Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spectral Moment Embedding: A canonical generalisation of Characteristic Response Analysis for covariance-based signal classification

Jeet Bandhu Lahiri

d23146@students.iitmandi.ac.in

Indian Institute of Technology Mandi

Siddharth Panwar

Indian Institute of Technology Mandi

Article

Keywords:

Posted Date: May 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9286518/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spectral Moment Embedding: A canonical generalisation of Characteristic Response Analysis for covariance-based signal classification

Jeet Bandhu Lahiri[1,*] **and Siddharth Panwar**[1]

[1]School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, Mandi, 175005, India
[*]d23146@students.iitmandi.ac.in (Corresponding author)

ABSTRACT

Covariance matrices underpin multivariate signal classifiers in brain-computer interfaces, clinical neurophysiology, and condition monitoring. Characteristic Response Analysis (CRA) and its Positive-Orthant Characteristic Response (POCR) vector extract a sign-invariant descriptor from a covariance matrix via spherical coordinates of a weighted eigenvector sum. Despite practical utility, these representations are provably non-injective and carry implicit approximations that have not been formally characterised. Here we introduce Spectral Moment Embedding (SME), which subsumes and generalises CRA/POCR through the theory of group-invariant polynomial maps. We prove that any feature invariant to the sign-flip symmetry group on covariance eigenvectors must factor through the matrix moment sequence, that the upper-triangle projection is generically injective, and that POCR is a structurally distorted, non-injective approximation to the first-moment diagonal. Lipschitz stability and a Stone-Weierstrass density result are established. Across five publicly available datasets spanning scalp EEG, polysomnography, electrocardiography, rotating machinery vibration, and inertial measurement, SME outperforms POCR/CRA with accuracy improvements of up to 30.6 percentage points. A principled negative result is also reported: static epoch-level SME features are insufficient for motor imagery BCI decoding, where temporal covariance dynamics dominate, and SME is significantly outperformed by all baselines. Feature importance and kurtosis analyses confirm moment-order interpretations predicted by theory.

Introduction

Covariance matrices encode the second-order statistical structure of multivariate signals and are central to classification pipelines in brain-computer interface (BCI) research, clinical neurophysiology, and industrial condition monitoring. Their geometric structure has been exploited through Riemannian approaches that embed the positive semi-definite (PSD) matrix cone $\mathscr{S}_+(N)$ into a tangent space via the matrix logarithm[1,2], and through eigendecomposition-based methods that extract compact feature vectors from the spectral factorisation $\boldsymbol{C} = \sum_i \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$.

Panwar et al. introduced CRA and the POCR vector for EEG classification[3]. CRA characterises the distribution of the last spherical coordinate θ_{N-1} of a time-series of covariance matrices via its Bhattacharyya distance from a reference Gaussian. POCR provides a per-epoch descriptor via the map $\boldsymbol{u} = \sum_i \lambda_i |\boldsymbol{v}_i|$, converted to $N-1$ spherical coordinates. The absolute value is required because eigenvectors are determined only up to a sign flip.

Despite practical utility, POCR carries three structural limitations. First, it is provably non-injective: distinct covariance matrices can map to the same POCR vector. Second, the positive-orthant projection discards sign patterns distinguishing synchronised from anti-synchronised channel pairs. Third, spherical coordinates carry pole singularities that do not reflect the intrinsic geometry of $\mathscr{S}_+(N)$.

The natural symmetry group acting on eigenvectors is $\mathscr{G}_N = \{-1,+1\}^N$: for any $\boldsymbol{s} \in \mathscr{G}_N$, the map $\boldsymbol{v}_i \mapsto s_i \boldsymbol{v}_i$ leaves $\boldsymbol{P}_i = \boldsymbol{v}_i \boldsymbol{v}_i^\top$ and $\boldsymbol{C} = \sum_i \lambda_i \boldsymbol{P}_i$ unchanged. The theory of invariant polynomial maps shows that the canonical $\mathscr{G}_N$-invariants at order k are the entries of the k-th matrix moment $\boldsymbol{C}^k = \sum_i \lambda_i^k \boldsymbol{P}_i$, computable without eigendecomposition by repeated matrix multiplication. We formalise this as Spectral Moment Embedding (SME):

$$\mathrm{SME}_{P,p}(\boldsymbol{C}) = \left[P(\boldsymbol{C}^1),\, P(\boldsymbol{C}^2),\, \ldots,\, P(\boldsymbol{C}^p)\right], \tag{1}$$

where P is a linear projection (diagonal entries, upper triangle, or row sums) and p is the moment order. We prove five theorems situating SME as the canonical generalisation of CRA/POCR, validate the framework across five heterogeneous public datasets, and report a principled negative result motivating trajectory-based extensions. Code is available at https://github.com/jeetblahiri/sme.

Results

Theoretical framework

Setting. Let $\boldsymbol{X} \in \mathbb{R}^{T\times N}$, let $\hat{\boldsymbol{C}}_0 = \boldsymbol{X}_c^\top \boldsymbol{X}_c/(T-1)$ be the unregularised sample covariance with $\boldsymbol{X}_c$ mean-centred, and let $\boldsymbol{C} = \hat{\boldsymbol{C}}_0 + \varepsilon \frac{\mathrm{tr}(\hat{\boldsymbol{C}}_0)}{N}\boldsymbol{I}$, $\varepsilon = 10^{-6}$. Fix $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ with $\boldsymbol{V}$ orthogonal, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\dots,\lambda_N)$, $\lambda_i \geq 0$. Write v_{ij} for the j-th component of $\boldsymbol{v}_i$ and $\boldsymbol{P}_i = \boldsymbol{v}_i\boldsymbol{v}_i^\top$ for the rank-1 spectral projector.

