot)$ [9]. Many real-world similarities, however, are *indefinite*, violating the assumptions of kernel methods that require $s(\cdot,\cdot)$ to be symmetric and positive semi-definite (psd) within a Reproducing Kernel Hilbert Space (RKHS) [10]. Although psd kernels are well studied, indefinite kernels often yield superior empirical performance [9, 4, 11], but their use can disrupt optimization stability and convergence [9]. Existing remedies, such as spectral corrections enforcing psd constraints, often distort the geometry or lose discriminative information. We address this by introducing a projection scheme based on the *polar decomposition*[1], mapping indefinite similarity matrices to psd counterparts while preserving relational structure and enabling efficient out-of-sample (OOS) extensions. In contrast to prior work focused on (semi-) supervised settings [5, 4], we target unsupervised clustering, extending our prior formulations [5] to distance-based methods with non-vectorial inputs. The rarity of unsupervised methods operating directly on indefinite similarity measures renders our approach particularly attractive by enabling the use of highly successful methods such as kernel k-means..

*Funded in parts by Europ. Reg. Dev. Fund, grant FKZ:2404-003-1.2 (EU-EFRE GREEN-INNO), and in parts by the ProPere THWS program, and cyber-security center TTZ-WUE.

[1]In its core identical to the one introduced and analyzed in [5]

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 Similarity Functions Beyond Positive Semi-Definiteness

Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ denote a set of N objects from an arbitrary input domain, and let $s : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a similarity function yielding a matrix $S \in \mathbb{R}^{N \times N}$ with entries $S_{ij} = s(x_i, x_j)$. If s is symmetric and positive semi-definite, S serves as a valid kernel matrix inducing a reproducing kernel Hilbert space $\mathcal{H}$ with feature map $\phi : \mathcal{X} \to \mathcal{H}$ satisfying: $s(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$. This *kernel trick* enables nonlinear modeling in the input space via linear operations in $\mathcal{H}$ [10]. For all $c \in \mathbb{R}^N$, psd implies $c^\top Sc \geq 0$, guaranteeing convex optimization and a consistent geometric interpretation. In practice, many effective similarities are *indefinite*, leading to matrices S with negative eigenvalues [9]. Such violations of Mercers condition frequently arise from domain-specific similarity measures, normalization schemes, or learned operators [9, 3, 2]. Indefinite similarities no longer define an RKHS but can be embedded in a *Krein space* [9, 3], an indefinite inner-product space decomposing as

$$\mathcal{K} = \mathcal{H}_+ \oplus \mathcal{H}_-, \quad \langle f, g \rangle_{\mathcal{K}} = \langle f_+, g_+ \rangle_{\mathcal{H}_+} - \langle f_-, g_- \rangle_{\mathcal{H}_-},$$

where $\mathcal{H}_+$ and $\mathcal{H}_-$ correspond to the positive and negative spectral subspaces of S. Learning in $\mathcal{K}$ generalizes classical kernel methods to indefinite settings, as shown in indefinite support vector machine (SVM) and least-squares formulations [7, 3, 4]. We recap our former work, slightly more formal (depicted in Fig. 1, employing the *polar decomposition*. Given $S \in \mathbb{R}^{N \times N}$, by using the polar decomposition we get: $S = UH$, where U is a real, unitary (orthogonal) operator capturing directional structure, if S is invertible (otherwise a partial isometry), and $H \succeq 0$ is symmetric positive semi-definite, that preserves the geometric information of the similarity. This decomposition *always exists* for any square matrix S or K and can be construction from the singular value decomposition (SVD). Let $K = W\Sigma V^\top$ with W, V orthogonal and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_N)$, $\sigma_i \geq 0$. Define $U = WV^\top$, $H = V\Sigma V^\top$. Then U is orthogonal (when K is invertible), H is symmetric psd, and $K = W\Sigma V^\top = \underbrace{WV^\top}_{U}\,\underbrace{V\Sigma V^\top}_{H}$. By construction, H is psd since for any $c \in \mathbb{R}^N$, $c^\top Hc = \|\Sigma^{1/2} V^\top c\|_2^2 \geq 0$. This formulation enables mapping an arbitrary, possibly indefinite similarity matrix to a psd representation H while retaining orientation through U.

