Decomposition for symbolic compactness: complexity-penalised pursuit of parametric mode families

Dongtai Li

College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Abstract

Variational mode decomposition extracts modes by penalising spectral bandwidth, yet its modes need not correspond to a single analytic family, and their quality can depend on centre-frequency initialisation. We argue that for signals whose components are *a priori* describable by low-dimensional parametric families, decomposition should be posed directly as complexity-penalised approximation by a sum of symbolic components, dispensing with frequency-localisation machinery. The resulting pursuit over a library of parametric atoms (sinusoid, damped sinusoid, chirp, exponential, logarithm, low-order polynomial) uses data-driven candidate generation, BIC-governed greedy addition with a noise-floor stop, block-coordinate and joint nonlinear refinement, and compound split moves with co-removal of redundant components. On the five identifiable stationary synthetic cases, 20-seed full-structure recovery rates range from 0.65 to 1.00, with component MSE down to 10^{-6}; the method resolves a 0.6 Hz pair below the Rayleigh limit of a one-second record. In a 20-seed paired study on all eight cases it reduces periodic-component MSE over initialisation-corrected VMD by median factors of 1653 (30 dB), 129 (5 dB) and 49 (0 dB), every stationary case at the minimum attainable significance; on semi-synthetic bearing signals (CWRU with an injected component) it recovers the injected component two orders more accurately than VMD; and on measured rotating-machine records in three operating conditions the recovered frequencies admit stable, nominally constrained shaft-harmonic lattice fits and separate recurring off-lattice lines by the same test. We analyse the penalised objective's frequency bias under unmodelled exponential content.

Keywords: signal decomposition, variational mode decomposition, symbolic regression, model selection, sparse approximation, condition monitoring

1. Introduction

Many signals of engineering interest are superpositions of components with distinct generating mechanisms: sinusoidal oscillations from rotating machinery, exponential decays from relaxation processes, polynomial trends from sensor drift, chirps from accelerating machinery or Doppler sources. For diagnosis and interpretation one wants not merely separated waveforms but *analytically simple* components — each mode expressible by a short closed-form expression with few parameters.

Decomposition methods based on spectral localisation — Empirical Mode Decomposition [1], Variational Mode Decomposition (VMD) [2] and their numerous descendants [3, 4, 5] — answer the question "how should the mixture be separated by frequency?" They do not address whether the returned modes are analytically describable. Two components with identical bandwidth, e.g. $\sin(2\pi ft)$ and $\sin(2\pi ft) + 0.01e^{-t}$, are treated as equivalent although only the first is a single symbolic primitive. Moreover, the result of spectral decompositions depends sensitively on initial centre frequencies: a mode whose centre is initialised far from its spectral content can remain essentially empty while another mode absorbs its content, and the dual update with $\tau = 0$ provides no mechanism to recover from such misallocation. We document this failure concretely in Section 4.2.

Symbolic regression (SR) [6, 7] recovers closed-form expressions from data, but its search space grows combinatorially with the complexity of its input; feeding it a mode that mixes heterogeneous analytical content produces expression bloat or degraded accuracy. Decomposition and symbolic regression have mostly been chained as independent pipeline stages, the decomposition optimising spectral criteria and SR operating on whatever it receives [8]. Neither side of this pipeline solves the problem directly: spectral decompositions control bandwidth but not analytic form, while parametric estimators (Prony, matrix pencil) recover a single homogeneous family with prescribed order, and generic sparse pursuits select fixed vectors rather than closed-form physical families. The gap we address is the unified, complexity-controlled selection of *heterogeneous* analytic families and their component decomposition in one objective.

This paper takes the position that when the components of interest are *a priori* describable by a catalogue of low-dimensional parametric families — a situation common in condition monitoring and physical measurement — the decomposition should be posed directly on the space of such families, with the reconstruction error as the data term and an explicit

Email address: LDongtai@hotmail.com (Dongtai Li)

complexity penalty as the regulariser. We develop a concrete algorithm, *symbolic matching pursuit*, named DTHy-SMP — *downstream-task-guided hybrid symbolic matching pursuit* — and show that this change of formalisation avoids the centre-initialisation and empty-mode failure examined here: the pursuit has no centre frequencies to initialise, and every accepted component belongs to an explicit analytic family. Specifically:

1. We characterise and mitigate two failure modes of greedy penalised decomposition: (a) collectively redundant atoms block structurally favourable moves, which we resolve by *compound* split moves evaluated jointly with co-removal (Section 3.7); and (b) in the presence of an unmodelled exponential component, the penalised objective's minimum shifts away from true sub-Rayleigh sinusoid frequencies, and partial orthogonalisation against a nuisance basis distorts the landscape because tonal content overlaps that basis by a non-negligible amount — we resolve this by jointly fitting the candidate pair with the nuisance basis (Section 3.8).
2. We formulate decomposition as complexity-penalised approximation by a sum of parametric symbolic components (Section 3), estimated by *symbolic matching pursuit* (DTHy-SMP), and show that model order, analytic families and parameters can be recovered from the data by the Bayesian information criterion (BIC) with a noise-floor stop in the tested stationary cases, while mapping family-confusion and budget failures through 20-seed controls and component- budget sensitivity (Section 4.6).
3. We provide an evaluation (Section 4): eight synthetic cases spanning stationary, sub-Rayleigh, transient and non-stationary regimes, with a 20-seed paired statistical study on all of them; a forensic reproduction of the initialisation sensitivity of VMD; semi-synthetic validation on real bearing recordings from CWRU; measured rotating-machine records in healthy, misaligned and unbalanced conditions, where the recovered components form a shaft-harmonic lattice; comparison throughout against recent strong decompositions (successive VMD, empirical Fourier decomposition, nonlinear chirp mode decomposition) in addition to VMD, EMD and matrix pencil; and an ablation of each algorithmic component.

The method is an offline estimator for records whose content admits a parametric description — free-decay and impulse-response measurements (impact modal testing), quasi-stationary harmonic and trend content, and records containing isolated chirp components — and delivers symbolic models for physical interpretation rather than real-time monitoring waveforms. The real-data application focus of this paper is the quasi-stationary regime: separating shaft-harmonic lattices from constant-frequency structural lines on rotating-machine records, as studied in Section 4.8. Signals dominated by trains of delayed impulses, as in faulty rolling-element bearings, are outside the current catalogue; independent analyses caution that such signals are not validly decomposable into mono-components at all [9]. The present evidence therefore concerns analytic compactness in the stated quasi-stationary and isolated-component settings.

Relation to VMD. In the name DTHy-SMP — *downstream-task-guided hybrid symbolic matching pursuit — downstream-task-guided* means that the decomposition is guided by the requirement of the downstream task, symbolic modeling of the components, rather than by spectral bandwidth alone; and *hybrid* means that closed-form parametric atoms are hybridised with the data term inside the estimation loop, instead of being fitted only after a spectral decomposition has frozen the modes. The forensic study of Section 4.2 shows that VMD accuracy can depend strongly on centre initialisation, and the ablation of Section 4.9 evaluates the natural repair-from-inside route — a data-fitted parametric prior injected into VMD's Wiener update, representative of the prior-repair line of work — finding that its apparent gains do not survive correct initialisation of the plain baseline. Correct initialisation repairs that numerical failure but does not impose analytic compactness on the returned modes. DTHy-SMP therefore changes the decomposition variable from spectral modes to symbolic components while retaining the objective of separating a machinery mixture into interpretable parts.

2. Related work

Spectral decomposition. EMD [1] and noise-assisted variants [10, 11] separate by sifting; VMD [2] minimises an explicit bandwidth functional with alternating direction minimisation, with mode number K and centre initialisation as user choices. Successive VMD [12] extracts modes sequentially with data-driven centres; VME [13] targets one mode around a desired centre frequency; generalized VMD [14] fixes the centre frequencies a priori. All of these guide decomposition by frequency information; DTHy-SMP guides it by analytic-form information and dispenses with centre frequencies altogether. Nonlinear chirp mode decomposition [15] handles non-stationary components by variational means, again with spectral penalties. This line remains very active: impulsive mode decomposition [16], adaptive extraction of fault components from mixtures [17], automatically parameterised VMD [18], empirical Fourier decomposition [19] and harmogram-based periodic-impulse detection [20] all refine spectral or impulsive objectives. DTHy-SMP differs from all of these in what is being optimised: not spectral bandwidth, impulsiveness, or sparsity of a time-frequency representation, but the complexity of an explicit symbolic model.

Parametric and sparse decompositions. The classical line of Prony, matrix pencil [21] and ESPRIT estimates sums of damped sinusoids directly, without decomposition; it is a special case of our atom library (damped sinusoid only, fixed model order) and we use it as both a candidate generator and a baseline. Matching pursuit [22] decomposes signals over structured, often parametrically generated dictionaries. Parametric

atomic models based on damped sinusoids [23] and composite feature dictionaries for machinery signals [24, 25] already provide signal-adaptive component descriptions. Most closely, Cui et al. combine Fourier and impulse dictionaries and select multiple atoms per iteration with coefficient thresholding [24]; the later CD-SaMP replaces that step by single-atom matching and an attenuation-based termination rule [25]. Those methods target gear-fault feature extraction from fixed subdictionaries. Here each returned component is itself a fitted analytic family, heterogeneous families and order are selected jointly by a global complexity score, and the difficult move is a penalised replacement of an existing structure rather than simply another dictionary projection. Chirplets [26], adaptive chirplet transforms [27], and continuous-dictionary refinement further optimise nonlinear atom parameters. DTHy-SMP therefore does not claim parametrised atoms as a new idea. Its contribution is the joint selection of heterogeneous analytic families and model order under one complexity score, together with compound split-with-co-removal moves and nuisance-aware pair fitting. Morphological component analysis [28, 29] instead separates components through fixed dictionaries and convex sparsity. Bayesian line-spectral estimation [30] and convex off-the-grid formulations [31, 32] recover sinusoid sets with principled model order control, but assume a single homogeneous family (stationary tones) and do not address heterogeneous analytic content such as decays, trends or chirps. Finite rate of innovation theory [33] analyses sampling and recovery of streams of parametric pulses under known form and essentially noiseless conditions; here the family itself is chosen from data. Stochastic model selection frameworks such as single best replacement [34] explore model spaces by replacement moves; our compound split-with-co-removal move addresses a specific dead-lock of greedy addition that standard replacement does not, as analysed in Section 3.7.

Model selection and sparse recovery. BIC/MDL [35, 36] serves as our complexity gate; related Bayesian sparsity approaches include sparse Bayesian learning [37]. We stress that the BIC is used here as a *complexity-regularised model selection criterion*, not as a procedure with guaranteed selection consistency: the classical BIC theory does not automatically extend to a library of nonlinear parametric families with data-adaptive candidate generation and correlated atoms, and we do not claim otherwise. The evidence behind the score is mechanism-level (Section 3.8) and empirical (Section 4).

