Deconvolution of Electron-Ionization Mass Spectra of Mixtures Without Chromatography

Deconvolution of Electron-Ionization Mass Spectra of Mixtures Without Chromatography

Raffaele Correale*
NanoTech Analysis Srl
Torino, Italy
r.correale@nanotechanalysis.com

Evelyne Lutton*
UMR 518 MIA-PS, INRAE, Université Paris-Saclay
Palaiseau, France
UAR 3611 ISC-PIF, CNRS
Paris, France
evelyne.lutton@inrae.fr

Giorgio Mongardi*
Politecnico di Torino
Torino, Italy
giorgio.mongardi@polito.it

Giovanni Squillero*
Politecnico di Torino
Torino, Italy
giovanni.squillero@polito.it

Raffaella Todino*
NanoTech Analysis Srl
Torino, Italy
r.todino@nanotechanalysis.com

Alberto Tonda*
UMR 518 MIA-PS, INRAE, Université Paris-Saclay
Palaiseau, France
UAR 3611 ISC-PIF, CNRS
Paris, France
alberto.tonda@inrae.fr

Abstract

We explain a single electron-ionization mass spectrum as a sparse nonnegative combination of reference spectra from a library without any chromatographic separation. The result has practical value for compact volatile-organic-compound analysis workflows. The deconvolution of a spectrum into its constituent compounds is a challenging task: the support is discrete, parameters are continuous, and objectives are nonconvex in correlated libraries. On top of that, we handle small integer mass-to-charge shifts, moderate variability in relative fragment intensities, and instrument-induced noise. The proposed workflow first screens the library to produce a high-recall shortlist using ion-informed heuristics; then, a greedy builder adds one candidate at a time, with CMA-ES with margin and a local refinement used before evaluating the residual fitness. Candidate supports are scored with a replicate-calibrated likelihood of the observed spectrum under the reconstruction, augmented with penalties that discourage structured residuals and peak-shaped mismatch. Experiments on real spectra and synthetic mixtures show effective identification with practical runtime.

CCS Concepts

• **Mathematics of computing** → **Evolutionary algorithms**; • **Theory of computation** → *Mixed discrete-continuous optimization*; • **Hardware** → *Signal integrity and noise analysis.*

Keywords

EI mass spectrometry, spectral deconvolution, library mixtures, mixed-integer optimization, CMA-ES with margin, greedy selection

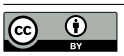
ACM Reference Format:
Raffaele Correale, Evelyne Lutton, Giorgio Mongardi, Giovanni Squillero, Raffaella Todino, and Alberto Tonda. 2026. Deconvolution of Electron-Ionization Mass Spectra of Mixtures Without Chromatography. In *Genetic and Evolutionary Computation Conference (GECCO Companion '26), July 13–17, 2026, San Jose, Costa Rica.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3795101.3805392

1 Introduction

Electron ionization mass spectrometry (EI–MS) obtains reproducible fragmentation patterns and is supported by large curated libraries [7, 9, 10]. A unit-resolution "stick" spectrum is commonly identified by library matching using dot-product similarity. Mixtures break this approach: when multiple compounds contribute fragments to the same integer m/z bins, similarity scores degrade for true constituents and can increase for incorrect references. Chromatographic separation (e.g., AMDIS [1, 8]) resolves components before matching, but requires additional equipment, increasing cost and complexity. The EI-only setting provides a single spectrum with no chromatographic context; mixture reasoning must be done directly against the library, resembling sparse nonnegative approximation with a redundant dictionary [4, 6].

In our prior work [3], we addressed this with a diagnostic-aware library filter and a PSO-driven greedy builder, demonstrating feasibility but highlighting library redundancy and strong fragment overlap as persistent challenges. This paper extends that pipeline with: (i) a shift-tolerant conjunctive filter tuned by leave-one-experiment-out CV; (ii) a greedy selector driven by CMA-ES with margin [5] (CMA-ESwM) with a two-level shift model; and (iii) a residual fitness combining replicate-calibrated log-scale likelihood with structure and shape penalties, using a label-free noise model.

*All authors contributed equally to this work and their names are listed in alphabetical order.