Theorem 1 (Factorisation through spectral projectors)**.** *Any polynomial $\varphi : \mathscr{S}_+(N) \to \mathbb{R}$ that is well-defined on $\boldsymbol{C}$ (independent of eigendecomposition choice) is a polynomial in the entries of $\{\boldsymbol{C}^k\}_{k\geq 1}$.*

Proof sketch. $\mathscr{G}_N$-invariance forces each monomial in $\{v_{ij}\}$ to have even total degree in each row index i, so the $\mathscr{G}_N$-invariant ring is generated by quadratic factors $v_{ij}v_{ij'} = P_i^{jj'}$. Permutation-invariance then requires φ to be symmetric in the atoms $\{(\lambda_i, \boldsymbol{P}_i)\}$. A multivariate power-sum induction expands products $\prod_r (\boldsymbol{C}^{k_r})_{j_r j'_r}$ into injective and collision parts, expressing every such symmetric polynomial in $\{(\boldsymbol{C}^k)_{jj'} : k \geq 1\}$. Full proof in Supplementary Note S1. □

Theorem 2 (Generic injectivity)**.** *The map $\Phi : \boldsymbol{C} \mapsto [\boldsymbol{I}, \boldsymbol{C}, \boldsymbol{C}^2, \dots, \boldsymbol{C}^{N-1}]$ is injective on the Zariski-open dense set $\mathscr{U}$ of matrices with strictly distinct eigenvalues.*

Proof sketch. The power sums $p_k = \mathrm{tr}(\boldsymbol{C}^k)$ determine $\{\lambda_i\}$ via Newton's recurrences; for $\boldsymbol{C} \in \mathscr{U}$ the characteristic polynomial has N distinct roots. Each $P_i^{jj'}$ is recovered from the $N \times N$ Vandermonde system $\sum_i \lambda_i^k P_i^{jj'} = (\boldsymbol{C}^k)_{jj'}$, $k = 0,\dots,N-1$, which is invertible since $\det \boldsymbol{W} = \prod_{i<i'}(\lambda_{i'} - \lambda_i) \neq 0$. (The $k = 0$ row uses $\sum_i \boldsymbol{P}_i = \boldsymbol{I}$, requiring no extra computation.) Full proof in Supplementary Note S2. □

Theorem 3 (POCR as a distorted upper bound on SME)**.** *With* $\mathrm{POCR}_j(\boldsymbol{C}) = \sum_i \lambda_i |v_{ij}|$ *and* $\mathrm{SME}_j^{(1)}(\boldsymbol{C}) = (\boldsymbol{C})_{jj} = \sum_i \lambda_i v_{ij}^2$*:*

$$\mathrm{POCR}_j(\boldsymbol{C})^2 \;\leq\; \mathrm{tr}(\boldsymbol{C}) \cdot \mathrm{SME}_j^{(1)}(\boldsymbol{C}), \tag{2}$$

with equality iff $|v_{ij}|$ is constant across all i. POCR is non-injective: there exist distinct $\boldsymbol{C}, \boldsymbol{C}' \in \mathscr{S}_+(N)$ with $\mathrm{POCR}(\boldsymbol{C}) = \mathrm{POCR}(\boldsymbol{C}')$*, while* $\mathrm{SME}_{\mathrm{vech},1}(\boldsymbol{C}) = \mathrm{vech}(\boldsymbol{C}) \neq \mathrm{vech}(\boldsymbol{C}')$.

Proof sketch. Cauchy-Schwarz with $\boldsymbol{\alpha} = (\sqrt{\lambda_i})$ and $\boldsymbol{\beta} = (\sqrt{\lambda_i}|v_{ij}|)$ gives equation (2) directly. Non-injectivity follows from an explicit 2×2 construction: orthonormal frames $\boldsymbol{v}_1 = (c,s)$, $\boldsymbol{v}_2 = (-s,c)$ versus $\boldsymbol{v}'_1 = (c,-s)$, $\boldsymbol{v}'_2 = (s,c)$ for $c,s > 0$, $c \neq s$, share $|v_{ij}| = |v'_{ij}|$ for all i,j, giving identical POCR, yet produce opposite off-diagonal signs $\pm(\lambda_1 - \lambda_2)cs \neq 0$ so $\boldsymbol{C} \neq \boldsymbol{C}'$. Full proof in Supplementary Note S3. □

Theorem 4 (Lipschitz stability)**.** *For bounded $\mathscr{B} \subset \mathscr{S}_+(N)$ with $M = \sup_{\mathscr{B}} \|\boldsymbol{C}\|_{\mathrm{op}}$ and projection P with norm $\|P\|$:*

$$\|\mathrm{SME}_{P,p}(\boldsymbol{C}) - \mathrm{SME}_{P,p}(\boldsymbol{C}')\| \;\leq\; \|P\| \left(\textstyle\sum_{k=1}^{p} k M^{k-1}\right) \|\boldsymbol{C} - \boldsymbol{C}'\|_{\mathrm{op}}. \tag{3}$$

Proof sketch. The matrix telescoping identity $\boldsymbol{A}^k - \boldsymbol{B}^k = \sum_{r=0}^{k-1} \boldsymbol{A}^r(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{B}^{k-1-r}$ and sub-multiplicativity yield $\|\boldsymbol{C}^k - \boldsymbol{C}'^k\| \leq kM^{k-1}\|\boldsymbol{C} - \boldsymbol{C}'\|$; equation (3) follows by linearity of P and the triangle inequality. Full proof in Supplementary Note S4. □

Theorem 5 (Approximation density)**.** *The algebra $\mathscr{A}$ of polynomials in $\{(\boldsymbol{C}^k)_{jj'}\}_{k\geq 1}$, viewed as functions on compact $K \subset \mathscr{S}_+(N)$, is uniformly dense in $C(K)$.*

Proof sketch. Stone-Weierstrass applies: $\mathscr{A}$ contains constants, is closed under products, and separates points because for any $\boldsymbol{C} \neq \boldsymbol{C}' \in K$ the function $f : \boldsymbol{C} \mapsto (\boldsymbol{C}^1)_{jj'} = C_{jj'} \in \mathscr{A}$ witnesses separation without any genericity assumption. Full proof in Supplementary Note S5. □

Remark 1. *Theorem 1 gives algebraic completeness. Theorem 2 gives information-theoretic completeness on a generic set. Theorem 3 situates POCR as a specific, provably lossy first-order instance. Theorem 4 guarantees stability under finite-sample covariance estimation. Theorem 5 gives universal approximation over any compact set of covariance matrices.*

Walk-counting interpretation. The j-th diagonal entry of $\boldsymbol{C}^k$ expands as:

$$(\boldsymbol{C}^k)_{jj} = \sum_{i_1,\dots,i_{k-1}=1}^{N} C_{ji_1}C_{i_1i_2}\cdots C_{i_{k-1}j}, \tag{4}$$

the total weight of all k-step closed walks originating at channel j in the covariance graph. At $k = 1$: per-channel variance. At $k = 2$: total squared inter-channel coupling $\sum_{j'} C_{jj'}^2$. At $k = 3$: covariance triangle products closing at j.