3. Method

3.1. Problem formulation

Let $f \in \mathbb{R}^N$ be a uniformly sampled observation and let $\mathcal{L} = \{\text{exp, log, poly}_1, \text{poly}_2, \text{sinusoid, damped sinusoid, chirp}\}$ be a catalogue of parametric families. Each family $\lambda \in \mathcal{L}$ has p_λ real parameters $\theta \in \Theta_\lambda \subseteq \mathbb{R}^{p_\lambda}$ and generates signals $s_\lambda(\theta) \in \mathbb{R}^N$; amplitudes and phases enter linearly in all families. We seek a *component list* $\mathcal{M} = \{(\lambda_r, \theta_r)\}_{r=1}^K$ solving

$$\min_{\mathcal{M}} \left\| f - \textstyle\sum_r s_{\lambda_r}(\theta_r) \right\|_2^2 \text{ subject to complexity control on } \mathcal{M}, \tag{1}$$

where complexity is controlled explicitly through a penalised score

$$\text{score}(\mathcal{M}) = N \log \frac{\text{SSE}(\mathcal{M})}{N} + \Big(\sum_r p_{\lambda_r} \Big) \text{pen}, \qquad \text{pen} = \log N, \tag{2}$$

i.e. the Gaussian BIC. Note what Eq. (1) does *not* contain: no centre frequencies, no bandwidth functionals, no dual variables, no constraint that modes be band-limited. Modes are parametric objects; misallocation-by-initialisation and mode death are undefined by construction. The model order K is not an input: it emerges from when (2) stops improving (Section 3.4).

3.2. Atom families

With normalised time $\tau = (t - t_1)/(t_N - t_1) \in [0, 1]$ and a chirp phase $\phi(t) = 2\pi(f_0 t + \frac{1}{2} f_1 t^2)$:

family	form	p_λ
exp	$a\, e^{-b\tau}$	2
log	$a\, \log(1 + b\tau)$	2
poly_d	$\sum_{j \le d} c_j \tau^j$	d+1
sinusoid	$a \cos 2\pi f t + c \sin 2\pi f t$	3
damped	$e^{-b\tau}(a \cos 2\pi f t + c \sin 2\pi f t)$	4
chirp	$e^{-b\tau}(a \cos \phi + c \sin \phi)$	5

Families are composable by addition at the component level; composite exponential-plus-polynomial shapes, which nested single templates tend to fit with spurious terms, are represented here as *separate* components whose coexistence must be justified by (2) — the complexity penalty is the only source of spurious-term suppression, and Section 4 verifies it works. The chirp family is admitted only with a strong-evidence margin: a constant $\Delta_{\text{chirp}} = 20$ (equivalent to $20/\log(1024) = 2.89$ additional BIC parameters at $N = 1024$) is added to its score in (2). This is *not* part of the BIC; it is an additional empirical penalty guarding against a specific structural confusion — a five-parameter chirp mimicking a close sinusoid pair — of the same nature as the complexity penalty itself. Its sensitivity to record length, SNR and the two difficult catalogue cases is reported in Section 4.9. All bounds Θ_λ are box constraints ($f \in [0.05, 480]$ Hz, $b \in [10^{-3}, 80]$, etc.).

3.3. Candidate generation

All candidates derive from the observation only. For the observed f or the current residual: positive-spectrum peaks (scipy `find_peaks`) seed sinusoid and damped-sinusoid atoms with closed-form linear amplitudes; the low band seeds exponential, logarithmic and polynomial atoms on a decay/degree grid; a matrix-pencil pass of order six [21] adds damped-sinusoid candidates that survive at low SNR. Chirp candidates come from a two-stage procedure we found necessary: a coarse (f_0, f_1) grid anchored at the dominant spectral peak f_{dom} of the signal being fitted ($f_0 \in [0.5, 2]\, f_{\text{dom}}$, $f_1 \in \pm 1.5\, f_{\text{dom}}/T$, 41×25

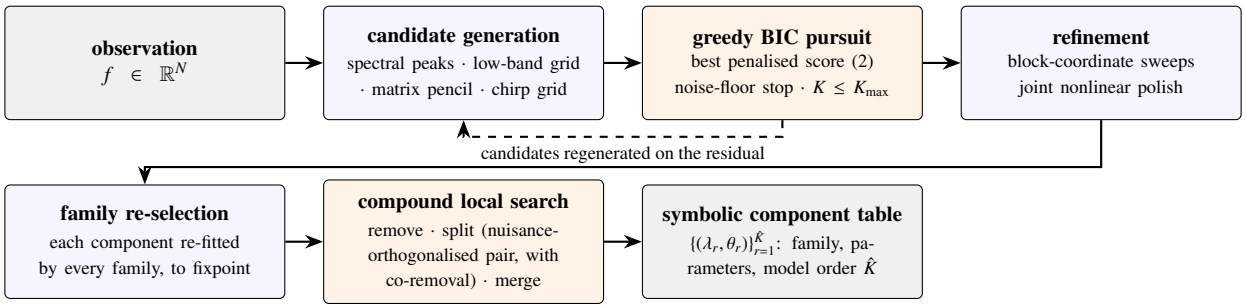


Figure 1: Overview of DTHy-SMP. Decomposition is posed directly as complexity-penalised approximation of the observation by a sum of parametric symbolic atoms (no centre frequencies, no bandwidth functionals). Candidates are generated from the data; a BIC-governed greedy pursuit with a noise-floor stop builds the model; block-coordinate refinement, family re-selection and a compound local-search fixpoint (split evaluated jointly with co-removal of components it renders redundant) correct the greedy solution; the deliverable is a closed-form component table with an estimated model order.

points, T the record length), so that the search scale adapts to the record's own spectral and temporal scale rather than to absolute Hz values. On this grid the damping is held at $b = 0$ (an undamped chirp: the grid is a seeding device, not the final estimate), amplitudes are obtained by a 2×2 linear solve per grid point, and the best three grid points are passed to nonlinear polish, which releases all five parameters (a, c, b, f_0, f_1) simultaneously. Drift-based heuristics (spectral peak drift between half-windows) proved unreliable — the centroid is pinned by any coexisting tonal content — and were abandoned.

3.4. Greedy BIC pursuit and noise-floor stop

Components are added greedily. At each step every candidate is scored by jointly least-squares-fitting it together with the current components, and the candidate with the best penalised score (2) is accepted if it improves on the current model; otherwise the pursuit stops. Nonlinear parameters of the enlarged model are refined by a short block-coordinate sweep after each addition. Because candidates are regenerated on the current residual, collinearity between an atom and already-selected components is handled naturally (an exponential riding on a linear trend is found from the residual, not from the mixture).

Two stopping safeguards complement the BIC comparison. First, a *noise-floor stop*: let $S_k = |X_k|^2$ be the unnormalised one-sided real-FFT power of the current residual and let $\mathcal{I}$ exclude DC and Nyquist. For real white noise, interior bin powers are exponential and the one-sided Parseval factor doubles their contribution, giving

$$\mathrm{SSE}_{\mathrm{floor}} = \frac{2|\mathcal{I}|}{N \log 2} \operatorname*{median}_{k \in \mathcal{I}} S_k. \tag{3}$$

Across 1000 white-noise records, the median estimated/true SSE ratios are 0.997, 0.996 and 1.001 for $N = 256, 1024, 4096$, respectively. If $\mathrm{SSE}(\mathcal{M}) \le \mathrm{SSE}_{\mathrm{floor}}$ the pursuit stops — further atoms would fit background structure, not signal. Second, the component budget $K \le K_{\max}$ (default 6). On synthetic Gaussian-noise data either safeguard terminates the pursuit at the true order (Section 4); on coloured broadband machinery noise a model-free one-shot floor estimate is ill-posed — both mean-of-top-band and median estimators under-estimate the in-band level — and $K_{\max}$ acts as the operative budget (Section 4.7). We return to this in the discussion.

3.5. Refinement

After the pursuit, each component's parameters are refined by block-coordinate descent: component r is fitted to its *own signal* $f - \sum_{j \neq r} s_j$ by boxed Levenberg–Marquardt. Several sweeps are run after each greedy addition and after the pursuit. Because block-coordinate steps can stall when several chirp-like components share a band, a final *joint polish* refines all parameters simultaneously against the reconstruction objective, accepted only if the SSE improves.

3.6. Family re-selection

Each component, isolated on its own signal, is re-fitted by every family (with candidates regenerated from that isolated signal) and keeps the family with the best penalised score, repeated to a fixpoint. This guards against family mismatch introduced by the initial candidate set.

3.7. Local search: compound structure moves

Greedy addition followed by single-atom deletion is not sufficient: we observed a systematic dead-lock. When a single wide-band atom (e.g. a damped sinusoid) has absorbed the content of two close tones, subsequent atoms are added to mop up the residual; these "junk" atoms collectively explain the two-tone structure so well that replacing the wide-band atom by the correct pair no longer improves the penalised score by enough to pay for the added parameters — the move is rejected and the wrong structure is frozen.

We resolve this with a fixpoint loop of *compound* moves, each accepted only if the global penalised score improves:

remove delete one component;

split replace one tonal component (sinusoid/damped/chirp) by the best sinusoid pair, fitted as described in Section 3.8;

merge replace two tonal components closer than 2 Hz by one sinusoid.

The split move is compound: the replacement is evaluated together with the removal of up to two existing tonal components, capturing the fact that the split may render them redundant. In a development version that lacked the post-pursuit family re-selection (Section 3.6), this dead-lock occurred on S3 itself:

without the co-removal term the split was rejected (penalised-score change +3.7) and the wrong structure was frozen. In the final algorithm the re-selection pass removes the junk atoms before local search in most conditions, and at 30 dB the compound term is neutral on S3 (Table 12); it becomes the discriminating mechanism at lower SNR, where the junk atoms survive re-selection: at 20 dB, over 10 paired seeds, full DTHy-SMP recovers the $K = 3$ structure in 10/10 runs (median worst-tone MSE 1.6×10^{-4}), while without co-removal the pursuit freezes into K=5–6 junk-atom configurations with the tone pair unresolved in 9/10 runs (median worst-tone MSE 0.18). A ridge-guarded solver ($\lambda = 10^{-8}\,\mathrm{tr}(\mathbf{G})/k$ added to the Gram matrix of each joint fit) is essential: without it, a sinusoid pair at nearly identical frequencies with amplitudes of order 10^{10} can emulate a low-band shape and game the SSE numerically.