2.1 Out-of-sample (OOS) extension

For indefinite similarity operators, extending a trained model to unseen data requires mapping new samples into the same positive semi-definite space used during training. Given the polar decomposition $K = UH$ of the training similarity, with orthogonal U and $H \succeq 0$, the mapping is obtained via $H' = U^\top K'$, where $K' \in \mathbb{R}^{N \times M}$ denotes the cross-similarity between training and test samples. H' thus represents the new samples in the psd space consistent with training, enabling direct use in kernel k-means, or SVM inference. We next address estimation of *self-similarities*, required for distance-based methods such as kernel k-means. While trivial in stationary psd kernels ($s(x_i, x_i) = c$), these values must

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

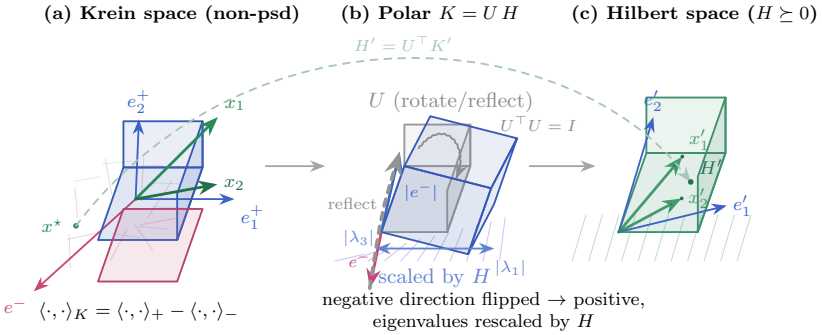

Fig. 1: Polarization of an indefinite kernel via polar decomposition. The orthogonal component U ($U^\top U = I$) re-orients negative directions, while the symmetric positive semi-definite factor $H = (K^\top K)^{1/2} \succeq 0$ performs stretching (no skew).

be explicitly estimated for indefinite operators under the polar decomposition, as detailed in the following section.

2.2 Self-similarity estimation

Let $K \in \mathbb{R}^{n\times n}$ denote the (possibly indefinite) training kernel, and $K_{\text{new}} \in \mathbb{R}^{n\times m}$ the test-to-train similarities mapped into the positive semidefinite space via the polar decomposition. Let $H \in \mathbb{R}^{n\times n}$ be the regularized psd matrix (e.g. $H = P^\top P + \lambda I$), $\mathbf{1} \in \mathbb{R}^n$ the all-ones vector, and $\Delta = \operatorname{diag}(K)$. For each test column $f_j = K^\top k_{\text{new},j}$, the weights $w_j \in \mathbb{R}^n$ are obtained as the solution of

$$\min_{w\in\mathbb{R}^n} \tfrac{1}{2}\, w^\top H w \;-\; f_j^\top w \quad \text{s.t.} \quad \mathbf{1}^\top w = 1. \tag{1}$$

