A Unified Maximum-Likelihood Framework for Optimal Filtering and Subspace Signal Reconstruction

Dowling Wong

Feb 2026

Contents

1 Statement of the Problem

1.1 LAMCAL and cryogenic detector response

1.2 Quasi-particle signal formation

1.3 Simulation implementation

Summary of Section 1.3: Simulation Implementation

1.3.1 Noise Simulation Implementation. The noise simulation is implemented through a dedicated `NoiseGenerator` class, designed to provide frequency-domain colored noise in a compact and modular way. The workflow separates configuration from generation: spectral definitions and normalization are handled during setup, while each call to the generator produces a time-domain trace consistent with the selected power spectral density (PSD). Built-in models include standard noise colors (white, blue, violet, pink, brownian), each corresponding to a simple analytic frequency dependence. The implementation first normalizes the spectral shape so that it controls only how power is distributed across frequency, and then applies discrete FFT scaling to ensure the total noise power matches the configured level. Noise traces are synthesized in frequency space by assigning magnitudes from the PSD and random phases, followed by an inverse real FFT to obtain the time-domain waveform. A custom PSD mode allows user-supplied spectra to be interpolated and used directly. Overall, the method ensures physically consistent mapping between continuous PSD definitions and discrete sampled waveforms.

1.3.2 TraceSimulator Architecture and Signal-Formation Workflow. The `TraceSimulator` package follows a modular architecture in which detector response maps, energy partitioning, geometry handling, and noise generation are clearly separated. For each simulated event, deposited recoil energy is first partitioned into phonon, triplet, UV, and IR components using the `DELightSignalFormation` model. These energy components are converted into discrete quanta and distributed across detector channels according to position-dependent efficiencies derived from response maps. Arrival times are sampled from precomputed transport templates, histogrammed onto digitizer bins, and converted into ADC-equivalent amplitudes.

The quasi-particle (QP) signal is formed primarily through the phonon branch in vacuum channels. The number of phonon quanta is computed from the phonon energy component, scaled by the local phonon collection efficiency (PCE), and fluctuated with Poisson statistics. Arrival times are sampled from the phonon arrival-time distribution, histogrammed in time, scaled by detector gain and energy-to-ADC conversion factors, and finally shaped with the detector pulse template (shifted-sum convolution). Conceptually, the chain is:

Energy → Quanta → Channel Acceptance (PCE) → Arrival-Time Sampling → Histogramming → Template S

After signal construction, colored noise from the `NoiseGenerator` is added, and optional digitization effects are applied. A standard simulator produces baseline waveforms, while an extended long-trace version additionally provides clean/noisy pairs and signal masks for analysis studies. The overall design emphasizes clear separation between physics modeling, geometry handling, and stochastic waveform generation.

Energy-to-Quanta Conversion. After energy partitioning by `DELightSignalFormation`, the deposited recoil energy E is decomposed into four components

$$E = E_{\mathrm{ph}} + E_{\mathrm{tr}} + E_{\mathrm{uv}} + E_{\mathrm{ir}},$$

corresponding to phonon, triplet, UV, and IR channels.

Each component is converted into discrete quanta using the corresponding mean excitation energy scale. In the implementation this is approximated as

$$N_{\mathrm{uv}} = \left\lfloor \frac{E_{\mathrm{uv}}}{E_{\mathrm{UV}}} \right\rfloor, \qquad N_{\mathrm{tr}} = \left\lfloor \frac{E_{\mathrm{tr}}}{E_{\mathrm{triplet}}} \right\rfloor,$$

$$N_{\mathrm{ir}} = \left\lceil \frac{E_{\mathrm{ir}}}{E_{\mathrm{IR,avg}}} \right\rceil, \qquad N_{\mathrm{ph}} \approx \frac{E_{\mathrm{ph}}}{\epsilon_{\mathrm{ph}}},$$

where in the current code convention $\epsilon_{\mathrm{ph}} \simeq 1$ meV per phonon quantum.

Thus, the continuous deposited energy is mapped onto integer excitation numbers, which are the fundamental stochastic objects propagated through the detector model.

Channel Distribution and Detection Statistics. For optical (UV/IR) and triplet channels, quanta are distributed across LAMCAL channels according to position-dependent light-collection efficiencies (LCE) or triplet-collection efficiencies (TCE). This is implemented as a multinomial draw conditioned on the local efficiency vector $\{p_i(x, y, z)\}$ with

$$\sum_i p_i(x, y, z) \leq 1.$$

For phonons (relevant for QP formation in vacuum channels), the expected number of detected phonons in channel i is

$$\lambda_i(x, y, z) = \mathrm{PCE}_i(x, y, z)\, N_{\mathrm{ph}},$$

where PCE_i is the phonon collection efficiency from the response map. The observed phonon count is then drawn as

$$N_{\mathrm{ph},i}^{\mathrm{obs}} \sim \mathrm{Poisson}(\lambda_i)\,,$$

explicitly modeling counting fluctuations at the quasi-particle level.