3.8. Nuisance-orthogonalised pair fitting

Splitting a tonal component into two sinusoids requires solving a two-frequency estimation problem that may lie below the Rayleigh limit $1/T$ of the record. Two facts make this delicate. First, *any unmodelled low-band content biases the landscape*: for a mixture of tones at 5.0 and 5.6 Hz (1 s record) with an exponential component, the shifted pair partially absorbs the exponential's Lorentzian leakage, and the global minimum of the pair SSE can move far from the true frequencies (in a noisier variant of the context below we observed it at (5.09, 5.47) Hz). Second, *removing the nuisance the wrong way is worse than not removing it*: projecting the residual off the nuisance basis $\mathcal{B} = \{1, \tau, \tau^2, e^{-b\tau}\}$ but fitting the pair on the projected signal distorts the landscape in the opposite direction, because tonal content overlaps $\mathcal{B}$ by a non-negligible amount ($\|P_{\mathcal{B}} s_{\text{tones}}\|^2 / \|s_{\text{tones}}\|^2 \approx 14\%$ in our case).

The correct operation is the *joint* fit: fit the pair and $\mathcal{B}$ together. We implement this exactly by orthogonalising *both* the candidate pair columns and the target signal against $\mathcal{B}$ before the Gram scoring, i.e. applying the projector $P^{\perp}_{\mathcal{B}} = I - B(B^{\top}B)^{-1}B^{\top}$, with $B \in \mathbb{R}^{N\times 4}$ the nuisance basis matrix $[\,1,\ \tau,\ \tau^2,\ e^{-b\tau}\,]$, to every column of the pair dictionary and to the residual in question. Because $\|P^{\perp}_{\mathcal{B}}(s - \hat{s})\|^2$ equals the SSE of the joint fit of the pair together with $\mathcal{B}$, the operation is exact (not an approximation of the joint fit) and costs one projection of the grid against $\mathcal{B}$ per split candidate.

Fig. 2 visualises the three landscapes on the S3 dead-lock context (tonal content plus residual low-band leakage after the coarse low-band atom is subtracted). Under the joint fit (panel a) the basin is centred on the true pair and the grid winner (5.00, 5.60) Hz polishes to (5.000, 5.597) Hz; the tones are recovered with MSE of order 10^{-5}. Fitting the pair alone (panel b) displaces the basin along the diagonal: the argmin still corresponds to the true pair on this realisation, but the minimum region straddles it, and in noisier contexts the global minimum moves outright. The asymmetric projection (panel c) distorts the landscape in the opposite direction, as predicted by the overlap argument, and its argmin (4.84, 5.70) Hz misses the true pair by more than the joint-fit error by an order of magnitude.

Algorithm 1 DTHy-SMP: one run

```
generate candidates from f (peaks, low band, matrix pencil, chirp grid)
while |M| < K_max do
    regenerate candidates on current residual
    add best-scoring candidate if (2) improves and residual above noise floor
    refine (block-coordinate, 2 sweeps)
end while
refine (block-coordinate, full sweeps) + joint polish
family re-selection to fixpoint
local search fixpoint: remove / split (nuisance-orthogonalised, compound) / merge
final refinement pass + family re-selection
```

3.9. Summary and complexity

Let M be the candidate count per iteration and K the final order. Each greedy iteration costs M joint least-squares fits of width $O(K)$ on N samples plus one candidate regeneration (one FFT and one matrix-pencil SVD), and $O(K_{\max})$ iterations are run; refinement and re-selection cost $O(K^2)$ boxed least-squares solves of a few parameters each. Observed runtimes are 2–17 s for $N = 1024$ and 19–24 s for $N = 4096$ at 12 kHz (single thread, laptop; Section 4).

4. Experiments

4.1. Setup

All synthetic signals use $N = 1024$ samples at $f_s = 1024$ Hz. SNR is defined as $10\log_{10}$ of the ratio of clean-signal power to noise power. Unless stated, one Gaussian noise realisation at the scenario SNR (seed 42) is used; statistical studies use 20 paired seeds. Cases S1–S8 (Table 2, Fig. 3) cover a basic mixture, multitone with trend, overlapping frequencies (0.6 Hz spacing, sub-Rayleigh for a 1 s record), the basic mixture at 5 dB and 0 dB, chirp and harmonic, an AM–FM component, and a non-periodic-only case. Methods compared: EMD [1] (plain sifting, up to four IMFs plus residue; no noise-assisted variant); VMD with uniform centre initialisation (excluding DC and Nyquist) and with data-driven (spectral-peak) initialisation, the latter optionally with a pinned DC mode; matrix pencil [21]; and DTHy-SMP. We additionally compare against three recent strong decompositions, all given the same order K as the ground truth: successive VMD [12] (SVMD), which extracts modes one at a time with data-driven centre frequencies; empirical Fourier decomposition [19] (EFD), an adaptive zero-phase filter-bank segmentation of the Fourier spectrum; and nonlinear chirp mode decomposition [15] (NCMD), a variational demodulation method designed for time-varying frequency content. All three are used through PySDKit version 0.5.0 [38], a documented port of the official reference implementations (MATLAB File Exchange 64292 for NCMD and 97747 for EFD), and the faithfulness of the ports was checked by reproducing the published reference simulations (NCMD on the crossing-IF example of [15]: IF relative errors below 0.1%;

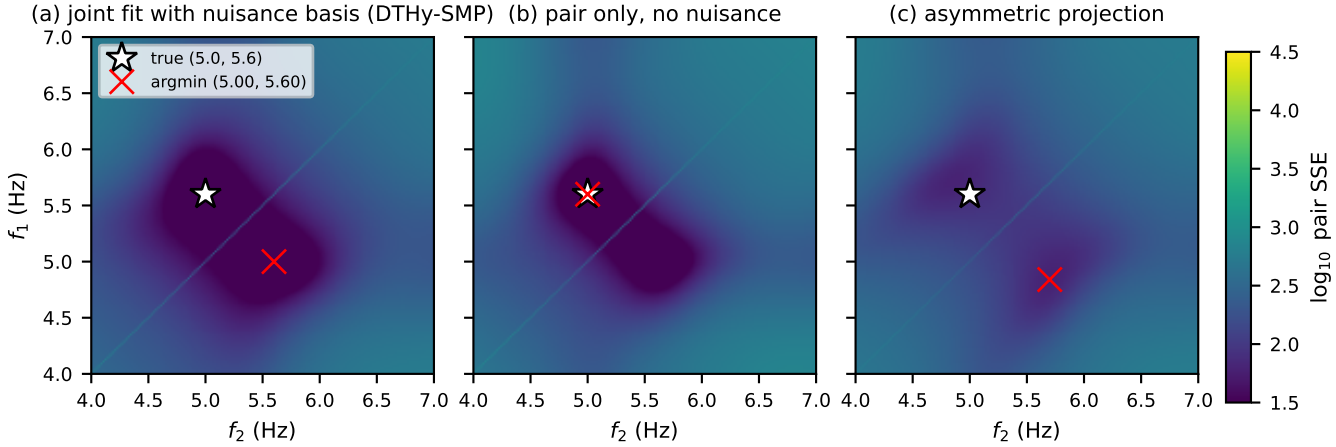


Figure 2: Pair-fitting landscapes on the S3 dead-lock context ($\log_{10}$ SSE of the two-sinusoid fit over the frequency grid; white star: true pair (5.0, 5.6) Hz; red cross: grid argmin). (a) Joint fit with the nuisance basis $\mathcal{B}$ (the DTHy-SMP estimator): the basin is centred on the true pair. (b) Pair fit without nuisance modelling: the basin is displaced along the diagonal and the minimum region straddles the true pair. (c) Asymmetric projection (signal projected off $\mathcal{B}$, pair columns raw): the landscape distorts in the opposite direction and the argmin (4.84, 5.70) Hz misses the true pair.

EFD on the fSig2 segmentation test of [19]: clean separation of the closely spaced pair), and our VMD core was cross-checked against the reference package vmdpy [39]. NCMD is run in the reference configuration of the official implementation ($\alpha = 5 \times 10^{-6}$, $\beta = 10^{-6}$, var = 0, tol = 10^{-8}) with initial IFs at the K strongest spectral lines including DC. A prior-guided VMD variant (quadratic prior injected into the Wiener update, representative of the prior-repair line of work) is evaluated only in the ablation (Section 4.9). Component errors are one-to-one matched to ground truth by minimising total MSE (Hungarian assignment).

4.2. A forensic note: what the VMD baseline measures

A widely used VMD configuration initialises the centre frequencies uniformly from DC to Nyquist — the default of the reference code distributed with the original paper [2]. Reproducing this configuration on S1, we obtain a periodic-mode MSE of 5.72×10^{-2} with the second mode essentially empty (energy 4.7×10^{-5}): the DC-initialised mode absorbs the exponential and part of the tonal content while the Nyquist-initialised mode never acquires energy. Any initialisation that does not place a centre at DC and Nyquist — uniform without DC, spectral peaks, peaks with a pinned DC mode — lets plain VMD reach a periodic-mode MSE of 2.4×10^{-3}, 23 times better than the DC/Nyquist-initialised default. Large "gains over VMD" reported in the literature for methods evaluated against the default-initialised baseline can therefore be an artefact of the baseline's initialisation rather than of the proposed mechanism. All VMD baselines in this paper use the corrected, initialisation-robust configuration, making the remaining comparisons conservative.