GECCO Companion '26, San Jose, Costa Rica

ACM ISBN 979-8-4007-2488-6/2026/07
https://doi.org/10.1145/3795101.3805392

2 Problem Formulation

Mixture model. Let $y \in \mathbb{R}^M_{\geq 0}$ be an observed EI spectrum on an integer m/z grid, and let $R = [r_1, \ldots, r_N] \in \mathbb{R}^{M\times N}_{\geq 0}$ be a library matrix (in our tests, $N = 2{,}451$ entries derived from the NIST/EPA/NIH Mass Spectral Library, base-peak normalized). We model the observation as a sparse nonnegative superposition:

$$y \approx \hat{y}(S, a, \boldsymbol{\beta}, \mathbf{s}) = \sum_{j\in S} a_j \, T_{s_j} \tilde{r}_j(\beta_j), a_j \geq 0, \quad \beta_j \in [\beta_{\min}, \beta_{\max}], \quad (1)$$

where $S \subset \{1, \ldots, N\}$ is the support set to be recovered, a_j are relative amplitudes, T_{s_j} shifts by s_j integer bins, and $\tilde{r}_j(\beta_j)$ is the library spectrum warped by a bounded per-component stretch:

$$[\tilde{r}_j(\beta)]_i \;=\; r_{j,\max}\left(\frac{r_{j,i}}{r_{j,\max}}\right)^{\beta}, \qquad \beta \in [\beta_{\min}, \beta_{\max}], \qquad (2)$$

with $r_{j,\max} = \max_i r_{j,i}$, $\beta_{\min} = 0.8$, $\beta_{\max} = 1.2$, and $\beta = 1$ recovering the original spectrum. Values $\beta > 1$ emphasize dominant ions; $\beta < 1$ flattens the spectrum. The base peak is preserved, but the total ion count changes; a_j absorbs this normalization shift.

Integer shift tolerance. Real instruments may introduce small integer m/z drift. We handle this at two levels: the filter evaluates candidates under shifts $s \in \{-s_{\max}, \ldots, s_{\max}\}$ ($s_{\max} = 2$) without stretches, while the solver estimates a global instrument offset before the greedy loop and searches a penalized per-component correction $\delta s_j \in \{-1, 0, +1\}$ during selection.

Truth groups. Libraries contain near-duplicates and isomers. We merge entries sharing the same CAS number, and entries sharing the same molecular formula if their cosine similarity exceeds 0.95. Evaluation credits a prediction as correct when it hits any member of a truth group.

3 Pipeline

Each query spectrum is mapped to an integer m/z grid and base-peak normalized.

3.1 Ion-Informed Library Filtering

The filtering stage reduces the candidate set using fast per-entry heuristics. A library entry is retained if it passes a conjunction of necessary conditions, evaluated under shifts $s \in \{-s_{\max}, \ldots, s_{\max}\}$. Let $\mathrm{supp}(y)$ and $\mathrm{supp}(r)$ be the nonzero m/z supports:

- **Overlap:** $|\mathrm{supp}(y) \cap \mathrm{supp}(r)| \geq \tau_{\mathrm{ovlp}}$, with a coverage requirement for entries with ≥ 15 nonzero peaks.
- **Cosine similarity:** $\cos(y, r) \geq \tau_{\cos}$, with a stricter threshold when both the query is dense (≥ 20 peaks) and the library entry is large.
- **Intensity checks:** overlap ratio, overlap intensity fraction, library-side recall $I_\cap / \|T_s r_j\|_1 \geq \tau_{\mathrm{rec}}$ (where $I_\cap$ is the summed intensity on shared peaks), missing-fraction bound, and excess/overshoot limit for dense mixtures.

Seven thresholds are tuned on replicate spectra by LOOCV over a grid of ~30,000 combinations, maximizing $U = 100 \cdot TP - FP$ with a recall floor of 0.25 per experiment.

3.2 Greedy Selection

Given a filtered candidate set C, we build a compact support by adding one component at a time. Before the greedy loop, a global m/z offset s_{global} is estimated by comparing the top-10 cosine-ranked candidates under all valid shifts and selecting the shift minimizing the NNLS reconstruction residual. At each greedy step, cosine similarity with the current residual re-ranks and retains the top-$K_{\mathrm{pre}} = 200$ candidates from C.