Arrival-Time Sampling and Histogram Formation. For each detected quantum, an arrival time is sampled from the precomputed transport distribution

$$t_{\mathrm{ph},i}^{(k)} \sim P_{\mathrm{arr},i}(t \mid x, y, z),$$

obtained from the phonon arrival-time maps. The resulting discrete arrival times are histogrammed into digitizer bins of width $\Delta t = 1/f_s$:

$$H_i[n] = \sum_k \mathbf{1}\Big(t_{\mathrm{ph},i}^{(k)} \in [n\Delta t, (n+1)\Delta t)\Big).$$

This produces a discrete time-series representation of deposited phonon quanta per channel.

Template Shaping and QP Pulse Formation. The histogrammed counts are converted into ADC-equivalent amplitudes via gain and calibration constants G_i:

$$A_i[n] = G_i\, H_i[n].$$

Detector response shaping is then applied through convolution with the channel-specific pulse template $T_i[n]$:

$$s_i[n] = (A_i * T_i)[n] = \sum_m A_i[m]\, T_i[n-m].$$

Conceptually, the QP signal formation chain can therefore be written as

$$E_{\mathrm{ph}} \to N_{\mathrm{ph}} \to \lambda_i \to N_{\mathrm{ph},i}^{\mathrm{obs}} \to H_i[n] \to A_i[n] \to s_i[n].$$

Finally, colored noise $\eta_i[n] \sim \mathcal{N}(0, \Sigma)$ generated by `NoiseGenerator` is added:

$$x_i[n] = s_i[n] + \eta_i[n],$$

optionally followed by digitization to integer ADC counts.

This formalizes the simulation as a sequence of stochastic mappings from continuous recoil energy to discrete quasi-particle statistics, then to deterministic template-shaped waveforms, and finally to noisy measured traces.

2 Optimal Filter

2.1 Basic Optimal Filter

Following Kurinsky Appendix E.1, assume one pulse per trace with known template and stationary Gaussian noise in frequency space:

$$X_k = AS_k + N_k, \qquad \mathbb{E}[N_k] = 0, \qquad \mathbb{E}[|N_k|^2] = J_k. \tag{1}$$

For uncorrelated frequency bins, maximizing the Gaussian likelihood is equivalent to minimizing

$$\chi^2(A) = \sum_k \frac{|X_k - A S_k|^2}{J_k}. \tag{2}$$

Expanding the quadratic form,

$$\chi^2(A) = \sum_k \frac{X_k^* X_k - A S_k X_k^* - A^* S_k^* X_k + |A|^2 |S_k|^2}{J_k}. \tag{3}$$

Taking the Wirtinger derivative with respect to A^* and setting it to zero gives the normal equation

$$\frac{\partial \chi^2}{\partial A^*} = 0 \implies \sum_k \frac{S_k^* X_k}{J_k} = \hat{A} \sum_k \frac{|S_k|^2}{J_k}, \tag{4}$$

hence

$$\hat{A} = \frac{\sum_k \dfrac{S_k^* X_k}{J_k}}{\sum_k \dfrac{|S_k|^2}{J_k}}. \tag{5}$$

Define the unnormalized optimal filter and normalization

$$\Phi_k = \frac{S_k^*}{J_k}, \qquad N_\Phi = \sum_k \Phi_k S_k = \sum_k \frac{|S_k|^2}{J_k}, \tag{6}$$

and the normalized filter

$$\Phi_k^{(0)} = \frac{\Phi_k}{N_\Phi}. \tag{7}$$

Then the estimator is the filtered projection

$$\hat{A} = \sum_k \Phi_k^{(0)} X_k. \tag{8}$$

This is the discrete counterpart of Kurinsky Eq. (E.9). If A is constrained to be real, use $\mathrm{Re}\,\hat{A}$ after filtering.

2.2 Optimal Filter with Time Shift

Following Kurinsky Appendix E.2, include an unknown arrival-time shift t_0 via

$$S_k(t_0) = e^{-i\omega_k t_0} S_k, \qquad X_k = A\, S_k(t_0) + N_k. \tag{9}$$

The weighted objective is

$$\chi^2(A, t_0) = \sum_k \frac{|X_k - A e^{-i\omega_k t_0} S_k|^2}{J_k}. \tag{10}$$

For fixed t_0, the stationarity equation $\partial\chi^2/\partial A^* = 0$ yields

$$\hat{A}(t_0) = \frac{\sum_k \dfrac{e^{i\omega_k t_0} S_k^* X_k}{J_k}}{\sum_k \dfrac{|S_k|^2}{J_k}} = \sum_k e^{i\omega_k t_0} \Phi_k^{(0)} X_k. \tag{11}$$

Therefore $\hat{A}(t_0)$ is the inverse DFT of the filtered spectrum, matching the Kurinsky E.2 interpretation.