Connection to random matrix theory. Under the Marchenko-Pastur (MP) null model[4] for i.i.d. Gaussian noise with ratio $\gamma = N/T < 1$:

$$\frac{\mathrm{tr}(\boldsymbol{C}^k)}{N} \to m_k(\gamma) = \frac{1}{k}\sum_{j=0}^{k-1}\binom{k}{j}\binom{k}{j+1}\gamma^j. \tag{5}$$

Verification at $\gamma = 0.1$ yields $m_1 = 1.000$, $m_2 = 1.100$, $m_3 = 1.310$, matching free-cumulant calculations[5]. The MP deviation $\Delta_k = \mathrm{tr}(\boldsymbol{C}^k)/N - m_k(\gamma)$ vanishes for pure noise and departs for structured signals, providing a theoretically grounded null model.

The SME algorithm

Three-step pipeline. Given $\boldsymbol{X} \in \mathbb{R}^{T\times N}$, order p, and projection P:

Step 1 (covariance). $\boldsymbol{C} = \boldsymbol{X}_c^\top \boldsymbol{X}_c/(T-1) + \varepsilon \frac{\mathrm{tr}(\hat{\boldsymbol{C}}_0)}{N}\boldsymbol{I}_N$, $\varepsilon = 10^{-6}$. The regularisation is scale-invariant; no eigendecomposition is performed at any stage.

Step 2 (moments). $\boldsymbol{C}^1 = \boldsymbol{C}$, $\boldsymbol{C}^k = \boldsymbol{C}^{k-1}\cdot\boldsymbol{C}$ for $k = 2,\ldots,p$. Cost: $O(N^3 p)$.

Step 3 (project and concatenate). $\boldsymbol{f} = \mathrm{SME}_{P,p}(\boldsymbol{C}) = [P(\boldsymbol{C}^1) \mid \cdots \mid P(\boldsymbol{C}^p)] \in \mathbb{R}^{d_P p}$.

Canonical projections. *Diagonal* ($d_P = N$): $P_{\mathrm{diag}}(\boldsymbol{A}) = (A_{11},\ldots,A_{NN})^\top$. Entry $(\boldsymbol{C}^k)_{jj}$ counts closed k-walks at channel j via equation (4).

Vech ($d_P = N(N+1)/2$): $P_{\mathrm{vech}}(\boldsymbol{A}) = \mathrm{vech}(\boldsymbol{A})$, lossless for symmetric $\boldsymbol{C}^k$. At $p = 1$ this is the full covariance and is the unique injective $p = 1$ representation on $\mathscr{U}$ (Theorem 2).

Row-sum ($d_P = N$): $P_{\mathrm{row}}(\boldsymbol{A}) = \boldsymbol{A}\mathbf{1}_N$, encoding total inter-channel coupling; combined with diagonal in E5.

Feature vector lengths: Np (diagonal or row-sum); $N(N+1)p/2$ (vech); $2Np$ (combined). For E3 ($N = 2$, $p = 1$): vech gives 3 features. For E1 ($N = 22$, $p \le 6$): diagonal gives at most 132 features.

Trajectory variant. For W consecutive windows with SME vectors $\boldsymbol{f}^{(1)},\ldots,\boldsymbol{f}^{(W)}$, the trajectory feature stacks five statistics per component:

$$\boldsymbol{f}_{\mathrm{traj}} = [\mu(\boldsymbol{f}),\ \sigma^2(\boldsymbol{f}),\ \gamma_1(\boldsymbol{f}),\ \kappa(\boldsymbol{f}),\ \mathrm{BD}(\boldsymbol{f})] \in \mathbb{R}^{5d_P p}, \tag{6}$$

where μ, σ^2, γ_1, κ denote mean, variance, skewness, excess kurtosis, and BD is the per-component Bhattacharyya distance from a fitted Gaussian. Used in E4 with $W = 10$.

SME subsumes CRA/POCR. From Theorem 3, $\mathrm{POCR}_j \le \sqrt{\mathrm{tr}(\boldsymbol{C})}\cdot\sqrt{\mathrm{SME}_j^{(1)}}$: POCR is a square-root-distorted upper bound on the first-moment diagonal, followed by spherical coordinate conversion. It discards the $N^2 - N$ off-diagonal entries of $\boldsymbol{C}^1$ and never accesses $\boldsymbol{C}^k$ for $k \ge 2$. SME subsumes CRA/POCR at the $p = 1$ diagonal special case, resolves its non-injectivity via the vech projection, and extends to arbitrary p.

Experimental validation

Table 1 summarises five experiments. All use subject-level cross-validation (CV); hyperparameters are selected on training folds only. SME, POCR, and Log-Euclidean features are computed per epoch without fitting, giving unbiased CV estimates. The CSP baseline uses proper nested-CV (spatial filters fitted on training folds only; see Methods). Significance is assessed by the two-sided Wilcoxon signed-rank test at $\alpha = 0.05$ on fold-level accuracy, with 95% bootstrap confidence intervals (2,000 resamples).

Table 1. Experimental datasets. N: channels; F_s: sampling rate; SKF: stratified k-fold; LOSO: leave-one-subject-out. All datasets are freely available without registration.