3.2.1 Per-addition search with CMA-ESwM.. At step t with current support $S^{(t)}$, we seek the candidate addition yielding the largest fitness decrease:

$$\Delta\mathcal{L}(j) = \max_{\beta_j \in [\beta_{\min}, \beta_{\max}]} \left[\mathcal{L}\left(S^{(t)}, \boldsymbol{\beta}^{(t)}\right) - \mathcal{L}\left(S^{(t)} \cup \{j\}, (\boldsymbol{\beta}^{(t)}, \beta_j)\right)\right], \quad (3)$$

evaluated after a short local refinement initialized with the proposed (α, β_j). CMA-ESwM [5] optimizes a continuous proposal vector $u = (u_{\mathrm{id}}, u_\alpha, u_\beta) \in \mathbb{R}^3$: the first coordinate is mapped to a discrete candidate index via logistic squashing and rounding; the remaining coordinates map to admissible initialization parameters:

$$\alpha(u) = \alpha_{\max}\, \varsigma(u_\alpha), \qquad \beta_j(u) = \beta_{\min} + (\beta_{\max} - \beta_{\min})\, \varsigma(u_\beta), \quad (4)$$

with $\alpha_{\max} = 2$ and ς denoting the logistic function. Each proposal additionally searches local shifts $\delta s_j \in \{-1, 0, +1\}$, selecting the shift minimizing a penalized fitness $\mathcal{L}_{\mathrm{pen}} = \mathcal{L} + \lambda\,|\delta s_j|$ ($\lambda = 30$). A seed scan of the top-$K_{\mathrm{seed}} = 20$ cosine-ranked candidates is evaluated before CMA-ESwM to provide a warm-start; below, $\tilde{R}_S(\boldsymbol{\beta}, \mathbf{s})$ denotes the matrix of stretched-and-shifted columns for S, with $s_j = s_{\mathrm{global}} + \delta s_j$ (shift vectors are selected during search and omitted from objective arguments for readability).

Amplitude refits under multiplicative noise. When the fidelity term is in log space, the objective is no longer quadratic in amplitudes. We refit using an iteratively reweighted NNLS proxy: at each refinement step we form weights $W(\mu) = \mathrm{diag}\big(1/\max(\mu_i, \varepsilon')\big)$ where ε' is a relative floor (1% of the current reconstruction maximum), and solve

$$a^\star(S, \boldsymbol{\beta}) = \arg\min_{a\geq 0} \; \tfrac{1}{2}\left\| W(\mu)\left(y - \tilde{R}_S(\boldsymbol{\beta}, \mathbf{s})\, a\right)\right\|_2^2, \qquad (5)$$

updating $\mu = \tilde{R}_S(\boldsymbol{\beta}, \mathbf{s})a$ and repeating for a small number of alternating steps with bounded stretch updates. Eq. (5) is a full-grid proxy; in the final fitness, the likelihood term is evaluated on $\Omega_+ = \{i : y_i > 0\}$ while penalty terms are computed on the full grid so that spurious predicted peaks are penalized.

3.2.2 Residual-based fitness. Given a support S with fitted amplitudes $a^\star$, stretches $\boldsymbol{\beta}$, and shifts $\mathbf{s}$, we compute $\hat{y} = \tilde{R}_S(\boldsymbol{\beta}, \mathbf{s})\, a^\star$. To prevent bins where $\hat{y} \approx 0$ and $y > 0$ from dominating the log-NLL, we define $\mu_{\mathrm{eff}} = \hat{y} + b$, where b is an instrument background level estimated as the median (across training-fold replicas) of the 10th percentile of intensities exceeding 1.0 on the base-peak-normalized scale. Supports are scored by:

$$\mathcal{L}(S, \boldsymbol{\beta}) = \mathcal{L}_{\mathrm{het}}(y, \mu_{\mathrm{eff}}) + \eta\, \mathcal{L}_{\mathrm{struct}}(\tilde{e}) + \kappa\, \mathcal{L}_{\mathrm{shape}}(\tilde{e}), \qquad (6)$$

where $\mathcal{L}_{\mathrm{het}}(y, \mu_{\mathrm{eff}}) = \sum_{i\in\Omega_+} -\log p(y_i \mid \mu_{\mathrm{eff},i})$ under a noise model learned from replicates (Section 3.3). The standardized residual uses

a bounded denominator:

$$\tilde{e}_i = \frac{y_i - \mu_{\text{eff},i}}{\max(\mu_{\text{eff},i},\ y_i) + \varepsilon}, \tag{7}$$

which saturates $|\tilde{e}_i| \le 1$, preventing blow-up on background peaks not in the library. The fidelity term $\mathcal{L}_{\text{het}}$ is unaffected as it operates on log-ratios directly. The structure penalty is the sum of squared autocorrelations up to lag $L = 5$:

$$\mathcal{L}_{\text{struct}}(\tilde{e}) = \sum_{\ell=1}^{L} \rho(\ell)^2, \rho(\ell) = \frac{\sum_{i=1}^{M-\ell}(\tilde{e}_i - \bar{\tilde{e}})(\tilde{e}_{i+\ell} - \bar{\tilde{e}})}{\sum_{i=1}^{M}(\tilde{e}_i - \bar{\tilde{e}})^2},\ \bar{\tilde{e}} = \frac{1}{M}\sum_{i=1}^{M}\tilde{e}_i. \tag{8}$$