To profile out amplitude, define

$$\chi_0^2 = \sum_k \frac{|X_k|^2}{J_k}, \qquad B(t_0) = \sum_k \frac{e^{i\omega_k t_0} S_k^* X_k}{J_k}. \tag{12}$$

Then

$$\chi^2(A, t_0) = \chi_0^2 - A^* B(t_0) - A B^*(t_0) + |A|^2 N_\Phi. \tag{13}$$

At $A = \hat{A}(t_0) = B(t_0)/N_\Phi$,

$$\chi^2(t_0) = \chi_0^2 - \frac{|B(t_0)|^2}{N_\Phi} = \chi_0^2 - |\hat{A}(t_0)|^2 N_\Phi. \tag{14}$$

Hence the optimal delay is

$$\hat{t}_0 = \arg\min_{t_0} \chi^2(t_0) = \arg\max_{t_0} |\hat{A}(t_0)|. \tag{15}$$

For real-valued traces with Hermitian-symmetric spectra, this reduces to the usual "maximize filtered amplitude in time" criterion used by Kurinsky.

3 Principal Component Analysis

3.1 Classic PCA

Following Bailey, let the mean-subtracted dataset be

$$\mathbf{X} = [\mathbf{X}_{:,1}, \ldots, \mathbf{X}_{:,M}] \in \mathbb{C}^{N \times M}, \qquad \frac{1}{M} \sum_{j=1}^{M} \mathbf{X}_{:,j} = 0. \tag{16}$$

Classic PCA seeks a rank-K model

$$\mathbf{X} \approx \mathbf{U}\mathbf{C}, \qquad \mathbf{U} \in \mathbb{C}^{N \times K},\ \mathbf{C} \in \mathbb{C}^{K \times M}, \qquad \mathbf{U}^\dagger \mathbf{U} = \mathbf{I}_K, \tag{17}$$

by minimizing the reconstruction error

$$\chi^2_{\mathrm{PCA}} = \|\mathbf{X} - \mathbf{U}\mathbf{C}\|_F^2 = \sum_{j=1}^{M} \|\mathbf{X}_{:,j} - \mathbf{U} c_j\|_2^2. \tag{18}$$

For fixed $\mathbf{U}$, the least-squares coefficients are

$$c_j = \mathbf{U}^\dagger \mathbf{X}_{:,j}, \qquad \mathbf{C} = \mathbf{U}^\dagger \mathbf{X}. \tag{19}$$

Substituting back gives

$$\chi^2_{\mathrm{PCA}} = \|\mathbf{X}\|_F^2 - \|\mathbf{U}^\dagger \mathbf{X}\|_F^2, \tag{20}$$

so minimizing χ^2_{PCA} is equivalent to maximizing captured variance. Therefore $\mathbf{U}$ is given by the top-K eigenvectors of the sample covariance

$$\mathbf{R}_x = \frac{1}{M} \mathbf{X}\mathbf{X}^\dagger. \tag{21}$$

For each event, the reconstruction is $\hat{\mathbf{X}}_{:,j} = \mathbf{U}c_j$; if the global mean was removed, add $\boldsymbol{\mu}$ back after reconstruction.

Statistical optimality and noise assumptions

It is important to emphasize that *standard* PCA admits a maximum-likelihood (ML) interpretation only under a very restrictive noise model. Specifically, consider the generative model

$$\mathbf{x} = \mathbf{s} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma). \tag{22}$$

Standard PCA corresponds to ML estimation of a low-dimensional signal subspace only when the noise covariance is proportional to the identity,

$$\Sigma = \sigma^2 \mathbf{I}. \tag{23}$$

In this case, minimizing the Euclidean reconstruction error is equivalent to minimizing the negative log-likelihood.

Any deviation from this assumption—such as colored noise, frequency-dependent noise power spectral density, or channel-dependent noise variances—breaks the optimality of standard PCA. Under such conditions, Euclidean-distance-based PCA no longer corresponds to a statistically optimal estimator.

Expectation–maximization PCA (EMPCA) restores maximum-likelihood optimality by explicitly incorporating the inverse noise covariance Σ^{-1} into the objective function.

3.2 Expectation maximization PCA

Bailey's EMPCA formulation solves the same low-rank problem iteratively and avoids explicitly forming the full covariance matrix when N is large. Given current basis $\mathbf{U}^{(r)}$, iteration $r \to r+1$ alternates:

E-step (solve coefficients). For each observation $\mathbf{X}_{:,j}$,

$$c_j^{(r+1)} = \arg\min_c \|\mathbf{X}_{:,j} - \mathbf{U}^{(r)} c\|_2^2 = \left((\mathbf{U}^{(r)})^\dagger \mathbf{U}^{(r)} \right)^{-1} (\mathbf{U}^{(r)})^\dagger \mathbf{X}_{:,j}. \tag{24}$$

If $\mathbf{U}^{(r)}$ is orthonormal, this reduces to $c_j^{(r+1)} = (\mathbf{U}^{(r)})^\dagger \mathbf{X}_{:,j}$.

M-step (solve basis). With coefficients fixed,

$$\mathbf{U}^{(r+1)} = \arg\min_{\mathbf{U}} \|\mathbf{X} - \mathbf{U}\mathbf{C}^{(r+1)}\|_F^2 = \mathbf{X}(\mathbf{C}^{(r+1)})^\dagger \left(\mathbf{C}^{(r+1)}(\mathbf{C}^{(r+1)})^\dagger\right)^{-1}. \quad (25)$$

After each M-step, columns of $\mathbf{U}^{(r+1)}$ are re-orthonormalized (e.g. QR), and iterations continue until $\Delta\chi^2$ or parameter updates are below tolerance.