The constrained LS normal equations are: $\begin{bmatrix} H & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} w_j \\ \lambda_j \end{bmatrix} = \begin{bmatrix} f_j \\ 1 \end{bmatrix}$,the KKT conditions yield the closed-form solution $w_j = H^{-1} f_j \;-\; H^{-1}\mathbf{1}\, \frac{\mathbf{1}^\top H^{-1} f_j - 1}{\mathbf{1}^\top H^{-1}\mathbf{1}}$. Stacking all m test samples, with $F = K^\top K_{\text{new}}$, we obtain $W = H^{-1}F \;- H^{-1}\mathbf{1}\, \frac{\mathbf{1}^\top H^{-1} F - \mathbf{1}^\top}{\mathbf{1}^\top H^{-1}\mathbf{1}}, \qquad W \in \mathbb{R}^{n\times m}$. In practice, $H^{-1}(\cdot)$ is computed via the Cholesky factorization. The estimated self-similarities for the test samples are then $\widehat{k}_{\text{self}} = W^\top \Delta, \qquad \widehat{k}_{\text{self},j} = \sum_{i=1}^n w_{ij} K_{ii}$. An equivalent approximation can be expressed in Nyström form as $\widehat{k}_{\text{self}} = \operatorname{diag}\big(K_{\text{new}}^\top (UK_0)^+ K_{\text{new}}\big)$, where UK_0 denotes the psd-corrected kernel operator obtained from the polar decomposition of the original (possibly indefinite) kernel $K = UH$, and $(\cdot)^+$ is the Moore-Penrose pseudoinverse, yielding similar results when the kernel is centered or well-conditioned, but without the affine normalization.

3 Clustering Methods on General Similarity Operators

To assess our framework, we employ three typical distance-based cluster methods operating directly on psd similarity matrices. For non metric inputs the methods are expected to fail.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Kernel k-means (KM). Kernel k-means [13] partitions the data $\mathcal{C} = \{C_1, \ldots, C_K\}$ by minimizing within-cluster squared distances in the implicit feature space $\phi(\cdot)$: $J(\mathcal{C}) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|\phi(x_i) - \mu_k\|^2, \qquad \mu_k = \frac{1}{|C_k|} \sum_{x_j \in C_k} \phi(x_j)$. Using $h(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, this expands to

$$J(\mathcal{C}) = \sum_{k=1}^{K} \left(\sum_{x_i \in C_k} h(x_i, x_i) - \frac{2}{|C_k|} \sum_{x_i, x_j \in C_k} h(x_i, x_j) + \frac{1}{|C_k|^2} \sum_{x_j, x_\ell \in C_k} h(x_j, x_\ell) \right).$$

For stationary psd kernels the diagonal $h(x_i, x_i)$ is constant and omitted, whereas in nonstationary or indefinite cases it is estimated as in Section 2.2.

Spectral clustering (SC). SC applies directly, as graph laplacians are psd in standard formulations [6], but negative similarities (not contradicting psd-ness), must be addressed. Following [6], the input S_o is symmetrized and its diagonal entries removed. We apply the polar decomposition $S_o = UH$ and use the positive semi-definite component H as a stabilized similarity matrix. H is regularized by a small diagonal offset, symmetrized, and truncated to nonnegative entries for numerical stability. The degree matrix $D = \mathrm{diag}(d_i)$ with $d_i = \sum_j H_{ij}$ defines the normalized affinity $L = D^{-1/2} H D^{-1/2}$. The top k eigenvectors of L form the spectral embedding $V \in \mathbb{R}^{N \times k}$ (rows ℓ_2-normalized), on which standard k-means is applied. Here we combine [6, 12] and our approach of polarized similarities.

Medoid k-means (MED). k-medoids performs prototype-based clustering directly on pairwise similarities, without prior embeddings [8]. Unlike kernel k-means, cluster representatives are restricted to data points. Given a (possibly indefinite) similarity matrix S, pairwise dissimilarities $d_{ij} = s_{ii} + s_{jj} - 2s_{ij}$ may become negative, complicating medoid selection. Using the psd component H from the polar decomposition of S ensures well-defined dissimilarities.Each sample x_i is assigned to the medoid x_{c_k} minimizing $d(x_i, x_{c_k})$.