Exp	Dataset	N	F_s (Hz)	Task (classes)	CV
E1	EEGMMIDB[7]	22	160	Motor imagery (4)	SKF-10
E2	Sleep-EDF-X[8]	4	100	Sleep staging (5)	SKF-10
E3	MIT-BIH[9]	2	360	Arrhythmia (5)	SKF-5
E4	CWRU Bearing[11]	2	12000	Fault type (4)	SKF-10
E5	UCI HAR v2.1[12]	6	50	Activity (6)	LOSO-30

E3 and E4: the two-channel case as a controlled injectivity test

With $N = 2$, the covariance has exactly three free entries (C_{11}, C_{12}, C_{22}). POCR yields one angular coordinate plus the BD scalar. SME-vech at $p = 1$ retains all three entries and is trivially injective, providing the most direct empirical test of Theorem 3.

E3 (ECG arrhythmia, MIT-BIH). We extracted 59,046 balanced beats across five AAMI classes from 48 records using 330×2 segments. SME-vech at $p = 1$ achieves 70.1% against POCR at 39.5% (+30.6 percentage points), quantifying the information discarded by POCR's single angular coordinate when $N = 2$ (Table 2, Figure 1). The lossless vech at $p = 1$ (70.1%) outperforms the diagonal (61.5%), confirming that C_{12} carries discriminative information across the two ECG leads. The SME-diagonal ablation over $p = 1, \ldots, 8$ decreases monotonically from 61.5% to 51.1%: with $\gamma = N/T \approx 0.006$, higher moment orders amplify finite-sample noise. The Wilcoxon $p = 0.062$ reflects only five CV folds, not a marginal effect size. Per-class F1 confirms SME-diagonal outperforms POCR on all five AAMI classes, with the largest gains for ventricular ectopic (V: +0.262), fusion (F: +0.244), and pacemaker beats (Q: +0.289).

Table 2. E3 results (5-fold patient-stratified CV). The 30.6 percentage-point gap between SME-vech and POCR quantifies the information discarded by POCR's single angular coordinate when $N = 2$. Values: mean ± s.d. across folds.

Method	Accuracy (%)	Macro-F1
Wavelet energy (db4, 4 levels)	82.4 ± 0.6	0.699 ± 0.008
SME-vech $p = 1$ (lossless)	70.1 ± 0.1	0.577 ± 0.002
SME-diagonal $p = 1$	61.5 ± 0.4	0.465 ± 0.003
RR-interval (HRV, 5 features)	48.5 ± 1.2	0.401 ± 0.006
POCR/CRA	39.5 ± 1.5	0.286 ± 0.012

E4 (Bearing fault diagnosis, CWRU). We segmented 12 kHz vibration into 1,024-sample windows (50% overlap), yielding 11,352 balanced segments. SME-diagonal at $p = 2$ achieves 93.0% (single segment) and 99.4% (10-segment trajectory; Wilcoxon $p = 0.002$), against POCR at 62.4% (+37.0 pp). Standard statistical features achieve 99.5%, indicating covariance moments and classical time-domain statistics carry overlapping information in this high-SNR dataset. Table 3 shows the most theoretically revealing result: trajectory excess kurtosis of SME-diagonal grows from 7.6 to 177.9 across $p = 1$ to $p = 6$ for Ball faults, against 0.4 to 13.2 for Normal bearings. This is equation (4) working as predicted: higher-order covariance walks raise bearing impacts to successively higher powers, amplifying non-Gaussian character monotonically.

Table 3. Trajectory excess kurtosis of SME-diagonal components by fault class and moment order p (CWRU, $N = 2$, $T = 1024$, 10-segment super-windows). Growth for Ball faults directly confirms the walk-counting interpretation of equation (4). Values: means over at least 200 sampled windows per class.

Fault class	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
Normal	0.36	2.02	4.04	6.44	9.44	13.22
Inner Race	−1.05	0.04	6.47	17.67	24.75	31.33
Ball	7.57	23.33	36.60	57.39	149.87	177.92
Outer Race	−1.34	0.14	3.88	10.48	21.34	37.19

E5: feature importance confirms the walk-counting interpretation

Across 10,411 windows from 30 subjects, SME-combined ($p = 5$, diagonal and row-sum) achieves 80.6% LOSO accuracy against POCR at 63.0% (+17.6 pp; Wilcoxon $p < 0.001$). Per-channel statistics (91.6%) outperform SME because they include higher-order marginals beyond second-order covariance. The GBT feature importance decomposition (Figure 2) confirms equation (4): $p = 1$ features account for 66.4% of total importance and $p = 2$ for 23.6%, with the first two orders capturing 90.0%. Gravity-projected variance ($p = 1$) primarily discriminates static postures; inter-axis coupling ($p = 2$) is most informative for walking variants where accelerations across axes are phase-shifted. This validates the physical interpretation without supervised feature selection.

E2: multimodal covariance structure in sleep staging

With four channels (EEG Fpz-Cz, EEG Pz-Oz, EOG, EMG), SME at $p = 3$ achieves 63.5% and macro-F1 0.550, outperforming POCR (52.3%; Wilcoxon $p = 0.004$). Spectral power features (69.3%) outperform SME overall because sleep stages are clinically defined by frequency-domain criteria. However, SME uniquely outperforms spectral features on N3 slow-wave sleep

(F1: 0.806 vs 0.779; Table 4, Figure 3), where synchronised high-amplitude deflections generate cross-channel covariance structure at orders $p = 2, 3$ not captured by marginal spectral power per channel. REM detection is poor for all covariance-based methods (F1 < 0.15) because REM EEG is second-order indistinguishable from Wake EEG at 30-second resolution.

Table 4. E2 per-class F1 (10-fold subject-stratified CV, 5-subject subset of Sleep-EDF Expanded). SME outperforms POCR on all five stages and uniquely outperforms spectral power on N3. REM poor performance reflects second-order indistinguishability from Wake. Values: macro-F1 across pooled held-out folds.