Equivalent to Bailey's one-component deflation view, one may solve components sequentially and update

$$\mathbf{X} \leftarrow \mathbf{X} - u_m c_m, \quad (26)$$

before solving the next component, where u_m is the mth column of $\mathbf{U}$ and c_m is the mth row of $\mathbf{C}$.

3.3 EMPCA in frequency domain

Yu's microcalorimeter setting motivates a frequency-domain variant that is robust to random time shifts and non-white noise. For event j in Fourier space,

$$X_k^{(j)} = A_j e^{-i\omega_k t_{0,j}} S_k + N_k^{(j)}. \quad (27)$$

Direct PCA on $\{X_k^{(j)}\}$ is problematic because the effective coefficient $A_j e^{-i\omega_k t_{0,j}}$ depends on frequency bin k.

Following Yu, write $X_k^{(j)} = |X_k^{(j)}| e^{i\phi_k^{(j)}}$ and use phase differences

$$\Delta\phi_k^{(j)} = \phi_k^{(j)} - \phi_{k-1}^{(j)}. \quad (28)$$

Under a time shift, $\phi_k^{(j)} \mapsto \phi_k^{(j)} + \omega_k t_{0,j}$, hence

$$\Delta\phi_k^{(j)} \mapsto \Delta\phi_k^{(j)} + (\omega_k - \omega_{k-1}) t_{0,j}. \quad (29)$$

For uniform frequency spacing $\Delta\omega = \omega_k - \omega_{k-1}$, the shift term is independent of k, so the transformed coefficients behave as a single complex factor per event. This restores a linear subspace model in transformed frequency space:

$$\mathbf{X} \approx \mathbf{U}\mathbf{C}. \quad (30)$$

Here $\mathbf{X}$ denotes the transformed (phase-difference) frequency-domain data matrix.

To handle colored/heteroskedastic noise, use Bailey-style weighted EMPCA with Yu's weighted likelihood:

$$\chi^2_{\text{EMPCA}} = \sum_{k=1}^{N}\sum_{j=1}^{M} (\mathbf{W}_{\text{EMPCA}})_{k,j} \left|(\mathbf{X})_{k,j} - [\mathbf{U}\mathbf{C}]_{k,j}\right|^2. \quad (31)$$

Let $\mathbf{W}_j = \text{diag}(w_{1j}, \ldots, w_{Nj})$ for column j. The weighted E-step is

$$c_j = \arg\min_c \|\mathbf{W}_j^{1/2}(\mathbf{X}_{:,j} - \mathbf{U}c)\|_2^2, \quad (32)$$

which yields normal equations

$$(\mathbf{U}^\dagger \mathbf{W}_j \mathbf{U}) c_j = \mathbf{U}^\dagger \mathbf{W}_j \mathbf{X}_{:,j}. \tag{33}$$

For the M-step, with $\mathbf{C}$ fixed, each frequency row $\mathbf{X}_{k,:}$ solves

$$u_k^\dagger = \arg\min_{u^\dagger} \|\mathbf{W}_k^{1/2}(\mathbf{X}_{k,:} - u^\dagger \mathbf{C})\|_2^2, \tag{34}$$

with $\mathbf{W}_k = \mathrm{diag}(w_{k1}, \ldots, w_{kM})$, yielding

$$(\mathbf{C}\mathbf{W}_k\mathbf{C}^\dagger) u_k = \mathbf{C}\mathbf{W}_k \mathbf{X}_{k,:}^\dagger. \tag{35}$$

In the frequency domain, the EMPCA objective takes the form

$$\chi^2 = \sum_f \left| x(f) - \sum_{m=1}^{K} c_m u_m(f) \right|^2 \frac{1}{J(f)}, \tag{36}$$

where $J(f)$ is the noise power spectral density.

This expression makes explicit that EMPCA performs subspace estimation in a whitened space. Equivalently, one may view EMPCA as performing PCA after whitening the data by $\Sigma^{-1/2}$. Both interpretations are mathematically identical and correspond to a dual PCA formulation with diagonal noise covariance.

This is the Bailey weighted normal equation in row form and is equivalent to Yu's E/M linear systems in transformed frequency space.

This section matches the common structure in Yu and Bailey: transform to a linear representation, fit a low-rank model, and solve coefficients/basis by alternating weighted least-squares updates.

Dual and frequency-domain interpretation

PCA admits an equivalent *dual* formulation in which the optimization is expressed entirely in terms of inner products. This perspective is particularly natural in the frequency domain.

For stationary noise, the covariance matrix Σ is diagonal in the Fourier basis, with diagonal entries given by the noise power spectral density $J(f)$. In this basis, whitening corresponds to diagonalization of Σ and rescaling by $\Sigma^{-1/2}$.

From this viewpoint, frequency-domain PCA can be interpreted as dual PCA expressed in the Fourier basis. When noise between frequency bins is uncorrelated, the inverse covariance Σ^{-1} acts element-wise in frequency space. As a result, weighted inner products reduce to Hadamard (element-wise) products.