The shape-fidelity term measures mismatch after filtering by a discrete Gaussian kernel g ($\sigma_g = 1.0$, half-width 2, size 5):

$$\mathcal{L}_{\text{shape}}(\tilde{e}) = \|\tilde{e} * g\|_2^2. \tag{9}$$

Multiplicative (log-scale) likelihood. For multiplicative noise we model the log-residual $r_i^{\log} = \log(y_i + \varepsilon) - \log(\mu_{\text{eff},i} + \varepsilon)$. Our default is log-Student-t, $r_i^{\log} \sim t_\nu(0, \sigma^2)$:

$$-\log p_{\log t}(y_i \mid \mu_{\text{eff},i}) = \tfrac{1}{2}\log(\nu\pi\sigma^2) + \tfrac{\nu+1}{2}\log\left(1 + \frac{(r_i^{\log})^2}{\nu\sigma^2}\right) + \log(y_i + \varepsilon) + \text{const}(\nu). \tag{10}$$

3.2.3 Stopping criterion. The greedy loop terminates when the best per-step improvement satisfies $\Delta\mathcal{L} < \alpha_{\text{stop}} \cdot \Delta_1 \cdot |\Omega_+|$, where $\Delta_1 = \frac{\nu+1}{2}\log(1 + 1/\nu)$ is the per-bin NLL cost of a one-sigma log-residual (i.e., the NLL difference between $r^{\log} = \sigma$ and a perfect fit), and $|\Omega_+|$ is the number of scored bins. Stopping requires this condition to hold for a patience window of 3 consecutive iterations in which the best candidate fails the threshold. After the greedy loop terminates, a consolidation step prunes components with amplitude below 0.1% of $\max_j a_j$ and performs a final alternating refinement on the remaining support.

3.3 Noise Model from Replicate Spectra

We estimate noise parameters purely from replica-to-replica variability (label-free): for each experiment with ≥ 3 replicas, we form log-residuals $r_i^{\log} = \log(y_i^{(k)} + \varepsilon) - \log(\bar{y}_i + \varepsilon)$ on bins exceeding a minimum intensity, where $\bar{y}$ is the cleaned mean after removing outlier replicas (those with cosine to the experiment mean below $\mu_{\cos} - 3\sigma_{\cos}$, with $\mu_{\cos}, \sigma_{\cos}$ the mean/std of replica-to-mean cosine). We compare lognormal and log-Student-t models by 5-fold CV on experiment-level splits; log-Student-t consistently achieves lower predictive NLL (≈−0.45 vs. +0.11 per bin for lognormal), with selected $\nu = 2$ and stable σ in the 0.062–0.077 range across nested folds.

4 Experimental Design

Datasets. Synthetic mixtures are generated by combining library spectra with controlled size K, log-uniform amplitudes and stretches, and multiplicative noise $y_i = \max(0,\ \mu_i \cdot \exp(\epsilon_i))$, $\epsilon_i \sim \mathcal{N}(0, \sigma_{\log}^2)$, with detection threshold and rare background spikes. Real EI spectra were acquired with a NanoTech Analysis prototype (SRS CIS 300): 8 experiments (3–50 replicas each) spanning gas mixtures (N_2, O_2, Ar, CO_2), organics (limonene, o-xylene, cyclohexane, acetone, ethanol), and mixed compositions with 1–6 truth compounds. The organic compounds were deliberately chosen to have EI fragmentation patterns overlap substantially at unit resolution, creating a challenging identification setting.

Table 1: Ablation on 198 synthetic spectra ($K \in [2,10]$, $\sigma_{\log}$=0.15, seed 42).