Sleep stage	SME $p = 3$	POCR/CRA	Spectral power
Wake	0.648	0.513	0.760
N1	0.578	0.481	0.632
N2	0.607	0.534	0.662
N3	0.806	0.653	0.779
REM	0.138	0.140	0.541
Macro avg	0.550	0.456	0.663

E1: a principled negative result

On motor imagery EEG (50 subjects, 4,500 trials), static epoch-level SME-diagonal (29.4%) places last among all four methods (Table 5). With proper nested-CV CSP (spatial filters fitted on training folds only), the unbiased baselines are: CSP $34.1\% \pm 5.6\%$; POCR 33.4%; Log-Euclidean 36.0%. SME is significantly worse than all three (Wilcoxon $p = 0.027$ vs CSP, $p = 0.002$ vs POCR and Log-Euclidean). Notably, POCR and CSP are not significantly different ($p = 0.695$), nor are Log-Euclidean and CSP ($p = 0.391$), so the four methods form two statistically distinguishable groups: SME-diagonal alone at 29.4%, and POCR, CSP, Log-Euclidean clustered between 33–36%. Motor imagery decoding depends on the ERD/ERS pattern evolving within each trial[13]; a single epoch-level covariance averages over these dynamics and loses them. Accuracy decreases monotonically with p from 29.4% ($p = 1$) to 27.0% ($p = 6$; inset of Figure 1), indicating that higher moment orders add genuine but non-class-discriminative covariance structure.

Table 5. Grand summary across all five experiments. Δ: best SME accuracy minus POCR/CRA on identical CV splits. Two-sided Wilcoxon signed-rank test on fold-level accuracy pairs. In E1, SME is significantly worse than all three baselines. In E3, $p = 0.062$ reflects only five CV folds, not a marginal effect size.

Exp	Best SME method	SME (%)	POCR (%)	Δ (%)	p
E1	SME-diagonal $p = 1$	29.4	33.4	−4.0	0.002
E2	SME $p = 3$	63.5	52.3	+11.2	0.004
E3	SME-vech $p = 1$	70.1	39.5	+30.6	0.062
E4	SME-traj $p = 2$	99.4	62.4	+37.0	0.002
E5	SME-comb $p = 5$	80.6	63.0	+17.6	<0.001

Discussion

The five experiments form a coherent narrative, with each outcome interpretable from the theory rather than post-hoc. The most informative single result is in E3: with $N = 2$, POCR retains one scalar from a three-entry covariance matrix, and the 30.6 percentage-point advantage of SME-vech at $p = 1$ is a near-direct measurement of that information loss. Theorem 3 predicts this precisely, as POCR_j is bounded above by $\sqrt{\text{tr}(\boldsymbol{C}) \cdot (\boldsymbol{C})_{jj}}$, a square-root distortion of the diagonal SME. Wavelet features (82.4%) outperform SME because they operate on the full 330-sample beat waveform, a richer representation than the 2×2 covariance.

The kurtosis results in E4 provide mechanistic validation beyond accuracy. Ball fault kurtosis reaches 177.9 at $p = 6$, over ten times the Normal bearing value: equation (4) raises impulsive bearing impacts to successively higher powers, amplifying non-Gaussian character monotonically. That SME-trajectory (99.4%) and standard statistical features (99.5%) perform indistinguishably shows that covariance moments and classical time-domain statistics are informationally equivalent in the clean mechanical signal regime. The E5 feature importance result validates the walk-counting interpretation without supervised feature selection: the 90% cumulative importance captured by $p = 1$ and $p = 2$ follows from equation (4) and the physics of inertial measurement under gravity.

The E1 negative result is fully characterised with unbiased baselines. The four methods form two statistically distinguishable groups: SME-diagonal alone at 29.4%, and POCR (33.4%), CSP (34.1%), Log-Euclidean (36.0%) clustered together and mutually non-significant. The monotone accuracy decrease with p indicates that higher moment orders add genuine but spatially non-discriminative structure for motor imagery; the appropriate remedy is trajectory-based SME or combination with supervised spatial filtering. The E2 result reveals the natural boundary of the approach: second-order time-domain statistics are insufficient for signals whose class identity is primarily encoded in frequency content. The unique N3 result (SME outperforms spectral features at one key boundary) shows that cross-channel covariance adds information where temporal synchrony is the defining physiological characteristic.

Three limitations merit acknowledgement. The theoretical results concern population covariances; finite-sample concentration bounds are important future work. The projection P is fixed a priori; learning it end-to-end would likely improve performance. The MP deviation analysis showed that most ECG and vibration classes produce negative Δ_k because signal amplitudes are smaller than the unit-variance assumption of the standard MP law; signal-energy normalisation is needed for a calibrated null model. In summary, SME provides a principled, computationally efficient ($O(N^2 p)$ per window), and theoretically grounded generalisation of CRA/POCR. The five theorems resolve POCR's structural limitations; the experimental validation across four of five modalities, together with the fully characterised negative result in the fifth, demonstrates that the framework offers more information rather than merely more parameters.

Methods

SME computation

The three-step pipeline is described in the Results section, including equations (1)–(6). Matrix powers are computed by repeated multiplication without eigendecomposition. The specific projection, order p, and window count W for each experiment are given below.

POCR/CRA computation

POCR: $\mathrm{POCR}_j = \sum_i \lambda_i |v_{ij}|$ converted to $N-1$ spherical coordinates $\theta_k = \arccos(\mathrm{POCR}_k / \|\mathrm{POCR}_{k:N}\|)$. CRA: Bhattacharyya distance between the empirical distribution of θ_{N-1} and a Gaussian with matching mean and variance[3].

Marchenko-Pastur moments

$m_k(\gamma)$ is computed via equation (5), derived from the R-transform of free probability[5,6].

Datasets

E1 (EEGMMIDB[7]). 50 of 109 subjects; 22 sensorimotor EEG channels; 8–30 Hz FIR bandpass (MNE default); motor imagery runs R04/R08/R12 (left/right fist) and R06/R10/R14 (both fists/feet); 4-second epochs (640 samples at 160 Hz).