Therefore, the Hadamard-product implementation used in this work is not an approximation or heuristic, but the exact dual-PCA formulation under a diagonal noise covariance.

3.4 Multi-channel joint EMPCA (progressing)

4 Reconstruction of EMPCA

4.1 Signal amplitude reconstruction

4.1.1 Euclidean norm for height estimator

4.1.2 Complex coefficient reconstruction

5 PCA-Assisted OF

PCA-assisted optimal filtering (POF) uses the first principal component as the OF template instead of a hand-chosen mean pulse. With the notation of Section 3, let $\mathbf{u}_1$ be the first column of $\mathbf{U}$. Because PCA eigenvectors are defined up to a global complex phase, align $\mathbf{u}_1$ before using it as an OF template. Using the average spectrum

$$\bar{X}_k = \frac{1}{M}\sum_{j=1}^{M} X_k^{(j)}, \tag{37}$$

choose phase α by maximizing weighted correlation:

$$\alpha^\star = \arg\max_{\alpha} \mathrm{Re}\left[\sum_k \frac{\left(e^{i\alpha}[\mathbf{u}_1]_k\right)^* \bar{X}_k}{J_k}\right]. \tag{38}$$

Then define the POF template

$$S_k^{\mathrm{POF}} = e^{i\alpha^\star}[\mathbf{u}_1]_k. \tag{39}$$

Substitute S_k^{POF} into the OF-with-delay equations from Section 2:

$$\hat{A}_{\mathrm{POF}}(t_0) = \frac{\sum_k \frac{e^{i\omega_k t_0}(S_k^{\mathrm{POF}})^* X_k}{J_k}}{\sum_k \frac{|S_k^{\mathrm{POF}}|^2}{J_k}}, \qquad \hat{t}_0 = \arg\max_{t_0} \left|\hat{A}_{\mathrm{POF}}(t_0)\right|. \tag{40}$$

As in Yu, this keeps the OF statistical weighting while using a data-driven template that better tracks pulse-shape variation.

For comparison studies, Yu calibrates estimators with an affine map on calibration data:

$$A_{\mathrm{cal}} = \beta_0 + \beta_1 \hat{A}_{\mathrm{POF}}. \tag{41}$$

In his simulated example, the tuned POF estimator improved resolution relative to standard OF (reported $\sigma_{\mathrm{POF}} \approx 0.0393$ vs. $\sigma_{\mathrm{OF}} \approx 0.0469$).

Practical implementation recipe:

1. Train PCA/EMPCA on representative pulses (optionally per energy band).
2. Extract $\mathbf{u}_1$ and determine $\alpha^\star$ (global phase/sign alignment).

3. Build S_k^{POF} and run the Section 2 sliding/scanning OF equations.
4. Calibrate $\hat{A}_{\mathrm{POF}}$ and validate resolution on held-out data.
5. If needed, retrain by energy band and interpolate neighboring templates.

6 Equivalence Between Optimal Filtering and PCA Subspace Estimators

Notation and Problem Setup

We explicitly separate indices as follows:

- Event index: $i = 1, \ldots, N$
- Feature (time/frequency bin) index: $j = 1, \ldots, d$
- Channel index: $c = 1, \ldots, C$
- Frequency index (if working in Fourier domain): f

For clarity, we first treat the single-channel case and suppress the channel index c.

For each event i, we observe a vector

$$x_i \in \mathbb{C}^d,$$

representing either a time-domain trace or frequency-domain coefficients.

We assume the generative model

$$x_i = a_i s + n_i, \tag{42}$$

where

- $s \in \mathbb{C}^d$ is a fixed signal template,
- $a_i \in \mathbb{R}$ (or $\mathbb{C}$ in frequency domain) is the event amplitude,
- $n_i \sim \mathcal{N}(0, \Sigma)$ is zero-mean Gaussian noise with known covariance $\Sigma \in \mathbb{C}^{d \times d}$.

Define the whitening transform

$$\tilde{x}_i = \Sigma^{-1/2} x_i, \qquad \tilde{s} = \Sigma^{-1/2} s.$$

Stacking events row-wise gives

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1^\dagger \\ \vdots \\ \tilde{x}_N^\dagger \end{bmatrix} \in \mathbb{C}^{N \times d}.$$

We now distinguish three levels of equivalence.

Definitions of Equivalence

Estimator equivalence. Two estimators are equivalent if they produce identical amplitude estimates a_i for all events under identical preprocessing and weighting.

Subspace equivalence. Two methods are subspace-equivalent if they recover the same signal subspace $\mathcal{S} \subset \mathbb{C}^d$ up to unitary rotation.

Amplitude equivalence. Two methods are amplitude-equivalent if, after projection onto the same subspace basis and with consistent normalization, their reconstructed coefficients coincide.