4 Experiments

We evaluate the efficiency of polar decompositions as a preprocessing and OOS technique for indefinite kernel matrices in unsupervised similarity based learning. Our experimental setup incorporates various artificial and real life benchmark data sets, where information is encoded in the non-psd characteristics of the data. Data are split 80/20 for training and test. We restrict our experiments to popular, almost parameter free methods. The eigenspectrum's range is indicated by λ_{min} for the smallest and λ_{max} for the biggest eigenvalue. The distribution in percentage of $> 0, \approx 0, < 0$ eigenvalues is depicted by the signature in Tab 1. Additionally, we quantify the level of indefiniteness with $\tau = 0$ indicating a positive semi-definite and $\tau = 1$ a negative semi-definite matrix. We picked data with substantial indefiniteness, where the negative spectrum encodes relevant, meaningful information and is not just a numerical artifact.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Dataset	**N**	**G**	λ_{max}	λ_{min}	**signature**	τ
Anisotropic-I	900	3	6.25	−125	37%/61%/2%	0.91
Iris-I	150	3	0.01	−108.08	76%/10%/14%	0.99
Protein	213	4	72	−12.4	80%/1%/19%	0.2
Prodom	2604	53	311.07	−224.47	45%/16%/39%	0.21
Zongker	2000	10	1699.6	−13.7	52%/0%/48%	0.22

Table 1: *Signature*: percentage of positive, zero and negative eigenvalues v_i of S. τ is a measure to quantify indefiniteness $\left(\gamma = \sum_{v_i<0} |v_i| / \sum_i |v_i|\right)$.

Datasets (see also Table 1). To evaluate the proposed clustering method under controlled conditions, we generated artificial datasets from benchmark sources. **Anisotropic-I** was generated analogously from an anisotropic Gaussian mixture ($\gamma = 0.5$) to test robustness to directional variation. **Iris-I** is obtained from the normalized Iris data $\mathbf{X} \in \mathbb{R}^{150\times4}$. We formed an RBF kernel $K_{ij} = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $\gamma = 0.05$, symmetrized as $K = (K + K^\top)/2$. To obtain an indefinite kernel, we decomposed $K = E\Lambda E^\top$ and inverted the largest $m \approx 20$ eigenvalues, yielding $K^- = E\tilde{\Lambda}E^\top$, representing a pseudo-Hilbert space. The **Protein** dataset (213 sequences) [9] contains four globin classes (G, HA, HB, M) with strong indefiniteness. The **Zongker** dataset (2000 digits, 10 classes) provides dissimilarities from deformable-template matching on MNIST, characterized by many negative eigenvalues. The **Prodom** dataset (2604 sequences, 53 labels) [9] comprises pairwise structural protein alignments with high indefiniteness.

Results. The results presented in Table 2 clearly demonstrate that polarized kernels, which preserve metric dissimilarities, consistently outperform classical

	Original Kernel					**Polarized Kernel (ours)**				
Dataset	**Method**	**ARI**	**NMI**	**ACC**	**DUN**	**Method**	**ARI**	**NMI**	**ACC**	**DUN**
Anisotropic-I	KM	0.47	0.63	0.64	0.00	KM-P	**0.97**	**0.95**	**0.99**	**0.60**
	MED	0.25	0.43	0.59	0.00	MED-P	**0.47**	**0.63**	**0.67**	**0.11**
	SC	0.98	0.97	0.99	0.00	SC-P	0.98	0.97	0.99	**0.31**
Iris-I	KM	0.52	**0.75**	0.67	0.00	KM-P	**0.57**	0.65	**0.83**	**0.17**
	MED	0.45	0.58	0.73	0.00	MED-P	**0.56**	**0.71**	**0.80**	**0.12**
	SC	0.58	0.67	0.83	0.00	SC-P	**0.64**	**0.70**	**0.87**	0.00
Protein	KM	0.37	0.56	0.69	0.00	KM-P	**0.56**	**0.60**	**0.79**	0.00
	MED	0.14	0.35	0.52	0.00	MED-P	**0.39**	**0.60**	**0.71**	**0.08**
	SC	**0.63**	**0.72**	**0.86**	0.00	SC-P	0.44	0.64	0.76	0.00
Zongker	KM	0.11	0.31	0.32	0.00	KM-P	**0.75**	**0.82**	**0.88**	**0.17**
	MED	**0.06**	0.16	0.23	0.00	MED-P	0.02	**0.22**	**0.24**	**0.21**
	SC	0.70	0.75	0.86	0.00	SC-P	**0.80**	**0.84**	**0.90**	0.00
Prodom	KM	0.38	0.78	0.88	**0.17**	KM-P	**0.40**	**0.82**	**0.90**	0.00
	MED	0.27	0.75	0.86	0.00	MED-P	**0.42**	**0.78**	0.86	**0.08**
	SC	0.39	0.82	0.93	**0.15**	SC-P	**0.49**	**0.85**	**0.94**	0.00