4.3. Decomposition quality

Fig. 4 illustrates the contrast on S1. Table 3 gives the seed-42 snapshot. On S1–S4 and S7, DTHy-SMP recovers components to 10^{-6}–10^{-4} MSE with the correct analytic families and correct model order; in S3 it resolves the 0.6 Hz pair (Section 3.8); at 0 dB it retains a 49-fold median advantage in the statistical study below. Matrix pencil is strong at 0 dB (1.1e-2), as expected for a damped-sinusoid specialist; DTHy-SMP has lower error in the tested 30 and 5 dB cases. EMD, the classical reference, does not isolate the tonal content on these short (1 s) noisy records: the sifting order is dominated by the noise and the exponential, and the best-matching IMF carries errors of order 10^{-1}–10^{-2}; we report it as a sanity reference rather than a competitor for symbolic compactness. Among the recent decompositions, SVMD is the weakest of the three on the multitone cases (its successive extraction is sensitive to the stopping criterion) but competitive at 0 dB; EFD degrades at 0 dB (S7: 1.8×10^{-1}) and its spectral segmentation cannot resolve the sub-Rayleigh S3 pair; NCMD, in the reference configuration of the official implementation, is the strongest of the three recent methods on S1, S2 and S6 — its S6 AM–FM error (1.6×10^{-3}) is the best baseline result in the table — but fails outright on S3 and S5, where two initial IFs lie closer together than its tracking can separate, a documented limitation of the method [15]. Wherever the generating families lie in the catalogue (S1–S4, S7), the best baseline remains two to five orders of magnitude above DTHy-SMP, and none of these three configurations returns closed-form components or a model-order estimate. S8 is weakly identifiable over this one-second window — the exponential and quadratic conspire to look nearly linear — and all methods fit the mixture well while the family assignment is unstable. S5 is an in-catalogue but structurally difficult tone–chirp mixture, whereas S6 contains AM–FM content outside the catalogue. On S5 the genuine chirp is recovered by a chirp atom ($\hat{f}_0, \hat{f}_1 = (4.0, 6.0)$, true $(4, 6)$), the exponential is recovered by an exponential atom (MSE 7.6×10^{-5}), and K nevertheless rises to 6 because the tone–chirp overlap is over-resolved; on S6 the AM–FM content is represented by chirp atoms (matched MSE 4.6×10^{-2}) and the exponential by a de-

Table 1: Principal numerical settings used in all reported experiments unless explicitly varied.

item	setting
DTHy-SMP candidates	at most 12 FFT peaks; two-bin minimum separation; prominence $\max(10^{-4}$ of maximum power, $10^{-12})$; dampings $\{0.5, 1, 2, 4\}$; matrix-pencil order 6; chirp grid 41×25, best three seeds polished
parameter bounds	$f \in [0.05, 480]$ Hz; decay $b \in [10^{-3}, 80]$; log curvature $[10^{-3}, 50]$; chirp rate $[-200, 200]$ Hz/s; amplitudes and polynomial coefficients unbounded
selection/refinement	Gaussian BIC; $K_{\max} = 6$; chirp margin 20; ridge $10^{-8}\,\mathrm{tr}(X^T X)/p + 10^{-12}$; three refinement sweeps; boxed trust-region least squares, 200 evaluations per block and 400 in joint polish
VMD	peak initialisation unless labelled uniform; $\alpha = 2000$, $\tau = 0$, 300 iterations, relative tolerance 10^{-7}; true K supplied
recent baselines	PySDKit 0.5.0 defaults for SVMD/EFD; NCMD $\alpha = 5\times10^{-6}$, $\beta = 10^{-6}$, var = 0, tolerance 10^{-8}; true K supplied
compute environment	Apple M3 Pro (12 cores), 18 GB RAM; macOS 26.6.2; Python 3.13.1, NumPy 2.2.6, SciPy 1.16.3

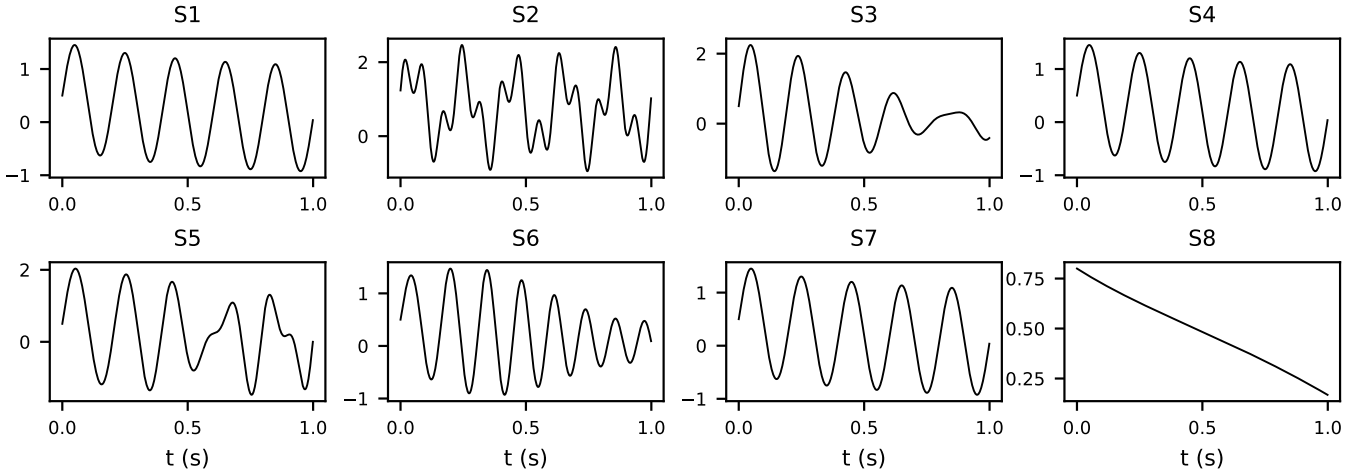


Figure 3: The eight synthetic cases (clean signals, $N = 1024$, $f_s = 1024$ Hz). S1–S5 and S7 use families present in the catalogue (S5 is non-stationary but contains an admitted chirp); S6 contains AM–FM content outside the catalogue; S8 is the weakly identifiable boundary case (see below).

Table 2: Synthetic cases ($t \in [0, 1)$ s). S2 contains two tones and a linear trend; unlike S1, it contains no exponential component. S4 and S7 replicate S1 at 5 dB and 0 dB.

case	components	K	SNR
S1	$\sin 2\pi 5t + 0.5e^{-2t}$	2	30 dB
S2	$\sin 2\pi 5t + 0.7\sin(2\pi 13t{+}0.5) + 0.4 + 0.3t$	3	30 dB
S3	$\sin 2\pi 5t + 0.8\sin 2\pi 5.6t + 0.5e^{-2t}$	3	30 dB
S4	= S1	2	5 dB
S5	$\sin 2\pi 5t + 0.6\,\mathrm{chirp}(4{+}6t) + 0.5e^{-2t}$	3	30 dB
S6	$0.8\,m(t) + 0.5e^{-2t}$, $m(t) = (1{+}0.5\sin 2\pi 0.8t)\ \sin 2\pi(6t{+}1.5t^2)$	2	30 dB
S7	= S1	2	0 dB
S8	$0.5e^{-2t} + 0.3 + 0.2t - 0.4t^2$	2	30 dB

caying atom (1.9×10^{-4}). On S6 the band-limited VMD modes achieve lower matched waveform MSE, whereas DTHy-SMP returns an analytic approximation with model-order overshoot ($\hat{K} = 6$). The contrast separates the two objectives: waveform fidelity favours VMD on this out-of-catalogue signal, while the overshoot exposes that the selected families do not provide a compact description. S5 shows that overshoot alone does not diagnose catalogue mismatch, so it is used only as a model-adequacy warning (Section 5).

4.4. Statistical study

Table 4 separates the value of the family catalogue from the value of the search. On S1/S2 the greedy control is already sufficient. On the close-tone S3 case, however, full DTHy-SMP recovers the complete structure in 19/20 runs, whereas the same-library greedy control never does and uses a median 22 rather than eight parameters. Thus the compound search is supported on the structural dead-lock it was designed to address. Fitting the same families after VMD is not equivalent to joint symbolic decomposition: it succeeds on S2 but misses complete structure on the remaining cases. The S5/S6 zero rates and the lower S7/S8 rates locate the remaining errors in tone–chirp over-selection, catalogue mismatch, low-SNR family confusion and weak identifiability. Against those boundaries, the S3 control provides a clean mechanism-level result: with the same catalogue, candidates and penalty, the compound search raises full-structure recovery from 0/20 to 19/20 while reducing the median parameter count from 22 to eight.

Table 5 and Fig. 5: on every stationary case (S1–S4, S7) DTHy-SMP wins all 20 seeds with the minimum attainable two-sided p; the advantage narrows gracefully with SNR but does not invert. S5 is a partial win (19/20): its in-catalogue tone–chirp overlap is often over-resolved. S6 and S8 expose different boundaries: oscillatory content outside the catalogue

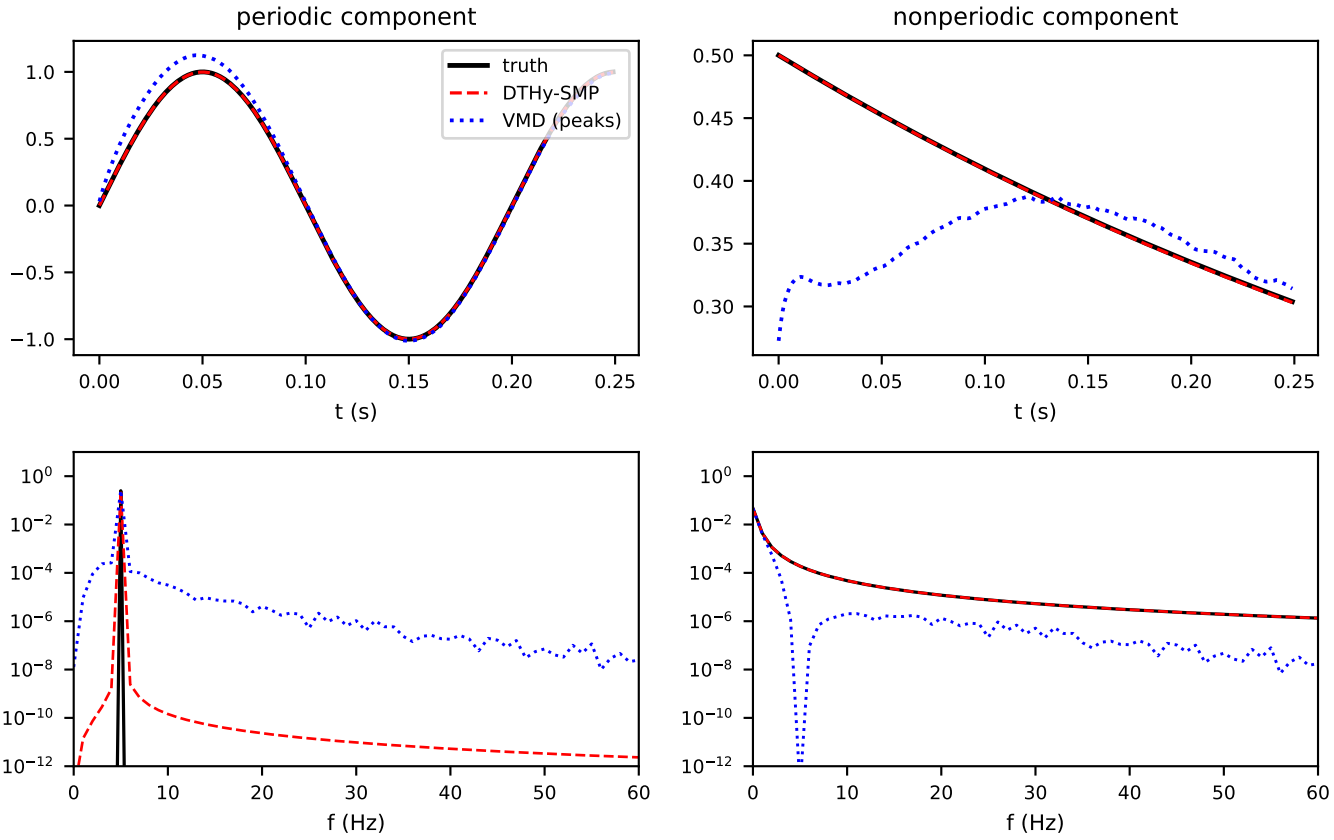


Figure 4: S1 (30 dB), matched components. Top row: time domain (first 0.25 s); bottom row: amplitude spectrum. DTHy-SMP components (red dashed) overlay the truth (black); the VMD mode matched to the periodic component carries a raised spectral floor, and the mode matched to the exponential contains tonal leakage at 5 Hz and its harmonics.