Main Theorem (Rank-1 Equivalence)

Theorem 6.1 (Rank-1 OF–EMPCA equivalence)**.** *Assume:*

1. *The generative model $x_i = a_i s + n_i$ with fixed template s.*
2. *Gaussian noise $n_i \sim \mathcal{N}(0, \Sigma)$ with known covariance.*
3. *Identical whitening and preprocessing conventions.*
4. *The true signal subspace is one-dimensional.*

Then:

1. *The rank-1 EMPCA subspace equals* $\mathrm{span}(\tilde{s})$.
2. *The resulting EMPCA amplitude estimator equals the generalized least-squares (Optimal Filter) estimator.*
3. *Therefore, OF and rank-1 EMPCA are estimator-equivalent under these assumptions.*

Proof. In whitened space,

$$\tilde{x}_i = a_i \tilde{s} + \tilde{n}_i, \quad \tilde{n}_i \sim \mathcal{N}(0, I).$$

The population covariance is

$$\mathbb{E}[\tilde{x}_i \tilde{x}_i^\dagger] = \mathbb{E}[a_i^2]\, \tilde{s}\tilde{s}^\dagger + I.$$

This matrix is a rank-1 signal perturbation of identity. By standard spiked-covariance theory, the leading eigenvector is proportional to $\tilde{s}$.

Thus rank-1 EMPCA recovers

$$u_1 \propto \tilde{s}.$$

Projection of $\tilde{x}_i$ onto u_1 gives

$$\hat{a}_i^{(\mathrm{EMPCA})} = u_1^\dagger \tilde{x}_i.$$

Since u_1 is proportional to $\tilde{s}$, and normalization can be chosen such that

$$u_1 = \frac{\tilde{s}}{\|\tilde{s}\|},$$

we obtain

$$\hat{a}_i^{(\mathrm{EMPCA})} = \frac{\tilde{s}^\dagger \tilde{x}_i}{\|\tilde{s}\|}.$$

Undoing whitening yields

$$\hat{a}_i = \frac{s^\dagger \Sigma^{-1} x_i}{s^\dagger \Sigma^{-1} s},$$

which is precisely the generalized least-squares (Optimal Filter) estimator.

Therefore, subspace equivalence implies amplitude equivalence, which implies estimator equivalence. □

6.1 Unified Maximum-Likelihood Formulation

Let $x \in \mathbb{C}^d$ denote one trace in frequency-domain representation, with noise model

$$x = s + n, \qquad n \sim \mathcal{N}_{\mathbb{C}}(0, \Sigma), \qquad \Sigma \succ 0. \tag{43}$$

The Gaussian negative log-likelihood (up to constants) is

$$\chi^2(\hat{s}) = (x - \hat{s})^\dagger \Sigma^{-1} (x - \hat{s}). \tag{44}$$

Optimal Filtering (OF). For a known template s_0 and unknown scalar amplitude A, $\hat{s} = A s_0$:

$$\hat{A}_{\mathrm{OF}} = \frac{s_0^\dagger \Sigma^{-1} x}{s_0^\dagger \Sigma^{-1} s_0}. \tag{45}$$

EMPCA. For a k-dimensional basis $U = [u_1, \ldots, u_k] \in \mathbb{C}^{d \times k}$:

$$\hat{s} = Uc, \qquad \hat{c} = \arg\min_c (x - Uc)^\dagger \Sigma^{-1} (x - Uc), \tag{46}$$

hence

$$\hat{c} = (U^\dagger \Sigma^{-1} U)^{-1} U^\dagger \Sigma^{-1} x. \tag{47}$$

6.2 Exact Degeneracy of EMPCA to OF at $k = 1$

Theorem 6.2 (Rank-1 equivalence)**.** *Assume:*

1. $k = 1$,
2. *the EMPCA basis spans the template direction, i.e.* $u = \gamma s_0$ *for some* $\gamma \neq 0$,
3. *OF and EMPCA use the same weighting* Σ^{-1},
4. *preprocessing/alignment conventions are matched.*

Then EMPCA and OF yield identical reconstructed signal $\hat{s}$*, and therefore identical weighted residual objective values.*

Proof. For rank-1 EMPCA:

$$\hat{c} = \frac{u^\dagger \Sigma^{-1} x}{u^\dagger \Sigma^{-1} u}. \tag{48}$$

With $u = \gamma s_0$:

$$\hat{c} = \frac{\gamma^* s_0^\dagger \Sigma^{-1} x}{|\gamma|^2 s_0^\dagger \Sigma^{-1} s_0} = \frac{1}{\gamma} \frac{s_0^\dagger \Sigma^{-1} x}{s_0^\dagger \Sigma^{-1} s_0} = \frac{\hat{A}_{\mathrm{OF}}}{\gamma}. \tag{49}$$

Hence

$$\hat{s}_{\mathrm{EMPCA}} = \hat{c}\, u = \frac{\hat{A}_{\mathrm{OF}}}{\gamma}(\gamma s_0) = \hat{A}_{\mathrm{OF}} s_0 = \hat{s}_{\mathrm{OF}}. \tag{50}$$

Therefore $(x - \hat{s})^\dagger \Sigma^{-1} (x - \hat{s})$ is identical for both estimators. □

Practical note on scale/sign/phase. The EMPCA basis has gauge freedom ($u \mapsto \alpha u$, $c \mapsto c/\alpha$), so amplitudes can differ by a global affine scale convention while reconstructed signals remain equivalent.