Table 2: Performance on indefinite similarities (OOS, 80/20). Left: Results without polar correction (KM, MED, SC). For original SC, negative values are flipped. Right: Results of *polarized* kernel variants (KM-P, MED-P, SC-P). ARI - Adj. Rand Index, NMI -Normalized Mutual Information, ACC - clustering accuracy. DUN - Dunn index,# clusters = # labels. Best results (highest values) per dataset across methods in bold.

.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

approaches. This advantage is particularly pronounced when the retained indefinite information carries meaningful structure. Conversely, when the indefinite component is disregarded or handled through heuristic approximations, performance metrics deteriorate significantly. In particular, negative distances may emerge in kernel k-means, and theoretical convergence guarantees can no longer be ensured.

5 Conclusion

We analyzed *polar decomposition* as a principled transformation for indefinite *unsupervised* kernel learning. Although indefinite kernels can outperform standard psd ones, they violate key assumptions of kernel-based methods. Our approach derives a psd component from an indefinite similarity while preserving the original Krein-space geometry and providing an explicit mapping to the associated Hilbert space, enabling efficient out-of-sample evaluation. Empirically, the method retains intrinsic structure and consistently yields competitive or superior clustering performance. These results position polar decomposition as a robust, theoretically grounded framework for indefinite kernel learning and motivate future work on scalable, memory-efficient variants [1].

References

[1] A. Gisbrecht and F.-M. Schleif. Metric and non-metric proximity transformations at linear costs. *Neurocomputing*, 167:643–657, 2015.

[2] S. Heilig, M. Münch, and F.-M. Schleif. Memory efficient kernel approximation for non-stationary and indefinite kernels. In *IJCNN'22*, pages 1–8. IEEE, 2022.

[3] G. Loosli, S. Canu, and C. Ong. Learning SVM in kreĭn spaces. *IEEE TPAMI*, 38(6):1204–1216, 2016.

[4] S. Mehrkanoon, X. Huang, and J. Suykens. Indefinite kernel spectral learning. *Pattern Recognit.*, 78:144–153, 2018.

[5] M. Münch, M. Röder, and F.-M. Schleif. Unlocking the potential of non-psd kernel matrices: A polar decomposition-based transformation for improved prediction models. In Ingo Frommholz, editor, *Proc. of the 32nd ACM CIKM'23*, pages 1867–1876. ACM, 2023.

[6] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In Thomas G. Dietterich, editor, *NIPS 2001*, pages 849–856. MIT Press, 2001.

[7] C. Ong, X. Mary, S. Canu, and A. Smola. Learning with non-positive kernels. In Carla E. Brodley, editor, *(ICML'04)*. ACM, 2004.

[8] H.-S. Park and C.-H. Jun. A simple and fast algorithm for k-medoids clustering. *Expert Syst. Appl.*, 36(2):3336–3341, 2009.

[9] F.-M. Schleif and P. Tiňo. Indefinite proximity learning: A review. *Neural Comput.*, 27(10):2039–2096, 2015.

[10] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[11] P. Väth, M. Münch, C. Raab, and F.-M. Schleif. Proval: A framework for comparison of protein sequence embeddings. *JCMDS*, 3:100044, 2022.

[12] J. Wang. Mathematical analysis on out-of-sample extensions. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(05):1850042, 2018.

[13] R. Yin, Y. Liu, W. Wang, and D. Meng. Randomized sketches for clustering: Fast and optimal kernel k-means. In Sanmi Koyejo, editor, *NeurIPS 2022*, 2022.