E2 (Sleep-EDF Expanded[8]). 5-subject sleep-cassette subset; channels EEG Fpz-Cz, EEG Pz-Oz, EOG horizontal, EMG submental; 30-second epochs; N3/N4 merged per AASM; Wake downsampled to 5× the rarest non-Wake stage count.

E3 (MIT-BIH Arrhythmia[9,10]). 48 records; 2-channel ECG; 330-sample beat segments; AAMI mapping N{N,L,R,e,j}, S{A,a,J,S}, V{V,E}, F{F}, Q{/,f,Q}; Normal beats downsampled to 5× the largest minority class.

E4 (CWRU Bearing[11]). Drive-end and fan-end 12 kHz vibration, $N = 2$ channels; 1,024-sample windows, 50% overlap; four fault classes balanced by undersampling; trajectory over $W = 10$ consecutive windows using mean, variance, skewness, excess kurtosis, and Bhattacharyya distance per SME-diagonal component.

E5 (UCI HAR v2.1[12]). 30 subjects; 6-channel IMU (3-axis accelerometer and 3-axis gyroscope) at 50 Hz; 128-sample windows (2.56 s), 50% overlap; activities 1–6; combined SME: diagonal and row-sum projections, $2 \times 6 \times p$ features.

Classifiers

E1: LDA (Ledoit-Wolf shrinkage). E2: Random Forest (200 trees, max depth 12, balanced class weights). E3, E4: SVM (RBF kernel, $C = 10$ and $C = 50$, `gamma=scale`, balanced class weights). E5: Gradient Boosted Trees (200 estimators, max depth 5, learning rate 0.1, subsample 0.8). All classifiers use a StandardScaler prefix within each fold. For the E1 CSP (OVR) baseline, a `CSPOVRTransformer` (one MNE `CSP` instance per class, 6 components) is embedded in a sklearn Pipeline with `StandardScaler` and LDA; a `StratifiedGroupKFold` loop calls `pipe.fit(X_train)` inside each fold so spatial filters never see test-fold data.

Statistical analysis

Fold-level accuracy pairs are compared by the two-sided Wilcoxon signed-rank test at $\alpha = 0.05$ (n values: E1 $n = 10$, E2 $n = 10$, E3 $n = 5$, E4 $n = 10$, E5 $n = 30$; non-parametric test appropriate for small fold counts). No multiplicity correction is applied to pairwise comparisons within each experiment; primary conclusions are based on effect sizes. Bootstrap 95% confidence intervals use 2,000 resamples.

Data availability

All five datasets are freely available without registration. EEGMMIDB: https://physionet.org/content/eegmmidb/1.0.0/. Sleep-EDF Expanded: https://physionet.org/content/sleep-edfx/1.0.0/. MIT-BIH Arrhythmia: https://physionet.org/content/mitdb/1.0.0/. CWRU Bearing: https://engineering.case.edu/bearingdatacenter. UCI HAR v2.1: https://archive.ics.uci.edu (dataset ID 341). No new datasets were generated. All experimental code is available at https://github.com/jeetblahiri/sme.

References

1. Barachant, A., Bonnet, S., Congedo, M. & Jutten, C. Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* **59**, 920–928 (2012).
2. Pennec, X., Fillard, P. & Ayache, N. A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66**, 41–66 (2006).
3. Panwar, S., Joshi, S. D., Gupta, A. & Agarwal, P. Automated epilepsy diagnosis using EEG with test set evaluation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 1106–1116 (2019).
4. Marčenko, V. A. & Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb.* **1**, 457–483 (1967).
5. Mingo, J. A. & Speicher, R. *Free Probability and Random Matrices* (Springer, New York, 2017).
6. Voiculescu, D. V., Dykema, K. J. & Nica, A. *Free Random Variables*, CRM Monograph Series Vol. 1 (American Mathematical Society, Providence, RI, 1992).
7. Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N. & Wolpaw, J. R. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **51**, 1034–1043 (2004).
8. Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. C. & Oberye, J. J. L. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **47**, 1185–1194 (2000).
9. Moody, G. B. & Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
10. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
11. Loparo, K. A. Bearing data center. Case School of Engineering, Case Western Reserve University. https://engineering.case.edu/bearingdatacenter (2012).
12. Reyes-Ortiz, J.-L. et al. Human activity and motion disorder recognition: towards smarter interactive cognitive environments. In *Proc. ESANN* (2013).
13. Pfurtscheller, G. & Lopes da Silva, F. H. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **110**, 1842–1857 (1999).

Author contributions statement

J.B.L. conceived the SME framework, developed the theoretical proofs, implemented all experiments, and wrote the manuscript. S.P. supervised the work, contributed the foundational CRA/POCR theory that this work generalises, guided the experimental design, and critically revised the manuscript. Both authors reviewed the manuscript.

Additional information

Funding: No funding was received for conducting this study.

Competing interests: The authors declare no competing interests.