6.3 When $k > 1$: Why EMPCA Can Improve Residual Fit

In whitened coordinates $\tilde{x} = \Sigma^{-1/2} x$, OF is projection onto a 1D subspace S_1, while rank-k EMPCA is projection onto a kD subspace U_k:

$$\hat{\tilde{x}}_{\mathrm{OF}} = P_{S_1} \tilde{x}, \qquad \hat{\tilde{x}}_k = P_{U_k} \tilde{x}. \tag{51}$$

If $S_1 \subseteq U_k$, then

$$\|\tilde{x} - P_{U_k} \tilde{x}\|_2^2 \leq \|\tilde{x} - P_{S_1} \tilde{x}\|_2^2, \tag{52}$$

equivalently

$$\chi^2_{\mathrm{EMPCA},k} \leq \chi^2_{\mathrm{OF}}. \tag{53}$$

So EMPCA with larger subspace can model shape variation beyond a fixed-template rank-1 model.

Metric	Value
$\max \lvert a_{\mathrm{OF}} - a_{\text{template GLS}} \rvert$	5.5339×10^{-4}
Weighted subspace cosine $\rho_w(U, s)$	0.9999999655
Amplitude correlation r	0.9942010763
Median relative error (origin-fit)	9.6782×10^{-5}
Residual KS statistic D	0.0302891
Residual KS p-value	0.2702893

Table 1: Real-data OF vs rank-1 EMPCA verification metrics.

6.4 Numerical Verification on Real K-alpha Data

Using the strict real-data notebook setup (same baseline correction, same PSD weighting, same train/test split):

Interpretation. These results provide strong empirical support that OF and rank-1 EMPCA are functionally equivalent on this dataset under matched assumptions.

7 EMPCA as a Noise-Aware Linear Autoencoder

7.1 Setup, assumptions, and notation

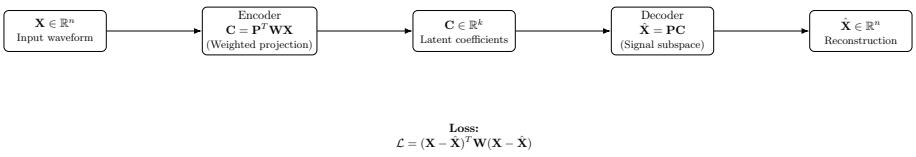

Figure 1: EMPCA as a noise-aware linear autoencoder.

7.2 Whitened complex formulation

Define whitened data $\tilde{X} \in \mathbb{C}^{N \times d}$ (rows are events), with $\tilde{X} = X\Sigma^{-1/2}$ in matrix notation. Use a tied linear autoencoder with decoder $W \in \mathbb{C}^{d \times k}$ and $W^\dagger W = I$:

$$Z = \tilde{X}W, \qquad \hat{\tilde{X}} = ZW^\dagger = \tilde{X}WW^\dagger. \tag{54}$$

The training objective is

$$\min_{W^\dagger W = I} \ \|\tilde{X} - \tilde{X}WW^\dagger\|_F^2. \tag{55}$$

7.3 Equivalence to Weighted PCA / EMPCA Objective

Theorem 7.1 (Linear AE in whitened space equals PCA projector)**.** *The minimizer of* $\|\tilde{X} - \tilde{X}WW^\dagger\|_F^2$ *under* $W^\dagger W = I$ *is given by the top-*k *right singular vectors of* $\tilde{X}$*.*

Proof. Expand:

$$\|\tilde{X} - \tilde{X}WW^\dagger\|_F^2 = \mathrm{Tr}\Big((\tilde{X} - \tilde{X}WW^\dagger)^\dagger(\tilde{X} - \tilde{X}WW^\dagger)\Big) \tag{56}$$

$$= \|\tilde{X}\|_F^2 - \mathrm{Tr}(W^\dagger \tilde{X}^\dagger \tilde{X} W). \tag{57}$$

So minimizing reconstruction error is equivalent to maximizing $\mathrm{Tr}(W^\dagger \tilde{X}^\dagger \tilde{X} W)$ with orthonormal columns. By Rayleigh-Ritz, optimum W spans the top-k eigenspace of $\tilde{X}^\dagger \tilde{X}$, i.e. top-$k$ right singular vectors of $\tilde{X}$. □

Connection to EMPCA. EMPCA with weight Σ^{-1} minimizes the same Gaussian ML objective as projection in whitened space. Therefore, at convergence, rank-k EMPCA and rank-k complex linear AE should recover the same subspace (up to unitary rotation/gauge).

7.4 Native weighted formulation

7.5 Bridge theorem

7.6 Numerical Equivalence Results

From the improved equivalence notebook (complex-whitened primary test):

Metric	Value
Principal-angle cosine (EMPCA vs exact complex AE)	0.9999999967
Principal angle (deg)	0.00463536°
Residual KS statistic D	0.00137678
Residual KS p-value	1.0
Relative mean residual difference	8.17299×10^{-7}

Table 2: Primary EMPCA vs exact complex linear AE equivalence metrics (real K-alpha split).

Diagnostic note. If one compares complex EMPCA to a real-feature rank-k model directly, apparent mismatch can increase (e.g. larger KS gap). This is a representation mismatch, not a contradiction of complex-whitened equivalence.