(S6) and weak family identifiability within a one-second window (S8); their aggregate matched MSE favours the waveform-fidelity baseline. The positive result is therefore specific but consistent across the intended stationary cases: S1–S4 and S7 each give 20/20 paired wins, including a 49-fold median gain at 0 dB. Runtimes per run are 1.7–18.5 s (DTHy-SMP).

4.5. Resolution boundary: separation, SNR, and the limits of estimation

The statistical study above fixes the tone geometry and varies the noise; we now fix the noise family and push the geometry to its breaking point. On the S3 family (tones of amplitude 1 and 0.8 at 5.0 and 5.0 + Δf Hz plus the additive component $0.5e^{-2t}$, $N = 1024$, $f_s = 1024$ Hz, twenty seeds per cell) we sweep $\Delta f \in \{0.2, \ldots, 1.0\}$ Hz against SNR $\in \{30, 20, 10, 0\}$ dB. Two rates are reported per cell: the *pair-resolution rate*, the fraction of runs in which both tones are recovered with $|f - \hat{f}|$ below one quarter of the separation, and the *exact-order-and-resolution rate*, the fraction that also returns the three-component generating order. Frequency errors are conditioned on successful resolution only for the descriptive medians printed during the sweep; Fig. 6c instead uses every run with at least two tonal outputs and therefore does not select on success.

The sweep does not define a monotone super-resolution threshold (Fig. 6). At 30 dB, the pair-resolution rates for $\Delta f = 0.2$–1.0 Hz, in 0.2 Hz steps, are respectively 0.90, 0, 0.95, 1.00, and 1.00. The corresponding exact-order-and-resolution rates are 0.85, 0, 0.95, 0, and 1.00. The failures at 0.4 Hz and the order inflation at 0.8 Hz show that candidate generation and local structural moves, rather than separation alone, control the operating boundary. At 20 dB the pair rates are $0.70, 1.00, 0.95, 1.00, 1.00$, while the joint rates are $0.70, 0.65, 0.90, 0.25, 1.00$. At 10 dB the joint rates rise from 0.05 at 0.2 Hz to 1.00 at 0.8–1.0 Hz; at 0 dB they rise from 0 to 0.95. Most strikingly, the 0.8-Hz joint rate is 0, 0.25, 1.00 and 0.85 as SNR decreases from 30 to 0 dB. This inversion is a path-dependent non-convex selection effect, not evidence that noise improves physical resolution.

At 30 dB every run returns at least two tonal outputs, so the all-output second-tone RMSE in Fig. 6c uses all 20 seeds per separation. The DTHy-SMP RMSEs are 0.020, 0.653, 0.083, 0.087 and 0.00169 Hz for increasing Δf; the 0.4-Hz value exposes a severe misassignment that a success-conditioned median would hide. VMD-peak resolves no pair under the $\Delta f/4$ rule at 30 dB, and its corresponding all-output RMSEs are 0.139, 0.984, 0.408, 0.494 and 0.655 Hz. Although DTHy-SMP does not define a monotone resolution limit, it has lower

Table 3: Seed 42: matched periodic-component MSE (best in bold; "–": no periodic component; the S6 column reports the matched MSE of the AM–FM component, which lies outside the family catalogue — no method is best there, see text). "unif." = uniform centres without DC/Nyquist (Section 4.2). SVMD, EFD and NCMD are given the true order K; all other settings are shared. "Families correct" counts matched truth families only and does not penalise extra atoms; full structural recovery requires both the correct order and all truth families.

	S1	S2	S3	S4	S5	S6	S7	S8
EMD (best IMF)	3.3e-1	7.2e-2	2.7e-1	2.9e-1	2.5e-1	8.6e-3	7.1e-2	–
VMD (unif.)	2.5e-3	3.1e-1	2.9e-1	6.0e-2	1.6e-1	8.3e-3	7.2e-2	–
VMD (peak)	2.4e-3	1.0e-3	2.9e-1	5.9e-2	7.6e-2	8.3e-3	7.0e-2	–
+ DC mode	2.5e-3	1.0e-3	2.9e-1	6.2e-2	7.6e-2	7.9e-3	7.0e-2	–
SVMD [12]	5.8e-2	3.2e-1	3.9e-1	5.5e-2	2.4e-1	1.6e-2	5.6e-2	–
EFD [19]	3.3e-2	4.5e-2	3.2e-1	8.7e-2	1.1e-1	5.7e-3	1.8e-1	–
NCMD [15]	1.2e-2	1.7e-3	1.0e0	1.1e-1	1.8e0	1.6e-3	1.7e-1	–
Matrix pencil	8.7e-3	8.2e-3	3.2e-1	9.4e-3	6.6e-2	1.5e-1	1.1e-2	–
DTHy-SMP	**5.7e-7**	**1.2e-6**	**9.1e-6**	**1.8e-4**	**2.0e-2**	4.6e-2	**5.6e-4**	–
DTHy-SMP K est. / true	2/2	3/3	3/3	2/2	6/3	6/2	2/2	2/2
DTHy-SMP families correct	2/2	3/3	3/3	2/2	3/3	0/2	2/2	0/2

Table 4: Direct symbolic-compactness controls over 20 noise seeds. "Full structure" requires the generating order and every matched truth family, with any extra atom counted as failure. Greedy BIC uses the same candidates and penalty, retains family re-selection and final joint polish, but disables the compound local-search loop; VMD→symbolic fits the best catalogue family to each of K VMD modes and is therefore given the true order. Values are full-structure rate and median number of fitted parameters p in parentheses.

case	full DTHy-SMP	greedy BIC	VMD→symbolic
S1	1.00 (5)	1.00 (5)	0.00 (8)
S2	1.00 (8)	1.00 (8)	0.95 (8)
S3	0.95 (8)	0.00 (22)	0.00 (13)
S4	0.90 (5)	0.90 (5)	0.00 (7)
S5	0.00 (22)	0.00 (22)	0.00 (14)
S6	0.00 (29)	0.00 (29)	0.00 (8)
S7	0.65 (5)	0.65 (5)	0.00 (7)
S8	0.05 (5)	0.00 (5)	0.00 (9)

all-output second-tone RMSE than VMD-peak at all five tested 30-dB separations and resolves the pair in four of the five cells. We report rates with Wilson intervals and all-output error rather than an efficiency bound.

4.6. Sensitivity to the component budget

The default budget $K_{\max} = 6$ is a safeguard, not a tuned parameter, but a reviewer may reasonably ask whether the estimated order merely tracks the budget. Table 6 and Fig. 7 answer this in both directions. For S1–S4 and S7, $\hat{K}$ equals the true order over the practical range $K_{\max} = 3$–6. The pursuit stops when the penalised score turns (Fig. 7a), rather than automatically spending the budget. This stability is not global. On S5/S6 the budget is spent, as reported in Section 4; this is a model-adequacy warning, not a stationarity classifier. Another sensitivity is S3 at $K_{\max} \geq 8$: with a budget about 2.7 times the true order, an early greedy mis-step can be absorbed instead of corrected, and the local search freezes a K=8–9 configuration. We therefore recommend $K_{\max} \leq 6$ for signals expected to contain about three components and recommend reporting a budget sweep whenever that expectation is uncertain.

On the rotor records of Section 4.8, where the noise floor defeats the stop and the budget is the operative limit (Section 4.7), the same sweep tells a consistency story: the shaft-harmonic lattice fitted to the recovered tonal lines is stable at $f_0 = 19.7875$–19.8095 Hz with maximum on-lattice deviation ≤ 4.34 mHz for every budget $K_{\max} \in \{6, 8, 12\}$ in all three conditions. The number of supported lattice lines rises from four or five at $K_{\max} = 6$ to nine at $K_{\max} = 12$. At $K_{\max} = 4$, however, the misalignment record has fewer than three supported lines and no lattice fit; the stability claim therefore begins at six, not four. The result shows that additional budget primarily recovers additional shaft orders, but it does not establish that every added atom is physical.

4.7. Real-background validation

Table 7: on real bearing backgrounds the injected component is recovered around 10^{-6} MSE by DTHy-SMP versus 4.6–6.8×10^{-4} for initialisation-corrected VMD (two orders of magnitude). The K estimate pins the budget $K_{\max} = 6$ on every segment: real broadband machinery noise is coloured, and a model-free one-shot noise-floor estimate from a single spectrum under-estimates the in-band level (Section 3.4), so the BIC keeps accepting small structural fits. The matched injected-component MSE is a target-recovery metric, not a measure of complete background decomposition; the repeated $K_{\max}$ saturation shows that the latter is not achieved. In practice the explicit component cap, rather than the white-noise floor, is therefore operative on these records. Runtime per segment is 19–24 s at $N = 4096$. The broader sweep in Table 8 varies frequency from 20 to 80 cycles per segment (58.6–234.4 Hz) and amplitude from 0.5 to 1.0 times the background RMS. DTHy-SMP has lower target NMSE in all 48 paired runs, with median gains of 356–1270, while returning $K = 6$ in every run. Thus the sweep supports target recovery across record, segment, fre-

Table 5: 20 paired noise seeds per case, all synthetic cases: median matched periodic-component MSE (IQR in parentheses for DTHy-SMP), median gain vs. DTHy-SMP, two-sided paired Wilcoxon p (1.9×10^{-6} is the smallest attainable for $n = 20$), and number of seeds where DTHy-SMP is smaller. The S6 row reports the AM–FM component (as in Table 3); the S8 row reports the aggregate over its two non-periodic truths. Listed p values are raw; Holm correction across the eight tests gives adjusted $p \leq 1.53 \times 10^{-5}$ for every row.