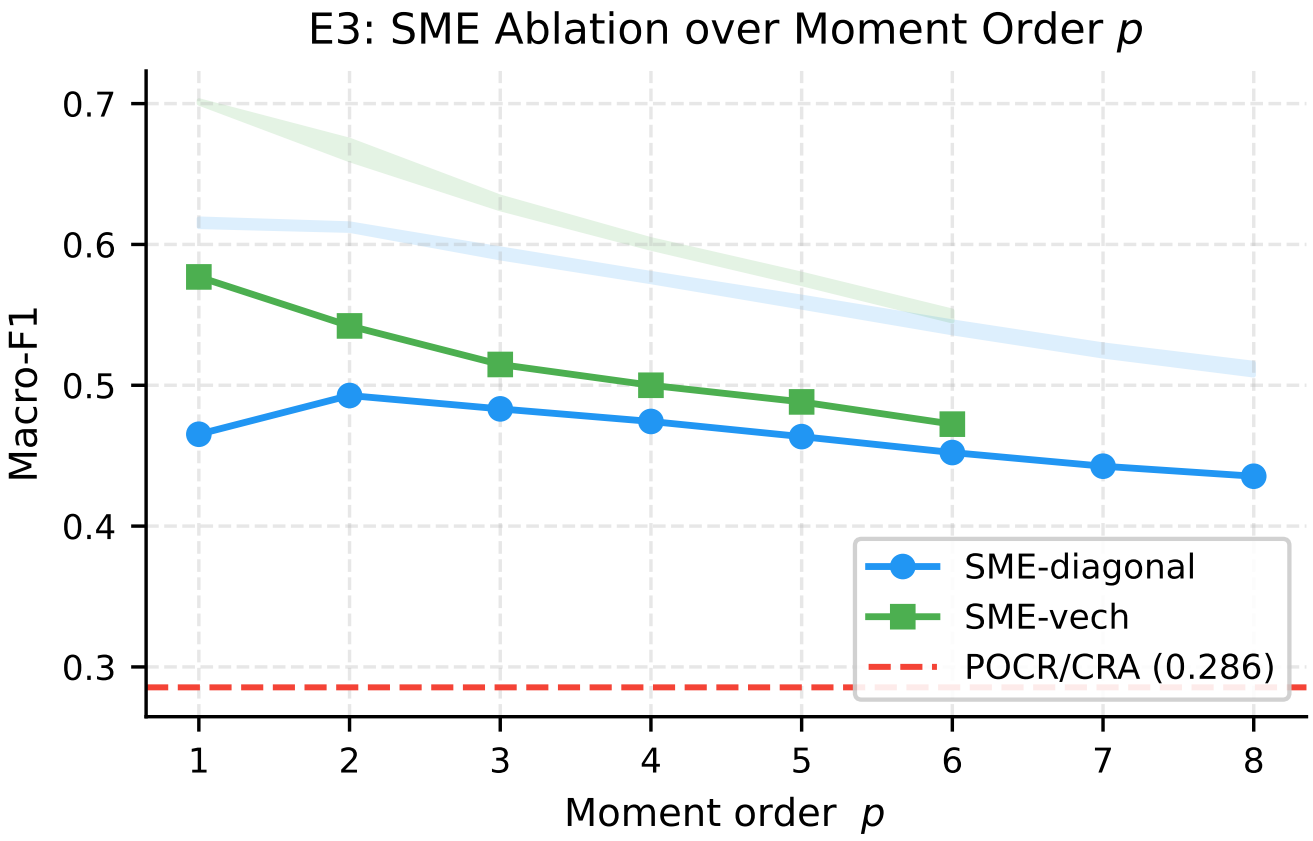


Figure 1. SME moment-order ablation on E3 (MIT-BIH ECG) with E1 (motor imagery EEG) inset. Main panel: Accuracy versus moment order p for SME-diagonal (solid blue) and SME-vech (dashed green) on E3 ($N = 2$ channels, $T = 330$ samples per beat), with shaded 95% bootstrap confidence intervals (2,000 resamples). Red dashed line: POCR/CRA baseline (39.5%). The lossless vech at $p = 1$ achieves 70.1%, a 30.6 percentage-point advantage over POCR, providing direct empirical evidence of Theorem 3: with $N = 2$, POCR discards C_{12} entirely. Both curves decrease monotonically, confirming noise amplification dominates in the thin-matrix regime ($\gamma = 0.006$). **Inset:** E1 (22-channel motor imagery EEG) SME-diagonal ablation. Accuracy decreases monotonically from 29.4% ($p = 1$) to 27.0% ($p = 6$), remaining below POCR (orange, 33.4%), CSP (green, 34.1%), and Log-Euclidean (grey, 36.0%) throughout. Error bars: s.d. across folds.