	$K \le 5$			$K > 5$		
Setup	R	P	F1	R	P	F1
New / Filter-only / none	1.00	0.00	0.01	1.00	0.00	0.01
Old / Filter-only / none	0.86	0.01	0.02	0.78	0.01	0.02
Old / CMA / log-Student-t	0.79	0.70	0.73	0.50	0.44	0.46
Old / CMA / lognormal	0.82	0.59	0.67	0.60	0.40	0.47
New / CMA / log-Student-t	0.88	0.74	0.79	0.60	0.49	0.53
New / CMA / lognormal	0.95	0.69	0.78	0.74	0.47	0.57
Old / PSO / old objective	0.65	0.28	0.32	0.46	0.16	0.21

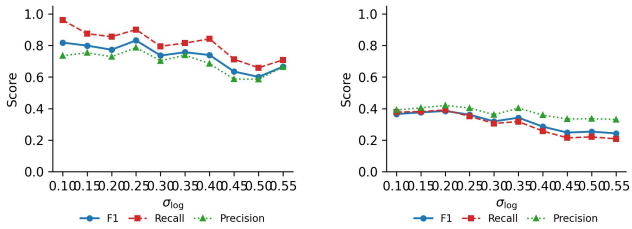


Figure 1: Noise robustness ($K \in [2, 19]$, 190 spectra per noise level). Left: $K \le 5$. Right: $K > 5$.

Ablation and metrics. We report a 7-configuration ablation on 198 synthetic spectra ($K \in [2, 10]$, $\sigma_{\log} = 0.15$). With truth groups $\{G_k\}$, a truth group is *hit* if $G_k \cap P \neq \emptyset$, where P is the predicted support; we report recall, precision, and F1.

5 Results

5.1 Ablation Results

Table 1 reports the ablation on 198 synthetic spectra. New/CMA/log-Student-t gives the best $K \le 5$ F1 (0.795), while New/CMA/lognormal gives the best $K > 5$ F1 (0.565). The new pipeline shows substantial improvement over the legacy PSO version, confirming the benefit of the new filter and of the CMA-ESwM with the log-scale fitness.

5.2 Synthetic Benchmark

Table 2 reports the fully-calibrated pipeline performance on synthetic mixtures by mixture size K ($K \in [2, 19]$, $n = 50$ per K, $\sigma_{\log} = 0.15$). The pipeline maintains high performance for small mixtures (aggregate $K \le 5$: R=0.895, P=0.721, F1=0.784) and degrades with mixture complexity (aggregate $K > 5$: R=0.418, P=0.424, F1=0.406).

Figure 1 shows noise robustness by sweeping the log-noise scale $\sigma_{\log}$ (10 levels, $n = 10$ per K per level). For $K \le 5$, F1 remains over 0.6; for $K > 5$, it shows lower absolute performance but similar stability.

Table 2: Pipeline performance on synthetic mixtures by K ($\sigma_{\log} = 0.15$, $n = 50$ per K, $K \in [2, 19]$).

K	R	P	F1	K	R	P	F1
2	0.98	0.84	0.89	11	0.43	0.42	0.41
3	0.92	0.74	0.80	12	0.50	0.49	0.49
4	0.84	0.66	0.72	13	0.33	0.37	0.34
5	0.84	0.65	0.72	14	0.33	0.36	0.33
6	0.80	0.63	0.70	15	0.28	0.37	0.31
7	0.73	0.58	0.64	16	0.28	0.35	0.31
8	0.63	0.49	0.54	17	0.27	0.36	0.30
9	0.56	0.47	0.50	18	0.24	0.36	0.28
10	0.47	0.45	0.45	19	0.20	0.31	0.24
$K \leq 5$	0.89	0.72	0.78	$K > 5$	0.42	0.42	0.41

Table 3: Per-experiment held-out results (nested LOOCV). Aggregate (mean): R=0.73, P=0.72, F1=0.69.

Experiment	K	n	R	P	F1
E1: Air	4	50	0.55	0.96	0.70
E2: N_2 bag	1	50	1.00	1.00	1.00
E3: limonene	1	3	1.00	0.47	0.64
E4: lim.+o-xyl.	2	3	0.83	0.39	0.53
E5: lim.+o-xyl.+cyc.	3	3	0.78	0.69	0.73
E6: lim.+o-xyl.+cyc.+acetone	4	3	0.58	0.78	0.67
E7: N_2+lim.	2	20	0.70	0.85	0.77
E8: N_2+full org.	6	10	0.43	0.63	0.51

5.3 Real Spectra Results

Table 3 reports results via nested leave-one-experiment-out CV: the outer loop holds out one experiment; the inner loop fits the noise model (σ, ν), background level b, and filter thresholds on the remaining 7 experiments (calibrated values in Table 4). In pure-organic experiments (E3–E6), recall is generally high but precision suffers, probably due to near-duplicate or isomeric library entries with overlapping fragments, producing persistent false positives. The air experiment (E1) achieves high precision (0.96) but only moderate recall (0.55), as argon's single peak at m/z 40 remains difficult for the filter to retain in a dense spectrum.