case	method	DTHy-SMP med	base med	gain	p (wins)
S1 (30 dB)	VMD-peak	1.5e-6 (1.1e-6)	2.5e-3	1653×	1.9e-6 (20/20)
S2 (30 dB)	VMD-peak	2.7e-6 (2.0e-6)	1.0e-3	364×	1.9e-6 (20/20)
S3 (30 dB)	VMD-peak	5.1e-6 (1.5e-5)	2.9e-1	6.1e4×	1.9e-6 (20/20)
S4 (5 dB)	VMD-peak	4.7e-4 (3.3e-4)	6.4e-2	129×	1.9e-6 (20/20)
S5 (30 dB)	VMD-peak	2.0e-2 (6.0e-4)	7.6e-2	3.8×	3.8e-6 (19/20)
S6 (30 dB)	VMD-peak	4.6e-2 (4.4e-5)	8.4e-3	0.18×	1.9e-6 (0/20)
S7 (0 dB)	VMD-peak	1.5e-3 (1.0e-3)	8.1e-2	49×	1.9e-6 (20/20)
S8 (30 dB)	VMD-peak	1.6e-1 (8.2e-1)	5.9e-2	0.46×	3.8e-6 (1/20)

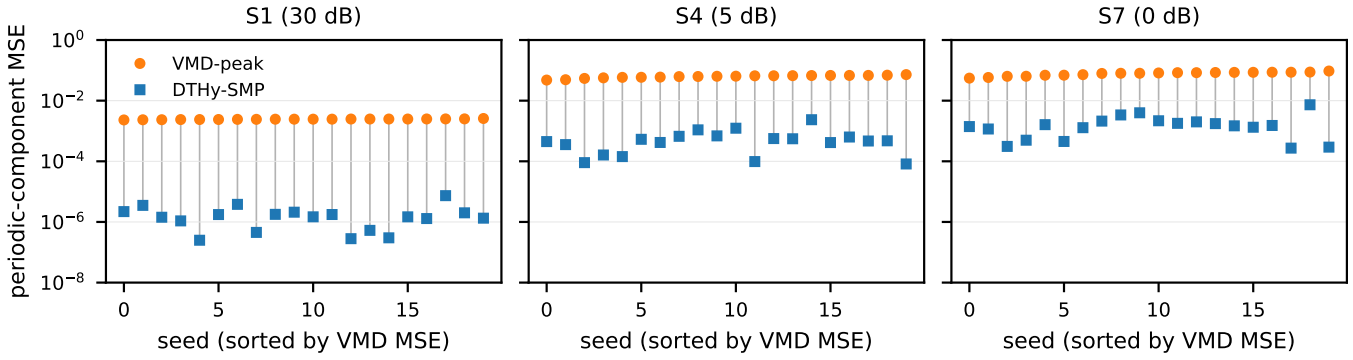


Figure 5: Per-seed paired periodic-component MSE on the three SNR conditions with a full 20-seed study (seeds sorted by VMD MSE within each panel). At every SNR every individual DTHy-SMP run sits below its paired VMD-peak run; the separation widens by orders of magnitude from 0 dB to 30 dB.

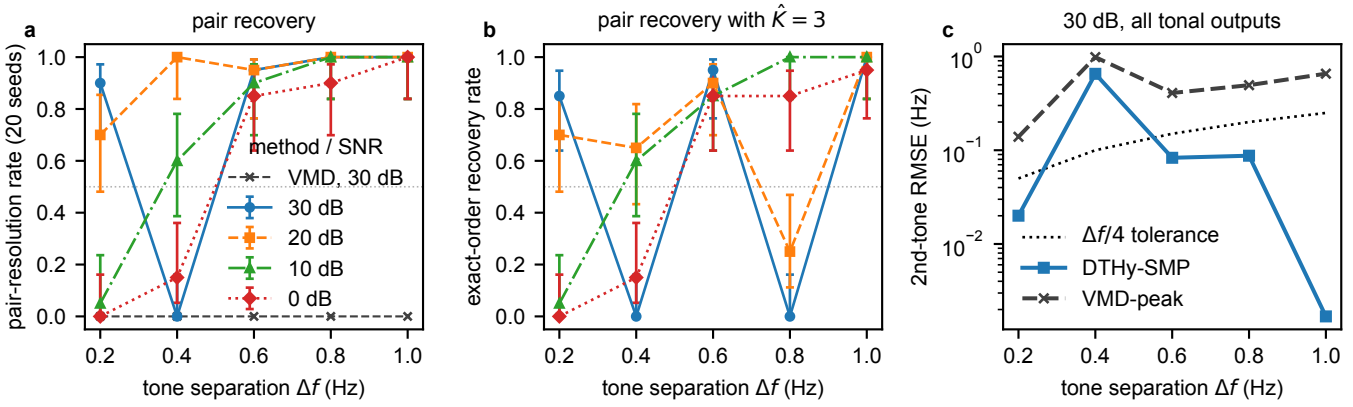


Figure 6: Resolution behaviour on the S3 family (20 seeds per cell). (a) Pair-resolution rate under the strict error threshold $\Delta f/4$; (b) the fraction that both resolves the pair and returns the generating order $\hat{K} = 3$. Error bars are 95% Wilson intervals. (c) Second-tone RMSE over all runs that return at least two tonal outputs at 30 dB, for DTHy-SMP and VMD-peak; the dotted line is the descriptive tolerance $\Delta f/4$.

quency and amplitude, but not complete decomposition of the coloured background.

4.8. Real rotating-machine records

Finally we apply DTHy-SMP to measured records from a generic rotating-machine benchmark [40]: a shaft running at nominal 20 Hz in three conditions — healthy, horizontal misalignment (NDE), and unbalance. We use the uncoupled horizontal accelerometer, a 2 s steady-state window (2–4 s), and restrict the analysis to the rotating-harmonic band (zero-phase decimation by 8 to $f_s = 748.5$ Hz, covering the rotating-harmonic band up to the bench's high-frequency resonance content); the bench's high-frequency resonance band is band-limited stochastic excitation outside the component catalogue,

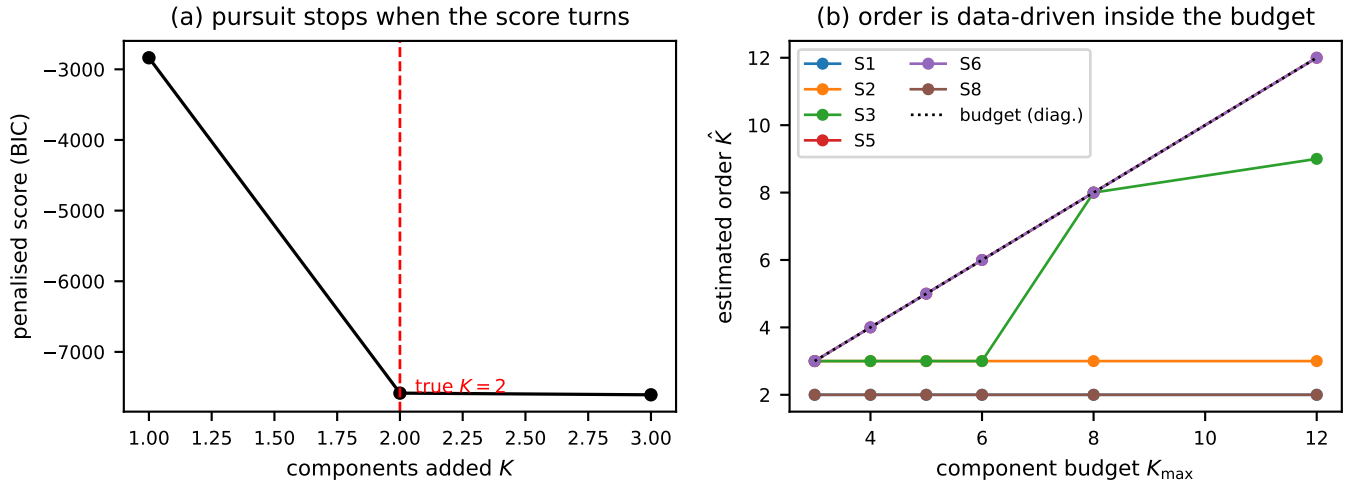


Figure 7: Budget sensitivity. (a) Penalised score of the augmented model on S1: the minimum sits at the true order $K = 2$ and the pursuit stops there. (b) Estimated order vs. budget for the synthetic cases: S1, S2, S4, S7 and S8 are flat at the true order; the difficult S5 mixture and out-of-catalogue S6 track the diagonal.

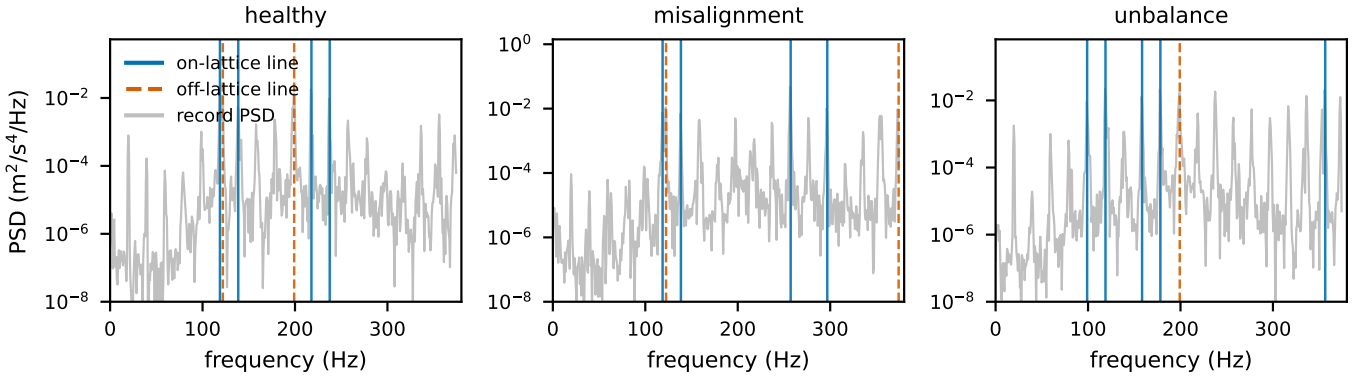


Figure 8: Analysis-band spectra of the three primary rotor records (Welch PSD, 2 s window). Blue solid lines: tonal components recovered by DTHy-SMP that lie on the shaft-harmonic lattice of Table 9; vermillion dashed lines: recovered constant-frequency lines rejected by the same lattice test (candidate structural, i.e. non-synchronous, lines). The two line families are separated by the fitted integer-order test.

consistent with the coloured-noise discussion of Section 3.4. The trigger channel yields 20.0 pulses/s (nominal).