7.7 Scope and Limits of the Claim

The equivalence claim in this chapter is rigorous for:

- linear subspace models,
- Gaussian noise likelihood,
- matched whitening/weighting and preprocessing conventions,
- identical rank k.

It does not automatically extend to nonlinear autoencoders or mismatched objectives.

8 Numerical performance of EMPCA

whitenoise, MMC noise and pink noise

9 Waveform reconstruction with EMPCA

10 Computation Enhancement of PCA

1. Sliding-window inference (for the next paper): fixed eigenvectors with fast coefficient solves; output amplitude, χ^2, and likelihood traces.
2. Optimizing EMPCA with cuBLAS / batched GEMM: higher ROI for acceleration. Example targets include batched linear solves ($Ac = b$), fused weighting/projection/reduction kernels, and improved data-flow/latency management.

Appendix A List of Symbols

This section summarizes the notation used throughout this work, unifying Optimal Filtering (OF), Principal Component Analysis (PCA), and Expectation–Maximization PCA (EMPCA) in both time and frequency domains.

Category	**Symbol**	**Meaning**
Indices	i	Discrete time-sample index within a trace, $i = 0, \ldots, N-1$.
	k	Frequency-bin index of the discrete Fourier transform.
	j	Trace (event) index in a dataset of size M.
	m	Principal component index, $m = 1, \ldots, K$.
Time / frequency	$t,\ \omega$	Continuous time and angular frequency, $\omega = 2\pi f$.
	$\Delta\omega$	Frequency-bin spacing, $\Delta\omega = \omega_k - \omega_{k-1}$.
Measured data	$x(t),\ x_i$	Measured detector trace in the time domain.
	$X(f),\ X_k$	Fourier transform of the measured trace.
	$\mathbf{x}$	Vectorized single-event trace (time or frequency representation, depending on context).
	$\bar{X}_k$	Mean spectrum over events, $\bar{X}_k = \frac{1}{M}\sum_{j=1}^{M} X_k^{(j)}$.
	$\phi_k,\ \Delta\phi_k$	Phase of X_k and phase difference $\Delta\phi_k = \phi_k - \phi_{k-1}$ used in shift-invariant frequency transforms.
Signal model	$s(t),\ s_i$	Signal template (pulse shape) in time domain.
	$S(f),\ S_k$	Fourier transform of the signal template.
	$\mathbf{s},\ \hat{\mathbf{s}}$	Vectorized signal template and reconstructed signal for a single event.
	S_k^{POF}	PCA-assisted OF template built from the aligned first principal component.
	A	Signal amplitude (pulse height).
	t_0	Signal arrival time (time offset).
	$\alpha^\star$	Global phase-alignment parameter used to fix the phase ambiguity of the first principal component.
Noise	$n(t),\ N(f)$	Additive detector noise in time and frequency domains.

Category	Symbol	Meaning
	$\mathbf{n}$	Vectorized additive noise for a single event.
	Σ	Noise covariance matrix; for stationary noise, Σ is diagonal in frequency space and related to J_k.
	$J(f),\ J_k$	Noise power spectral density (PSD).
	$W(f)$	Inverse-noise weight, $W(f) = 1/J(f)$.
Optimal Filtering	$\chi^2(A, t_0)$	Weighted least-squares objective defining the OF maximum-likelihood estimator.
	$\Phi_k,\ \Phi_k^{(0)}$	Unnormalized and normalized optimal-filter kernels.
	$N_\Phi,\ B(t_0)$	Filter normalization and delayed matched-filter response used in profiled $\chi^2(t_0)$.
	$\beta_0,\ \beta_1$	Affine calibration coefficients mapping raw estimator output to calibrated amplitude.
Dataset (PCA)	$\mathbf{X} \in \mathbb{C}^{N\times M}$	Dataset matrix whose columns are individual traces.
	$\boldsymbol{\mu}$	Mean trace of the dataset.
	$\tilde{\boldsymbol{\mu}},\ \mathbf{R}_{\tilde{s}}$	Whitened mean signal and corresponding covariance matrix used in the PCA/OF equivalence discussion.
PCA / EMPCA	$\mathbf{U} \in \mathbb{C}^{N\times K}$	Matrix of principal components (basis vectors).
	$\mathbf{C} \in \mathbb{C}^{K\times M}$	Coefficient matrix for PCA/EMPCA reconstruction.
	$\hat{\mathbf{X}}$	Low-rank reconstruction, $\hat{\mathbf{X}} = \boldsymbol{\mu}\mathbf{1}^\top + \mathbf{UC}$.
	$\mathbf{W}_{\mathrm{EMPCA}}$	Element-wise weight matrix used in weighted PCA / EMPCA (distinct from the spectral scalar weight $W(f) = 1/J(f)$).
Multi-channel	$x_c(t),\ X_{c,k}$	Time / frequency-domain signal for channel c.
	$J_{n,c}(f)$	Channel-dependent noise PSD.

Appendix B Implementation of Optimal Filter

B.1 Optimal Filter

B.2 Sliding Optimal Filter

B.3 Scanning Optimal Filter

Appendix C Implementation of EMPCA