5.4 Pipeline Parameters

Table 4 summarizes key pipeline parameters. All values are determined by label-free calibration, cross-validated tuning, or ablation.

Table 4: Key pipeline parameters and their provenance.

Parameter	Value	How determined
σ (noise scale)	0.062–0.077	Median of per-exp. $\hat{\sigma}$
ν (Student-t df)	2.0	5-fold CV on pooled residuals
b (background)	2.80–3.22	10th pctl. per training fold
η ($\mathcal{L}_{\text{struct}}$ weight)	1.0	Grid search (36 combos)
κ ($\mathcal{L}_{\text{shape}}$ weight)	0.5	Grid search (36 combos)
α_{stop}	0.05	Tuned on training set
Filter thresholds (7)	LOOCV grid (~30k combos)	

6 Conclusion

This paper presented a pipeline for deconvolving EI mass spectra of mixtures directly against a reference library, without chromatographic separation. The approach combines three contributions: a shift-tolerant ion-informed filter tuned by leave-one-experiment-out CV, a greedy support builder driven by CMA-ESwM over a mixed discrete–continuous search space, and a residual fitness grounded in a label-free replicate-calibrated noise model. Experiments on both synthetic and real spectra confirm that the pipeline delivers reliable identification for small mixtures and degrades gracefully with complexity. Two systematic weaknesses persist: false positives from near-duplicate or isomeric library entries that share most fragments at unit resolution, and reduced recall when weak organic signals are overshadowed by a dominant carrier gas. Both limitations stem from the same structural bottleneck: the greedy one-at-a-time builder cannot revise earlier selections in light of later evidence. Future work will explore cooperative or coevolutionary schemes [2, 11] that evolve component sets jointly, which could mitigate sensitivity to greedy ordering. Additional directions include duplicate-aware support pruning to reduce isomer confusion, and intensity-discriminative screening to improve recall on minor components.

References

[1] A. A. Aksenov, I. Laponogov, Z. Zhang, P. C. Dorrestein, K. Veselkov, et al. 2021. Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. *Nature Biotechnology* 39, 2 (2021), 169–173. doi:10.1038/s41587-020-0700-3

[2] Pierre Collet, Evelyne Lutton, Frédéric Raynal, and Marc Schoenauer. 2000. Polar ifs+ parisian genetic programming= efficient ifs inverse problem solving. *Genetic Programming and Evolvable Machines* 1, 4 (2000), 339–361.

[3] Raffaele Correale, Evelyne Lutton, Giorgio Mongardi, Giovanni Squillero, Raffaella Todino, and Alberto Tonda. 2026. Deconvolution of Mass Spectra through Particle Swarm Optimization: An Industrial Experience. Conference paper / preprint.

[4] T. Gautier, J. Serigano, J. Bourgalais, S. M. Hörst, and M. G. Trainer. 2020. Decomposition of electron ionization mass spectra for space application using a Monte-Carlo approach. *Rapid Communications in Mass Spectrometry* 34, 8 (2020), e8684. doi:10.1002/rcm.8684

[5] Ryoki Hamano, Shota Saito, Masahiro Nomura, and Shinichi Shirakawa. 2022. CMA-ES with Margin: Lower-Bounding Marginal Probability for Mixed-Integer Black-Box Optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*. 639–647. doi:10.1145/3512290.3528827

[6] S. G. Mallat and Z. Zhang. 1993. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41, 12 (1993), 3397–3415. doi:10.1109/78.258082

[7] National Institute of Standards and Technology. 2020. NIST MS Search Program User's Guide. Online manual.

[8] NIST Mass Spectrometry Data Center. 2020. AMDIS — Automated Mass Spectral Deconvolution and Identification System: User Guide. Online manual.

[9] Stephen E. Stein. 1994. Optimization and testing of mass spectral library search algorithms. *Journal of the American Society for Mass Spectrometry* 5, 9 (1994), 859–866. doi:10.1016/1044-0305(94)87009-8

[10] Stephen E. Stein. 1999. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry* (1999). Library matching and match factor variants.

[11] Alberto Tonda, Evelyne Lutton, and Giovanni Squillero. 2011. Lamps: A test problem for cooperative coevolution. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2011)*. Springer, 101–120.