With the default budget $K_{\max} = 6$ the estimated order pins the budget on all three records, and the pursuit returns six closed-form components each (Table 9). We post-process the recovered frequencies by searching a fundamental over 19.0–20.6 Hz, assigning integer orders, and requiring at least three lines within 50 mHz in fundamental-frequency units. The fitted lattices have maximum within-lattice deviations of 5 mHz, and the three fitted fundamentals span 19.788–19.806 Hz. These residuals measure internal consistency after the nominally constrained fit; they are not an independent shaft-speed accuracy estimate. The optical pulse rate is 19.88 Hz over the window, about 0.4% above the fitted lattice values, and the pulse-to-revolution convention is not documented. Lines that do not belong to the fitted lattice are separated by the same test: 121.9/122.5 Hz and 199.1 Hz recur across conditions at constant frequency and are therefore flagged as candidate structural (non-synchronous) lines rather than shaft orders (Fig. 8); confirming them as structural resonances in the strict modal sense would require a modal test, which is beyond the present records. The recovered order-dominance pattern is condition-specific and physically plausible: unbalance concentrates at mid orders 5–9 (1X–4X are comparatively weak in this channel and band), misalignment adds high-order content (13X, 15X), and the healthy record is dominated by the resonance-adjacent 6X–12X group. Runtime is 39–47 s per record.

The single-record analysis above could in principle be record-specific, so we replicate it: two further records per condition (003 and 009, with 008 retained) under identical settings (Table 10). Under the same nominally constrained search, all nine decompositions yield at least four on-lattice lines and maximum on-lattice deviation 5.1 mHz, and an estimated model order of six in every case. The nine implied shaft speeds span only 19.781–19.809 Hz (a 28 mHz spread, consistent with the 18 mHz spread of the three primary records). This

Table 6: Estimated order $\hat{K}$ vs. the component budget $K_{\max}$ (seed 42). True orders: S1/S4/S7/S8: 2; S2/S3/S5: 3; S6: 2. Cases S1–S4, S7 and S8 stop at the true order for $K_{\max} \leq 6$; the difficult tone–chirp S5 and out-of-catalogue S6 spend the budget (Section 5); S3 degrades from $K_{\max} = 8$, about 2.7 times its true order.

case	$K_{\max}$=3	4	5	6	8	12
S1	2	2	2	2	2	2
S2	3	3	3	3	3	3
S3	3	3	3	3	8	9
S4	2	2	2	2	2	2
S5	3	4	5	6	8	12
S6	3	4	5	6	8	12
S7	2	2	2	2	2	2
S8	2	2	2	2	2	2

Table 7: CWRU baseline records (12 kHz drive-end) with an injected sinusoid (f = 40 cycles per 4096-sample segment, amplitude = background RMS). Median injected-component MSE over 3 segments per record.

record	DTHy-SMP	VMD-peak	DTHy-SMP K est.
105	3.4e-6	6.7e-4	6,6,6
106	1.1e-6	4.7e-4	6,6,6
107	1.2e-6	5.1e-4	6,6,6
108	7.2e-7	4.6e-4	6,6,6

repetition establishes record-to-record stability under the same nominal-frequency constraint; it is not an independent test of absolute speed accuracy. Residual magnitudes and the DTHy-SMP/VMD gap are likewise stable across records (DTHy-SMP 0.19–0.38, VMD 0.09–0.22 m/s^2). The same lattice search applied to raw peaks and VMD centres (Table 11) confirms that the lattice is visible without DTHy-SMP: raw peaks fit all nine records and reach the same null-test floor. DTHy-SMP's contribution is the fitted representation: its closed-form atoms reduce median lattice deviation from 11.88 mHz for raw peaks and 27.19 mHz for VMD centres to 3.57 mHz, while providing the component equations directly. The independent-uniform null calibrates chance alignment rather than modelling machinery spectra.

For comparison, VMD with peak initialisation and the same order reaches a lower residual RMS on all three records (0.10–0.20 vs. 0.20–0.35 for DTHy-SMP), and the recent baselines behave likewise: SVMD 0.14–0.31, NCMD 0.19–0.23. (EFD's zero-phase filter bank partitions the Fourier spectrum among its modes, so its residual is near zero by construction and is not an accuracy metric.) The budget is spent, and broadband residual content is not DTHy-SMP's objective. What the DTHy-SMP output adds is the deliverable itself: a single closed-form table of components with estimated order, from which speed, order spectrum, and structural-line candidates are read off directly, whereas the spectral modes require a further per-mode fitting stage and provide no model-order estimate. We regard the two as complementary: spectral decomposition for waveform fidelity, symbolic pursuit for physical readout.

Table 8: CWRU robustness sweep over four records and two disjoint segments per record (eight runs per cell). Entries are median target NMSE, normalised by injected-tone power; gain is the median paired VMD/DTHy-SMP ratio.

ampl./RMS	cycles	DTHy-SMP	VMD-peak	gain
0.5	20	9.29e-5	3.47e-2	374×
0.5	40	6.75e-5	4.57e-2	634×
0.5	80	6.48e-5	7.87e-2	1270×
1.0	20	2.32e-5	8.28e-3	356×
1.0	40	1.68e-5	1.09e-2	602×
1.0	80	1.61e-5	1.81e-2	1202×

4.9. Ablation

Table 12 ablates the three structural ingredients of the local search and refinement stage, plus the prior-guided VMD variant as a baseline row. The *chirp margin* is the dominant safeguard on the seed-42 benchmark: without it, a five-parameter chirp absorbs the 0.6 Hz S3 tone pair (periodic MSE 3.0×10^{-1} instead of 9.1×10^{-6}). The *joint polish* gives a consistent mild improvement on the pair fit (9.1×10^{-6} vs. 1.1×10^{-5}). The *compound split-with-co-removal* move is neutral at 30 dB — at high SNR the family re-selection pass already clears the junk atoms before local search — but it is the discriminating mechanism at lower SNR, where the dead-lock described in Section 3.7 survives re-selection: at 20 dB, over 10 paired seeds, the full algorithm recovers the K=3 structure in 10/10 runs (median worst-tone MSE 1.6×10^{-4}), while without co-removal the pursuit freezes into K=5–6 junk-atom configurations with the pair unresolved in 9/10 runs (median worst-tone MSE 0.18).

Table 13 prevents a stronger interpretation of the margin. At the benchmark length $N = 1024$, 20 is effective for S3 and remains useful at 20 dB, but 40 already degrades the 30-dB recovery rate. At $N = 512$ no tested margin repairs S3, while at $N = 2048$ no extra margin is needed. More importantly, none of the tested margins resolves the S5 tone–chirp over-selection at 30 dB. The margin is therefore a narrow empirical guard against chirp/close-tone confusion, not a generally validated prior or a selection-consistency device; its value should be reported and rechecked when record length or component geometry changes.

The prior-guided VMD row is an in-house baseline representative of the prior-repair line of work: VMD whose Wiener update is modulated by a quadratic prior fitted to manifold templates,

$$\hat{u}_k(\omega) = \frac{\hat{f}(\omega) - \sum_{i\neq k} \hat{u}_i(\omega) + \eta_k\, \hat{s}_k(\omega)}{1 + \alpha(\omega - \omega_k)^2 + \eta_k},$$

where s_k is the current template fit (sinusoid for modes ranked periodic, a member of the non-periodic catalogue otherwise, selected by a training-MSE or BIC rule) and η_k an annealed prior strength. It underperforms initialisation-corrected plain VMD on all three cases, consistent with the forensic analysis of Section 4.2: gains that a prior appears to deliver against a mis-initialised VMD baseline do not transfer once the baseline is initialised correctly.

Three further ingredients do not move any synthetic number but are kept as correctness safeguards: the nuisance-orthogonalised pair fit of Section 3.8 (it removes the leakage-induced bias of the pair-fitting objective itself; on these seeds

Table 9: Generic rotating machine (nominal 20 Hz), 2 s steady-state window, rotating-harmonic band (f_s = 748.5 Hz after zero-phase decimation). Recovered tonal lines with assigned shaft order h and lattice deviation $|f/h - f_0|$. The lattice fit f_0 is the fundamental selected by a nominally constrained post-processing search.

condition	shaft f_0 (Hz)	on-lattice lines (h: f Hz, dev mHz)	off-lattice lines (Hz)
healthy	19.799	6: 118.80 (1.1); 7: 138.62 (4.5); 11: 217.77 (2.0); 12: 237.56 (2.4)	121.93; 199.12
misalignment	19.788	6: 118.72 (2.0); 7: 138.52 (0.1); 13: 257.25 (0.7); 15: 296.84 (1.2)	122.46; 374.03
unbalance	19.806	5: 99.03 (0.5); 6: 118.86 (4.7); 8: 158.45 (0.0); 9: 178.24 (1.1); 18: 356.56 (2.9)	199.19

Table 10: Replication across records: three records per condition (records 003/008/009 at 20 Hz; record 008 is the primary record of Table 9). The same nominally constrained lattice analysis is applied to all nine decompositions (K_{est} = 6 throughout); residual RMS in m/s^2.

condition	rec.	f_0 (Hz)	dev (mHz)	#lat.	DTHy-SMP resid	VMD resid
healthy	003	19.7975	3.92	4	0.200	0.099
	008	19.8000	3.55	4	0.199	0.104
	009	19.8010	2.13	4	0.192	0.101
misalignment	003	19.7810	0.41	4	0.219	0.085
	008	19.7875	1.69	4	0.243	0.115
	009	19.7865	2.89	4	0.233	0.105
unbalance	003	19.8065	3.35	5	0.380	0.216
	008	19.8080	3.11	5	0.352	0.203
	009	19.8090	5.08	6	0.350	0.200

Table 11: Same lattice post-processing applied to six raw spectral peaks, six peak-initialised VMD centres, and DTHy-SMP tonal frequencies on all nine rotor records. The null draws six independent uniform frequencies over the analysis band (20,000 trials); p_{null} is the median record-wise tail probability.

source	fits/9	med. #lines	med. dev. (mHz)	med. p_{null}
raw spectral peaks	9	4	11.88	$< 5.0 \times 10^{-5}$
VMD centres	8	4	27.19	1.2×10^{-3}
DTHy-SMP atoms	9	4	3.57	$< 5.0 \times 10^{-5}$

the subsequent joint polish masks its absence numerically), the ridge guard on all normal equations (without it, nearly-collinear atom pairs can game the SSE with cancelling coefficients of order 10^{10}, a failure observed during development), and the white-noise floor stop, whose calibration is verified above but which under-estimates coloured machinery backgrounds; in those records the explicit $K_{\max}$ cap, not the floor estimate, is operative (Section 4.7).

5. Discussion

What the symbolic objective adds. Section 4.2 shows that correcting VMD centre initialisation is necessary for a fair comparison but does not turn spectral modes into compact analytic components. DTHy-SMP instead selects the family, order and parameters inside the decomposition objective. This distinction is supported most cleanly on S3: the full search recovers the generating structure in 19/20 runs, whereas the same-library greedy control and VMD followed by symbolic fitting both recover it in 0/20 runs.

Model-adequacy warning. When a record is not compactly represented by the selected families and search path, DTHy-SMP may spend its component budget. This behaviour is visible in the returned order, but it is not a specific stationarity test: S6 has out-of-catalogue AM–FM content, whereas S5 is an admitted chirp mixture and still over-selects. Budget saturation should therefore be read only as a model-adequacy warning that motivates catalogue or search refinement.