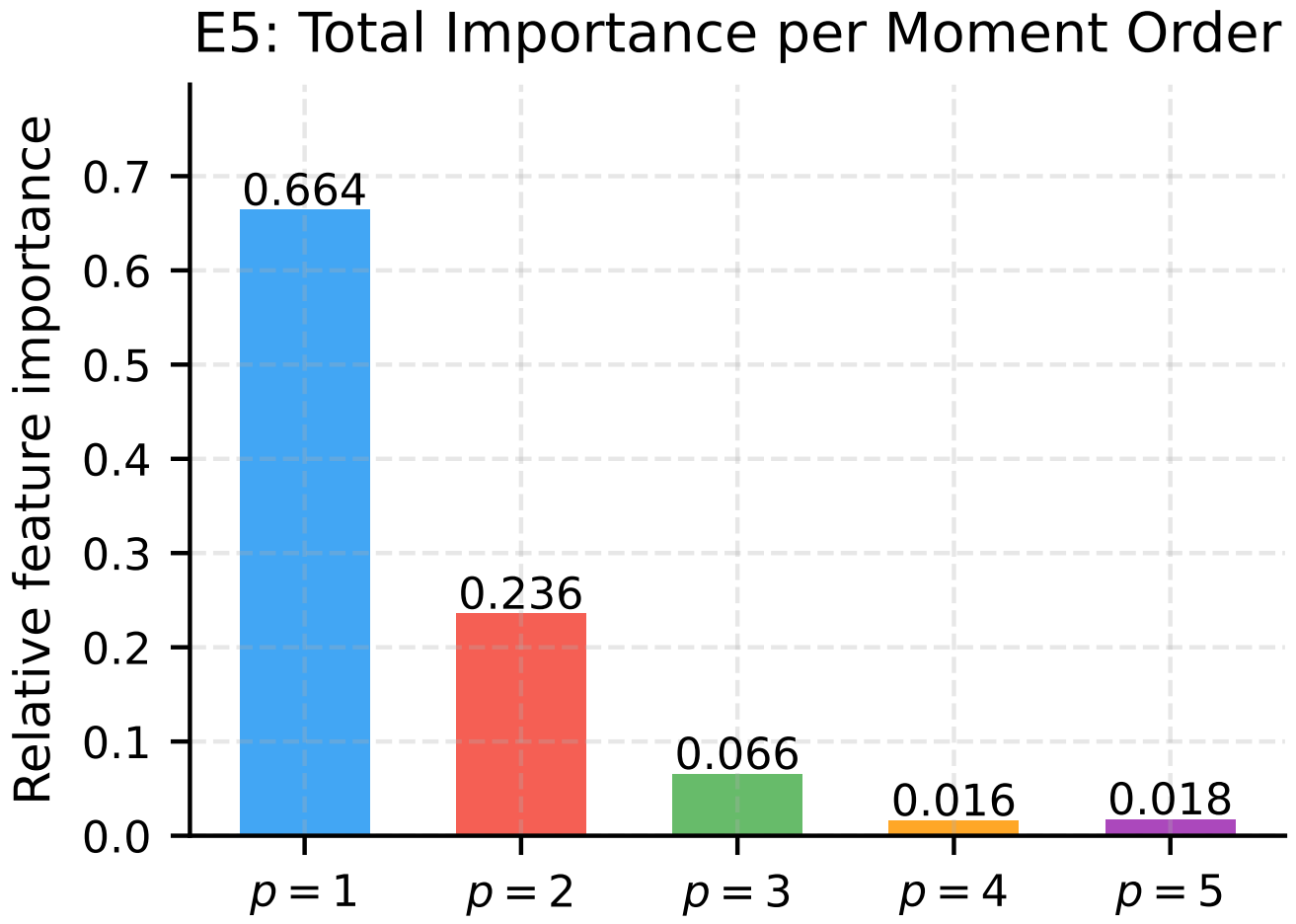


Figure 2. GBT feature importance by moment order (E5, UCI HAR v2.1, 30-subject LOSO CV). Bar heights show the fraction of total Gradient Boosted Tree importance attributed to SME features at each order (diagonal and row-sum combined; 200 trees). The $p = 1$ features (per-channel variance, equation (4)) account for 66.4% of discriminative power. The $p = 2$ features (squared inter-axis coupling $(\boldsymbol{C}^2)_{jj} = \sum_{j'} C_{jj'}^2$) contribute 23.6%. Together the first two orders capture 90.0% of total importance, consistent with the walk-counting interpretation for near-Gaussian inertial signals. This result independently validates the physical interpretation without supervised feature selection.

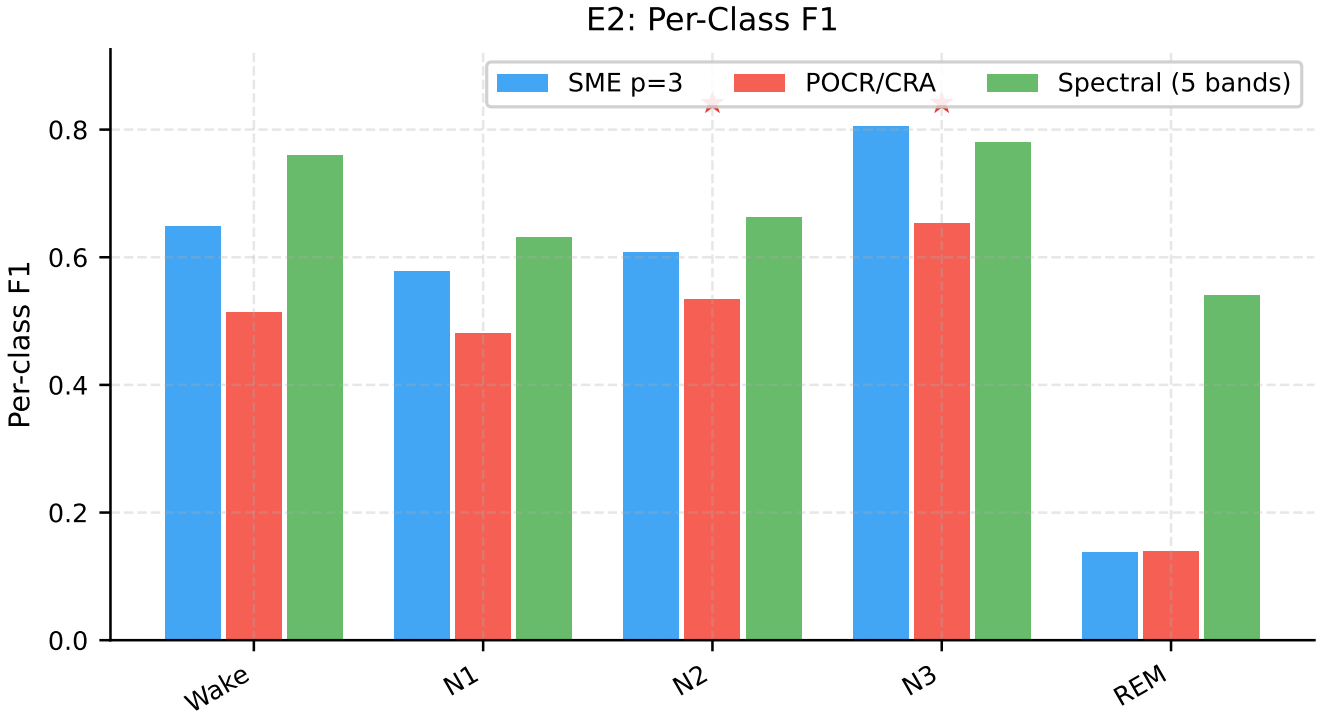


Figure 3. Per-class F1 for sleep staging methods (E2, Sleep-EDF Expanded, 5-subject subset, 10-fold subject-stratified CV). Grouped bars show F1 for SME $p = 3$ (blue), POCR/CRA (red), and spectral power (green). SME outperforms POCR on all five stages and uniquely outperforms spectral power on N3 slow-wave sleep (0.806 vs 0.779), where synchronised high-amplitude deflections generate cross-channel covariance structure at orders $p = 2, 3$ not captured by marginal spectral power per channel. REM detection is poor for all covariance-based methods because REM EEG is second-order indistinguishable from Wake EEG at 30-second epoch resolution. Error bars: inter-fold s.d.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SMESupplementary.pdf