Operating boundaries. (i) Coloured-noise floors defeat the one-shot noise estimator (Section 4.7); estimating the floor under a noise model (e.g. ARMA) is natural future work. (ii) The penalty (2) is Gaussian BIC; heavy-tailed machinery noise may need robustified scores. (iii) The family catalogue is fixed; letting it grow from data (e.g. learned dictionaries of parametric forms) would widen applicability. In particular, trains of delayed impulses — the dominant structure in faulty rolling-element bearing signals [9] — require atoms with free onset times and are out of scope here. (iv) No convergence or selection-consistency theory is provided; the noise-floor/threshold interplay and the compound-move fixpoint are supported by mechanism-level analysis and extensive experiments instead. (v) In S3, the selection begins to degrade at $K_{\max}$ = 8, about 2.7 times the true order: an early greedy mis-step can be absorbed rather than corrected and the pursuit locks a supersaturated structure (Table 6). In the tested compact stationary cases, $K_{\max}$ = 3–6 is the empirically stable range; behaviour outside that range is not supported as budget-invariant, and S5/S6 are not stable within it. These boundaries limit universality but not the tested compact-stationary result: over S1–S4 and S7 the method wins every paired VMD comparison across 20 seeds, and the real-record lattice fits repeat over all nine records with 3.57 mHz median deviation.

Table 12: Ablation on seed 42 (periodic-component MSE; S3/S5 are the discriminative cases). The last row is an in-house prior-guided VMD baseline (see text), shown here only.

variant	S1	S3	S5	comment
full DTHy-SMP	5.7e-7	9.1e-6	2.0e-2	
w/o compound split	5.7e-7	9.1e-6	2.0e-2	neutral at 30 dB; discriminating at 20 dB (text)
w/o chirp margin	5.7e-7	3.0e-1	2.1e-2	chirp absorbs close tones
w/o joint polish	5.7e-7	1.1e-5	2.0e-2	mild degradation of the pair fit
prior-guided VMD (BIC)	8.5e-3	1.1e-1	6.0e-2	Wiener-update prior; Section 4.2

Table 13: Chirp-margin sensitivity (five seeds per cell): full-structure rate for $\Delta_{\text{chirp}} = 0/20/40$. At $N = 512, 1024, 2048$, a margin of 20 equals 3.21, 2.89 and 2.62 BIC parameters, respectively.

case	N	SNR	0	20	40
S3	512	30	0.00	0.00	0.00
S3	1024	30	0.00	1.00	0.40
S3	2048	30	1.00	1.00	1.00
S3	1024	20	0.40	0.80	0.80
S5	512	30	0.00	0.00	0.00
S5	1024	30	0.00	0.00	0.00
S5	2048	30	0.00	0.00	0.00
S5	1024	20	0.20	0.00	0.00

6. Conclusion

We replaced the spectral compactness objective of VMD-style decomposition with explicit complexity-penalised approximation by parametric symbolic components, and built a pursuit algorithm whose model order, analytic families and parameters are estimated from the data. Across the five identifiable stationary synthetic cases, 20-seed full-structure rates range from 0.65 to 1.00 and matched target errors are below peak-initialised VMD in every paired run. The resolution sweep exposes non-monotone selection failures, and S5, S6 and coloured machinery backgrounds show that budget saturation is a model-adequacy warning rather than proof of a physical decomposition. Within the intended compact-stationary regime, however, the evidence is consistent: the method wins all 20 paired comparisons on each of S1–S4 and S7, the same-library compound search raises close-tone full-structure recovery from 0/20 to 19/20, and the measured-record atoms reduce median harmonic-lattice deviation to 3.57 mHz while returning the component equations directly. Together, these results support symbolic decomposition as a distinct capability within the stated regime, not universal waveform-decomposition superiority.

Data availability

CWRU Bearing Data Center baseline recordings are public. The rotating-machine records are public (Zenodo, DOI:10.5281/zenodo.14388093).

References

[1] N. E. Huang, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454 (1971) (1998) 903–995.

[2] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544.

[3] J. Lian, Z. Liu, H. Wang, X. Dong, Adaptive variational mode decomposition method for signal processing based on mode characteristic, Mech. Syst. Signal Process. 107 (2018) 53–77. doi:10.1016/j.ymssp.2018.01.019.

[4] Q. Chen, J. Chen, X. Lang, L. Xie, N. ur Rehman, H. Su, Self-tuning variational mode decomposition, J. Franklin Inst. 358 (15) (2021) 7825–7862.

[5] N. Rehman, H. Aftab, Multivariate variational mode decomposition, IEEE Trans. Signal Process. 67 (23) (2019) 6039–6052.

[6] M. Cranmer, et al., Discovering symbolic models from deep learning with inductive biases, in: NeurIPS, 2020.

[7] S.-M. Udrescu, M. Tegmark, AI Feynman: a physics-inspired method for symbolic regression, Sci. Adv. 6 (16) (2020).

[8] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85.

[9] R. B. Randall, J. Antoni, Why EMD and similar decompositions are of little benefit for bearing diagnostics, Mech. Syst. Signal Process. 192 (2023) 110207.

[10] Z. Wu, N. E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (1) (2009) 1–41.

[11] M. E. Torres, M. A. Colominas, G. Schlotthauer, P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: ICASSP, 2011, pp. 4144–4147.

[12] M. Nazari, S. M. Sakhaei, Successive variational mode decomposition, Signal Processing 174 (2020) 107610.

[13] M. Nazari, S. M. Sakhaei, Variational mode extraction: a new efficient method to derive respiratory signals from ECG, IEEE J. Biomed. Health Inform. 22 (4) (2018) 1059–1067.

[14] Y. Guo, Z. Zhang, Generalized variational mode decomposition: a multiscale and fixed-frequency decomposition algorithm, IEEE Trans. Instrum. Meas. 70 (2021) 1–13.

[15] S. Chen, X. Dong, Z. Peng, W. Zhang, G. Meng, Nonlinear chirp mode decomposition: a variational method, IEEE Trans. Signal Process. 65 (22) (2017) 6024–6037.

[16] B. Hou, M. Xie, H. Yan, D. Wang, Impulsive mode decomposition, Mech. Syst. Signal Process. 211 (2024) 111227.

[17] B. Hou, M. Xie, Z. Peng, D. Wang, Enspectrumix: novel adaptive methodology for fault components extraction from vibration mixtures, Mech. Syst. Signal Process. 219 (2024) 111611.

[18] A. Wang, P. Qin, X.-M. Sun, Y. Li, An automatic parameter setting variational mode decomposition method for vibration signals, IEEE Trans. Ind. Informat. 20 (2) (2024) 2053–2062.

[19] W. Zhou, Z. Feng, Y. F. Xu, X. Wang, H. Lv, Empirical Fourier decomposition: an accurate signal decomposition method for nonlinear and non-stationary time series analysis, Mech. Syst. Signal Process. 163 (2022) 108155.

[20] K. Zhang, P. Chen, M. Yang, L. Song, Y. Xu, The harmogram: a periodic impulses detection method and its application in bearing fault diagnosis, Mech. Syst. Signal Process. 165 (2022) 108374.

[21] Y. Hua, T. K. Sarkar, Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise, IEEE Trans. Acoust. Speech Signal Process. 38 (5) (1990) 814–824.

[22] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.

[23] M. M. Goodwin, M. Vetterli, Matching pursuit and atomic signal models based on recursive filter banks, IEEE Transactions on Signal Processing 47 (7) (1999) 1890–1902. doi:10.1109/78.771038.

[24] L. Cui, C. Kang, H. Wang, P. Chen, Application of composite dictionary multi-atom matching in gear fault diagnosis, Sensors 11 (6) (2011) 5981–6002. doi:10.3390/s110605981.

[25] L. Cui, N. Wu, W. Wang, C. Kang, Sensor-based vibration signal feature extraction using an improved composite dictionary matching pursuit algorithm, Sensors 14 (9) (2014) 16715–16739. doi:10.3390/s140916715.

[26] S. Mann, S. Haykin, Adaptive chirplet transform: an adaptive generalization of the wavelet transform, Opt. Eng. 31 (6) (1992) 1243–1256. doi:10.1117/12.57676.

[27] S. Mann, S. Haykin, The chirplet transform: physical considerations, IEEE Trans. Signal Process. 43 (11) (1995) 2745–2761. doi:10.1109/78.482123.

[28] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, D. L. Donoho, Morphological component analysis, in: Wavelets XI, Proc. SPIE, Vol. 5914, San Diego, CA, USA, 2005, pp. 209–223. doi:10.1117/12.615237.

[29] M. Elad, J.-L. Starck, P. Querre, D. L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis, Appl. Comput. Harmon. Anal. 19 (2005) 340–358.

[30] G. L. Bretthorst, Bayesian Spectrum Analysis and Parameter Estimation, Springer, 1988.

[31] B. N. Bhaskar, G. Tang, B. Recht, Atomic norm denoising with applications to line spectral estimation, IEEE Trans. Signal Process. 61 (23) (2013) 5987–5999.

[32] G. Tang, B. N. Bhaskar, B. Recht, Near minimax line spectral estimation, IEEE Trans. Inf. Theory 61 (1) (2015) 499–512.

[33] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, Sparse sampling of signal innovations: Theory, algorithms, and performance bounds, IEEE Signal Process. Mag. 25 (2) (2008) 31–40. doi:10.1109/MSP.2007.914998.

[34] C. Soussen, J. Idier, D. Brie, J. Duan, From Bernoulli–Gaussian deconvolution to sparse signal restoration, IEEE Trans. Signal Process. 59 (10) (2011) 4572–4584.

[35] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (2) (1978) 461–464.

[36] J. Rissanen, Modeling by shortest data description, Automatica 14 (5) (1978) 465–471.

[37] D. Wipf, B. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.

[38] W. Wang, the PySDKit contributors, PySDKit: a Python library for signal decomposition algorithms, `github.com/wwhenxuan/PySDKit`, version 0.5.0 (2026).

[39] V. R. Carvalho, vmdpy: Python implementation of variational mode decomposition, PyPI package, `pypi.org/project/vmdpy`, companion paper: V. R. Carvalho, M. F. D. Moraes, A. P. Braga, E. M. A. M. Mendes, Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification, Biomed. Signal Process. Control 62 (2020) 102073, doi:10.1016/j.bspc.2020.102073 (2020).

[40] A. Lucatto Marra, C. Garcia, Vibration signals of different fault conditions in a generic rotating machine, Zenodo dataset (2024). doi:10.5281/zenodo.14388093.
URL `https://zenodo.org/records/14388093`